Uncertainty-Aware Transparency in Time Series Forecasting

Richard Franck [1]

Abstract

Machine learning can improve decision-making in high-stakes domains such as healthcare and finance, but is hindered by limited transparency and interpretability. While established transparent models exist for classification and regression tasks, achieving comparable transparency in time series forecasting remains an open challenge. For time series forecasting based on static features, Kacprzyk et al. (2024) introduce TIMEVIEW as a transparent ML model that produces deterministic trajectories. Conditional on a fixed set of learned model parameters, TIMEVIEW enables understanding how changes in inputs affect the predicted trajectory and its properties. We show that epistemic uncertainty means that a single feature vector may be assigned to qualitatively different trajectories, making explanations that ignore uncertainty potentially misleading. In high-stakes settings, this may have serious consequences. We therefore augment the bi-level transparency framework of Kacprzyk et al. (2024) with a third layer capturing uncertainty over qualitative trajectory behaviour.

Code: https://github.com/richardfranck/uncertainty-ts-forecasting

1. Introduction

Machine learning promises to improve decision-making by uncovering relationships in historical evidence that humans may overlook. This has led to the adoption of machine learning models to improve or even automate decision-making in high-stakes domains, including healthcare and finance. The high-stakes nature of these domains means that predictions greatly impact human lives and poor modelling practices can cause significant societal harm (Barocas et al., 2023).

To mitigate these risks, practitioners must be able to understand how and why model predictions change in response to inputs. However, many modern machine learning models operate as black boxes, meaning the decision-making process that generates predictions from inputs is indecipherable to humans. The field of explainable AI seeks to address this limitation by promoting transparency and interpretability (Rudin, 2019). Following Kacprzyk et al. (2024), we define a machine learning model as transparent when the relationship between changes in one or multiple feature values and the resulting predictions is interpretable. Established transparent models exist for classification and regression tasks, including decision trees and linear regression. Achieving similar transparency for time series forecasting, which seeks to predict a trajectory rather than a single categorical or real-valued label, remains an open challenge.

Kacprzyk et al. (2024) introduce TIMEVIEW, a transparent ML model for time series forecasting based on static features. Rather than explaining forecasts at the level of individual time points, TIMEVIEW characterises a trajectory holistically through its trend (i.e., its overall shape) and trend properties (e.g., the existence and location of minima and maxima), which together constitute the two levels of the author's bi-level transparency framework. Based on this framework, TIMEVIEW produces a deterministic trajectory forecast and provides an interactive visualisation tool that allows users to explore how trends and their properties change in response to perturbations of the static input features.

TIMEVIEW implicitly assumes that a given feature vector is associated with a single predicted trajectory. We show that epistemic uncertainty can result in a single feature vector being associated with qualitatively different trajectories, making explanations that ignore uncertainty potentially misleading. For example, a medical doctor may use TIMEVIEW trained on historical patient data to forecast a patent's tumor volume trajectory based on baseline covariates and proposed drug dosages. We show that epistemic uncertainty means that multiple qualitatively distinct future trajectories, such as monotonic tumour shrinkage versus shrinkage followed by regrowth, may be consistent with the data. Ignoring this uncertainty can create a false sense of certainty in model predictions and their explanations. Model uncertainty must therefore be treated as a first-class component of model transparency. We therefore augment the bi-level transparency framework of Kacprzyk et al. (2024) with a third layer that captures uncertainty over qualitative

[1]Department of Computing, Imperial College London, London, United Kingdom. Correspondence to: Richard Franck <richard.franck25@imperial.ac.uk>.

trajectory behaviour.

2. Methodology I: Deterministic Prediction

We build on TIMEVIEW (Kacprzyk et al., 2024), a transparent machine learning model for time series forecasting based on static features, whose core elements are therefore briefly outlined below.

2.1. Time Series Forecasting Problem Formulation

Let $T \in \mathbb{R}$ denote the time horizon. We observe a dataset $\{(\boldsymbol{x}^{(d)}, \boldsymbol{y}^{(d)}, \boldsymbol{t}^{(d)})\}_{d=1}^{D}$ of $D \in \mathbb{N}$ samples. Each sample d is represented by a triple $(\boldsymbol{x}^{(d)}, \boldsymbol{y}^{(d)}, \boldsymbol{t}^{(d)})$, where $\boldsymbol{x}^{(d)} \in \mathbb{R}^M$ is a vector of static features, with $M \in \mathbb{N}$ denoting the number of features. The corresponding outcome is a trajectory $\boldsymbol{y}^{(d)} \in \mathbb{R}^{N_d}$ recorded at observation times $\boldsymbol{t}^{(d)} \in \mathbb{R}^{N_d}$, where $N_d \in \mathbb{N}$ denotes the number of measurements for sample d.

The data-generating process of $\boldsymbol{y}^{(d)}$ is assumed to be a continuous time process such that true trajectory for sample d is given by $y_*^{(d)} : [0, T] \mapsto \mathbb{R}$. The goal is to learn a continuous trajectory $\hat{y} : [0, T] \mapsto \mathbb{R}$ based on discrete time observations that minimises the expectation over all samples of the mean squared error loss

$$\frac{1}{T} \int_0^T \left(\hat{y}(t) - y_*(t)\right)^2 dt. \tag{1}$$

2.2. The TIMEVIEW Model

Let $\phi_b : [0, T] \mapsto \mathbb{R}$ for $b \in 1\{, \ldots, B\}$ be a cubic B-spline basis function that only depends on time. Given a fixed set of B such functions, TIMEVIEW predicts the entire trajectory $\hat{y}(t)$ as a linear weighted sum

$$\hat{y}(t) = \sum_{b=1}^{B} c_b \phi_b(t), \tag{2}$$

where c_b for $b = \{1, ..., B\}$ denote coefficients that determine the shape of the trajectory. The coefficients depend only on the static features $\boldsymbol{x}$ and are estimated form a feed-forward neural network parametrised by θ,

$$c(\boldsymbol{x}) = h_\theta(\boldsymbol{x}). \tag{3}$$

The network parameters are learned via gradient descent by minimising the l_2-regularised mean squared error loss function

$$\mathcal{L} = \frac{1}{D} \sum_{d=1}^{D} \left(\frac{1}{N_d} \sum_{j=1}^{N_d} \left(y_j^d - \hat{y}_j^d\right)^2 \right) + \lambda \|\theta\|_2^2. \tag{4}$$

2.3. Bi-level Transparency

The spline-based formulation of TIMEVIEW enables understanding how changes in the inputs affect both the global trend of the predicted trajectory (level 1 transparency) and the properties of a given trend (level 2 transparency). Following Kacprzyk et al. (2024), we introduce motifs and compositions to formalise the conditions under which these two levels of transparency are achieved.

A motif formalises the idea of the *local* shape (i.e., a qualitative property) of a function over a specific subinterval. Motifs are required to satisfy two assumptions: well-definedness, which ensures that a motif can only be assigned on intervals where the function is defined, and translation invariance, which guarantees that motifs describe shapes independently of the absolute position of the interval. The global trend of a function is described by a composition, which is the shortest ordered sequence of motifs assigned to a partition of the function's domain, where each subinterval corresponds to a region of consistent local shape (e.g. increasing, constant or decreasing). Transition points are the boundaries between consecutive motifs and, for the dynamical motifs considered in TIMEVIEW, correspond to minima, maxima, and inflection points.[1]

A set of motifs is said to be compatible with a class of trajectories if every trajectory in the class admits a unique composition, and if this composition can be computed deterministically. Under this assumption, a composition map can assign each input feature vector to a single, well-defined composition. In TIMEVIEW, this requirement is met by restricting the model output to cubic splines. Thus, the trajectory in equation 2 induces a well-defined composition map and enables bi-level transparency.

3. Methodology II: Uncertainty in Prediction

While the interpretability guarantees of TIMEVIEW hold conditional on a trained model, different plausible models may be obtained under epistemic uncertainty. As a result, the same input features may induce different predicted trajectories across model instances. Understanding and quantifying this variability is essential for transparent reasoning in high-stakes settings.

3.1. Predictive Uncertainty

Uncertainty in machine learning predictions is commonly decomposed into aleatoric and epistemic components. Aleatoric uncertainty arises from inherent noise in the observations and cannot be reduced through additional data. Epis-

[1] A graphical illustration for a composition is provided in Appendix A.1.

temic uncertainty arises from uncertainty about the model parameters and, in principle, can be reduced away with sufficient training data (Kendall & Gal, 2017).

Epistemic uncertainty is particularly important for transparency in time series forecasting as it implies that multiple models with varying parameterisation may be consistent with the observed data. For time series forecasting, this implies that the same input feature vector can induce qualitatively different predicted trajectories across model instances. In high-stakes domains, ignoring this variability can lead to overconfident and potentially harmful decisions, making the explicit quantification of epistemic uncertainty a necessary component of transparent machine learning models.

3.2. Synthetic Tumour Growth Data

To explore the significance of epistemic uncertainty for TIMEVIEW, we generate a dataset of synthetic tumour growth data based on Wilkerson et al. (2017) and Kacprzyk et al. (2024). Details of the data-generation procedure are provided in Appendix A.2.

The dataset comprises 2000 samples, each characterised by four static features drawn independently from uniform distributions with ranges specified in Table 1. For each sample, the tumour volume is computed at twenty equally spaced time points in the interval $[0, 1]$. No observational noise is added to the generated trajectories, such that aleatoric uncertainty is effectively zero and all predictive uncertainty is epistemic.

Table 1. Ranges of static features sampled from independent uniform distributions.

VARIABLE	MINIMUM	MAXIMUM
AGE	20	80
WEIGHT	40	100
INITIAL TUMOR VOLUME	0.1	0.5
DRUG DOSAGE	0.0	1.0

3.3. Bootstrap Estimation of Epistemic Uncertainty

To approximate epistemic uncertainty in the trajectory predictions of TIMEVIEW, we employ a bootstrap approach (Efron, 1979).

Let $\mathcal{D} = \{(\boldsymbol{x}^{(d)}, \boldsymbol{y}^{(d)}, \boldsymbol{t}^{(d)})\}_{d=1}^{D}$ denote the tumour growth dataset generated as described in Section 3.2. A bootstrap replicate $\mathcal{D}^{(b)}$ is constructed by sampling D elements from $\mathcal{D}$ with replacement. We repeat this procedure B times to obtain a set of bootstrap datasets $\{\mathcal{D}^{(b)}\}_{b=1}^{B}$. For each bootstrap dataset, we train a separate TIMEVIEW model, yielding a set of fitted models $\{\hat{y}^{(b)}(\boldsymbol{x})(t)\}_{b=1}^{B}$.

To ensure that variability across bootstrap models arises only from differences in the resampled training data, each model is trained using a fixed random seed for neural network initialisation, and all Bootstrap models share the same internal knot positions for the B-spline basis functions.[2]

For a fixed vector of input features $\boldsymbol{x}$, each bootstrap model produces a predicted trajectory. The resulting collection of trajectories $\{\hat{y}_{\boldsymbol{x}}^{(b)}(t)\}_{b=1}^{B}$ defines an empirical distribution over plausible predictions induced by epistemic uncertainty.

For each individual bootstrap model, the predicted trajectory admits a unique composition, such that the bi-level transparency guarantees of TIMEVIEW hold at the level of individual model instances. Across bootstrap models, for a fixed feature vector $\boldsymbol{x}$, epistemic uncertainty induces variability in the predicted trajectories and locations of transition points. Thus, bootstrapping enables an uncertainty-aware interpretation of model outputs while preserving the bi-level transparency of each individual prediction.

4. Empirical Analysis

We implement an ensemble of TIMEVIEW models trained on $B = 30$ bootstrap resamples of the original dataset obtained as described in Section 3.3.

For illustrative purposes, we consider a single individual, *patient zero*, drawn from the original simulated dataset described in Section 3.2. Patient zero's static covariates are summarised in Table 2. Given the set of fixed covariates of patient zero, all variability in predictions across ensembles can be attributed to epistemic uncertainty.

Table 2. Static covariates and drug dosage of patient zero.

VARIABLE	VALUE
AGE	58.22
WEIGHT	98.64
INITIAL TUMOR VOLUME	0.34
DRUG DOSAGE	0.66

Figure 1 shows the observed tumour volume trajectory for patient zero as well as two predicted trajectories obtained from TIMEVIEW models trained on different bootstrap samples. The predicted trajectories exhibit qualitatively different long-term behaviour. While both trajectories initially predict tumour shrinkage, one model forecasts a monotonic decrease over the entire time horizon, whereas the other predicts tumour regrowth after an initial decline. This illustrates that, even in the absence of aleatoric noise, a single feature vector may correspond to qualitatively contradictory

[2]The internal knots are computed once from the original dataset $\mathcal{D}$ using the knot placement algorithm from Kacprzyk et al. (2024), then held fixed across all bootstrap models.

trajectory predictions.

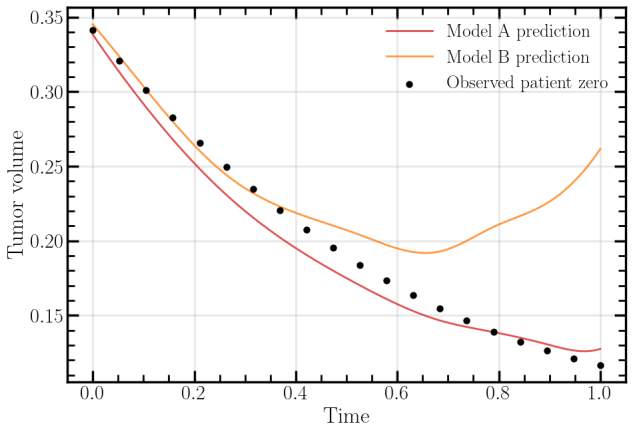

Figure 1. Qualitatively contradictory trajectory predictions for patient zero from models trained on different bootstrap samples.

To quantify the epistemic uncertainty illustrated for two selected bootstrap models in Figure 1 more systematically across all thirty trained models, we characterise the distribution of trajectory behaviours across the bootstrap ensemble. For each model $b \in \{1, \ldots, B\}$, we compute two summary statistics. First, the maximum decline, representing how much the tumour shrinks from its predicted volume at $t = 0$,

$$d^{(b)} = \hat{y}^{(b)}(\boldsymbol{x})(0) - \min_{t \in [0,1]} \hat{y}^{(b)}(\boldsymbol{x})(t), \tag{5}$$

and, second, its regrowth from its minimum to its predicted final volume at $t = 1$,

$$r^{(b)} = \hat{y}^{(b)}(\boldsymbol{x})(1) - \min_{t \in [0,1]} \hat{y}^{(b)}(\boldsymbol{x})(t). \tag{6}$$

Figure 2 visualises the empirical distribution of these $(d^{(b)}, r^{(b)})$ pairs across all B bootstrap models as a two-dimensional histogram. The (d, r) space is partitioned into an 8×8 grid of uniformly sized cells with axes spanning $[0, d_{\max}]$ where $d_{\max} = \max(\max_b d^{(b)}, \max_b r^{(b)})$. Each cell displays the fraction of the $B = 30$ bootstrap models falling within that bin's boundaries.

Points near the horizontal axis correspond to monotonic decline (minimal regrowth), points along the diagonal represent trajectories that return to initial tumour volume, and points above the diagonal indicate net growth despite treatment. For patient zero, 25 of the 30 bootstrap models predict monotonic decline (regrowth ≤ 0.03), while the remaining 5 models predict modest regrowth that remains below the initial tumour volume.

In a medical context, a doctor may be interested in forecasting tumour trajectories for a patient under different hypothetical drug dosages. We therefore examine how epistemic

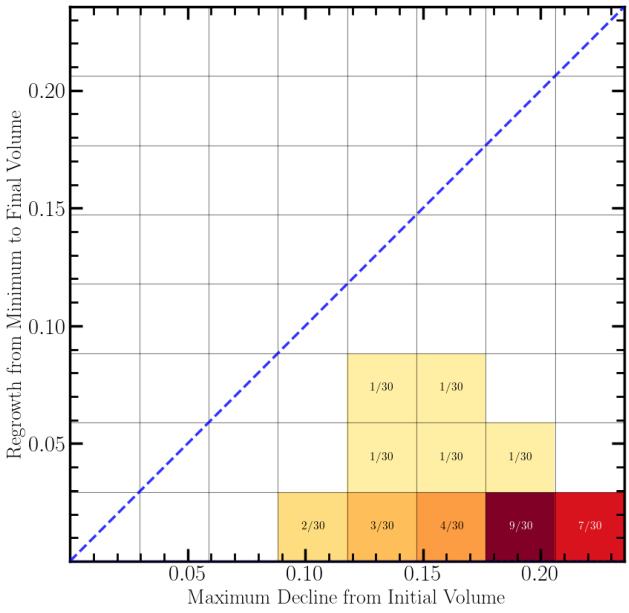

Figure 2. Empirical distribution of $(d^{(b)}, r^{(b)})$ pairs across all B bootstrap models for patient zero.

uncertainty varies with treatment intensity. Figure 3 summarises the variability of predicted tumour volume across the bootstrap ensemble as a function of time for multiple hypothetical drug dosage levels for patient zero. For each dosage level, we compute the standard deviation of the predicted tumour volume across bootstrap models at each time point as

$$\sigma(t) = \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left(\hat{y}^{(b)}(\boldsymbol{x})(t) - \bar{\hat{y}}(\boldsymbol{x})(t) \right)^2}, \tag{7}$$

where $\bar{\hat{y}}(\boldsymbol{x})(t) = \frac{1}{B} \sum_{b=1}^{B} \hat{y}^{(b)}(\boldsymbol{x})(t)$ is the mean predicted trajectory.

We observe that uncertainty is time-dependent, with increasing uncertainty toward the end of the forecasting horizon and varies systematically with drug dosage. Lower dosages are associated with substantially higher uncertainty at later time points, whereas higher dosages yield more tightly concentrated predictions.

5. Uncertainty-Aware Transparency

Section 4 demonstrates that epistemic uncertainty can lead to significant disagreement between equally plausible TIMEVIEW models. Bi-level transparency, which is conditional on a single trained model with a fixed set of learned parameters, does not address the range of possible

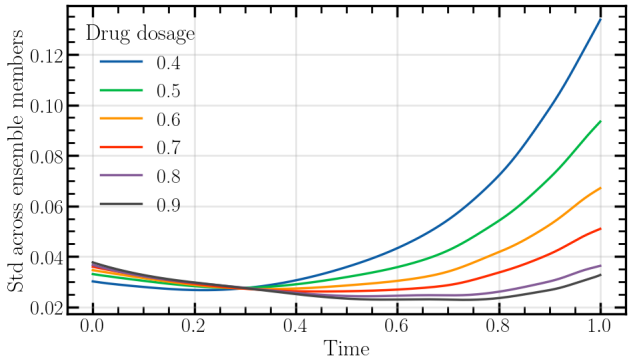

Figure 3. Epistemic uncertainty of patent zero's tumour growth under different drug dosages.

predictions supported by the data. To address this, we augment bi-level transparency with a third, uncertainty-aware layer. We treat the predictions of a bootstrap ensemble of TIMEVIEW models as an empirical distribution over trajectories for a fixed feature vector. Because each ensemble member is itself a TIMEVIEW predictor, bi-level transparency continues to hold at the level of individual models. The third layer therefore does not replace the framework by Kacprzyk et al. (2024), but acts as a complement that exposes how stable or unstable the transparent explanation is under uncertainty.

We implement this third layer as an interactive visualisation tool that allows practitioners to explore uncertainty for arbitrary feature configurations via sliders. Figure 4 illustrates this for the synthetic tumour dataset employed in the previous sections. The first panel visualises point-wise epistemic uncertainty over the forecasting horizon by plotting the point-wise ensemble median trajectory together with a 90% confidence band given by the 5th and 95th point-wise percentiles across bootstrap predictions. This representation highlights the empirically observed fact that epistemic uncertainty can vary substantially across time.

While point-wise bands convey where the model is uncertain about predicted values, they do not capture uncertainty over qualitative trajectory behaviour. The second panel, therefore, visualises epistemic uncertainty at the trajectory level by displaying a distribution over clinically meaningful tumour growth behaviours, obtained by summarising each ensemble trajectory using the decline–regrowth statistics introduced in Section 4. Plotting these statistics as a two-dimensional histogram yields a heatmap whose mass describes how likely different qualitative futures are under bootstrap-induced epistemic uncertainty.

Together, the two panels introduce uncertainty-aware transparency as a third layer that complements TIMEVIEW's original visualisation framework. In a clinical decision-support setting, a practitioner may use the tool to explore predicted tumour growth trajectories for a given patient under different treatment dosages by adjusting the feature sliders, while simultaneously inspecting how epistemic uncertainty evolves over the forecasting horizon. For some dosages, the median prediction may suggest sustained tumour shrinkage, yet the uncertainty band and trajectory behaviour distribution may reveal substantial probability mass on trajectories exhibiting regrowth. In this case, the practitioner may decide that more frequent follow-up measurements may be warranted.

6. Discussion of Limitations

The proposed framework is useful for reasoning about predictive robustness under epistemic uncertainty, but it is subject to several important limitations.

First, all experiments are conducted on a simulated tumour growth dataset, which facilitates isolating epistemic uncertainty from characteristics of real-world data such as inherent noise and measurement error. Future work should validate the utility of uncertainty-aware transparency on real patient datasets.

Second, the analysis of qualitative trajectory behaviour is restricted to a decline–regrowth pattern. While this is sufficient to describe tumour size dynamics under treatment, it cannot capture more complex temporal motifs (e.g. oscillations). Future work should extend the presented heatmap as one instantiation of a broader pipeline that quantifies uncertainty over qualitative trajectory properties more generally.

Third, although the proposed interface exposes the effects of epistemic uncertainty, it does not provide uncertainty attribution. In particular, the current tool does not indicate whether model disagreement is driven by sparse data support in regions of the covariate space or by competing explanations supported by different model instances. Incorporating uncertainty attribution would strengthen transparency by connecting uncertainty estimates to interpretable causes.

7. Related Work

This work lies at the intersection of uncertainty-aware explainability and transparent time series forecasting.

Uncertainty in Explanations. Recent work argues that uncertainty is largely overlooked in explainability research and that its incorporation can improve the reliability of predictive models. Sokol & Hüllermeier (2025) argue that transparency research largely overlooks uncertainty quantification and that its introduction promises to yield more understandable and reliable predictive models, while Salvi et al. (2025) study how integrating uncertainty quantification in explainable AI methods can enhance the reliability

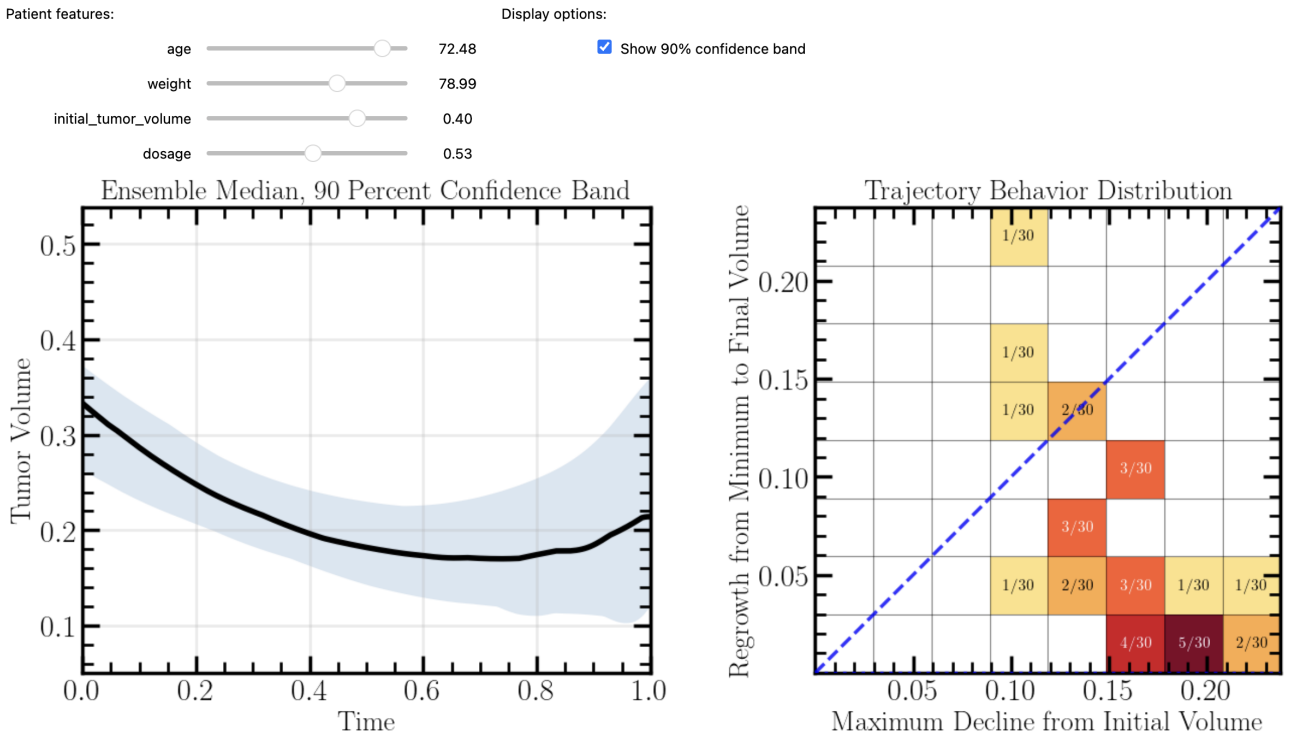

Figure 4. Snapshot of the dynamical uncertainty visualisation tool.

of deep learning models in healthcare. Slack et al. (2021) propose a Bayesian framework for generating post-hoc local explanations, including their associated uncertainty for black-box machine learning models. In contrast to this line of work, which focuses primarily on uncertainty in explanations of single-labelled outputs, our work considers epistemic uncertainty over entire forecast trajectories and their qualitative temporal behaviours.

Transparency in Time Series Forecasting. Kacprzyk et al. (2024) introduce TIMEVIEW, a transparent spline-based forecasting model that provides bi-level explanations in terms of trajectory composition and transition points, forming the foundation for our approach. Subsequent work extends TIMEVIEW to incorporate time-varying exogenous covariates beyond the static feature setting of the original formulation (Qin, 2025). While these methods provide transparency conditional on a single trained model, they do not account for epistemic uncertainty. Our work complements this line of research by extending TIMEVIEW with an uncertainty-aware transparency layer that highlights model disagreement over trajectories and qualitative behaviours.

8. Conclusion

We examined a key limitation of bi-level transparency in TIMEVIEW, which arises from the assumption that a single feature vector can be deterministically mapped to a single forecast trajectory. We showed that under epistemic uncertainty, equally plausible parameterisations of the same model can yield qualitatively different predictions for identical feature configurations, motivating the need for transparency mechanisms that explicitly account for model disagreement.

We proposed uncertainty-aware transparency as a third layer that leverages bootstrap resampling to represent predictions as distributions over trajectories and qualitative behaviours. Through an interactive visualisation, this approach enables practitioners to assess not only predicted outcomes but also the robustness of those predictions under epistemic uncertainty. Our findings suggest that transparency in time series forecasting should explicitly incorporate uncertainty to avoid overconfident explanations in high-stakes applications such as healthcare.

9. Code for Reproduction

The code to reproduce the results is available at: https://github.com/richardfranck/uncertainty-ts-forecasting

References

Barocas, S., Hardt, M., and Narayanan, A. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.

Efron, B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

Kacprzyk, K., Liu, T., and van der Schaar, M. Towards transparent time series forecasting. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*, 2024.

Kendall, A. and Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

Qin, J. Robust multi-modal forecasting: Integrating static and dynamic features. *arXiv preprint arXiv:2505.15083*, 2025.

Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

Salvi, M., Seoni, S., Campagner, A., Gertych, A., Acharya, U. R., Molinari, F., and Cabitza, F. Explainability and uncertainty: Two sides of the same coin for enhancing the interpretability of deep learning models in healthcare. *International Journal of Medical Informatics*, 197:105846, 2025.

Slack, D., Hilgard, A., Singh, S., and Lakkaraju, H. Reliable post hoc explanations: Modeling uncertainty in explainability. *Advances in neural information processing systems*, 34:9391–9404, 2021.

Sokol, K. and Hüllermeier, E. All you need for counterfactual explainability is principled and reliable estimate of aleatoric and epistemic uncertainty. *arXiv preprint arXiv:2502.17007*, 2025.

Wilkerson, J., Abdallah, K., Hugh-Jones, C., Curt, G., Rothenberg, M., Simantov, R., Murphy, M., Morrell, J., Beetsch, J., Sargent, D. J., et al. Estimation of tumour regression and growth rates during treatment in patients with advanced prostate cancer: a retrospective analysis. *The Lancet Oncology*, 18(1):143–154, 2017.

A. Appendix

A.1. Illustration of a Composition

Figure 5 illustrates $sin(x)$ as composition of motifs and their separating transition points.

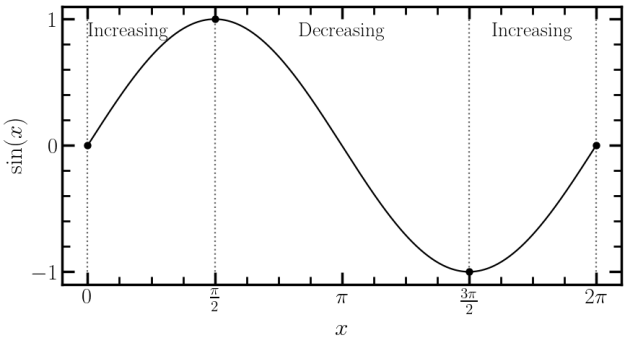

Figure 5. Illustration of $sin(x)$ as a composition.

A.2. Details on Synthetic Tumour Data Generation

We generate synthetic tumour growth trajectories using the model proposed by Wilkerson et al. (2017). For an individual with initial tumor volume v_0, tumor volume at time t is given by

$$y(t) \;=\; v_0\Big(\phi\exp(-dt) + (1-\phi)\exp(gt)\Big), \qquad (8)$$

where g controls tumor growth, d controls tumor decay due to treatment, and $\phi \in (0,1)$ is a mixing weight between the two exponentials.

Following Kacprzyk et al. (2024), the parameters in equation 8 are related to the static covariates age, weight and drug dosage as

$$\begin{aligned} g &= g_0\left(\frac{\text{age}}{20}\right)^{0.5}, \\ d &= d_0\frac{\text{dosage}}{\text{weight}}, \\ \phi &= \frac{1}{1+\exp(-\text{dosage}\,\phi_0)}, \end{aligned} \qquad (9)$$

where $(g_0, d_0, \phi_0) = (2.0, 180, 10)$.

To generate the dataset, we create 2000 samples. For each sample, static covariates are drawn independently from uniform distributions with ranges specified in Table 1 of the main text.

For each sample, tumour volume is evaluated at 20 equally spaced time points on the interval $[0, 1]$. No observational noise is added to trajectories, effectively setting aleatoric uncertainty equal to zero.