INRELCAST: IN-CONTEXT RELATION LEARNING FOR TIME SERIES FORECASTING WITH EXOGENOUS VARIABLES

Anonymous authors
Paper under double-blind review

ABSTRACT

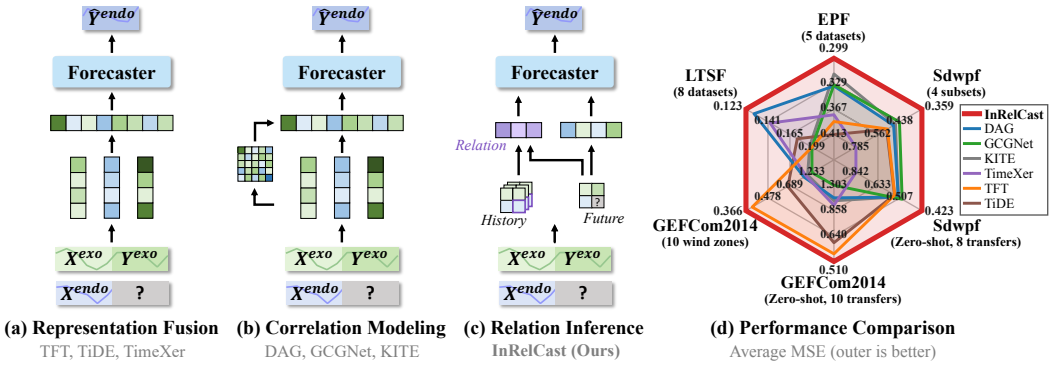
Exogenous variables such as weather and schedules are essential for forecasting a target series, and their values over the forecast horizon are often known or predicted in advance. Prevailing methods learn the forecast end to end from the look-back window and the future exogenous variables by fusing their representations or modeling temporal and channel correlations. However, these two inputs play different roles. The future exogenous variables specify the conditions ahead, whereas how the target responds to them depends on the current state of the series, so the forecast has to judge the current relation between the exogenous conditions and the target. The evidence for this judgment lies in the look-back window, where each transition from one period to the next pairs its conditions with the response that followed. Thus, these observed transitions act as answered examples and the future transitions, whose responses are unknown, act as new cases, which is precisely the setting of in-context learning. Nevertheless, prevailing methods encode both inputs together and leave this pairing and its use to implicit learning. Therefore, we propose **InRelCast**, an **In**-context **Rel**ation learning framework for fore**Cast**ing with exogenous variables, which infers the relation of each future transition in context from the observed transitions. Specifically, each observed transition serves as a demonstration and each future transition as a query, whose inferred relation generates low-rank parameter increments of a shared residual forecaster that corrects a baseline forecast from the target history. Since future responses are unobserved, training uses later observed transitions as pseudo-queries, whose hidden responses supervise the relations inferred for them. Experiments on multiple real-world datasets show that InRelCast outperforms state-of-the-art baselines, both in-domain and in zero-shot transfer to unseen series. The code is available at `https://anonymous.4open.science/r/InRelCast`.

1 INTRODUCTION

Time series forecasting estimates the future values of a target series and supports decisions in power dispatch and energy management (Hong et al., 2016; Lago et al., 2021), as well as in industrial production (Wang et al., 2024a; 2026; Cheng et al., 2026a). Most forecasting methods rely on the past values of the target series, whose variables are called endogenous (Zeng et al., 2023; Wu et al., 2023; Nie et al., 2023; Liu et al., 2024; Wang et al., 2024b; Qiu et al., 2025). However, many applications also provide exogenous variables that inform the evolution of the target, such as weather, calendar features, and operating schedules. Forecasting the target from its history and these exogenous variables is termed time series forecasting with exogenous variables (TSF-X) (Qiu et al., 2026; Li et al., 2026). Moreover, some of these variables are known or predicted in advance for the forecast horizon. As these future exogenous variables directly characterize the conditions of the forecast period, exploiting them effectively is central to TSF-X.

A direct approach is representation fusion, which incorporates the exogenous variables as additional inputs and builds a joint representation for forecasting through input encoding and feature interaction (Fig. 1 (a)). TFT (Lim et al., 2021) integrates static, historical, and known future inputs through variable selection and attention, TiDE (Das et al., 2023) combines the target history and the exogenous variables through an encoder-decoder structure, and TimeXer (Wang et al., 2024c)

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.



(a) Representation Fusion TFT, TiDE, TimeXer **(b) Correlation Modeling** DAG, GCGNet, KITE **(c) Relation Inference** **InRelCast (Ours)** **(d) Performance Comparison** Average MSE (outer is better)

Figure 1: Comparison of (a) representation fusion, (b) correlation modeling, and (c) relation inference paradigms, and (d) average MSE with future exogenous variables available.

connects temporal representations of the target history with representations of the exogenous variables through attention. In these models, how the exogenous variables and the target depend on each other is learned inside the joint representation. More recently, correlation modeling methods make this dependence an explicit object of modeling (Fig. 1 (b)). DAG (Qiu et al., 2026) discovers temporal and channel correlations and injects them into the corresponding forecasting pathways, GCGNet (Li et al., 2026) jointly models the two through graph structure alignment, and KITE (Cheng et al., 2026b) regularizes the channel correlations with statistical priors. Both approaches take the look-back window and the future exogenous variables as input and learn the forecast end to end.

However, the two parts of this input play different roles. The future exogenous variables specify the conditions ahead, whereas how the target responds to them depends on the series and on its current state, so the forecast has to judge the current relation between the exogenous conditions and the target. The evidence for this judgment lies in the look-back window, the other part of the input, where each transition from one period to the next records the conditions of the series and the response that followed. Thus, the observed transitions act as answered examples that pair conditions with their responses, while the future transitions present conditions of the same form and act as new cases whose responses are to be determined. This is precisely the setting of in-context learning (ICL), in which a model infers a rule from the answered examples in its input and applies it to new cases (Brown et al., 2020; Xie et al., 2022; Garg et al., 2022; von Oswald et al., 2023; Hendel et al., 2023). Nevertheless, representation fusion and correlation modeling both encode the two parts by variable and by time step as inputs of one network, and the modeled correlations link variables and time steps rather than each past condition with its response, leaving this pairing and its use to implicit learning. Therefore, exploiting the in-context structure of the task explicitly, by inferring the relation of each future transition from the observed transitions, remains an open design problem.

To this end, we propose **InRelCast**, an **In**-context **Rel**ation learning framework for fore**Cast**ing with exogenous variables, which infers the relation of each future transition in context from the observed transitions and conditions the forecaster on it (Fig. 1 (c)). A causal Transformer (Vaswani et al., 2017) first produces a baseline forecast from the target history alone, and each observed transition becomes a demonstration that pairs its endogenous state and exogenous conditions with the residual response of the target relative to this baseline, so that demonstrations carry only what the history does not predict. Each future transition, with its estimated endogenous state and known exogenous conditions, forms a query, and a context encoder over the demonstrations together with a relation decoder over the queries infers a relation for each query through attention. Each relation generates low-rank parameter increments (Hu et al., 2022) of a shared residual forecaster, adapting the forecasting map itself to the transition at hand. Since future responses are unobserved, relation inference is trained with historical pseudo-queries, in which later observed transitions serve as queries with their responses hidden, and an ICL loss matches the relation inferred for each hidden transition to its own demonstration, alongside the forecast loss. As shown in Fig. 1 (d), InRelCast achieves the lowest average MSE in all six evaluation groups, which span several dataset families under in-domain training and zero-shot transfer. Our main contributions are as follows.

- We formulate TSF-X as in-context relation learning, in which each observed transition serves as a demonstration that pairs its conditions with the response that followed, and each future transition forms a query whose relation is inferred from the demonstrations.

- We develop InRelCast, which infers the relation of each query from the demonstrations by attention, adapts a shared residual forecaster to each future transition through low-rank parameter increments generated from this relation, and, since future responses are unobserved, trains the inference with historical pseudo-queries.
- Extensive experiments on multiple real-world datasets show that InRelCast outperforms state-of-the-art baselines, both in-domain and in zero-shot transfer to unseen series. The code and experiment scripts are available at `https://anonymous.4open.science/r/InRelCast`.

2 Related work

2.1 Time series forecasting with exogenous variables

Classical statistical methods such as ARIMAX (Williams, 2001) and SARIMAX (Vagropoulos et al., 2016) add the exogenous variables to autoregressive models as regressors under a fixed parametric relation to the target. Instead, deep learning methods learn this relation from data and differ mainly in how the exogenous variables enter the network. Specifically, NBEATSx (Olivares et al., 2023) processes them in dedicated branches of neural basis expansion, CrossLinear (Zhou et al., 2025) captures dependencies between the endogenous and exogenous variables through cross-correlation embeddings, and TimeXer (Wang et al., 2024c), ExoTST (Tayal et al., 2024), and CITRAS (Yamaguchi et al., 2026) fuse exogenous representations into the target representation through attention. More recently, DAG (Qiu et al., 2026) models this relation explicitly as temporal and channel correlations, discovering those of the historical exogenous variables with the future exogenous and historical endogenous variables in separate modules and injecting them into the forecast. In contrast, GCGNet (Li et al., 2026) captures both in one graph over time and channels, aligns the graphs of generated and true sequences in training, and refines the forecast with this graph. Meanwhile, KITE (Cheng et al., 2026b) applies correlation modeling to probabilistic forecasting by flow matching conditioned on the exogenous variables, and guides its attention over them with statistical priors against spurious correlations. However, whether they fuse representations or model correlations between variables and time steps, these methods map the look-back window and the future exogenous variables directly to the forecast, so the pairing of each past condition with the response that followed is learned implicitly.

2.2 In-context learning for time series

For time series, ICL has mainly been used to forecast a series with its look-back window as the context (Das et al., 2024; Ansari et al., 2024; Woo et al., 2024; Auer et al., 2025b). When this context also contains the exogenous variables, it holds pairs of past conditions and the responses that followed, which can serve as in-context examples. Existing models differ in whether and how they form such examples. Specifically, COSMIC (Auer et al., 2025a) concatenates the exogenous and target patches and Chronos-2 (Ansari et al., 2025) connects them through group attention, both without forming explicit examples. In contrast, TabPFN-TS (Hoo et al., 2026) and ApolloPFN (Potapczynski et al., 2026) treat each time step as an example pairing its features with its target value. At a coarser scale, ICTSP (Lu et al., 2025) and Baguan-TS (Yang et al., 2026) pair a look-back segment with the segment that followed it, where ICTSP treats all channels alike and Baguan-TS includes the aligned exogenous variables. Beyond ICL, DeepTime (Woo et al., 2023) fits a ridge regressor to each look-back window on meta-learned time-index features, and Proceed (Zhao & Shen, 2025) turns the concept drift between recent training samples and each test sample into parameter adjustments. However, these models forecast without explicitly inferring the relation between the exogenous conditions and the target. In our work, the observed transitions serve as examples, from which the model infers how the target responds to the exogenous conditions of each future transition before forecasting.

3 Methodology

Problem formulation. Given the endogenous history $X^{\text{endo}} \in \mathbb{R}^{T\times N}$ and the exogenous history $X^{\text{exo}} \in \mathbb{R}^{T\times D}$, TSF-X aims to forecast the future target $Y^{\text{endo}} \in \mathbb{R}^{F\times N}$, where T and F are the look-back and horizon lengths and N and D are the numbers of target and exogenous channels. We

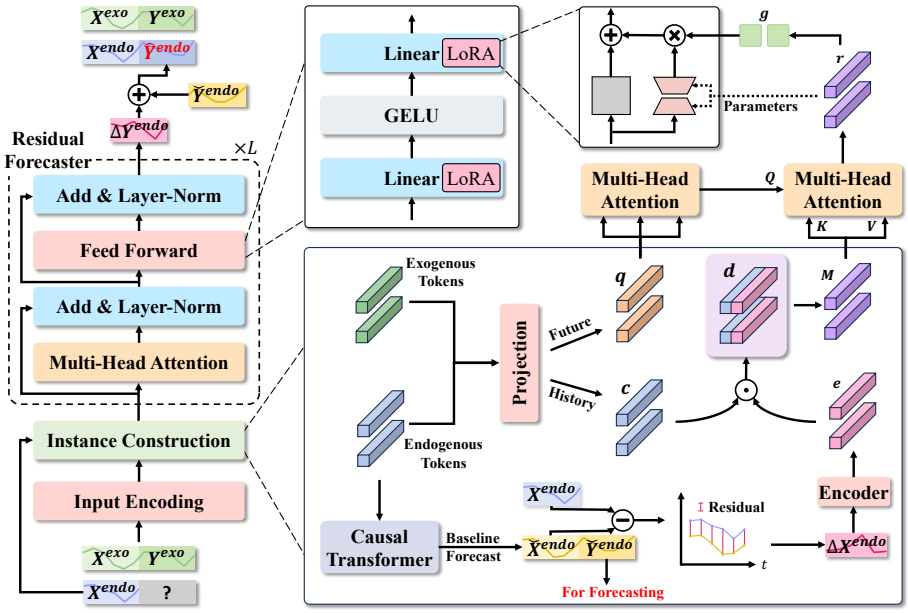


Figure 2: Overview of InRelCast. A causal Transformer forecasts the baseline from the target history, and its residuals form demonstrations d. For each query q, a relation r inferred from them generates increments of the residual forecaster, which adds the exogenous effect to the baseline.

consider the setting in which the future exogenous variables $Y^{\text{exo}} \in \mathbb{R}^{F\times D}$ are also available and learn a forecaster $\widehat{Y}^{\text{endo}} = \mathcal{F}_\theta(X^{\text{endo}}, X^{\text{exo}}, Y^{\text{exo}})$. When Y^{exo} is unavailable at inference, a predictor G trained on the training windows supplies $\widehat{Y}^{\text{exo}} = G(X^{\text{exo}})$ in its place.

Model overview. InRelCast infers within each window how the target responds to its exogenous conditions and applies the inferred relation to the future part of the same window. To this end, we divide the window into nonoverlapping patches of length p, with $K = T/p$ historical and $J = F/p$ future patches, and normalize the target by the mean μ and standard deviation s of its history (Kim et al., 2022). Let $x_i \in \mathbb{R}^{p\times N}$ and $v_i \in \mathbb{R}^{p\times D}$ be target and exogenous patch i over history and horizon, $i = 1, \ldots, K+J$. We treat the window as a sequence of transitions $i \to i+1$, each with a condition, formed by the endogenous state at patch i and the exogenous patches v_i and v_{i+1}, and a response of the target at patch $i+1$. Thus, the $K-1$ observed transitions serve as demonstrations and the J future transitions, whose responses are unknown, serve as queries, which casts TSF-X as ICL. We represent the relation of a transition, namely how its response depends on its condition, by a vector. For an observed transition, this vector is its demonstration, and for a future transition, InRelCast infers it from the demonstrations by attention. The vector adapts a shared residual forecaster through low-rank parameter increments, and the inference is trained with historical pseudo-queries. We summarize this flow in Fig. 2.

3.1 Endogenous Baseline and Demonstrations

Building a demonstration requires a state for the transition and a baseline against which its response is measured, both provided by a causal Transformer over the target history. Specifically, each target patch is embedded into a token of width d_{model}, the K tokens are followed by learned tokens that stand for the future patches, and a linear head maps each output state $h_i \in \mathbb{R}^{d_{\text{model}}}$ to a baseline forecast of the next patch:

$$\check{x}_{i+1} = \text{Linear}(h_i), \quad i = 1, \ldots, K-1, \qquad \check{y}_j = \text{Linear}(h_{K+j-1}), \quad j = 1, \ldots, J. \tag{1}$$

We write $\check{X}^{\text{endo}} = [\check{x}_2; \ldots; \check{x}_K]$ and $\check{Y}^{\text{endo}} = [\check{y}_1; \ldots; \check{y}_J]$ for the historical and future baselines.

For each observed transition $i \to i+1$, $i = 1, \ldots, K-1$, the condition c_i combines the state h_i with the exogenous patches on both sides of the transition. The response is the residual of the next target patch relative to its baseline, so it keeps only what the target history does not predict. The

demonstration d_i pairs the two:

$$c_i = \mathcal{C}(h_i, v_i, v_{i+1}), \quad \Delta x_{i+1} = x_{i+1} - \check{x}_{i+1},$$
$$e_i = E_r(\Delta x_{i+1}), \qquad d_i = \mathcal{D}(c_i, e_i). \tag{2}$$

Here the condition network $\mathcal{C}$ and the demonstration network $\mathcal{D}$ are multilayer perceptrons (MLPs) over their concatenated inputs, drawn as Projection and $\odot$ in Fig. 2, and the response encoder E_r is a linear layer, drawn as Encoder.

3.2 In-context relation inference

A future transition has no observed response, so we infer its relation from the demonstrations, with its condition $q_j = \mathcal{C}(h_{K+j-1}, v_{K+j-1}, v_{K+j})$, $j = 1, \ldots, J$, as the query. A context encoder turns the demonstrations into a memory M, from which a relation decoder infers one relation per query:

$$M = \mathrm{Enc}_{\mathrm{ctx}}(d_1, \ldots, d_{K-1}), \qquad (r_1, \ldots, r_J) = \mathrm{Dec}_{\mathrm{rel}}(q_1, \ldots, q_J; M). \tag{3}$$

We illustrate these inputs for $K = 3$ and $J = 2$ in Fig. 3. The context encoder attends over all demonstrations. The relation decoder self-attends across the J queries and cross-attends to M, so each relation $r_j \in \mathbb{R}^{d_{\mathrm{model}}}$ can draw on the whole future exogenous trajectory, whereas observed responses enter only via the demonstrations. Both modules also embed the offset of each input from the last demonstration, which places demonstrations and queries on one time axis. They are shared by all windows, and no parameter is updated at inference. Thus, as in neural processes (Garnelo et al., 2018; Kim et al., 2019), a new window adapts its relations only through its own demonstrations and queries, but the relations then modulate the weights of the forecaster instead of entering it as an input.

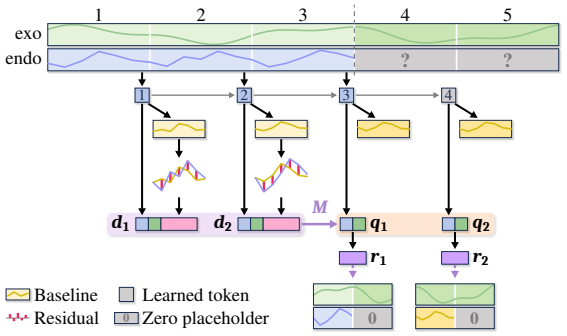


Figure 3: Input construction at inference. Observed transitions form demonstrations, and future transitions form queries whose relations condition the forecaster.

3.3 Relation-conditioned forecasting

The residual forecaster $\Psi(u, v, v'; z)$ is a Transformer that predicts the residual of the next target patch from the current target patch u, the exogenous patches v and v' on both sides of the transition, and a relation vector $z \in \mathbb{R}^{d_{\mathrm{model}}}$, with the unknown next target patch set to zero. The vector z is the demonstration d_i of an observed transition, whose response is this residual, or the inferred relation r_j of a future one. For future patch $K + j$, the current target patch is $u_1 = x_K$ for $j = 1$ and its baseline $u_j = \check{y}_{j-1}$ for $j > 1$, so all J residuals are computed in parallel:

$$\Delta y_j = \Psi(u_j, v_{K+j-1}, v_{K+j}; r_j), \qquad \widehat{Y}^{\mathrm{endo}} = \mathrm{Denorm}\big(\check{Y}^{\mathrm{endo}} + \Delta Y^{\mathrm{endo}}\big), \tag{4}$$

where Denorm inverts the normalization of the target and $\Delta Y^{\mathrm{endo}} = [\Delta y_1; \ldots; \Delta y_J]$.

The relation modulates Ψ through generated parameter increments (Ha et al., 2017; Phang et al., 2023), one set per future transition, so that each relation changes the forecasting map itself rather than only its inputs. Specifically, for every feedforward weight $W \in \mathbb{R}^{d^{\mathrm{out}} \times d^{\mathrm{in}}}$ of Ψ, affine generators produce rank-R factors $B(r_j) \in \mathbb{R}^{d^{\mathrm{out}} \times R}$ and $A(r_j) \in \mathbb{R}^{R \times d^{\mathrm{in}}}$ as in LoRA (Hu et al., 2022), and the effective weight is

$$W^{(j)} = W + \frac{g_j}{R} B(r_j) A(r_j), \qquad g_j = \mathrm{sigmoid}(\mathrm{Linear}(r_j)), \tag{5}$$

where the gate g_j scales the increments of all layers.

3.4 Training with pseudo-queries

Since future responses are unobserved, the forecast loss $\mathcal{L}_{\mathrm{pred}} = \ell(\widehat{Y}^{\mathrm{endo}}, Y^{\mathrm{endo}})$, with the mean absolute error ℓ, trains the relation inference only through the forecasts of Ψ. Therefore, we add

three losses on the history, where responses are observed. Two auxiliary losses train the baselines and the forecaster with each demonstration as its relation:

$$\mathcal{L}_{\text{endo}} = \frac{1}{K-1}\sum_{i=1}^{K-1} \ell(\check{x}_{i+1}, x_{i+1}), \qquad \mathcal{L}_{\text{rec}} = \frac{1}{K-1}\sum_{i=1}^{K-1} \ell(\Psi(x_i, v_i, v_{i+1}; d_i), \Delta x_{i+1}). \quad (6)$$

To supervise the inference itself, in each batch we split the $K-1$ demonstrations after $m \in \{1, \ldots, K-2\}$ entries and treat the remaining transitions as pseudo-queries whose responses are hidden. The causal Transformer is rerun on the first $m+1$ target patches followed by learned tokens, so that pseudo-queries receive their states in the same way as future queries at inference. The ICL loss then trains the relations r_j^{ps} inferred from $d_1, \ldots, d_m$ to match the demonstrations of the hidden transitions, held fixed as targets by the stop-gradient operator sg:

$$\mathcal{L}_{\text{icl}} = \frac{1}{K-1-m}\sum_{j=1}^{K-1-m} \left\| r_j^{\text{ps}} - \text{sg}(d_{m+j}) \right\|_2^2. \quad (7)$$

The total objective weights the four terms,

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda_{\text{endo}}\mathcal{L}_{\text{endo}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{icl}}\mathcal{L}_{\text{icl}}, \quad (8)$$

and includes the prediction loss $\lambda_{\text{exo}}\mathcal{L}_{\text{exo}}$ of G when G is used. Since $\mathcal{L}_{\text{rec}}$ trains Ψ to use demonstrations and $\mathcal{L}_{\text{icl}}$ trains relations to match them, r_j stands in for the missing demonstration at inference.

4 Experiments

4.1 Experimental setup

Datasets. To evaluate the forecasting performance of InRelCast, we conduct experiments on three groups of datasets. 1) Following prior work (Qiu et al., 2026; Li et al., 2026), we evaluate on the 12 real-world TSF-X datasets, namely five electricity price forecasting (EPF) datasets (Lago et al., 2021; Wang et al., 2024c) and seven datasets collected by DAG (Energy, Colbun, Rapel, and Sdwpf with 4 subsets). Their future exogenous variables are day-ahead forecasts or recorded values. 2) To evaluate forecasting of a weather-driven target whose future exogenous variables are available only as forecasts, we use the GEFCom2014 wind dataset (Hong et al., 2016), which pairs the hourly power output of ten wind farms with numerical weather predictions issued every day. 3) To examine robustness beyond the TSF-X datasets, we follow DAG and evaluate on eight common multivariate datasets that are widely used as forecasting benchmarks (ETT with 4 subsets, Weather, Exchange, Electricity, and Traffic), where the last variable is the forecasting target and the remaining variables serve as exogenous variables whose future values are provided. Details of all datasets are given in Appendix A.1.

Baselines. We compare InRelCast with ten baselines in two categories. 1) Methods that natively support future exogenous variables include DAG (Qiu et al., 2026), GCGNet (Li et al., 2026), KITE (Cheng et al., 2026b), TimeXer (Wang et al., 2024c), TFT (Lim et al., 2021), and TiDE (Das et al., 2023). 2) Methods without such native support include DUET (Qiu et al., 2025), CrossLinear (Zhou et al., 2025), TimeKAN (Huang et al., 2025), and PatchTST (Nie et al., 2023). For the second category, we follow DAG and incorporate future exogenous variables through an MLP fusion module. The adaptation procedure is described in Appendix A.2.

Implementation details. Following prior work (Qiu et al., 2026; Li et al., 2026), we adopt MSE and MAE as evaluation metrics. Colbun and Rapel use look-back lengths of 60 and 180 with horizons of 10 and 30. The other TSF-X datasets and GEFCom2014 wind use look-back lengths of 168 and 720 with horizons of 24 and 360. The eight common multivariate datasets fix the look-back length at 96 with horizons of 96, 192, 336, and 720. Data splits and preprocessing follow each benchmark, with normalization statistics computed from the training set only. We implement InRelCast in PyTorch (Paszke et al., 2019) and run all experiments on a single NVIDIA A800 GPU. Baseline results on the TSF-X and common multivariate datasets are taken from DAG, GCGNet, and KITE where available, whereas all results on GEFCom2014 wind and in zero-shot transfer are

obtained from our own experiments, with every model tuned along the same hyperparameter axes under the same budget (Appendix A.2) and selected on the validation split. For a fair comparison, all experiments that we run keep every test sample, including the final incomplete batch (Qiu et al., 2024; Li et al., 2025), and report the mean over three random seeds.

4.2 Main results

Forecasting on TSF-X datasets. Table 1 reports the results on the 12 TSF-X datasets. InRelCast achieves the best average MSE and MAE, with a broad lead across application domains. Compared with DAG, GCGNet, and KITE, the most recent state-of-the-art methods, it reduces the average MSE by 16.9%, 14.0%, and 14.3%, respectively. When future exogenous variables are not available and InRelCast forecasts them from their history, it also achieves the best average MSE and MAE (Table 9). This lead benefits from in-context relation learning, which infers from the look-back window of each series how its target currently responds to the exogenous variables.

Table 1: Average results over two horizons on the 12 TSF-X datasets with future exogenous variables available. **Red**: best, Blue: second best. Full results are provided in Table 8 of Appendix B.

Models	InRelCast (Ours)		DAG (2026)		GCGNet (2026)		KITE (2026b)		TimeXer (2024c)		TFT (2021)		TiDE (2023)		DUET (2025)		CrossLinear (2025)		TimeKAN (2025)		PatchTST (2023)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	**0.267**	**0.297**	0.362	0.344	0.346	0.337	0.325	0.323	0.418	0.371	0.379	0.375	0.443	0.400	0.411	0.408	0.371	0.387	0.405	0.419	0.390	0.396
PJM	**0.085**	0.181	0.093	0.180	0.093	0.186	0.096	**0.178**	0.108	0.198	0.114	0.207	0.142	0.246	0.102	0.197	0.112	0.223	0.139	0.262	0.133	0.259
BE	**0.404**	**0.271**	0.423	0.279	0.423	0.287	0.427	0.286	0.452	0.290	0.454	0.291	0.498	0.325	0.515	0.354	0.479	0.337	0.548	0.407	0.577	0.432
FR	0.411	0.228	0.414	**0.219**	0.405	0.230	**0.387**	0.225	0.427	0.241	0.504	0.257	0.484	0.281	0.496	0.327	0.483	0.298	0.547	0.374	0.588	0.410
DE	**0.328**	**0.354**	0.370	0.370	0.387	0.387	0.350	0.370	0.475	0.418	0.489	0.446	0.499	0.447	0.482	0.430	0.485	0.452	0.473	0.445	0.501	0.455
Energy	**0.105**	**0.253**	0.124	0.267	0.122	0.262	0.111	0.257	0.163	0.315	0.130	0.283	0.153	0.302	0.203	0.367	0.239	0.402	0.218	0.381	0.226	0.377
Sdwpfm1	**0.334**	**0.378**	0.423	0.461	0.416	0.449	0.417	0.451	0.701	0.609	0.482	0.474	0.483	0.507	0.599	0.570	0.426	0.502	0.447	0.534	0.435	0.502
Sdwpfm2	**0.398**	**0.421**	0.477	0.485	0.458	0.479	0.475	0.483	0.803	0.653	0.476	0.488	0.486	0.516	0.514	0.490	0.533	0.573	0.497	0.564	0.510	0.547
Sdwpfh1	**0.331**	**0.396**	0.448	0.486	0.414	0.468	0.441	0.491	0.746	0.643	0.479	0.491	0.453	0.508	0.539	0.516	0.557	0.593	0.577	0.638	0.468	0.527
Sdwpfh2	**0.372**	**0.409**	0.523	0.530	0.472	0.501	0.500	0.526	0.891	0.719	0.566	0.521	0.599	0.583	0.647	0.566	0.538	0.574	0.647	0.672	0.549	0.580
Colbun	**0.085**	**0.154**	0.098	**0.154**	0.098	**0.154**	0.088	0.172	0.145	0.235	0.238	0.297	0.164	0.227	0.198	0.266	0.126	0.195	0.128	0.175	0.239	0.309
Rapel	**0.188**	**0.248**	0.230	0.305	0.212	0.259	0.244	0.294	0.344	0.362	0.305	0.333	0.320	0.351	0.269	0.326	0.252	0.313	0.249	0.311	0.269	0.332
Average	**0.276**	**0.299**	0.332	0.340	0.321	0.333	0.322	0.338	0.473	0.421	0.385	0.372	0.394	0.391	0.415	0.401	0.383	0.404	0.406	0.432	0.407	0.427
1st Count	**12**	**11**	0	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Forecasting with weather predictions. Table 2 reports the results on GEFCom2014 wind, whose future exogenous variables are numerical weather predictions. InRelCast achieves the best results in most zones, followed by TFT, the strongest baseline on this dataset. We attribute the lead to the demonstrations, which pair the predicted wind of each past transition with the power output that followed and from which InRelCast infers how the output of each wind farm responds to the predictions it receives.

Table 2: Average results over two horizons on the 10 zones of GEFCom2014 wind with future exogenous variables available. Full results are provided in Table 10 of Appendix B.

Models	InRelCast (Ours)		DAG (2026)		GCGNet (2026)		KITE (2026b)		TimeXer (2024c)		TFT (2021)		TiDE (2023)		DUET (2025)		CrossLinear (2025)		TimeKAN (2025)		PatchTST (2023)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Zone 1	**0.413**	**0.464**	1.019	0.751	1.201	0.866	1.029	0.774	1.019	0.792	0.514	0.528	0.700	0.653	1.106	0.799	0.498	0.537	0.820	0.735	0.487	0.534
Zone 2	0.560	**0.512**	1.047	0.791	1.401	0.928	1.246	0.849	1.149	0.842	0.569	0.525	0.729	0.658	0.907	0.764	**0.545**	0.535	0.806	0.737	0.572	0.553
Zone 3	0.277	**0.399**	1.021	0.805	1.532	0.950	1.091	0.823	1.094	0.839	**0.269**	0.403	0.788	0.721	0.855	0.785	0.300	0.424	0.789	0.764	0.305	0.433
Zone 4	**0.284**	**0.358**	0.818	0.693	1.089	0.847	0.891	0.728	0.933	0.761	0.304	0.382	0.529	0.572	0.864	0.776	0.354	0.428	0.666	0.701	0.373	0.450
Zone 5	0.273	0.375	0.868	0.739	1.165	0.868	0.976	0.770	0.949	0.781	**0.268**	**0.373**	0.537	0.580	0.840	0.797	0.335	0.427	0.715	0.737	0.385	0.476
Zone 6	**0.309**	**0.380**	0.851	0.735	1.163	0.863	0.995	0.774	0.938	0.779	0.329	0.401	0.601	0.609	0.830	0.788	0.397	0.438	0.715	0.724	0.439	0.473
Zone 7	**0.299**	**0.394**	1.045	0.777	1.207	0.864	1.119	0.812	1.111	0.839	0.344	0.434	0.627	0.617	1.111	0.827	0.351	0.439	0.852	0.757	0.377	0.457
Zone 8	**0.553**	**0.503**	1.110	0.800	1.293	0.892	1.212	0.844	1.161	0.841	0.568	0.522	0.780	0.680	1.120	0.831	0.590	0.547	0.912	0.780	0.579	0.542
Zone 9	**0.326**	**0.408**	0.930	0.738	1.372	0.907	1.018	0.775	0.909	0.751	0.400	0.457	0.564	0.584	0.906	0.761	0.372	0.446	0.722	0.692	0.379	0.449
Zone 10	0.369	0.427	0.899	0.762	0.910	0.796	1.092	0.821	0.931	0.807	**0.353**	**0.423**	0.756	0.710	0.858	0.813	0.494	0.520	0.777	0.776	0.517	0.546
Average	**0.366**	**0.422**	0.961	0.759	1.233	0.878	1.067	0.797	1.019	0.803	0.392	0.445	0.661	0.638	0.940	0.794	0.424	0.474	0.777	0.740	0.441	0.491
1st Count	**7**	**9**	0	0	0	0	0	0	0	0	3	2	0	0	0	0	1	0	0	0	0	0

Zero-shot transfer across series. To evaluate generalization to unseen series, we conduct single-source transfer between Sdwpf subsets and multi-source transfer on GEFCom2014 wind, where a model trained jointly on five wind farms is evaluated on each of the other five, in both cases without training or fine-tuning on the target series. As shown in Table 3, InRelCast performs best in both settings, and its advantage holds on most source–target pairs and in both directions of multi-source transfer.

Table 3: Average results over two horizons of zero-shot transfer with future exogenous variables available, between the Sdwpf subsets (top) and from five zones of GEFCom2014 wind to each of the other five (bottom). Full results are provided in Tables 11 and 12 of Appendix B.

Models		InRelCast (Ours)		DAG (2026)		GCGNet (2026)		KITE (2026b)		TimeXer (2024c)		TFT (2021)		TiDE (2023)		DUET (2025)		CrossLinear (2025)		TimeKAN (2025)		PatchTST (2023)	
Source	Target	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Sdwpf, single-source																							
Sdwpfh1	Sdwpfh2	**0.433**	**0.443**	0.518	0.524	0.508	0.522	0.516	0.534	0.862	0.691	0.579	0.537	0.518	0.549	0.632	0.570	0.653	0.627	0.644	0.658	0.540	0.554
	Sdwpfm2	**0.445**	**0.451**	0.509	0.516	0.506	0.511	0.549	0.523	0.915	0.680	0.628	0.558	0.495	0.525	0.643	0.575	0.660	0.631	0.650	0.658	0.538	0.552
Sdwpfh2	Sdwpfh1	**0.322**	**0.366**	0.461	0.499	0.431	0.474	0.446	0.494	0.781	0.674	0.488	0.479	0.530	0.548	0.530	0.536	0.476	0.558	0.591	0.659	0.472	0.552
	Sdwpfm2	**0.432**	**0.426**	0.519	0.523	0.457	0.487	0.514	0.515	0.898	0.697	0.610	0.538	0.581	0.557	0.642	0.576	0.533	0.566	0.647	0.670	0.526	0.563
Sdwpfm1	Sdwpfh2	0.514	**0.480**	0.617	0.567	0.582	0.545	0.605	0.547	0.897	0.694	0.501	0.497	0.623	0.599	0.483	0.492	**0.453**	0.518	0.545	0.573	0.501	0.537
	Sdwpfm2	**0.441**	**0.438**	0.494	0.503	0.481	0.487	0.476	0.482	0.801	0.653	0.507	0.488	0.543	0.544	0.484	0.493	0.478	0.524	0.518	0.557	0.495	0.526
Sdwpfm2	Sdwpfh2	**0.444**	**0.463**	0.576	0.548	0.587	0.557	0.595	0.546	0.864	0.694	0.488	0.485	0.539	0.557	0.483	0.494	0.555	0.587	0.534	0.579	0.485	0.547
	Sdwpfm1	**0.352**	**0.392**	0.421	0.456	0.426	0.457	0.423	0.448	0.717	0.617	0.454	0.457	0.433	0.484	0.414	0.466	0.472	0.554	0.455	0.544	0.447	0.526
Average		**0.423**	**0.432**	0.514	0.517	0.497	0.505	0.516	0.511	0.842	0.675	0.532	0.505	0.533	0.545	0.539	0.525	0.535	0.571	0.573	0.612	0.501	0.545
1st Count		**8**	**9**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
GEFCom2014 wind, multi-source																							
Zones 1–5	Zone 6	0.467	**0.484**	0.933	0.780	1.039	0.823	0.860	0.723	0.840	0.732	**0.451**	0.489	0.535	0.571	0.593	0.628	0.584	0.618	0.706	0.672	0.490	0.578
	Zone 7	**0.424**	**0.471**	1.086	0.826	1.214	0.872	1.038	0.784	0.965	0.774	0.489	0.523	0.553	0.571	0.681	0.659	0.650	0.633	0.679	0.652	0.607	0.645
	Zone 8	**0.689**	**0.562**	1.142	0.847	1.245	0.878	1.113	0.807	1.024	0.795	0.727	0.595	0.694	0.625	0.761	0.685	0.731	0.669	0.738	0.677	0.748	0.700
	Zone 9	**0.462**	**0.494**	0.955	0.779	1.051	0.815	0.969	0.760	0.874	0.738	0.502	0.532	0.495	0.540	0.584	0.611	0.578	0.603	0.600	0.621	0.517	0.589
	Zone 10	0.552	**0.530**	0.969	0.804	1.053	0.839	0.986	0.788	0.917	0.778	0.579	0.558	0.582	0.596	0.646	0.665	0.612	0.635	0.751	0.696	**0.525**	0.595
Zones 6–10	Zone 1	**0.584**	**0.556**	1.108	0.827	1.656	0.973	1.100	0.805	1.007	0.795	0.657	0.618	0.635	0.619	0.761	0.691	0.707	0.674	0.754	0.706	0.734	0.737
	Zone 2	0.781	**0.627**	1.120	0.838	1.562	0.945	1.170	0.821	1.007	0.789	0.777	0.640	0.833	0.696	0.781	0.691	0.775	0.689	0.774	0.703	**0.771**	0.710
	Zone 3	**0.445**	**0.510**	1.169	0.865	1.403	0.921	0.949	0.776	0.929	0.783	0.477	0.547	0.638	0.637	0.630	0.663	0.622	0.653	0.658	0.679	0.547	0.625
	Zone 4	**0.363**	**0.420**	0.946	0.769	1.437	0.908	0.878	0.717	0.860	0.742	0.416	0.485	0.526	0.569	0.627	0.641	0.565	0.609	0.633	0.655	0.551	0.630
	Zone 5	**0.331**	**0.405**	0.942	0.770	1.370	0.900	0.838	0.714	0.841	0.745	0.354	0.446	0.510	0.564	0.599	0.640	0.569	0.620	0.616	0.657	0.490	0.591
Average		**0.510**	**0.506**	1.037	0.811	1.303	0.887	0.990	0.770	0.926	0.767	0.543	0.543	0.600	0.599	0.666	0.657	0.639	0.640	0.691	0.672	0.598	0.640
1st Count		**8**	**11**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0

Common multivariate datasets. We further evaluate InRelCast on the eight common multivariate datasets. The number of exogenous variables on these datasets ranges from 6 on ETT to 861 on Traffic. As shown in Table 4, InRelCast achieves the best performance. Compared with DAG, the strongest baseline on these datasets, it reduces the MSE by 15.5% on Traffic and the MAE by 26.9% on Weather. These results show that the advantage of InRelCast also holds on the widely used multivariate benchmarks when the future values of the remaining variables are provided.

Table 4: Average results over four horizons on the eight common multivariate datasets with future exogenous variables available. Full results are provided in Table 13 of Appendix B.

Models	InRelCast (Ours)		DAG (2026)		GCGNet (2026)		KITE (2026b)		TimeXer (2024c)		TFT (2021)		TiDE (2023)		DUET (2025)		CrossLinear (2025)		TimeKAN (2025)		PatchTST (2023)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT (Avg)	**0.106**	**0.241**	**0.106**	0.242	0.159	0.300	0.157	0.270	0.116	0.253	0.152	0.293	0.112	0.249	0.193	0.358	0.193	0.360	0.178	0.337	0.194	0.361
Weather	**0.001**	**0.019**	0.002	0.026	0.002	0.030	**0.001**	0.025	**0.001**	0.027	0.002	0.031	0.002	0.030	**0.001**	0.029	0.008	0.074	0.007	0.067	0.009	0.076
Electricity	0.208	0.351	**0.194**	**0.329**	–	–	0.408	0.420	0.208	0.337	0.375	0.471	0.313	0.417	0.281	0.411	0.264	0.399	0.301	0.429	0.269	0.404
Traffic	**0.109**	**0.199**	0.129	0.203	–	–	0.120	0.212	0.142	0.238	0.132	0.220	0.154	0.255	0.263	0.322	0.232	0.291	0.211	0.285	0.233	0.295
Exchange	**0.190**	**0.324**	0.212	0.353	0.506	0.515	0.290	0.388	0.250	0.380	0.334	0.438	0.293	0.398	0.270	0.455	0.312	0.494	0.425	0.565	0.464	0.618
Average	**0.123**	**0.227**	0.129	0.231	–	–	0.195	0.263	0.143	0.247	0.199	0.291	0.175	0.270	0.202	0.315	0.202	0.324	0.224	0.337	0.234	0.351
1st Count	**5**	**5**	2	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0

4.3 Model Analysis

Table 5: Ablation of InRelCast with future exogenous variables available. Sdwpf and GEFCom2014 are averaged over their subsets and zones, and Sdwpf† denotes zero-shot transfer between its subsets.

Variant	NP		DE		Energy		Sdwpf		Sdwpf†		GEFCom2014		Weather		Traffic		Average	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
InRelCast	**0.267**	**0.297**	**0.328**	**0.354**	**0.105**	**0.253**	**0.359**	**0.401**	**0.423**	**0.432**	**0.366**	**0.422**	**0.0008**	**0.0193**	**0.109**	**0.199**	**0.245**	**0.297**
w/o relation	0.312	0.315	0.449	0.417	0.122	0.273	0.430	0.442	0.457	0.458	0.406	0.443	0.0013	0.0249	0.133	0.229	0.289	0.325
w/o $\mathcal{L}_{\text{endo}}$ and $\mathcal{L}_{\text{rec}}$	0.274	0.301	0.429	0.398	0.108	0.259	0.448	0.451	0.488	0.474	0.389	0.431	0.0010	0.0215	0.113	0.206	0.281	0.318
w/o $\mathcal{L}_{\text{icl}}$	0.300	0.320	0.371	0.372	0.197	0.345	0.442	0.445	0.483	0.469	0.389	0.436	0.0009	0.0212	0.113	0.205	0.287	0.327
Joint-encoded history	0.291	0.316	0.379	0.385	0.110	0.261	0.505	0.484	0.536	0.500	0.395	0.439	0.0009	0.0207	0.121	0.217	0.292	0.328
Mean-pooled relation	0.269	0.299	0.352	0.369	0.129	0.281	0.428	0.448	0.447	0.460	0.397	0.440	0.0010	0.0213	0.113	0.200	0.267	0.315

Ablation study. Each variant in Table 5 changes one component of InRelCast and is retrained with the same hyperparameters. Removing the relation causes the largest increase in MSE on most datasets, since the residual forecaster then keeps the same weights for every future transition. Mean-pooled relation, which applies the mean of the demonstrations as one relation to the whole horizon, closes only part of this gap, so each future transition needs its own relation. Moreover, Joint-encoded history, which replaces the demonstrations with joint tokens of the historical target and exogenous

patches, is the worst variant on Sdwpf and in zero-shot transfer, indicating that the relations should be inferred from the demonstrations of the window. Finally, removing the auxiliary losses $\mathcal{L}_{\text{endo}}$ and $\mathcal{L}_{\text{rec}}$ or the ICL loss $\mathcal{L}_{\text{icl}}$ degrades every dataset, so this inference also needs supervision on the history. Thus, the gain comes from building demonstrations from the window and inferring a relation for each future transition.

Relation visualization. We visualize the relations of InRelCast on a GEFCom2014 wind test window in Fig. 4. At initialization, the relations of all future transitions are nearly identical, and during training they become specific to each transition (a)–(d). After training, each relation resembles the demonstration that the same transition forms once observed (e), showing that the ICL objective learned on history holds on test windows, and relations under different wind conditions generate distinct weight increments of the forecaster (f). Thus, each future transition receives its own relation, which adapts the forecaster to its conditions.

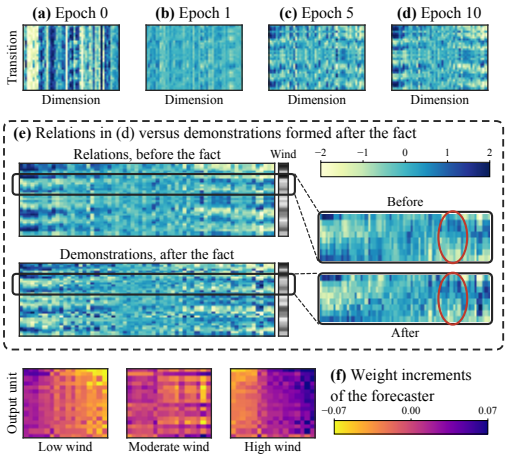


Figure 4: Visualization of the InRelCast relations on a test window of GEFCom2014 wind.

Model efficiency. As shown in Fig. 5, on all four datasets, InRelCast has the lowest MSE at a moderate inference cost, which stems from a single forward pass in which the relations only generate low-rank increments of a shared forecaster. In contrast, KITE, which averages many sampled forecasts, is the slowest model on all four datasets, and TFT and GCGNet exceed the 80 GB memory of the A800 GPU on Traffic. Although lightweight models such as CrossLinear, TiDE, and TimeXer are faster, their MSE is higher on every dataset. Moreover, InRelCast shows strong scalability, as its cost grows slowly when the number of exogenous variables increases from 2 on NP to 861 on Traffic, and Fig. 8 gives a more comprehensive analysis.

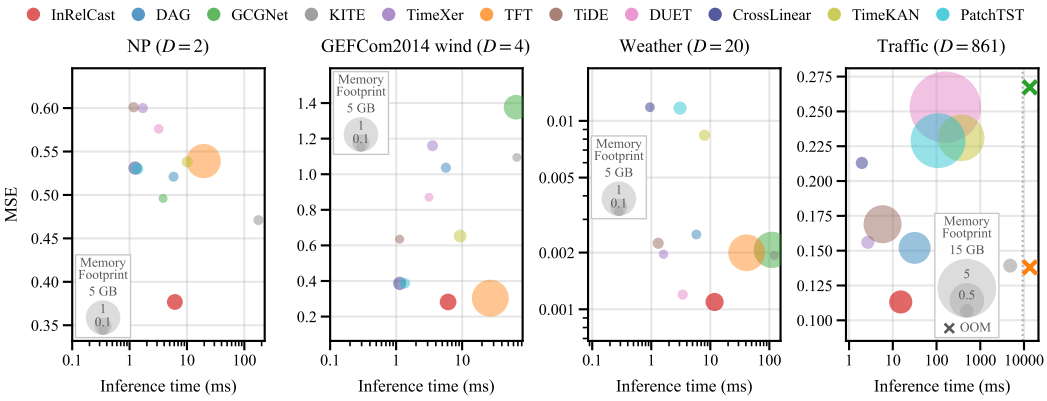


Figure 5: MSE against inference time at batch size 32, with future exogenous variables available.

5 Conclusion

This paper formulates TSF-X as in-context relation learning and proposes InRelCast, which infers a relation for each future transition from the observed transitions and adapts a shared residual forecaster to it through low-rank parameter increments. Experiments show that InRelCast outperforms state-of-the-art baselines both in-domain and in zero-shot transfer at a moderate computational cost. However, its demonstrations come only from the current look-back window, which may not cover the conditions ahead, and its advantage decreases as the noise in the future exogenous variables grows. Future work will retrieve demonstrations from beyond the current window and extend relation inference to probabilistic forecasting.

AI USE STATEMENT

We used a large language model (LLM) to check spelling, correct grammar, and refine the language of this paper. The LLM did not contribute to the research ideas, the method, or the experimental design. We take full responsibility for the content of this paper.

ETHICS STATEMENT

This work develops a forecasting method for time series with exogenous variables, with applications such as electricity price and wind power forecasting. It involves no human subjects, and all experiments use publicly available datasets released by prior work or forecasting competitions, which contain no personally identifiable information. We do not foresee direct societal harms from this research.

REPRODUCIBILITY STATEMENT

The anonymous repository `https://anonymous.4open.science/r/InRelCast` provides the code, the preprocessed datasets, and ready-to-run scripts for the in-domain, zero-shot, and ablation experiments, covering InRelCast and the baselines with the hyperparameters of every dataset and horizon. Appendix A describes the datasets, the baselines, and the tuning protocol.

REFERENCES

Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Türkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Bernie Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821*, 2025.

Andreas Auer, Raghul Parthipan, Pedro Mercado, Abdul Fatir Ansari, Lorenzo Stella, Bernie Wang, Michael Bohlke-Schneider, and Syama Sundar Rangapuram. Zero-shot time series forecasting with covariates via in-context learning. *arXiv preprint arXiv:2506.03128*, 2025a.

Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. TiRex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025b.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1877–1901, 2020.

Fang Cheng, Hui Liu, and Xinwei Lv. MetaGNSDformer: Meta-learning enhanced gated non-stationary informer with frequency-aware attention for point-interval remaining useful life prediction of lithium-ion batteries. *Advanced Engineering Informatics*, 69:103798, 2026a.

Hanyin Cheng, Jingrong Zhou, Yang Shu, and Chenjuan Guo. KITE: Knowledge-guided probabilistic modeling for time series forecasting with exogenous variables. In *International Conference on Machine Learning*, 2026b.

Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with TiDE: Time-series dense encoder. *Transactions on Machine Learning Research*, 2023.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 10148–10167, 2024.

Shivam Garg, Dimitris Tsipras, Percy Liang, and Gregory Valiant. What can transformers learn in-context? A case study of simple function classes. In *Advances in Neural Information Processing Systems*, volume 35, pp. 30583–30598, 2022.

Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Jimenez Rezende, and S. M. Ali Eslami. Conditional neural processes. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1704–1713, 2018.

David Ha, Andrew M. Dai, and Quoc V. Le. HyperNetworks. In *International Conference on Learning Representations*, 2017.

Roee Hendel, Mor Geva, and Amir Globerson. In-context learning creates task vectors. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 9318–9333, 2023.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, 2016.

Shi Bin Hoo, Samuel Müller, David Salinas, and Frank Hutter. From tables to time: Extending TabPFN-v2 to time series forecasting. *Transactions on Machine Learning Research*, 2026.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. TimeKAN: KAN-based frequency decomposition learning architecture for long-term time series forecasting. In *International Conference on Learning Representations*, 2025.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations*, 2024.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, S. M. Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In *International Conference on Learning Representations*, 2019.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 95–104, 2018.

Zhe Li, Xiangfei Qiu, Peng Chen, Yihang Wang, Hanyin Cheng, Yang Shu, Jilin Hu, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, and Bin Yang. TSFM-Bench: A comprehensive and unified benchmark of foundation models for time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5595–5606, 2025.

Zhengyu Li, Xiangfei Qiu, Yuhan Zhu, Xingjian Wu, Jilin Hu, Chenjuan Guo, and Bin Yang. GCGNet: Graph-consistent generative network for time series forecasting with exogenous variables. In *International Conference on Learning Representations*, 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.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

Jiecheng Lu, Yan Sun, and Shihao Yang. In-context time series predictor. In *International Conference on Learning Representations*, 2025.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Kin G. Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *International Journal of Forecasting*, 39(2):884–900, 2023.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Jason Phang, Yi Mao, Pengcheng He, and Weizhu Chen. HyperTuning: Toward adapting large language models without back-propagation. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27854–27875, 2023.

Andres Potapczynski, Ravi Kiran Selvam, Tatiana Konstantinova, Malcolm Wolff, Kin G. Olivares, Ruijun Ma, Michael W. Mahoney, Andrew Gordon Wilson, Boris N. Oreshkin, and Dmitry Efimov. Time-aware prior fitted networks for zero-shot forecasting with exogenous variables. *Transactions on Machine Learning Research*, 2026.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, and Bin Yang. TFB: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17 (9):2363–2377, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. DUET: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1185–1196, 2025.

Xiangfei Qiu, Yuhan Zhu, Zhengyu Li, Xingjian Wu, Bin Yang, and Jilin Hu. DAG: A dual correlation network for time series forecasting with exogenous variables. In *International Conference on Machine Learning*, 2026.

Kshitij Tayal, Arvind Renganathan, Xiaowei Jia, Vipin Kumar, and Dan Lu. ExoTST: Exogenous-aware temporal sequence transformer for time series prediction. In *IEEE International Conference on Data Mining (ICDM)*, pp. 857–862, 2024.

Stylianos I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In *IEEE International Energy Conference (ENERGYCON)*, pp. 1–6, 2016.

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

Johannes von Oswald, Eyvind Niklasson, Ettore Randazzo, João Sacramento, Alexander Mordvintsev, Andrey Zhmoginov, and Max Vladymyrov. Transformers learn in-context by gradient descent. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 35151–35174, 2023.

Hao Wang, Zhiyu Wang, Yunlong Niu, Zhaoran Liu, Haozhe Li, Yilin Liao, Yuxin Huang, and Xinggao Liu. An accurate and interpretable framework for trustworthy process monitoring. *IEEE Transactions on Artificial Intelligence*, 5(5):2241–2252, 2024a.

Hao Wang, Zhichao Chen, Licheng Pan, Xiaoyu Jiang, Yichen Song, Qunshan He, and Xinggao Liu. DeepFilter: A transformer-style framework for accurate and efficient process monitoring. *IEEE Transactions on Artificial Intelligence*, 7(7):3979–3989, 2026.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024b.

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. In *Advances in Neural Information Processing Systems*, volume 37, pp. 469–498, 2024c.

Billy M. Williams. Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 1776(1):194–200, 2001.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Learning deep time-index models for time series forecasting. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 37217–37237, 2023.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53140–53164, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Sang Michael Xie, Aditi Raghunathan, Percy Liang, and Tengyu Ma. An explanation of in-context learning as implicit Bayesian inference. In *International Conference on Learning Representations*, 2022.

Yosuke Yamaguchi, Issei Suemitsu, and Wenpeng Wei. CITRAS: Covariate-informed transformer for time series forecasting. *IEEE Access*, 14:77983–77998, 2026.

Linxiao Yang, Xue Jiang, Gezheng Xu, Tian Zhou, Min Yang, Zhaoyang Zhu, Linyuan Geng, Zhipeng Zeng, Qiming Chen, Xinyue Gu, Rong Jin, and Liang Sun. Baguan-TS: A sequence-native in-context learning model for time series forecasting with covariates. In *International Conference on Machine Learning*, 2026.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023.

Lifan Zhao and Yanyan Shen. Proactive model adaptation against concept drift for online time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2020–2031, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jian Tang, Jiantao Su, Yu Li, Ji Liu, Junfu Lyu, Yanjun Ma, and Dejing Dou. SDWPF: A dataset for spatial dynamic wind power forecasting over a large turbine array. *Scientific Data*, 11(1):649, 2024.

Pengfei Zhou, Yunlong Liu, Junli Liang, Qi Song, and Xiangyang Li. CrossLinear: Plug-and-play cross-correlation embedding for time series forecasting with exogenous variables. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4120–4131, 2025.

A EXPERIMENTAL DETAILS

A.1 DATASETS

Table 6 summarizes the datasets, all of which are split chronologically.

TSF-X datasets. The five EPF datasets (Lago et al., 2021), from the Nord Pool (NP), PJM, Belgian (BE), French (FR), and German (DE) markets, follow the version of TimeXer. The other seven were collected by DAG from the Chilean power grid (Energy), two Chilean reservoirs (Colbun and Rapel), and two wind turbines (Zhou et al., 2024) at two sampling rates (four Sdwpf subsets). Their future exogenous variables are day-ahead forecasts on the EPF datasets and recorded values on the others.

Common multivariate datasets. ETT (Zhou et al., 2021), Weather (Wu et al., 2021), and Exchange, Electricity, and Traffic (Lai et al., 2018) follow the versions of TimeXer and DAG, in which the last variable is the target and the future values of the other variables are provided as exogenous variables.

GEFCom2014 wind. The wind track (Hong et al., 2016) of the Global Energy Forecasting Competition 2014 (GEFCom2014) covers ten wind farms (zones) in Australia with hourly power output normalized to $[0, 1]$ over 2012 and 2013. Each series concatenates the training data of the last competition task and its released solution. The exogenous variables are the numerical weather predictions (NWP) of the competition, issued every day at midnight for the next 24 hours. Horizon 24 matches one issue, and at horizon 360 the inputs concatenate successive daily issues, as the day-ahead forecasts of the EPF datasets do.

Table 6: Statistics of the datasets. Exo.: number of exogenous variables; Length: number of time steps; Split: train, validation, and test ratio. Sdwpf, ETT, and GEFCom2014 wind take one row for their subsets or zones.

Dataset	Exo.	Exogenous variables	Target	Frequency	Length	Split
NP	2	Grid load, wind power	Electricity price	1 hour	52,416	7:1:2
PJM	2	System load, zonal load	Electricity price	1 hour	52,416	7:1:2
BE	2	Generation, system load	Electricity price	1 hour	52,416	7:1:2
FR	2	Generation, system load	Electricity price	1 hour	52,416	7:1:2
DE	2	Wind power, zonal load	Electricity price	1 hour	52,416	7:1:2
Energy	5	Battery, geothermal, hydro, solar, wind	Thermoelectric generation	1 hour	13,064	7:1:2
Colbun	2	Precipitation, tributary inflow	Water level	1 day	2,958	7:1:2
Rapel	2	Precipitation, tributary inflow	Water level	1 day	3,366	7:1:2
Sdwpfh (2 subsets)	6	ERA5 weather variables (Hersbach et al., 2020)	Active power	1 hour	14,641	7:1:2
Sdwpfm (2 subsets)	6	ERA5 weather variables	Active power	30 min	29,281	7:1:2
GEFCom2014 (10 zones)	4	NWP wind components at 10 and 100 m	Wind power	1 hour	17,544	7:1:2
ETTh (2 subsets)	6	Power load features	Oil temperature	1 hour	14,400	6:2:2
ETTm (2 subsets)	6	Power load features	Oil temperature	15 min	57,600	6:2:2
Weather	20	Climate features	CO_2 concentration	10 min	52,696	7:1:2
Exchange	7	Exchange rates	Exchange rate	1 day	7,588	7:1:2
Electricity	320	Electricity consumption	Electricity consumption	1 hour	26,304	7:1:2
Traffic	861	Road occupancy rates	Road occupancy rate	1 hour	17,544	7:1:2

A.2 BASELINES

Sources of the reported results. All results of InRelCast come from our own experiments under the evaluation protocol of TFB (Qiu et al., 2024), as in DAG and GCGNet. Baseline results are taken from DAG, GCGNet, and KITE where available and are otherwise obtained by us with the released configurations of the baselines. On GEFCom2014 wind, which has no published configurations, every model, including InRelCast, is tuned with the same protocol and budget: starting from the base configuration in Table 7, each run changes one hyperparameter, a final run combines the changes that lower the validation loss, and the run with the lowest validation loss is selected. The same protocol tunes GCGNet and KITE on the common multivariate datasets and ICTSP in Appendix C. Accordingly, all baseline results in Tables 1, 8, 9, and 18 come from the DAG paper, except for GCGNet and KITE, which come from their own papers. Those in Tables 4 and 13 come from the DAG paper except for GCGNet and KITE, which are obtained by us. Those in Tables 2, 3, 10, 11, 12, 14, and 16 are obtained by us. In zero-shot transfer, each model is trained on the source with its configuration for that dataset and is applied to each target, whose variables match those of the source, without updating any parameter. For every model, the normalization statistics of a target are computed from its training split, as in the in-domain experiments.

MLP fusion. Following DAG and GCGNet, each baseline without native support for future exogenous variables first forecasts from the look-back window with its own architecture g, giving $Z = g(X^{\text{endo}}, X^{\text{exo}})$. At each future time step, an MLP with one hidden layer and ReLU activation maps the concatenation of Z and Y^{exo} along the channels to the final forecast $\widehat{Y}^{\text{endo}}$. Both parts are trained jointly. The hidden width of the MLP follows the configuration of each baseline.

Table 7: Uniform tuning protocol on GEFCom2014 wind.

Model	Loss	learning rate	d_{model}	patch	layers	dropout	batch
Search values	–	$10^{-4}, 5\times10^{-4}, 10^{-3}$	64, 128, 256	8, 12, 24	–	base $+0.1$	–
InRelCast	MAE	5×10^{-4}	128	12	2	0.1	64
DAG	MAE	10^{-3}	64	8	1	0	64
GCGNet	MAE	$10^{-3}/10^{-4}$	256	24/360	–	0	64
KITE	Flow	5×10^{-4}	64	–	2	0	128
TimeXer	Huber	10^{-4}	256	16	1	0.2	64
TFT	MSE	10^{-3}	128	–	2	0.1	8
TiDE	MSE	10^{-3}	256	–	2	0.1	512
DUET	Huber	2×10^{-2}	512	–	2	0.2	256
CrossLinear	MSE	10^{-5}	256	8	–	0.1	16
TimeKAN	MSE	10^{-3}	16	–	2	0.1	16
PatchTST	MSE	10^{-4}	512	16	1	0.1	32

A.3 FORECASTING AND TRAINING PROCEDURE

Algorithm 1 summarizes the forecasting and training procedure of InRelCast in the default configuration of Section 3.

Algorithm 1 Forecasting and training with InRelCast in the default configuration, where ρ is the minimum context fraction. When the future exogenous variables are unavailable, the predictor G supplies $\widehat{Y}^{\text{exo}} = G(X^{\text{exo}})$ in place of Y^{exo} at inference.

procedure FORECAST($X^{\text{endo}}, X^{\text{exo}}, Y^{\text{exo}}$)
  normalize the target history with μ and s; form $x_1, \dots, x_K$ and $v_1, \dots, v_{K+J}$
  $h_1, \dots, h_{K+J-1} \leftarrow$ causal Transformer; $\check{X}^{\text{endo}}, \check{Y}^{\text{endo}} \leftarrow$ Eq. (1)
  $d_1, \dots, d_{K-1} \leftarrow$ Eq. (2); $q_j \leftarrow \mathcal{C}(h_{K+j-1}, v_{K+j-1}, v_{K+j}), j = 1, \dots, J$
  $r_1, \dots, r_J \leftarrow$ Eq. (3); $\Delta y_1, \dots, \Delta y_J \leftarrow$ Eqs. (4) and (5), in parallel
  return $\widehat{Y}^{\text{endo}} = \text{Denorm}(\check{Y}^{\text{endo}} + \Delta Y^{\text{endo}})$
procedure TRAINSTEP(a batch of training windows)
  if G is used **then**
    $\widehat{Y}^{\text{exo}} \leftarrow G(X^{\text{exo}})$; compute $\mathcal{L}_{\text{exo}}$
  $\widehat{Y}^{\text{endo}} \leftarrow$ FORECAST($X^{\text{endo}}, X^{\text{exo}}, Y^{\text{exo}}$); compute $\mathcal{L}_{\text{pred}}$, $\mathcal{L}_{\text{endo}}$, and $\mathcal{L}_{\text{rec}}$
  draw m uniformly from $\{\max(1, \lfloor \rho(K-1) \rfloor), \dots, K-2\}$
  $h^{\text{ps}}_{m+1}, \dots, h^{\text{ps}}_{K-1} \leftarrow$ causal Transformer on $x_1, \dots, x_{m+1}$ and $K-2-m$ learned tokens
  $q^{\text{ps}}_j \leftarrow \mathcal{C}(h^{\text{ps}}_{m+j}, v_{m+j}, v_{m+j+1}), j = 1, \dots, K-1-m$
  $r^{\text{ps}}_1, \dots, r^{\text{ps}}_{K-1-m} \leftarrow$ Eq. (3) with the context $d_1, \dots, d_m$ and the queries q^{ps}_j
  $\mathcal{L} \leftarrow$ Eq. (8), plus $\lambda_{\text{exo}}\mathcal{L}_{\text{exo}}$ if computed; one Adam (Kingma & Ba, 2015) step on $\mathcal{L}$

B FULL RESULTS

Tables 8–13 list the per-horizon results. In every table, Average is the mean over the datasets, zones, or source–target pairs, and 1st Count is the number of first places over all rows, Avg and Average included. The eight Sdwpf source–target pairs follow the ETT zero-shot convention of Time-LLM (Jin et al., 2024).

Table 8: Full results on the 12 TSF-X datasets with future exogenous variables available. **Red**: best, Blue: second best.

Models		InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.156**	**0.199**	0.202	0.237	0.197	0.233	0.179	0.223	0.236	0.266	0.219	0.249	0.284	0.301	0.246	0.287	0.210	0.266	0.273	0.310	0.249	0.294
	360	**0.377**	**0.395**	0.521	0.451	0.496	0.442	0.471	0.422	0.600	0.475	0.539	0.501	0.601	0.498	0.576	0.528	0.531	0.508	0.538	0.529	0.530	0.498
	Avg	**0.267**	**0.297**	0.362	0.344	0.346	0.337	0.325	0.323	0.418	0.371	0.379	0.375	0.443	0.400	0.411	0.408	0.371	0.387	0.405	0.419	0.390	0.396
PJM	24	0.062	**0.142**	0.057	0.143	0.058	0.148	**0.056**	0.143	0.075	0.166	0.095	0.195	0.106	0.214	0.072	0.166	0.088	0.191	0.115	0.244	0.116	0.239
	360	**0.107**	0.219	0.130	0.218	0.128	0.223	0.135	**0.214**	0.140	0.231	0.133	0.219	0.177	0.279	0.131	0.228	0.135	0.254	0.162	0.281	0.149	0.279
	Avg	**0.085**	0.181	0.093	0.180	0.093	0.186	0.096	**0.178**	0.108	0.198	0.114	0.207	0.142	0.246	0.102	0.197	0.112	0.223	0.139	0.262	0.133	0.259
BE	24	0.338	**0.221**	0.361	0.229	**0.323**	0.227	0.348	0.240	0.392	0.253	0.426	0.272	0.426	0.285	0.432	0.272	0.391	0.259	0.451	0.319	0.452	0.326
	360	**0.470**	0.320	0.485	0.330	0.524	0.347	0.507	0.332	0.512	0.327	0.482	**0.310**	0.571	0.364	0.597	0.436	0.568	0.416	0.645	0.495	0.702	0.538
	Avg	**0.404**	**0.271**	0.423	0.279	0.423	0.287	0.427	0.286	0.452	0.290	0.454	0.291	0.498	0.325	0.515	0.354	0.479	0.337	0.548	0.407	0.577	0.432
FR	24	0.358	0.180	0.355	**0.171**	0.332	0.179	**0.311**	0.178	0.366	0.208	0.543	0.253	0.418	0.255	0.384	0.251	0.390	0.226	0.454	0.296	0.518	0.368
	360	0.464	0.275	0.473	0.268	0.478	0.280	**0.462**	0.271	0.489	0.273	0.465	**0.261**	0.551	0.308	0.607	0.403	0.575	0.370	0.641	0.452	0.658	0.452
	Avg	0.411	0.228	0.414	**0.219**	0.405	0.230	**0.387**	0.225	0.427	0.241	0.504	0.257	0.484	0.281	0.496	0.327	0.483	0.298	0.547	0.374	0.588	0.410
DE	24	**0.252**	**0.307**	0.277	0.322	0.282	0.328	0.262	0.327	0.339	0.362	0.380	0.383	0.367	0.383	0.376	0.378	0.387	0.396	0.399	0.412	0.413	0.412
	360	**0.403**	**0.401**	0.462	0.418	0.491	0.445	0.437	0.413	0.610	0.474	0.599	0.509	0.630	0.511	0.589	0.482	0.583	0.507	0.547	0.479	0.589	0.498
	Avg	**0.328**	**0.354**	0.370	0.370	0.387	0.387	0.350	0.370	0.475	0.418	0.489	0.446	0.499	0.447	0.482	0.430	0.485	0.452	0.473	0.445	0.501	0.455
Energy	24	0.082	0.219	0.079	0.215	**0.065**	**0.194**	0.071	0.208	0.122	0.273	0.093	0.235	0.103	0.248	0.117	0.283	0.241	0.418	0.135	0.298	0.239	0.390
	360	**0.127**	**0.286**	0.169	0.320	0.179	0.330	0.151	0.305	0.204	0.357	0.167	0.331	0.202	0.355	0.288	0.452	0.237	0.385	0.302	0.464	0.214	0.363
	Avg	**0.105**	**0.253**	0.124	0.267	0.122	0.262	0.111	0.257	0.163	0.315	0.130	0.283	0.153	0.302	0.203	0.367	0.239	0.402	0.218	0.381	0.226	0.377
Sdwpf1	24	**0.311**	**0.355**	0.351	0.400	0.375	0.409	0.384	0.410	0.558	0.533	0.366	0.421	0.474	0.488	0.551	0.564	0.355	0.473	0.418	0.503	0.374	0.454
	360	**0.357**	**0.401**	0.495	0.522	0.458	0.490	0.450	0.491	0.845	0.684	0.597	0.528	0.492	0.526	0.646	0.577	0.497	0.532	0.476	0.566	0.497	0.550
	Avg	**0.334**	**0.378**	0.423	0.461	0.416	0.449	0.417	0.451	0.701	0.609	0.482	0.474	0.483	0.507	0.599	0.570	0.426	0.502	0.447	0.534	0.435	0.502
Sdwpf2	24	0.379	**0.414**	**0.372**	**0.414**	0.410	0.443	0.443	0.450	0.627	0.570	0.411	0.458	0.461	0.492	0.445	0.452	0.477	0.536	0.474	0.538	0.418	0.492
	360	**0.417**	**0.428**	0.583	0.556	0.505	0.514	0.506	0.515	0.978	0.736	0.541	0.519	0.511	0.540	0.584	0.528	0.589	0.611	0.520	0.589	0.602	0.603
	Avg	**0.398**	**0.421**	0.477	0.485	0.458	0.479	0.475	0.483	0.803	0.653	0.476	0.488	0.486	0.516	0.514	0.490	0.533	0.573	0.497	0.564	0.510	0.547
Sdwph1	24	**0.320**	**0.377**	0.408	0.438	0.381	0.441	0.424	0.464	0.651	0.587	0.401	0.460	0.434	0.489	0.527	0.513	0.548	0.585	0.511	0.582	0.422	0.505
	360	**0.341**	**0.415**	0.489	0.534	0.446	0.495	0.457	0.517	0.841	0.700	0.557	0.523	0.472	0.527	0.551	0.519	0.566	0.601	0.643	0.694	0.513	0.548
	Avg	**0.331**	**0.396**	0.448	0.486	0.414	0.468	0.441	0.491	0.746	0.643	0.479	0.491	0.453	0.508	0.539	0.516	0.557	0.593	0.577	0.638	0.468	0.527
Sdwph2	24	**0.377**	**0.423**	0.438	0.465	0.426	0.474	0.477	0.498	0.820	0.677	0.474	0.493	0.579	0.553	0.629	0.563	0.468	0.540	0.580	0.614	0.457	0.527
	360	**0.366**	**0.394**	0.608	0.595	0.519	0.529	0.524	0.555	0.962	0.761	0.657	0.549	0.619	0.614	0.665	0.569	0.608	0.609	0.713	0.729	0.641	0.632
	Avg	**0.372**	**0.409**	0.523	0.530	0.472	0.501	0.500	0.526	0.891	0.719	0.566	0.521	0.599	0.583	0.647	0.566	0.538	0.574	0.647	0.672	0.549	0.580
Colbun	10	0.056	**0.084**	0.061	0.094	0.057	**0.084**	**0.055**	0.101	0.113	0.172	0.092	0.135	0.089	0.131	0.089	0.134	0.071	0.102	0.061	0.101	0.062	0.122
	30	**0.113**	0.223	0.135	**0.215**	0.138	0.224	0.121	0.243	0.176	0.299	0.383	0.460	0.240	0.322	0.307	0.397	0.182	0.288	0.195	0.249	0.417	0.496
	Avg	**0.085**	**0.154**	0.098	**0.154**	0.098	**0.154**	0.088	0.172	0.145	0.235	0.238	0.297	0.164	0.227	0.198	0.266	0.126	0.195	0.128	0.175	0.239	0.309
Rapel	10	0.143	0.194	0.151	0.203	**0.140**	**0.191**	0.192	0.231	0.301	0.308	0.201	0.253	0.228	0.271	0.174	0.219	0.163	0.209	0.174	0.231	0.163	0.218
	30	**0.233**	**0.301**	0.309	0.408	0.284	0.327	0.296	0.358	0.387	0.416	0.409	0.414	0.411	0.432	0.365	0.432	0.340	0.417	0.325	0.390	0.374	0.445
	Avg	**0.188**	**0.248**	0.230	0.305	0.212	0.259	0.244	0.294	0.344	0.362	0.305	0.333	0.320	0.351	0.269	0.326	0.252	0.313	0.249	0.311	0.269	0.332
Average		**0.276**	**0.299**	0.332	0.340	0.321	0.333	0.322	0.338	0.473	0.421	0.385	0.372	0.394	0.391	0.415	0.401	0.383	0.404	0.406	0.432	0.407	0.427
1st Count		**28**	**28**	1	5	3	4	5	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0

Table 9: Full results on the 12 TSF-X datasets when future exogenous variables are not available. **Red**: best, Blue: second best.

Models		InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.219**	**0.248**	0.237	0.259	0.234	0.259	0.220	0.257	0.270	0.292	0.330	0.321	0.297	0.309	0.262	0.271	0.246	0.282	0.336	0.344	0.283	0.296
	360	0.622	0.487	0.601	0.500	0.617	0.495	**0.590**	**0.472**	0.609	0.474	0.964	0.655	0.638	0.523	0.626	0.494	0.655	0.505	0.632	0.527	0.630	0.505
	Avg	0.421	0.368	0.419	0.380	0.425	0.377	**0.405**	**0.365**	0.440	0.383	0.647	0.488	0.467	0.416	0.444	0.383	0.451	0.394	0.484	0.435	0.457	0.401
PJM	24	0.069	**0.160**	0.072	0.162	0.073	0.168	**0.068**	**0.160**	0.096	0.191	0.129	0.231	0.105	0.213	0.082	0.178	0.096	0.198	0.130	0.236	0.106	0.212
	360	0.183	0.260	**0.180**	**0.259**	0.192	0.266	0.185	0.260	0.186	0.267	0.270	0.310	0.210	0.292	0.198	0.273	0.197	0.283	0.219	0.304	0.189	0.273
	Avg	**0.126**	**0.210**	**0.126**	**0.210**	0.133	0.217	0.127	**0.210**	0.141	0.229	0.200	0.270	0.158	0.253	0.140	0.226	0.147	0.241	0.175	0.270	0.148	0.243
BE	24	0.370	**0.241**	0.376	0.244	0.363	0.251	**0.355**	0.255	0.392	0.253	0.400	0.267	0.492	0.322	0.376	0.253	0.389	0.253	0.424	0.280	0.405	0.264
	360	**0.542**	**0.333**	0.547	0.350	0.553	0.351	0.550	0.349	0.563	0.349	0.727	0.442	0.602	0.375	0.571	0.357	0.564	0.346	0.552	0.350	0.564	0.354
	Avg	0.456	**0.287**	0.462	0.297	0.458	0.301	**0.453**	0.302	0.477	0.301	0.563	0.354	0.547	0.348	0.473	0.305	0.477	0.300	0.488	0.315	0.485	0.309
FR	24	0.378	0.194	0.345	**0.190**	0.338	0.194	**0.337**	0.197	0.366	0.195	0.445	0.231	0.415	0.254	0.359	0.208	0.397	0.204	0.448	0.250	0.397	0.223
	360	0.520	**0.291**	0.526	0.307	**0.519**	0.296	0.521	0.297	0.542	0.300	0.624	0.345	0.574	0.325	0.577	0.317	0.555	0.310	0.534	0.303	0.542	0.306
	Avg	0.449	**0.243**	0.435	0.249	**0.428**	0.245	0.429	0.247	0.454	0.247	0.535	0.288	0.494	0.290	0.468	0.262	0.476	0.257	0.491	0.276	0.470	0.264
DE	24	**0.415**	0.402	0.422	0.411	**0.415**	**0.400**	0.418	0.402	0.501	0.445	0.576	0.452	0.465	0.433	0.424	0.403	0.438	0.423	0.504	0.462	0.503	0.450
	360	0.791	0.567	0.784	**0.563**	0.808	0.594	**0.783**	0.566	0.818	0.568	0.792	0.578	0.823	0.604	0.895	0.623	0.832	0.592	0.834	0.590	0.889	0.604
	Avg	0.603	0.485	0.603	0.487	0.611	0.497	**0.601**	**0.484**	0.659	0.507	0.684	0.515	0.644	0.519	0.660	0.513	0.635	0.508	0.669	0.526	0.696	0.527
Energy	24	0.112	0.258	0.112	0.257	**0.102**	**0.246**	0.103	0.247	0.138	0.293	0.352	0.463	0.117	0.265	0.108	0.254	**0.102**	**0.246**	0.109	0.254	0.108	0.254
	360	0.192	0.344	0.189	0.342	0.197	0.352	0.188	0.340	0.206	0.360	0.399	0.497	0.207	0.358	0.211	0.362	0.299	0.425	**0.186**	**0.338**	0.298	0.428
	Avg	0.152	0.301	0.150	0.300	0.150	0.299	**0.146**	**0.294**	0.172	0.326	0.376	0.480	0.162	0.311	0.160	0.308	0.201	0.336	0.147	0.296	0.203	0.341
Sdwpfin1	24	**0.500**	**0.481**	0.532	0.512	0.536	0.493	0.534	0.485	0.558	0.533	0.596	0.531	0.561	0.540	0.550	0.495	0.549	0.553	0.598	0.579	0.557	0.557
	360	0.863	0.675	**0.845**	0.695	0.864	0.712	0.862	**0.669**	**0.845**	0.684	1.372	0.925	0.864	0.725	0.898	0.672	1.070	0.834	0.853	0.764	0.910	0.745
	Avg	**0.682**	0.578	0.689	0.604	0.700	0.602	0.698	**0.577**	0.701	0.609	0.984	0.728	0.713	0.633	0.724	0.583	0.809	0.694	0.725	0.672	0.733	0.651
Sdwpfin2	24	**0.600**	**0.506**	0.609	0.548	0.614	0.548	0.623	0.527	0.627	0.570	0.749	0.657	0.636	0.582	0.612	0.537	0.635	0.586	0.684	0.621	0.648	0.610
	360	0.985	**0.711**	**0.962**	0.766	1.036	0.717	0.977	0.717	0.978	0.736	1.340	0.877	0.995	0.783	1.028	0.745	0.965	0.788	0.987	0.780	1.020	0.819
	Avg	0.793	**0.609**	**0.786**	0.657	0.825	0.633	0.800	0.622	0.803	0.653	1.044	0.767	0.816	0.682	0.820	0.641	0.800	0.687	0.836	0.701	0.834	0.714
Sdwpfh1	24	**0.612**	0.587	0.647	0.592	0.668	0.583	0.650	**0.570**	0.651	0.587	0.754	0.670	0.719	0.641	0.673	0.575	0.687	0.652	0.755	0.679	0.669	0.629
	360	0.822	0.683	**0.820**	0.695	0.846	**0.674**	0.869	0.683	0.841	0.700	0.831	0.727	0.897	0.757	0.886	0.712	0.849	0.752	0.853	0.762	0.939	0.804
	Avg	**0.717**	0.635	0.733	0.643	0.757	0.628	0.760	**0.627**	0.746	0.643	0.793	0.698	0.808	0.699	0.779	0.644	0.768	0.702	0.804	0.720	0.804	0.717
Sdwpfh2	24	**0.734**	0.642	0.739	0.635	0.797	0.644	0.801	**0.632**	0.820	0.677	0.811	0.714	0.783	0.674	0.975	0.694	0.810	0.718	0.892	0.740	0.769	0.680
	360	0.985	**0.714**	1.000	0.755	0.968	0.739	**0.958**	0.749	0.962	0.761	1.042	0.779	1.056	0.829	1.039	0.735	1.102	0.830	0.990	0.825	1.280	0.925
	Avg	**0.860**	**0.678**	0.870	0.695	0.882	0.691	0.880	0.691	0.891	0.719	0.926	0.746	0.919	0.751	1.007	0.715	0.956	0.774	0.941	0.782	1.025	0.802
Colbun	10	**0.064**	0.086	0.072	0.103	**0.064**	**0.081**	0.069	0.098	0.084	0.146	0.449	0.279	0.095	0.140	0.073	0.106	0.068	0.101	0.091	0.129	0.083	0.125
	30	0.191	0.272	**0.161**	**0.208**	0.168	0.246	0.203	0.287	0.181	0.293	0.663	0.494	0.282	0.341	0.221	0.290	0.190	0.289	0.311	0.377	0.217	0.316
	Avg	0.128	0.179	0.117	**0.155**	**0.116**	0.164	0.136	0.193	0.132	0.219	0.556	0.386	0.188	0.240	0.147	0.198	0.129	0.195	0.201	0.253	0.150	0.221
Rapel	10	**0.187**	0.218	0.195	0.215	0.191	**0.204**	0.212	0.229	0.301	0.308	0.238	0.274	0.238	0.270	0.221	0.225	0.201	0.224	0.226	0.238	0.215	0.232
	30	0.319	0.357	0.327	0.347	0.325	**0.343**	0.392	0.382	0.387	0.416	0.377	0.394	0.409	0.444	0.387	0.386	**0.278**	0.374	0.458	0.437	0.436	0.436
	Avg	0.253	0.288	0.261	0.281	0.258	**0.273**	0.302	0.306	0.344	0.362	0.308	0.334	0.323	0.357	0.304	0.306	**0.240**	0.299	0.342	0.337	0.325	0.334
Average		**0.470**	**0.405**	0.471	0.413	0.479	0.411	0.478	0.410	0.497	0.433	0.635	0.505	0.520	0.458	0.511	0.424	0.507	0.449	0.525	0.465	0.528	0.460
1st Count		**14**	**15**	7	6	6	7	10	11	1	0	0	0	0	0	0	0	3	1	1	1	0	0

Table 10: Full results on the 10 zones of GEFCom2014 wind with future exogenous variables available. **Red**: best, Blue: second best.

Models		InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Zone 1	24	**0.425**	**0.467**	0.767	0.639	0.818	0.693	0.694	0.620	0.722	0.645	0.439	0.485	0.570	0.576	1.040	0.781	0.463	0.518	0.853	0.736	0.470	0.524
	360	**0.400**	**0.461**	1.270	0.863	1.583	1.038	1.364	0.927	1.316	0.939	0.589	0.570	0.830	0.729	1.171	0.817	0.533	0.555	0.786	0.733	0.503	0.543
	Avg	**0.413**	**0.464**	1.019	0.751	1.201	0.866	1.029	0.774	1.019	0.792	0.514	0.528	0.700	0.653	1.106	0.799	0.498	0.537	0.820	0.735	0.487	0.534
Zone 2	24	**0.482**	**0.479**	0.775	0.662	0.959	0.753	0.885	0.692	0.866	0.720	0.495	0.495	0.574	0.570	0.955	0.781	0.503	0.518	0.859	0.756	0.553	0.541
	360	0.638	**0.545**	1.319	0.920	1.843	1.102	1.606	1.006	1.431	0.964	0.642	0.554	0.883	0.745	0.858	0.747	**0.586**	0.552	0.753	0.718	0.590	0.565
	Avg	0.560	**0.512**	1.047	0.791	1.401	0.928	1.246	0.849	1.149	0.842	0.569	0.525	0.729	0.658	0.907	0.764	**0.545**	0.535	0.806	0.737	0.572	0.553
Zone 3	24	**0.276**	**0.403**	0.831	0.707	0.816	0.706	0.828	0.701	0.841	0.725	0.278	0.414	0.441	0.528	0.936	0.821	0.288	0.413	0.820	0.774	0.307	0.434
	360	0.278	0.395	1.211	0.902	2.248	1.193	1.353	0.944	1.347	0.952	**0.260**	**0.391**	1.135	0.914	0.773	0.749	0.311	0.435	0.758	0.754	0.303	0.431
	Avg	0.277	**0.399**	1.021	0.805	1.532	0.950	1.091	0.823	1.094	0.839	**0.269**	0.403	0.788	0.721	0.855	0.785	0.300	0.424	0.789	0.764	0.305	0.433
Zone 4	24	**0.286**	**0.364**	0.599	0.584	0.801	0.691	0.687	0.622	0.704	0.647	0.304	0.386	0.423	0.497	0.857	0.782	0.320	0.406	0.678	0.702	0.358	0.442
	360	**0.282**	**0.352**	1.037	0.801	1.377	1.003	1.095	0.834	1.161	0.875	0.303	0.378	0.635	0.647	0.870	0.769	0.387	0.449	0.653	0.700	0.387	0.458
	Avg	**0.284**	**0.358**	0.818	0.693	1.089	0.847	0.891	0.728	0.933	0.761	0.304	0.382	0.529	0.572	0.864	0.776	0.354	0.428	0.666	0.701	0.373	0.450
Zone 5	24	0.278	0.382	0.663	0.638	0.754	0.689	0.750	0.659	0.692	0.654	**0.269**	**0.372**	0.413	0.501	0.829	0.798	0.319	0.417	0.727	0.743	0.363	0.462
	360	0.268	**0.367**	1.073	0.839	1.575	1.046	1.202	0.880	1.206	0.907	**0.267**	0.374	0.660	0.658	0.850	0.795	0.351	0.436	0.702	0.731	0.407	0.489
	Avg	0.273	0.375	0.868	0.739	1.165	0.868	0.976	0.770	0.949	0.781	**0.268**	**0.373**	0.537	0.580	0.840	0.797	0.335	0.427	0.715	0.737	0.385	0.476
Zone 6	24	**0.296**	**0.377**	0.659	0.633	0.794	0.697	0.752	0.658	0.694	0.654	0.325	0.396	0.502	0.540	0.833	0.797	0.364	0.418	0.720	0.728	0.414	0.461
	360	**0.322**	**0.382**	1.043	0.836	1.531	1.028	1.237	0.889	1.181	0.904	0.332	0.405	0.700	0.677	0.827	0.779	0.429	0.458	0.710	0.720	0.464	0.485
	Avg	**0.309**	**0.380**	0.851	0.735	1.163	0.863	0.995	0.774	0.938	0.779	0.329	0.401	0.601	0.609	0.830	0.788	0.397	0.438	0.715	0.724	0.439	0.473
Zone 7	24	**0.319**	**0.404**	0.800	0.666	0.792	0.674	0.779	0.667	0.840	0.703	0.360	0.434	0.462	0.515	1.056	0.820	0.340	0.431	0.915	0.775	0.374	0.454
	360	**0.278**	**0.383**	1.290	0.888	1.621	1.054	1.458	0.956	1.382	0.974	0.328	0.433	0.792	0.719	1.165	0.833	0.361	0.447	0.788	0.738	0.380	0.459
	Avg	**0.299**	**0.394**	1.045	0.777	1.207	0.864	1.119	0.812	1.111	0.839	0.344	0.434	0.627	0.617	1.111	0.827	0.351	0.439	0.852	0.757	0.377	0.457
Zone 8	24	**0.526**	**0.503**	0.817	0.674	0.862	0.706	0.864	0.697	0.882	0.702	0.576	0.521	0.594	0.578	1.083	0.830	0.555	0.532	0.977	0.801	0.550	0.532
	360	0.579	**0.502**	1.402	0.925	1.724	1.077	1.560	0.991	1.439	0.979	**0.560**	0.523	0.965	0.781	1.157	0.832	0.624	0.561	0.846	0.759	0.607	0.551
	Avg	**0.553**	**0.503**	1.110	0.800	1.293	0.892	1.212	0.844	1.161	0.841	0.568	0.522	0.780	0.680	1.120	0.831	0.590	0.547	0.912	0.780	0.579	0.542
Zone 9	24	**0.321**	**0.402**	0.752	0.641	0.766	0.676	0.777	0.650	0.692	0.642	0.411	0.456	0.401	0.483	0.916	0.765	0.350	0.437	0.740	0.700	0.353	0.433
	360	**0.331**	**0.413**	1.107	0.835	1.977	1.137	1.258	0.900	1.126	0.860	0.388	0.457	0.726	0.684	0.896	0.756	0.394	0.454	0.703	0.684	0.405	0.465
	Avg	**0.326**	**0.408**	0.930	0.738	1.372	0.907	1.018	0.775	0.909	0.751	0.400	0.457	0.564	0.584	0.906	0.761	0.372	0.446	0.722	0.692	0.379	0.449
Zone 10	24	0.360	0.424	0.732	0.665	0.801	0.727	0.848	0.707	0.778	0.709	**0.350**	**0.422**	0.465	0.527	0.930	0.859	0.462	0.504	0.818	0.800	0.493	0.535
	360	0.378	0.430	1.065	0.858	1.018	0.865	1.335	0.934	1.083	0.905	**0.356**	**0.424**	1.046	0.892	0.785	0.766	0.525	0.536	0.736	0.752	0.540	0.557
	Avg	0.369	0.427	0.899	0.762	0.910	0.796	1.092	0.821	0.931	0.807	**0.353**	**0.423**	0.756	0.710	0.858	0.813	0.494	0.520	0.777	0.776	0.517	0.546
Average		**0.366**	**0.422**	0.961	0.759	1.233	0.878	1.067	0.797	1.019	0.803	0.392	0.445	0.661	0.638	0.940	0.794	0.424	0.474	0.777	0.740	0.441	0.491
1st Count		**20**	**25**	0	0	0	0	0	0	0	0	9	6	0	0	0	0	2	0	0	0	0	0

Table 11: Full results of single-source zero-shot transfer between the Sdwpf subsets. **Red**: best, Blue: second best.

Models			InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Source	Target	Horizon	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Sdwpfh1	Sdwpfh2	24	**0.377**	**0.412**	0.462	0.473	0.431	0.475	0.481	0.498	0.742	0.629	0.444	0.487	0.484	0.525	0.608	0.565	0.636	0.612	0.573	0.600	0.458	0.517
		360	**0.488**	**0.473**	0.573	0.575	0.585	0.568	0.550	0.570	0.981	0.752	0.713	0.587	0.552	0.572	0.655	0.574	0.670	0.642	0.714	0.715	0.621	0.591
		Avg	**0.433**	**0.443**	0.518	0.524	0.508	0.522	0.516	0.534	0.862	0.691	0.579	0.537	0.518	0.549	0.632	0.570	0.653	0.627	0.644	0.658	0.540	0.554
	Sdwpfm2	24	**0.411**	**0.428**	0.417	0.443	0.485	0.476	0.442	0.461	0.811	0.622	0.502	0.498	0.472	0.498	0.620	0.574	0.633	0.614	0.572	0.598	0.442	0.517
		360	**0.479**	**0.473**	0.601	0.588	0.526	0.545	0.655	0.585	1.019	0.738	0.754	0.618	0.518	0.552	0.666	0.575	0.687	0.648	0.727	0.717	0.633	0.587
		Avg	**0.445**	**0.451**	0.509	0.516	0.506	0.511	0.549	0.523	0.915	0.680	0.628	0.558	0.495	0.525	0.643	0.575	0.660	0.631	0.650	0.658	0.538	0.552
Sdwpfh2	Sdwpfh1	24	**0.334**	**0.373**	0.389	0.434	0.386	0.444	0.437	0.473	0.728	0.634	0.430	0.467	0.518	0.522	0.521	0.543	0.424	0.522	0.525	0.603	0.414	0.511
		360	**0.310**	**0.359**	0.532	0.564	0.475	0.504	0.455	0.514	0.833	0.713	0.545	0.491	0.541	0.573	0.539	0.529	0.528	0.593	0.656	0.714	0.529	0.593
		Avg	**0.322**	**0.366**	0.461	0.499	0.431	0.474	0.446	0.494	0.781	0.674	0.488	0.479	0.530	0.548	0.530	0.536	0.476	0.558	0.591	0.659	0.472	0.552
	Sdwpfm2	24	0.470	0.447	**0.399**	**0.431**	0.438	0.465	0.460	0.472	0.761	0.654	0.613	0.545	0.538	0.514	0.626	0.580	0.440	0.517	0.581	0.613	0.429	0.506
		360	**0.394**	**0.405**	0.638	0.614	0.475	0.508	0.568	0.558	1.034	0.739	0.606	0.531	0.624	0.599	0.658	0.572	0.626	0.615	0.713	0.726	0.623	0.619
		Avg	**0.432**	**0.426**	0.519	0.523	0.457	0.487	0.514	0.515	0.898	0.697	0.610	0.538	0.581	0.557	0.642	0.576	0.533	0.566	0.647	0.670	0.526	0.563
Sdwpfm1	Sdwpfh2	24	0.565	0.498	0.550	0.516	0.532	0.516	0.543	0.517	0.813	0.637	0.433	**0.488**	0.652	0.604	0.474	0.494	**0.423**	0.508	0.626	0.621	0.483	0.517
		360	**0.463**	**0.461**	0.684	0.617	0.631	0.573	0.666	0.577	0.981	0.750	0.569	0.506	0.593	0.594	0.492	0.490	0.482	0.528	**0.463**	0.524	0.519	0.556
		Avg	0.514	**0.480**	0.617	0.567	0.582	0.545	0.605	0.547	0.897	0.694	0.501	0.497	0.623	0.599	0.483	0.492	**0.453**	0.518	0.545	0.573	0.501	0.537
	Sdwpfm2	24	0.474	0.444	0.402	**0.436**	0.421	0.442	0.437	0.442	0.643	0.579	0.425	0.461	0.527	0.524	0.464	0.492	**0.385**	0.487	0.567	0.593	0.421	0.478
		360	**0.407**	**0.431**	0.586	0.570	0.541	0.531	0.514	0.521	0.958	0.727	0.588	0.515	0.559	0.564	0.503	0.493	0.571	0.560	0.469	0.521	0.569	0.574
		Avg	**0.441**	**0.438**	0.494	0.503	0.481	0.487	0.476	0.482	0.801	0.653	0.507	0.488	0.543	0.544	0.484	0.493	0.478	0.524	0.518	0.557	0.495	0.526
Sdwpfm2	Sdwpfh2	24	0.426	**0.452**	0.477	0.491	0.542	0.525	0.548	0.524	0.790	0.636	**0.397**	0.468	0.527	0.543	0.474	0.499	0.534	0.561	0.607	0.623	0.464	0.529
		360	**0.461**	**0.473**	0.675	0.604	0.631	0.588	0.642	0.567	0.938	0.752	0.578	0.501	0.551	0.570	0.491	0.488	0.575	0.612	**0.461**	0.535	0.506	0.564
		Avg	**0.444**	**0.463**	0.576	0.548	0.587	0.557	0.595	0.546	0.864	0.694	0.488	0.485	0.539	0.557	0.483	0.494	0.555	0.587	0.534	0.579	0.485	0.547
	Sdwpfm1	24	0.351	0.391	**0.339**	**0.390**	0.374	0.415	0.403	0.416	0.563	0.536	0.389	0.439	0.419	0.463	0.403	0.475	0.430	0.523	0.489	0.570	0.383	0.477
		360	**0.352**	**0.393**	0.503	0.521	0.478	0.498	0.443	0.480	0.870	0.697	0.519	0.474	0.447	0.504	0.424	0.456	0.514	0.585	0.421	0.517	0.511	0.574
		Avg	**0.352**	**0.392**	0.421	0.456	0.426	0.457	0.423	0.448	0.717	0.617	0.454	0.457	0.433	0.484	0.414	0.466	0.472	0.554	0.455	0.544	0.447	0.526
Average			**0.423**	**0.432**	0.514	0.517	0.497	0.505	0.516	0.511	0.842	0.675	0.532	0.505	0.533	0.545	0.539	0.525	0.535	0.571	0.573	0.612	0.501	0.545
1st Count			**19**	**21**	2	3	0	0	0	0	0	0	1	1	0	0	0	0	3	0	2	0	0	0

Table 12: Full results of multi-source zero-shot transfer on GEFCom2014 wind, with models trained jointly on five source zones. **Red**: best, Blue: second best.

Models			InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Source	Target	Horizon	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Zones 1–5	Zone 6	24	0.476	0.486	0.795	0.707	0.696	0.658	0.739	0.653	0.685	0.640	**0.415**	**0.464**	0.457	0.524	0.598	0.617	0.588	0.611	0.604	0.638	0.460	0.558
		360	**0.458**	**0.482**	1.071	0.853	1.381	0.987	0.980	0.793	0.995	0.824	0.486	0.513	0.613	0.617	0.587	0.639	0.579	0.625	0.807	0.706	0.519	0.598
		Avg	0.467	**0.484**	0.933	0.780	1.039	0.823	0.860	0.723	0.840	0.732	**0.451**	0.489	0.535	0.571	0.593	0.628	0.584	0.618	0.706	0.672	0.490	0.578
	Zone 7	24	**0.380**	**0.447**	0.868	0.724	0.751	0.674	0.755	0.654	0.723	0.656	0.495	0.532	0.470	0.520	0.683	0.634	0.653	0.618	0.650	0.646	0.533	0.596
		360	**0.468**	**0.495**	1.304	0.927	1.677	1.069	1.321	0.914	1.207	0.892	0.482	0.514	0.636	0.622	0.679	0.684	0.647	0.647	0.707	0.657	0.681	0.693
		Avg	**0.424**	**0.471**	1.086	0.826	1.214	0.872	1.038	0.784	0.965	0.774	0.489	0.523	0.553	0.571	0.681	0.659	0.650	0.633	0.679	0.652	0.607	0.645
	Zone 8	24	**0.592**	**0.533**	0.949	0.753	0.809	0.690	0.847	0.684	0.836	0.697	0.723	0.600	0.598	0.573	0.714	0.645	0.709	0.646	0.712	0.671	0.663	0.650
		360	0.786	0.591	1.335	0.940	1.681	1.065	1.378	0.930	1.211	0.893	**0.730**	**0.590**	0.790	0.677	0.808	0.724	0.752	0.691	0.763	0.682	0.832	0.750
		Avg	**0.689**	**0.562**	1.142	0.847	1.245	0.878	1.113	0.807	1.024	0.795	0.727	0.595	0.694	0.625	0.761	0.685	0.731	0.669	0.738	0.677	0.748	0.700
	Zone 9	24	**0.420**	**0.473**	0.784	0.696	0.723	0.650	0.723	0.636	0.682	0.635	0.484	0.530	0.443	0.504	0.590	0.593	0.598	0.603	0.578	0.603	0.484	0.565
		360	**0.503**	**0.514**	1.125	0.861	1.379	0.979	1.215	0.884	1.066	0.840	0.520	0.534	0.546	0.576	0.577	0.629	0.557	0.602	0.622	0.638	0.549	0.613
		Avg	**0.462**	**0.494**	0.955	0.779	1.051	0.815	0.969	0.760	0.874	0.738	0.502	0.532	0.495	0.540	0.584	0.611	0.578	0.603	0.600	0.621	0.517	0.589
	Zone 10	24	0.540	**0.511**	0.803	0.722	0.767	0.705	0.798	0.696	0.747	0.684	0.559	0.548	**0.495**	0.545	0.684	0.678	0.602	0.620	0.666	0.671	0.505	0.580
		360	0.564	**0.548**	1.134	0.886	1.338	0.972	1.174	0.880	1.086	0.872	0.599	0.567	0.669	0.647	0.608	0.651	0.621	0.649	0.835	0.721	**0.544**	0.610
		Avg	0.552	**0.530**	0.969	0.804	1.053	0.839	0.986	0.788	0.917	0.778	0.579	0.558	0.582	0.596	0.646	0.665	0.612	0.635	0.751	0.696	**0.525**	0.595
Zones 6–10	Zone 1	24	**0.542**	**0.535**	0.779	0.668	0.831	0.692	0.761	0.650	0.769	0.681	0.631	0.606	0.558	0.561	0.734	0.654	0.670	0.642	0.722	0.693	0.643	0.677
		360	**0.625**	**0.576**	1.437	0.985	2.480	1.253	1.438	0.959	1.245	0.908	0.682	0.629	0.712	0.677	0.788	0.728	0.743	0.705	0.786	0.719	0.825	0.797
		Avg	**0.584**	**0.556**	1.108	0.827	1.656	0.973	1.100	0.805	1.007	0.795	0.657	0.618	0.635	0.619	0.761	0.691	0.707	0.674	0.754	0.706	0.734	0.737
	Zone 2	24	0.688	**0.579**	0.863	0.714	0.841	0.678	0.860	0.677	0.773	0.669	**0.638**	0.588	0.711	0.633	0.741	0.664	0.685	0.640	0.737	0.694	0.701	0.677
		360	0.874	**0.675**	1.376	0.962	2.282	1.212	1.479	0.965	1.241	0.909	0.916	0.691	0.955	0.759	0.821	0.718	0.864	0.738	**0.810**	0.711	0.840	0.743
		Avg	0.781	**0.627**	1.120	0.838	1.562	0.945	1.170	0.821	1.007	0.789	0.777	0.640	0.833	0.696	0.781	0.691	0.775	0.689	0.774	0.703	**0.771**	0.710
	Zone 3	24	**0.435**	**0.501**	0.836	0.720	0.789	0.694	0.753	0.674	0.737	0.684	0.450	0.526	0.475	0.549	0.659	0.661	0.618	0.639	0.661	0.673	0.528	0.605
		360	**0.454**	**0.519**	1.501	1.009	2.016	1.147	1.145	0.877	1.121	0.881	0.504	0.568	0.800	0.725	0.601	0.664	0.625	0.666	0.655	0.685	0.566	0.645
		Avg	**0.445**	**0.510**	1.169	0.865	1.403	0.921	0.949	0.776	0.929	0.783	0.477	0.547	0.638	0.637	0.630	0.663	0.622	0.653	0.658	0.679	0.547	0.625
	Zone 4	24	**0.360**	**0.411**	0.721	0.658	0.636	0.613	0.620	0.587	0.594	0.605	0.372	0.458	0.419	0.492	0.645	0.631	0.527	0.577	0.637	0.650	0.509	0.593
		360	**0.365**	**0.429**	1.171	0.879	2.237	1.202	1.136	0.847	1.125	0.878	0.460	0.511	0.633	0.646	0.609	0.650	0.602	0.641	0.629	0.659	0.592	0.666
		Avg	**0.363**	**0.420**	0.946	0.769	1.437	0.908	0.878	0.717	0.860	0.742	0.416	0.485	0.526	0.569	0.627	0.641	0.565	0.609	0.633	0.655	0.551	0.630
	Zone 5	24	**0.320**	**0.391**	0.727	0.673	0.651	0.631	0.642	0.613	0.643	0.636	0.326	0.427	0.408	0.500	0.621	0.632	0.558	0.605	0.625	0.652	0.491	0.581
		360	**0.342**	**0.419**	1.157	0.866	2.088	1.168	1.034	0.815	1.039	0.854	0.382	0.465	0.611	0.628	0.577	0.648	0.579	0.634	0.606	0.661	0.489	0.600
		Avg	**0.331**	**0.405**	0.942	0.770	1.370	0.900	0.838	0.714	0.841	0.745	0.354	0.446	0.510	0.564	0.599	0.640	0.569	0.620	0.616	0.657	0.490	0.591
Average			**0.510**	**0.506**	1.037	0.811	1.303	0.887	0.990	0.770	0.926	0.767	0.543	0.543	0.600	0.599	0.666	0.657	0.639	0.640	0.691	0.672	0.598	0.640
1st Count			**22**	**29**	0	0	0	0	0	0	0	0	4	2	1	0	0	0	0	0	1	0	3	0

Table 13: Full results on the eight common multivariate datasets with future exogenous variables available. '–': out of memory on an 80 GB GPU. **Red**: best, Blue: second best.

Models		InRelCast		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.053**	0.177	0.054	0.176	0.065	0.197	0.054	**0.175**	0.056	0.179	0.081	0.214	0.056	0.179	0.163	0.338	0.152	0.321	0.118	0.273	0.216	0.403
	192	0.076	0.217	0.072	0.208	0.087	0.231	**0.071**	**0.203**	0.073	0.207	0.073	0.213	0.075	0.210	0.206	0.382	0.199	0.374	0.166	0.330	0.174	0.336
	336	0.086	0.232	**0.079**	**0.219**	0.107	0.259	0.088	0.231	0.089	0.234	0.099	0.246	0.090	0.236	0.248	0.419	0.218	0.383	0.140	0.295	0.253	0.429
	720	0.084	0.232	**0.079**	**0.221**	0.118	0.272	0.118	0.270	0.104	0.254	0.124	0.282	0.096	0.244	0.283	0.446	0.295	0.475	0.284	0.440	0.302	0.462
	Avg	0.075	0.215	**0.071**	**0.206**	0.094	0.240	0.083	0.220	0.081	0.219	0.094	0.239	0.079	0.217	0.225	0.396	0.216	0.388	0.177	0.335	0.236	0.407
ETTh2	96	0.125	0.273	0.127	0.273	0.209	0.360	0.169	0.309	0.141	0.289	0.217	0.364	0.127	**0.271**	**0.121**	0.275	0.135	0.289	0.122	0.278	0.134	0.293
	192	0.168	**0.322**	0.177	0.327	0.259	0.408	0.220	0.363	0.194	0.346	0.256	0.405	0.179	0.327	0.234	0.391	**0.154**	0.323	0.186	0.354	0.158	0.326
	336	0.204	0.344	0.215	0.369	0.323	0.461	0.268	0.407	0.226	0.380	0.287	0.435	0.223	0.375	0.228	0.379	0.179	0.343	0.181	0.342	**0.172**	**0.338**
	720	**0.199**	**0.357**	0.214	0.370	0.348	0.480	0.445	0.514	0.234	0.385	0.283	0.431	0.253	0.405	0.209	0.365	0.319	0.463	0.337	0.473	0.337	0.475
	Avg	**0.174**	**0.324**	0.183	0.335	0.285	0.427	0.276	0.398	0.199	0.350	0.261	0.409	0.195	0.344	0.198	0.352	0.197	0.354	0.207	0.362	0.200	0.358
ETTm1	96	0.029	0.129	0.029	0.127	0.037	0.149	**0.028**	**0.126**	0.030	0.129	0.030	0.131	0.030	0.130	0.063	0.204	0.087	0.247	0.067	0.205	0.092	0.253
	192	**0.044**	**0.148**	**0.044**	0.159	0.050	0.174	**0.044**	0.159	**0.044**	0.161	0.048	0.168	0.049	0.168	0.217	0.413	0.164	0.351	0.156	0.331	0.132	0.304
	336	0.059	**0.176**	**0.057**	0.184	0.063	0.196	0.058	0.185	0.060	0.189	0.068	0.200	0.069	0.203	0.252	0.446	0.239	0.426	0.191	0.363	0.255	0.440
	720	0.079	0.217	0.079	0.216	0.085	0.227	0.077	0.215	0.082	0.222	0.095	0.240	**0.073**	**0.211**	0.265	0.436	0.286	0.462	0.261	0.428	0.251	0.434
	Avg	0.053	**0.168**	**0.052**	0.172	0.059	0.187	**0.052**	0.171	0.054	0.175	0.060	0.185	0.055	0.178	0.199	0.375	0.194	0.372	0.169	0.332	0.183	0.358
ETTm2	96	0.066	0.186	**0.063**	**0.183**	0.145	0.294	0.084	0.196	0.072	0.195	0.117	0.253	0.070	0.190	0.107	0.260	0.156	0.332	0.146	0.309	0.173	0.351
	192	0.104	0.239	**0.098**	**0.230**	0.184	0.333	0.123	0.258	0.110	0.249	0.179	0.324	0.101	0.235	0.200	0.366	0.203	0.366	0.139	0.293	0.142	0.295
	336	0.129	0.275	**0.127**	**0.270**	0.206	0.355	0.158	0.294	0.139	0.284	0.206	0.356	0.128	0.272	0.136	0.292	0.136	0.292	0.182	0.338	0.151	0.311
	720	0.182	0.318	0.179	0.328	0.255	0.402	0.503	0.414	0.190	0.340	0.271	0.417	0.179	0.328	**0.147**	**0.310**	0.156	0.316	0.169	0.332	0.164	0.326
	Avg	0.120	0.255	**0.117**	**0.253**	0.198	0.346	0.217	0.291	0.128	0.267	0.193	0.338	0.119	0.256	0.148	0.307	0.163	0.326	0.159	0.318	0.157	0.321
Weather	96	**0.001**	**0.015**	**0.001**	0.023	**0.001**	0.028	**0.001**	0.018	**0.001**	0.023	**0.001**	0.028	**0.001**	0.026	0.002	0.028	0.012	0.093	0.008	0.072	0.006	0.062
	192	**0.001**	**0.018**	**0.001**	0.022	**0.001**	0.028	**0.001**	0.022	**0.001**	0.026	0.002	0.030	**0.001**	0.028	0.002	0.030	0.005	0.056	0.009	0.078	0.008	0.073
	336	**0.001**	**0.021**	0.002	0.028	0.002	0.030	**0.001**	0.026	**0.001**	0.027	0.002	0.030	0.002	0.031	**0.001**	0.025	0.009	0.080	0.009	0.074	0.011	0.087
	720	**0.001**	**0.023**	0.002	0.030	0.002	0.034	0.002	0.033	0.002	0.032	0.002	0.035	0.002	0.036	0.002	0.033	0.007	0.067	0.003	0.045	0.012	0.084
	Avg	**0.001**	**0.019**	0.002	0.026	0.002	0.030	**0.001**	0.025	**0.001**	0.027	0.002	0.031	0.002	0.030	**0.001**	0.029	0.008	0.074	0.007	0.067	0.009	0.076
Electricity	96	0.194	0.337	**0.158**	0.298	–	–	0.233	0.342	0.160	**0.296**	0.365	0.456	0.312	0.414	0.248	0.385	0.281	0.416	0.276	0.413	0.274	0.406
	192	0.186	0.328	**0.156**	**0.295**	–	–	0.336	0.373	0.186	0.320	0.339	0.448	0.274	0.391	0.290	0.421	0.248	0.384	0.309	0.435	0.264	0.399
	336	0.218	0.363	**0.217**	0.353	–	–	0.384	0.419	0.218	**0.346**	0.367	0.471	0.313	0.417	0.295	0.426	0.260	0.396	0.292	0.419	0.271	0.408
	720	**0.234**	0.374	0.246	**0.370**	–	–	0.677	0.545	0.267	0.386	0.429	0.510	0.353	0.445	0.290	0.414	0.268	0.401	0.327	0.448	0.268	0.402
	Avg	0.208	0.351	**0.194**	**0.329**	–	–	0.408	0.420	0.208	0.337	0.375	0.471	0.313	0.417	0.281	0.411	0.264	0.399	0.301	0.429	0.269	0.404
Traffic	96	0.109	0.198	0.119	**0.187**	–	–	**0.106**	0.198	0.135	0.233	0.139	0.221	0.149	0.251	0.247	0.316	0.261	0.322	0.221	0.292	0.241	0.307
	192	**0.109**	0.198	0.121	**0.191**	–	–	0.112	0.204	0.135	0.230	0.116	0.212	0.149	0.246	0.240	0.297	0.232	0.289	0.181	0.254	0.246	0.305
	336	**0.104**	**0.189**	0.125	0.200	–	–	0.121	0.215	0.143	0.237	0.134	0.228	0.150	0.251	0.310	0.356	0.223	0.280	0.210	0.283	0.215	0.280
	720	**0.113**	**0.209**	0.152	0.232	–	–	0.139	0.232	0.156	0.250	0.138	0.220	0.169	0.271	0.253	0.318	0.213	0.272	0.231	0.310	0.229	0.286
	Avg	**0.109**	**0.199**	0.129	0.203	–	–	0.120	0.212	0.142	0.238	0.132	0.220	0.154	0.255	0.263	0.322	0.232	0.291	0.211	0.285	0.233	0.295
Exchange	96	0.093	**0.222**	**0.087**	0.229	0.130	0.278	0.092	**0.222**	0.109	0.245	0.105	0.258	0.096	0.230	0.291	0.484	0.234	0.425	0.297	0.487	0.370	0.554
	192	**0.144**	**0.270**	0.147	0.298	0.242	0.391	0.195	0.332	0.197	0.350	0.258	0.393	0.205	0.340	0.236	0.421	0.231	0.430	0.300	0.461	0.427	0.590
	336	**0.171**	**0.333**	0.274	0.413	0.576	0.592	0.343	0.433	0.331	0.444	0.422	0.524	0.453	0.519	0.309	0.496	0.394	0.563	0.716	0.752	0.558	0.687
	720	0.351	0.472	0.338	0.470	1.077	0.797	0.531	0.565	0.362	0.482	0.552	0.577	0.419	0.502	**0.244**	**0.419**	0.390	0.558	0.387	0.559	0.503	0.640
	Avg	**0.190**	**0.324**	0.212	0.353	0.506	0.515	0.290	0.388	0.250	0.380	0.334	0.438	0.293	0.398	0.270	0.455	0.312	0.494	0.425	0.565	0.464	0.618
Average		**0.116**	**0.232**	0.120	0.235	–	–	0.181	0.266	0.133	0.249	0.181	0.291	0.151	0.262	0.198	0.331	0.198	0.337	0.207	0.337	0.219	0.355
1st Count		**18**	**19**	17	12	2	0	9	4	5	2	1	0	3	2	5	2	1	0	0	0	1	1

C Comparison with in-context forecasters

Table 14 compares InRelCast with in-context forecasters on the ten zones of GEFCom2014 wind, where we run every model on the same test windows. ICTSP (Lu et al., 2025) starts from its official configuration, follows the uniform tuning protocol with the training loop shared by all baselines, and receives the future exogenous variables through the MLP fusion used for baselines without native support. Chronos-2 (Ansari et al., 2025) is also fine-tuned on the training split with its official procedure and default settings (full fine-tuning, learning rate 10^{-6}, 1000 steps) and selected on the validation split. Chronos-2, TimesFM 3.0 (Das et al., 2024), TabPFN-TS (Hoo et al., 2026), and Moirai 1.1 (large, Woo et al., 2024) are applied zero-shot with the past and future exogenous variables and the same look-back lengths, and the medians of their predictive distributions serve as point forecasts. Fine-tuning substantially lowers the MSE of Chronos-2, but InRelCast remains the most accurate model.

Table 14: In-context forecasters on GEFCom2014 wind, averaged over horizons and zones. Time: per window at horizon 24. **Red**: best.

Setting		Trained on the target zones			Zero-shot			
Models		InRelCast	ICTSP	Chronos-2 (fine-tuned)	Chronos-2	TimesFM 3.0	TabPFN-TS	Moirai 1.1
Params		4.2 M	1.8 M	119 M	119 M	331 M	11.1 M	311 M
Time (ms)		0.6	0.7	2.7	2.7	17.5	764	1.5
Avg	MSE	**0.366**	0.418	0.433	0.659	0.662	0.720	1.429
	MAE	**0.422**	0.470	0.457	0.604	0.593	0.631	0.924

D ADDITIONAL MODEL ANALYSIS

D.1 PARAMETER MODULATION VERSUS INPUT CONDITIONING

Relation as input keeps all parts of the full model but, instead of generating low-rank increments, adds an affine map of r_j, scaled by the gate g_j, to every input token of the residual forecaster. The full model is better on every dataset (Table 15).

Table 15: Parameter modulation versus input conditioning. MSE and MAE are averaged over the horizons, and over the subsets and zones for Sdwpf and GEFCom2014 wind. **Red**: best.

Variant	DE		Energy		Sdwpf		GEFCom2014		Weather		Traffic		Average	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
InRelCast	**0.328**	**0.354**	**0.105**	**0.253**	**0.359**	**0.401**	**0.366**	**0.422**	**0.0008**	**0.0193**	**0.109**	**0.199**	**0.211**	**0.275**
Relation as input	0.389	0.385	0.112	0.264	0.369	0.405	0.384	0.429	0.0009	0.0204	0.118	0.210	0.229	0.286

D.2 RELATION VISUALIZATION

We visualize the relations of InRelCast on Sdwpfh1, Energy, NP, and DE in Fig. 6, each from one additional run at horizon 360. Each panel shows the evenly spaced test window whose relations vary most across transitions while matching the later demonstrations better than the median, cropped to the most varying dimensions and units. During training, the relations become specific to each transition and resemble the demonstrations that the same transitions form once observed. Moreover, on GEFCom2014 wind, NP, DE, and Sdwpfh1, the similarities among the relations of a window follow those among its own later demonstrations rather than those of another window.

D.3 ERROR ALONG THE HORIZON

We compute the MSE at each hour of the horizon over all test windows, with one run per model (Fig. 7). The error of InRelCast grows the most slowly along the horizon, whereas those of DAG, GCGNet, KITE, and TimeXer increase markedly with the distance from the forecast origin.

D.4 EFFICIENCY AND SCALABILITY

The efficiency comparison in Fig. 5 uses horizon 360 and look-back length $T = 720$ on NP and GEFCom2014 zone 4 and horizon 720 and $T = 96$ on Weather and Traffic, and reports the median inference time over 50 batches of 32 windows on one A800 GPU (PyTorch 2.4.1, 32-bit floating point) for every model and the peak training memory. To examine scalability, we first bound the cost of InRelCast in terms of the sizes defined in Section 3.

Proposition 1 (Cost)**.** *Fix p, the widths, the numbers of layers and heads, and R. In the default configuration,* FORECAST *and* TRAINSTEP *of Algorithm 1 take $O((K+J)^2 + (T+F)(N+D))$ time per window, and the model has $O(K+J+N+D)$ parameters. The predictor G and the optional linear maps of some configurations add time and parameters that grow at most linearly in D.*

Proof. The maps that read or write the variables act once per step on $O(N+D)$ values, and with at most $J+K-1$ calls of Ψ over $2p$ steps each they take $O((T+F)(N+D))$ time and hold $O(N+D)$ parameters. Self-attention acts over at most $K+J$ tokens in the causal Transformer, its rerun, and relation inference, and over $2p$ steps in each call of Ψ, which takes $O((K+J)^2)$ time. All other maps have fixed sizes and act once per patch, query, or call of Ψ, the two positional tables have $2K+J$ rows, and backpropagation adds a constant factor. Since the predictor G is an affine map or a Transformer run on each exogenous channel and each optional linear map acts on at most $N+D$ channels, their time and parameters grow at most linearly in D. □

Hence the time and the number of parameters of InRelCast grow at most linearly in D. To measure this growth and compare it with the baselines, we use Traffic, the dataset with the most exogenous variables, keep its first D exogenous variables, and increase D from 8 to 861, as shown in Fig. 8.

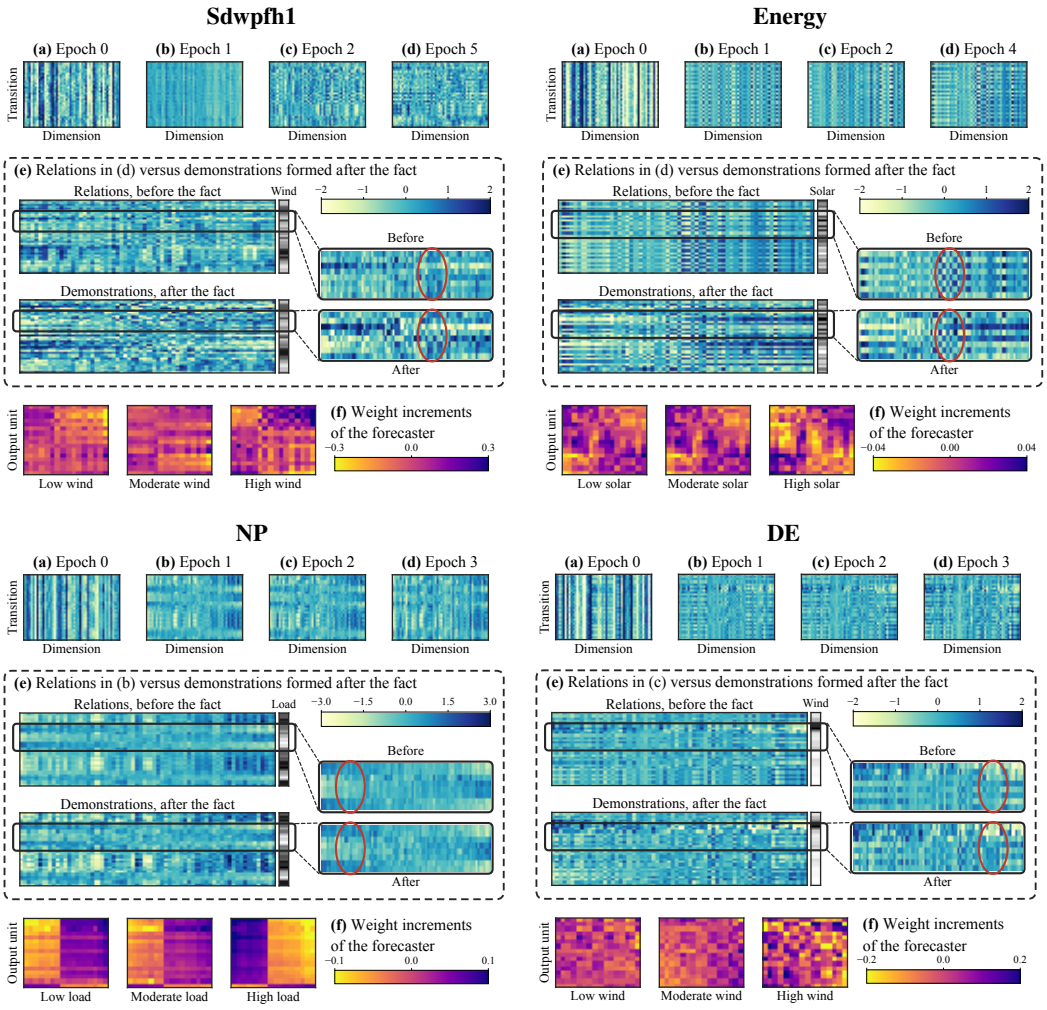


Figure 6: Visualization of the relations on test windows of Sdwpfh1, Energy, NP, and DE. The side strips show the wind speed, the solar generation, the grid load forecast, and the wind power forecast, respectively.

D.5 Wind components versus wind speed on GEFCom2014 wind

Table 16: GEFCom2014 wind with the NWP wind components (U/V) and with the wind speeds at 10 m and 100 m added (+ speed), averaged over the ten zones. Each model keeps its configuration of Table 10; Change: relative change of Avg.

Models		InRelCast		DAG		GCGNet		KITE		TimeXer		TFT	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	U/V	0.357	0.421	0.740	0.651	0.816	0.701	0.786	0.667	0.771	0.680	0.381	0.438
	+ speed	0.362	0.419	0.389	0.448	0.422	0.480	0.455	0.492	0.548	0.559	0.387	0.442
360	U/V	0.375	0.423	1.182	0.867	1.650	1.054	1.347	0.926	1.267	0.926	0.403	0.451
	+ speed	0.392	0.429	0.619	0.606	1.480	0.994	0.590	0.588	1.012	0.812	0.421	0.457
Avg	U/V	0.366	0.422	0.961	0.759	1.233	0.878	1.067	0.797	1.019	0.803	0.392	0.445
	+ speed	0.377	0.424	0.504	0.527	0.951	0.737	0.523	0.540	0.780	0.686	0.404	0.450
Change		+3.0%	+0.5%	−47.6%	−30.6%	−22.9%	−16.1%	−51.0%	−32.2%	−23.5%	−14.6%	+3.1%	+1.1%

Wind power depends mainly on the wind speed, whereas the NWP inputs of GEFCom2014 wind are its components. To test whether this explains the large errors of DAG, GCGNet, KITE, and TimeXer in Table 2, we add the wind speeds at 10 m and 100 m as exogenous variables. The speeds reduce the average MSE of these four models by 22.9% to 51.0%, whereas TFT and InRelCast,

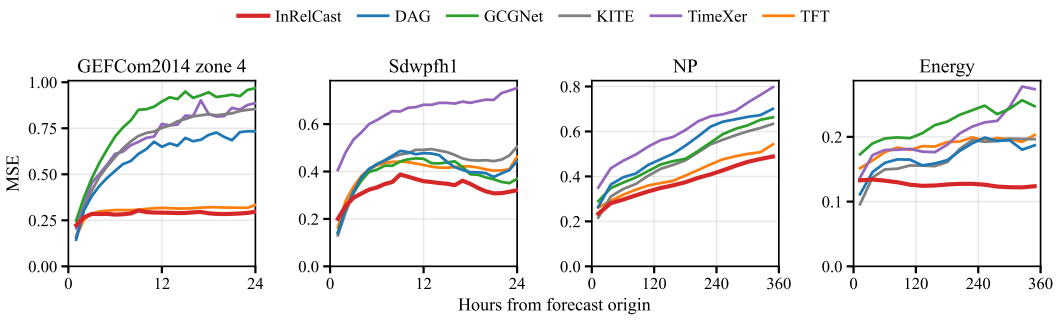


Figure 7: MSE at each hour of the horizon with future exogenous variables available, at horizon 24 on GEFCom2014 zone 4 and Sdwpfh1 and 360 on NP and Energy (daily means).

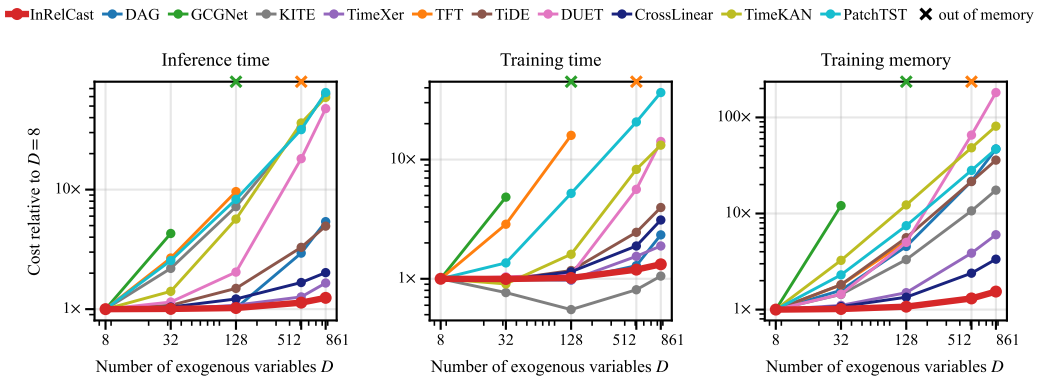


Figure 8: Cost as the number of exogenous variables increases, relative to the run with the first 8 exogenous variables of Traffic. ×: out of memory from that number on.

which process all exogenous variables of a time step jointly, do not improve (Table 16). InRelCast with the components alone remains more accurate than all five baselines with the speeds.

E Robustness and Sensitivity

E.1 Random Seed

Table 17 reports the mean and standard deviation of InRelCast over three random seeds, whose means are the values of the result tables.

E.2 Hyperparameter Sensitivity

We vary one hyperparameter at a time on NP, DE, Sdwpfh1, and GEFCom2014 zone 4 (Fig. 9). The MSE stays stable with respect to the rank R on all four datasets and with respect to every hyperparameter on NP and zone 4.

E.3 Effect of the Look-Back Length

We vary the look-back length T with all other hyperparameters fixed (Fig. 10). InRelCast is more accurate than DAG, GCGNet, KITE, and TimeXer at every look-back length and benefits from longer look-back windows.

Table 17: Mean ± standard deviation of InRelCast over three random seeds with future exogenous variables available.

Dataset	NP		PJM		DE		Energy	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	0.156±0.002	0.199±0.001	0.062±0.003	0.142±0.001	0.252±0.008	0.307±0.003	0.082±0.004	0.219±0.006
360	0.377±0.017	0.395±0.014	0.107±0.000	0.219±0.005	0.403±0.013	0.401±0.007	0.127±0.007	0.286±0.008
Dataset	GEFCom2014 zone 1		GEFCom2014 zone 4		GEFCom2014 zone 7		GEFCom2014 zone 10	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	0.425±0.005	0.467±0.003	0.286±0.003	0.364±0.001	0.319±0.004	0.404±0.003	0.360±0.005	0.424±0.001
360	0.400±0.001	0.461±0.001	0.282±0.000	0.352±0.000	0.278±0.002	0.383±0.002	0.378±0.000	0.430±0.001
Dataset	ETTh1		Weather		Traffic		Exchange	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.053±0.000	0.177±0.000	0.001±0.000	0.015±0.000	0.109±0.002	0.198±0.000	0.093±0.000	0.222±0.000
192	0.076±0.002	0.217±0.005	0.001±0.000	0.018±0.001	0.109±0.003	0.198±0.006	0.144±0.012	0.270±0.009
336	0.086±0.000	0.232±0.000	0.001±0.000	0.021±0.001	0.104±0.003	0.189±0.002	0.171±0.001	0.333±0.004
720	0.084±0.000	0.232±0.000	0.001±0.000	0.023±0.001	0.113±0.006	0.209±0.002	0.351±0.066	0.472±0.037

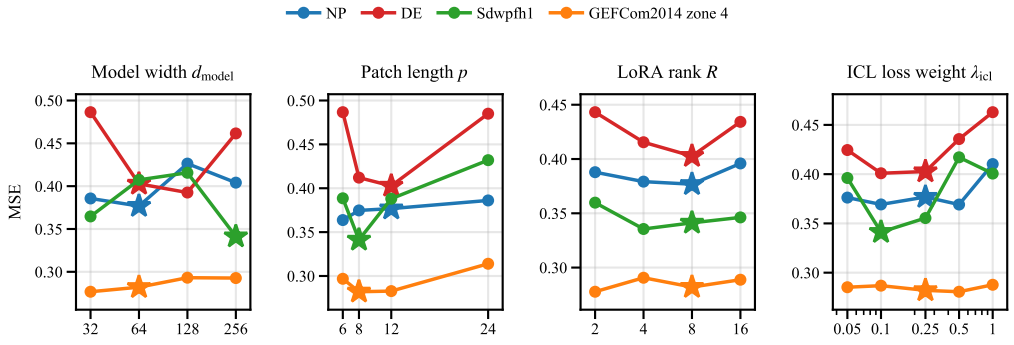


Figure 9: MSE at horizon 360 when one hyperparameter is varied and the others are fixed. Stars mark the values used in the main experiments.

E.4 One configuration per dataset across horizons

In Table 18, InRelCast (Fixed) uses the long-horizon configuration at both horizons and still achieves the best average MSE and MAE. The baseline columns repeat Table 8.

E.5 Imperfect exogenous variables

We add Gaussian noise to the future exogenous variables (Fig. 11) and, following GCGNet, mask exogenous entries in training and test data (Table 19). InRelCast is the most accurate model under mild noise and achieves the best average results under masking.

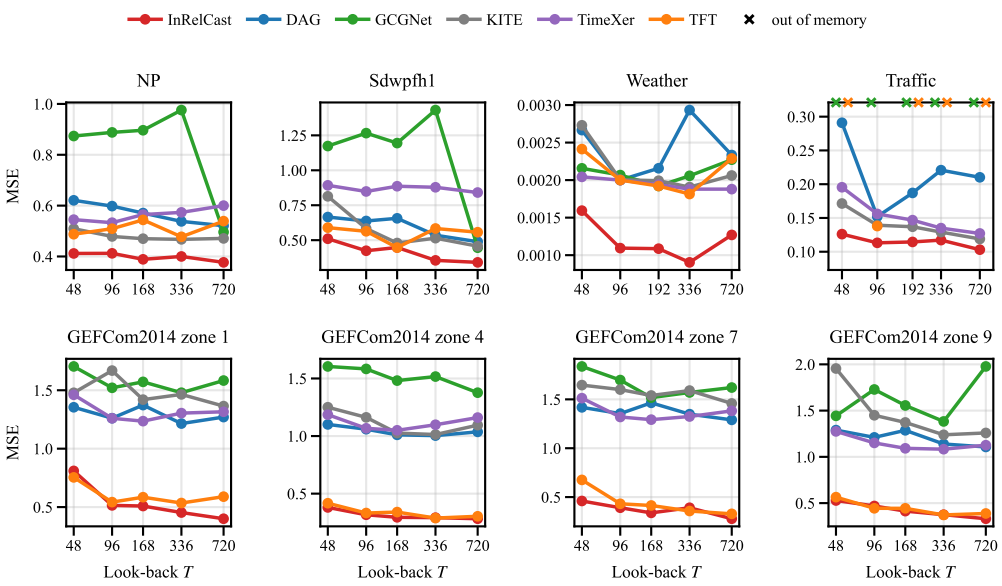


Figure 10: MSE against the look-back length T with future exogenous variables available, at horizon 360 (720 on Weather and Traffic). At the look-back length of the result tables, the points repeat Tables 8, 10, and 13.

Table 18: Full results on the 12 TSF-X datasets with future exogenous variables available and one configuration per dataset across both horizons. **Red**: best, Blue: second best.

Models		InRelCast (Fixed)		DAG		GCGNet		KITE		TimeXer		TFT		TiDE		DUET		CrossLinear		TimeKAN		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.167**	**0.204**	0.202	0.237	0.197	0.233	0.179	0.223	0.236	0.266	0.219	0.249	0.284	0.301	0.246	0.287	0.210	0.266	0.273	0.310	0.249	0.294
	360	**0.377**	**0.395**	0.521	0.451	0.496	0.442	0.471	0.422	0.600	0.475	0.539	0.501	0.601	0.498	0.576	0.528	0.531	0.508	0.538	0.529	0.530	0.498
	Avg	**0.272**	**0.300**	0.362	0.344	0.346	0.337	0.325	0.323	0.418	0.371	0.379	0.375	0.443	0.400	0.411	0.408	0.371	0.387	0.405	0.419	0.390	0.396
PJM	24	0.071	0.144	0.057	**0.143**	0.058	0.148	**0.056**	**0.143**	0.075	0.166	0.095	0.195	0.106	0.214	0.072	0.166	0.088	0.191	0.115	0.244	0.116	0.239
	360	**0.107**	0.219	0.130	0.218	0.128	0.223	0.135	**0.214**	0.140	0.231	0.133	0.219	0.177	0.279	0.131	0.228	0.135	0.254	0.162	0.281	0.149	0.279
	Avg	**0.089**	0.182	0.093	0.180	0.093	0.186	0.096	**0.178**	0.108	0.198	0.114	0.207	0.142	0.246	0.102	0.197	0.112	0.223	0.139	0.262	0.133	0.259
BE	24	0.362	0.228	0.361	0.229	**0.323**	**0.227**	0.348	0.240	0.392	0.253	0.426	0.272	0.426	0.285	0.432	0.272	0.391	0.259	0.451	0.319	0.452	0.326
	360	**0.470**	0.320	0.485	0.330	0.524	0.347	0.507	0.332	0.512	0.327	0.482	**0.310**	0.571	0.364	0.597	0.436	0.568	0.416	0.645	0.495	0.702	0.538
	Avg	**0.416**	**0.274**	0.423	0.279	0.423	0.287	0.427	0.286	0.452	0.290	0.454	0.291	0.498	0.325	0.515	0.354	0.479	0.337	0.548	0.407	0.577	0.432
FR	24	0.361	0.180	0.355	**0.171**	0.332	0.179	**0.311**	0.178	0.366	0.208	0.543	0.253	0.418	0.255	0.384	0.251	0.390	0.226	0.454	0.296	0.518	0.368
	360	0.464	0.275	0.473	0.268	0.478	0.280	**0.462**	0.271	0.489	0.273	0.465	**0.261**	0.551	0.308	0.607	0.403	0.575	0.370	0.641	0.452	0.658	0.452
	Avg	0.413	0.228	0.414	**0.219**	0.405	0.230	**0.387**	0.225	0.427	0.241	0.504	0.257	0.484	0.281	0.496	0.327	0.483	0.298	0.547	0.374	0.588	0.410
DE	24	**0.257**	**0.314**	0.277	0.322	0.282	0.328	0.262	0.327	0.339	0.362	0.380	0.383	0.367	0.383	0.376	0.378	0.387	0.396	0.399	0.412	0.413	0.412
	360	**0.403**	**0.401**	0.462	0.418	0.491	0.445	0.437	0.413	0.610	0.474	0.599	0.509	0.630	0.511	0.589	0.482	0.583	0.507	0.547	0.479	0.589	0.498
	Avg	**0.330**	**0.358**	0.370	0.370	0.387	0.387	0.350	0.370	0.475	0.418	0.489	0.446	0.499	0.447	0.482	0.430	0.485	0.452	0.473	0.445	0.501	0.455
Energy	24	0.094	0.234	0.079	0.215	**0.065**	**0.194**	0.071	0.208	0.122	0.273	0.093	0.235	0.103	0.248	0.117	0.283	0.241	0.418	0.135	0.298	0.239	0.390
	360	**0.127**	**0.286**	0.169	0.320	0.179	0.330	0.151	0.305	0.204	0.357	0.167	0.331	0.202	0.355	0.288	0.452	0.237	0.385	0.302	0.464	0.214	0.363
	Avg	**0.111**	0.260	0.124	0.267	0.122	0.262	**0.111**	**0.257**	0.163	0.315	0.130	0.283	0.153	0.302	0.203	0.367	0.239	0.402	0.218	0.381	0.226	0.377
Sdwpfm1	24	**0.345**	**0.381**	0.351	0.400	0.375	0.409	0.384	0.410	0.558	0.533	0.366	0.421	0.474	0.488	0.551	0.564	0.355	0.473	0.418	0.503	0.374	0.454
	360	**0.357**	**0.401**	0.495	0.522	0.458	0.490	0.450	0.491	0.845	0.684	0.597	0.528	0.492	0.526	0.646	0.577	0.497	0.532	0.476	0.566	0.497	0.550
	Avg	**0.351**	**0.391**	0.423	0.461	0.416	0.449	0.417	0.451	0.701	0.609	0.482	0.474	0.483	0.507	0.599	0.570	0.426	0.502	0.447	0.534	0.435	0.502
Sdwpfm2	24	0.393	**0.403**	**0.372**	0.414	0.410	0.443	0.443	0.450	0.627	0.570	0.411	0.458	0.461	0.492	0.445	0.452	0.477	0.536	0.474	0.538	0.418	0.492
	360	**0.417**	**0.428**	0.583	0.556	0.505	0.514	0.506	0.515	0.978	0.736	0.541	0.519	0.511	0.540	0.584	0.528	0.589	0.611	0.520	0.589	0.602	0.603
	Avg	**0.405**	**0.416**	0.477	0.485	0.458	0.479	0.475	0.483	0.803	0.653	0.476	0.488	0.486	0.516	0.514	0.490	0.533	0.573	0.497	0.564	0.510	0.547
Sdwpfh1	24	0.384	**0.416**	0.408	0.438	**0.381**	0.441	0.424	0.464	0.651	0.587	0.401	0.460	0.434	0.489	0.527	0.513	0.548	0.585	0.511	0.582	0.422	0.505
	360	**0.341**	**0.415**	0.489	0.534	0.446	0.495	0.457	0.517	0.841	0.700	0.557	0.523	0.472	0.527	0.551	0.519	0.566	0.601	0.643	0.694	0.513	0.548
	Avg	**0.363**	**0.416**	0.448	0.486	0.414	0.468	0.441	0.491	0.746	0.643	0.479	0.491	0.453	0.508	0.539	0.516	0.557	0.593	0.577	0.638	0.468	0.527
Sdwpfh2	24	**0.392**	**0.414**	0.438	0.465	0.426	0.474	0.477	0.498	0.820	0.677	0.474	0.493	0.579	0.553	0.629	0.563	0.468	0.540	0.580	0.614	0.457	0.527
	360	**0.366**	**0.394**	0.608	0.595	0.519	0.529	0.524	0.555	0.962	0.761	0.657	0.549	0.619	0.614	0.665	0.569	0.608	0.609	0.713	0.729	0.641	0.632
	Avg	**0.379**	**0.404**	0.523	0.530	0.472	0.501	0.500	0.526	0.891	0.719	0.566	0.521	0.599	0.583	0.647	0.566	0.538	0.574	0.647	0.672	0.549	0.580
Colbun	10	0.140	0.148	0.061	0.094	0.057	**0.084**	**0.055**	0.101	0.113	0.172	0.092	0.135	0.089	0.131	0.089	0.134	0.071	0.102	0.061	0.101	0.062	0.122
	30	**0.113**	0.223	0.135	**0.215**	0.138	0.224	0.121	0.243	0.176	0.299	0.383	0.460	0.240	0.322	0.307	0.397	0.182	0.288	0.195	0.249	0.417	0.496
	Avg	0.127	0.186	0.098	**0.154**	0.098	**0.154**	**0.088**	0.172	0.145	0.235	0.238	0.297	0.164	0.227	0.198	0.266	0.126	0.195	0.128	0.175	0.239	0.309
Rapel	10	0.190	0.220	0.151	0.203	**0.140**	**0.191**	0.192	0.231	0.301	0.308	0.201	0.253	0.228	0.271	0.174	0.219	0.163	0.209	0.174	0.231	0.163	0.218
	30	**0.233**	**0.301**	0.309	0.408	0.284	0.327	0.296	0.358	0.387	0.416	0.409	0.414	0.411	0.432	0.365	0.432	0.340	0.417	0.325	0.390	0.374	0.445
	Avg	**0.212**	0.261	0.230	0.305	**0.212**	**0.259**	0.244	0.294	0.344	0.362	0.305	0.333	0.320	0.351	0.269	0.326	0.252	0.313	0.249	0.311	0.269	0.332
Average		**0.289**	**0.306**	0.332	0.340	0.321	0.333	0.322	0.338	0.473	0.421	0.385	0.372	0.394	0.391	0.415	0.401	0.383	0.404	0.406	0.432	0.407	0.427
1st Count		**26**	**22**	1	5	5	6	7	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0

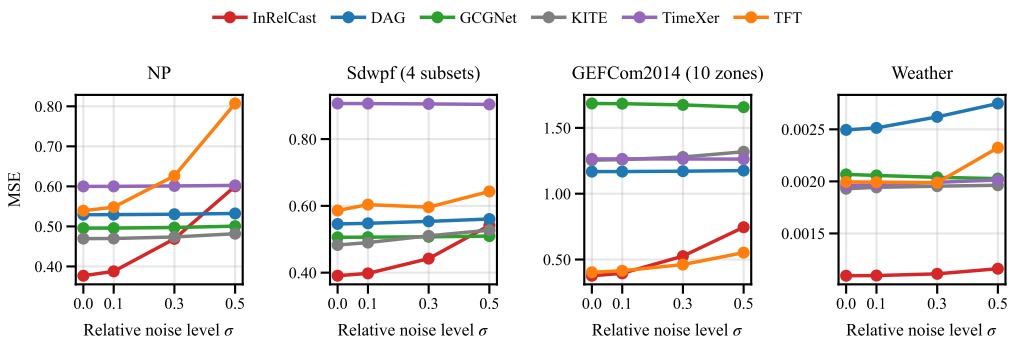


Figure 11: MSE with Gaussian noise of relative level σ added to the future exogenous variables at evaluation, at horizon 360 (720 on Weather). Sdwpf and GEFCom2014 wind are averaged over their subsets and zones.

Table 19: Results with masked exogenous variables (Zero: zeros; Random: standard normal draws), averaged over mask ratios of 10, 30, and 50 %; Average is the mean over the datasets and both mask types. **Red**: best, Blue: second best.

Models			InRelCast		DAG		TFT		GCGNet		KITE		TimeXer		CrossLinear		DUET		PatchTST	
Metrics			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Zero	Colbun	10	0.071	0.101	0.067	0.098	0.097	0.143	0.076	**0.093**	0.071	0.115	0.084	0.145	**0.065**	0.102	0.106	0.126	0.070	0.108
		30	0.278	0.374	**0.141**	**0.215**	0.367	0.430	0.223	0.300	0.194	0.299	0.197	0.313	0.213	0.313	0.993	0.780	0.494	0.538
		Avg	0.175	0.238	**0.104**	**0.157**	0.232	0.287	0.150	0.197	0.133	0.207	0.141	0.229	0.139	0.208	0.550	0.453	0.282	0.323
	Energy	24	0.098	0.240	0.089	0.231	0.120	0.268	**0.070**	**0.203**	0.087	0.228	0.122	0.273	0.097	0.241	0.100	0.244	0.108	0.257
		360	0.217	0.357	0.205	0.358	0.232	0.380	0.220	0.371	**0.188**	**0.342**	0.221	0.372	0.352	0.464	0.244	0.398	0.277	0.418
		Avg	0.158	0.299	0.147	0.295	0.176	0.324	0.145	0.287	**0.138**	**0.285**	0.172	0.323	0.225	0.353	0.172	0.321	0.193	0.338
	DE	24	0.304	**0.336**	0.349	0.361	0.431	0.414	**0.292**	**0.336**	0.324	0.358	0.418	0.403	0.426	0.416	0.428	0.407	0.480	0.445
		360	**0.440**	**0.425**	0.574	0.465	0.612	0.508	0.617	0.509	0.543	0.465	0.629	0.486	0.894	0.617	0.819	0.579	0.830	0.586
		Avg	**0.372**	**0.381**	0.462	0.413	0.522	0.461	0.455	0.423	0.434	0.412	0.524	0.445	0.660	0.517	0.624	0.493	0.655	0.516
	PJM	24	0.075	**0.152**	0.076	0.170	0.120	0.213	**0.064**	0.155	0.081	0.179	0.100	0.198	0.098	0.203	0.079	0.177	0.120	0.246
		360	**0.126**	**0.229**	0.164	0.246	0.156	0.244	0.163	0.253	0.167	0.260	0.164	0.250	0.148	0.253	0.157	0.243	0.183	0.296
		Avg	**0.101**	**0.191**	0.120	0.208	0.138	0.229	0.114	0.204	0.124	0.220	0.132	0.224	0.123	0.228	0.118	0.210	0.152	0.271
	NP	24	**0.187**	**0.219**	0.242	0.258	0.276	0.291	0.204	0.235	0.243	0.263	0.289	0.296	0.426	0.416	0.258	0.270	0.268	0.293
		360	**0.484**	**0.437**	0.635	0.513	0.502	0.460	0.609	0.494	0.555	0.458	0.612	0.479	0.625	0.525	0.609	0.519	0.592	0.516
		Avg	**0.336**	**0.328**	0.439	0.386	0.389	0.376	0.407	0.365	0.399	0.361	0.451	0.388	0.526	0.471	0.434	0.395	0.430	0.405
	Weather	96	**0.0006**	**0.0165**	0.0009	0.0205	0.0015	0.0283	0.0013	0.0267	0.0008	0.0187	0.0012	0.0257	0.0081	0.0729	0.0013	0.0264	0.0067	0.0637
		720	0.0015	0.0274	0.0024	0.0352	0.0020	0.0335	0.0020	0.0336	0.0020	0.0329	0.0020	0.0329	0.0097	0.0780	**0.0014**	**0.0273**	0.0090	0.0736
		Avg	**0.0011**	**0.0220**	0.0017	0.0279	0.0018	0.0309	0.0017	0.0302	0.0014	0.0258	0.0016	0.0293	0.0089	0.0755	0.0014	0.0269	0.0079	0.0687
Random	Colbun	10	0.066	0.098	0.064	**0.096**	0.110	0.153	0.070	**0.096**	**0.060**	0.108	0.112	0.169	0.066	0.108	0.084	0.118	0.078	0.120
		30	0.402	0.463	**0.147**	**0.218**	0.293	0.387	0.240	0.320	0.165	0.277	0.223	0.323	0.213	0.314	1.064	0.819	0.468	0.532
		Avg	0.234	0.281	**0.106**	**0.157**	0.202	0.270	0.155	0.208	0.113	0.193	0.168	0.246	0.140	0.211	0.574	0.469	0.273	0.326
	Energy	24	0.099	0.240	0.089	0.230	0.131	0.282	**0.081**	**0.217**	0.089	0.230	0.123	0.275	0.098	0.242	0.100	0.245	0.102	0.249
		360	0.181	0.338	0.228	0.380	**0.177**	**0.332**	0.215	0.367	0.200	0.353	0.205	0.359	0.354	0.465	0.242	0.397	0.274	0.416
		Avg	**0.140**	**0.289**	0.159	0.305	0.154	0.307	0.148	0.292	0.145	0.292	0.164	0.317	0.226	0.354	0.171	0.321	0.188	0.333
	DE	24	**0.304**	**0.334**	0.333	0.354	0.417	0.410	0.344	0.366	0.325	0.360	0.433	0.411	0.433	0.418	0.431	0.407	0.490	0.449
		360	**0.457**	**0.422**	0.562	0.461	0.632	0.517	0.573	0.490	0.541	0.462	0.641	0.492	0.932	0.628	0.801	0.573	0.830	0.586
		Avg	**0.381**	**0.378**	0.448	0.408	0.525	0.464	0.459	0.428	0.433	0.411	0.537	0.452	0.683	0.523	0.616	0.490	0.660	0.518
	PJM	24	**0.065**	**0.150**	0.076	0.170	0.129	0.226	0.066	0.158	0.085	0.183	0.102	0.199	0.101	0.201	0.079	0.176	0.112	0.224
		360	**0.135**	**0.227**	0.171	0.249	0.149	0.237	0.169	0.256	0.166	0.257	0.164	0.247	0.150	0.254	0.158	0.244	0.185	0.296
		Avg	**0.100**	**0.189**	0.124	0.210	0.139	0.232	0.118	0.207	0.126	0.220	0.133	0.223	0.126	0.228	0.119	0.210	0.149	0.260
	NP	24	**0.179**	**0.215**	0.239	0.257	0.299	0.302	0.208	0.236	0.242	0.260	0.293	0.297	0.248	0.285	0.260	0.270	0.269	0.291
		360	**0.442**	**0.420**	0.559	0.465	0.546	0.475	0.632	0.501	0.551	0.454	0.612	0.476	0.615	0.522	0.623	0.532	0.584	0.512
		Avg	**0.311**	**0.318**	0.399	0.361	0.423	0.389	0.420	0.369	0.397	0.357	0.453	0.387	0.432	0.404	0.442	0.401	0.427	0.402
	Weather	96	**0.0006**	**0.0173**	0.0008	0.0198	0.0013	0.0267	0.0013	0.0270	0.0008	0.0189	0.0012	0.0257	0.0084	0.0740	0.0014	0.0268	0.0065	0.0639
		720	**0.0014**	**0.0266**	0.0023	0.0346	0.0021	0.0347	0.0020	0.0337	0.0021	0.0333	0.0020	0.0328	0.0095	0.0772	0.0015	0.0276	0.0083	0.0722
		Avg	**0.0010**	**0.0220**	0.0016	0.0272	0.0017	0.0307	0.0017	0.0304	0.0015	0.0261	0.0016	0.0293	0.0090	0.0756	0.0015	0.0272	0.0074	0.0681
Average			**0.193**	**0.245**	0.209	0.246	0.242	0.283	0.215	0.253	0.204	0.251	0.240	0.274	0.275	0.304	0.319	0.318	0.285	0.319
1st Count			23	25	4	5	1	1	4	5	3	2	0	0	1	0	1	1	0	0

F SHOWCASES

On selected windows, the exogenous input moves the InRelCast forecast from the endogenous baseline toward the ground truth (Fig. 12) and shifts it when delayed or scaled (Fig. 13). Without the ICL loss or with a mean-pooled relation, the relations lose much of their daily structure and forecast peaks are distorted (Fig. 14). InRelCast, like TFT, follows the observed power curve (Fig. 15). On an NP window, it best tracks the lower price signaled by the rising wind power forecast (Fig. 16).

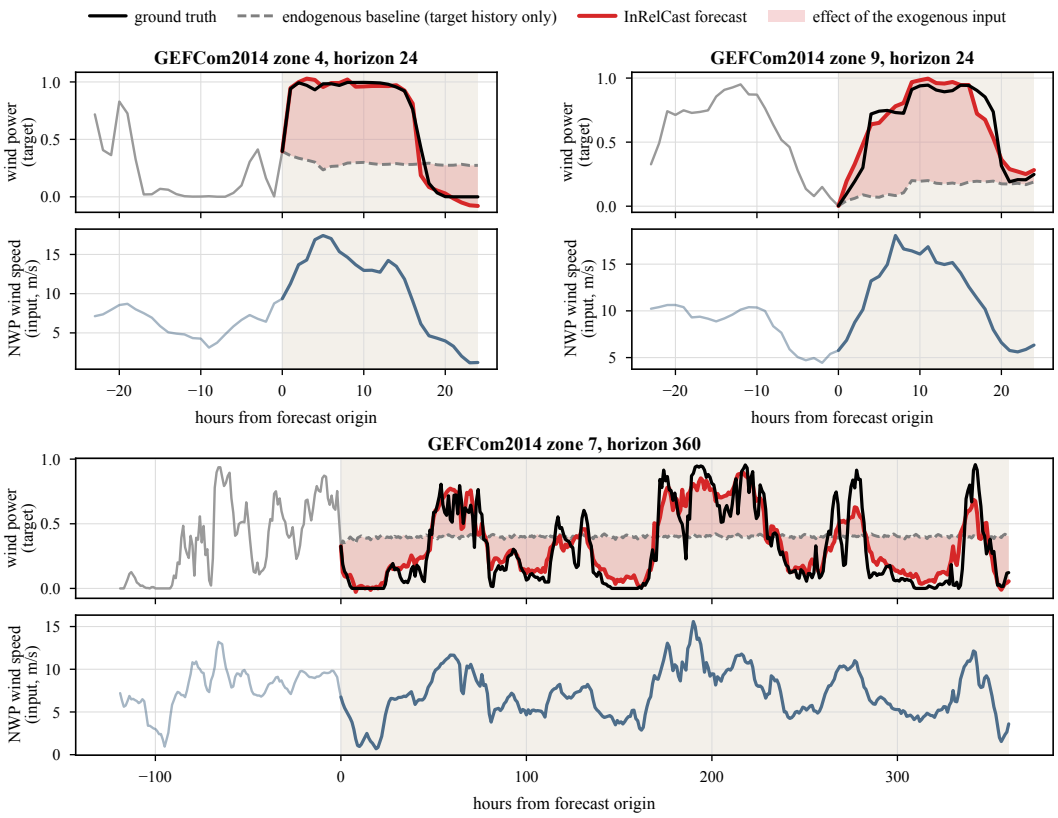


Figure 12: Decomposition of the InRelCast forecast into the endogenous baseline and the effect of the exogenous input on three windows of GEFCom2014 wind.

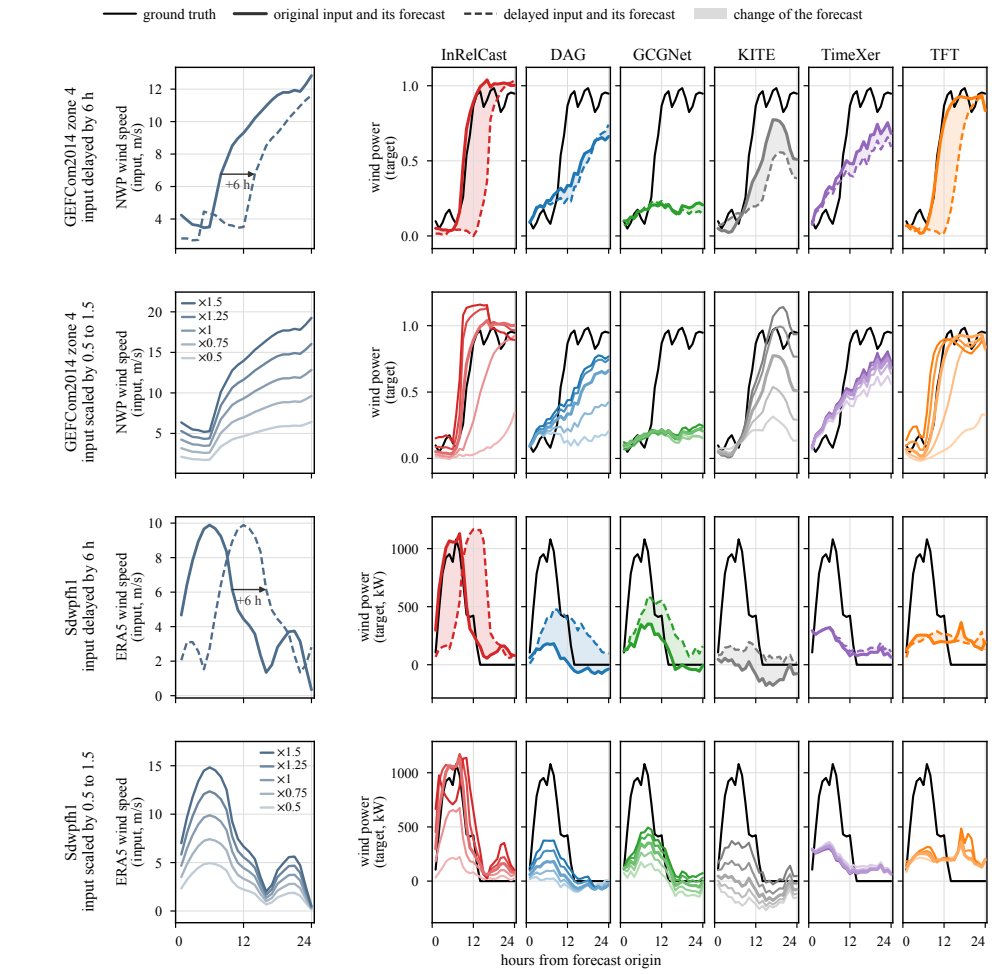


Figure 13: Forecasts of one window per dataset when the future exogenous input is delayed or scaled at inference, on zone 4 of GEFCom2014 wind (upper rows) and Sdwpfh1 (lower rows).

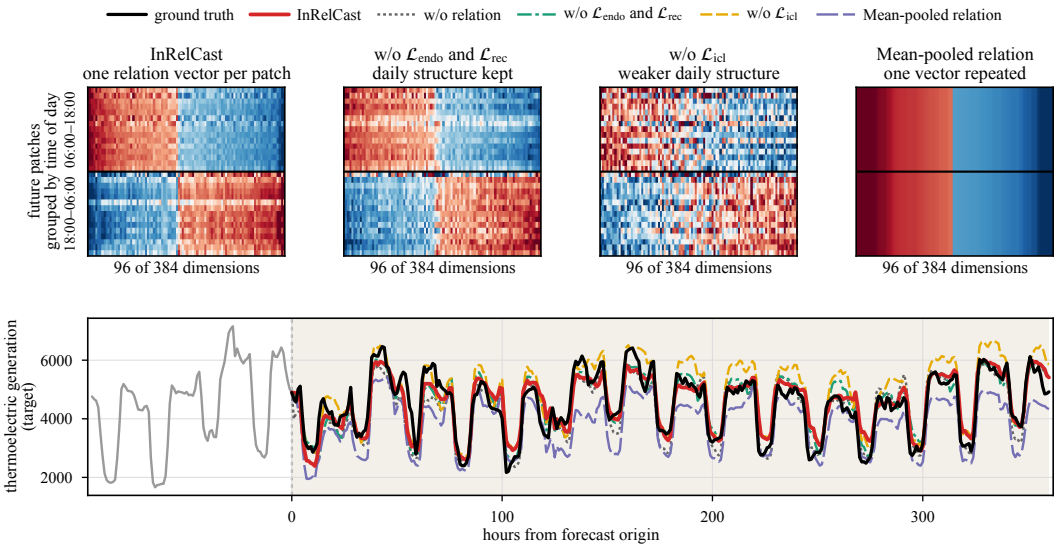


Figure 14: Relation vectors of InRelCast and of its ablations on one window of Energy at horizon 360, and the forecasts they produce.

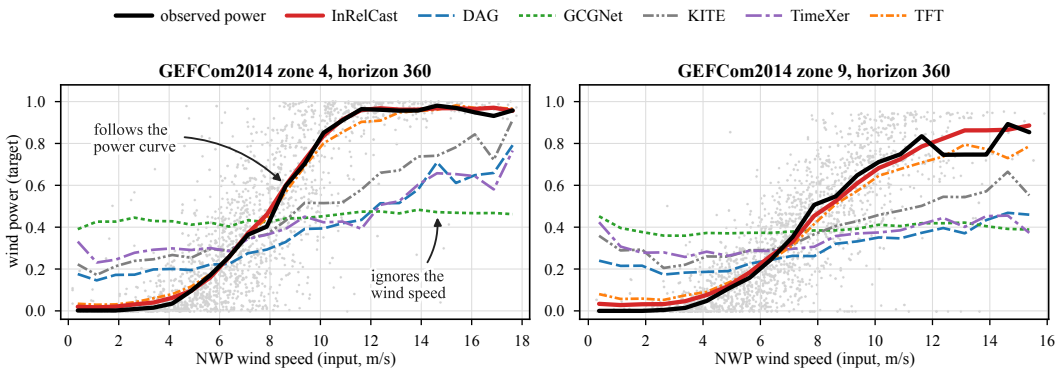


Figure 15: Forecast of each model against the NWP wind speed of the same hour at horizon 360 on two zones of GEFCom2014 wind, as the median per wind-speed bin over all test windows. Black curve and gray points: observed power.

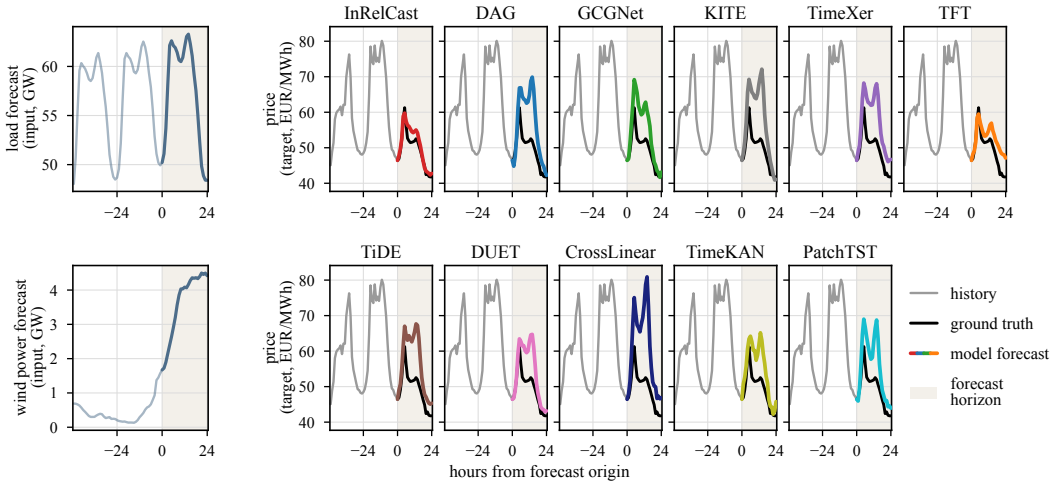


Figure 16: Forecasts of InRelCast and the ten baselines on one window of NP at horizon 24, with the two exogenous inputs on the left.