DECISION-AWARE TRAINING FOR SAMPLE-BASED GENERATIVE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Sample-based generative models are increasingly used for probabilistic forecasting in high-stakes decision settings, yet their training objectives are blind to the decision maker's cost structure. These models are commonly trained with strictly proper scoring rules, such as the energy score, which allocate their training signal in proportion to data density, with no awareness of where forecast errors are most costly for downstream decisions. We therefore propose *decision-aware training for sample-based generative models*, augmenting the energy score objective with a differentiable decision loss that directly penalises the cost incurred by acting on the model's forecast. This combined loss is theoretically grounded, as the decision loss is itself a proper scoring rule. We compute the decision loss via a differentiable optimisation layer. Its gradient concentrates in cost-sensitive regions of the output space, making the method's effects interpretable and predictable from the cost structure. We validate the method on one synthetic and two real-world tasks. In the synthetic task, the method corrects the mode weights of a learned bimodal distribution; in a wind power dispatch task, it concentrates improvements in the rare but costly tail region, and in a frost protection task, it improves how well the decision costs are anticipated. Our method yields generative models that retain full probabilistic forecasts while being better aligned with the decision maker's specific cost structure.

1 INTRODUCTION

Probabilistic forecasting models are increasingly deployed in high-stakes decision-making, for example, a wind farm operator committing to a power contract based on a wind forecast, risking costly shortfall penalties if they over-commit, or a farmer deciding whether to apply frost protection based on a temperature forecast, balancing asset loss against the cost of intervention. In the weather domain, state-of-the-art forecasting systems increasingly rely on sample-based generative models (Lang et al., 2024b; Bonev et al., 2025). These models are trained with strictly proper scoring rules such as the energy score (Gneiting and Raftery, 2007), which are uniquely minimised by the true conditional distribution. In practice, however, limited model capacity and empirical risk minimisation yield only local optimality (Blasiok et al., 2023), inducing tradeoffs between model error in different regions of the input/output space (Donti et al., 2017): proper scoring rules distribute the training gradient in proportion to data density, with no awareness of the decision maker's cost structure. The model's limited capacity is allocated globally, leaving decision-critical regions of the output space potentially underserved.

Given a forecast, a decision maker with cost function $c(a, y)$, of action a and outcome y, selects the action that minimises expected cost under the forecast distribution; a point forecast is insufficient to evaluate this expectation. A good forecast should yield low expected costs when acted upon and correctly anticipate those costs, the latter property being known as *decision calibration* (Zhao et al., 2021). Crucially, the observed cost of the optimal action is itself a proper scoring rule (Hartline et al., 2025; Kleinberg et al., 2023), placing it in the same family as the energy score which licenses their combination as a theoretically well-founded training objective for decision-aware training.

We augment the energy score with the optimal action cost as a differentiable decision loss. The decision loss alone is insufficient as a training objective: it only constrains the model in cost-sensitive regions, leaving the rest of the distribution unanchored and prone to degeneration. The energy score

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.

acts as that anchor, preventing the model from collapsing outside cost-sensitive regions. During training, samples drawn from the model are used to compute the optimal action a^* via a differentiable optimization layer; the decision loss then evaluates the cost incurred by a^* against the true outcome y, and gradients flow back through a^* to the model parameters via implicit differentiation. Our method is theoretically grounded and leads to better downstream decisions while retaining full probabilistic forecasts, as validated on synthetic and real-world forecasting tasks.

Contributions.

- An end-to-end *decision-aware* training framework for implicit generative models and distributional diffusion models, bridging decision-focused learning and probabilistic forecasting by retaining distributional accuracy via an energy score anchor.
- A theoretically principled combination of decision loss and energy score, exploiting the equivalence between optimal action costs and proper scoring rules.
- A gradient analysis showing which regions benefit from the decision loss and why, based on the cost function structure.
- Empirical validation of our method on synthetic and real-world forecasting tasks, considering both decision loss and decision calibration.

In the following, we start with summarising the necessary background in Sections 2 and 3 before introducing our method in Section 4 and validating it in Section 5.

2 Probabilistic Regression and Scoring Rules

We consider a probabilistic regression setting where input features $X \in \mathcal{X}$ and target $Y \in \mathcal{Y}$ are random variables distributed according to $P_{X,Y}$. Given a realisation x of X, the goal is to learn a model h_θ that produces a predictive distribution $\hat{F}_x = h_\theta(x)$ over Y which ideally recovers the true conditional $Y \mid X = x$. We drop the subscript x throughout for brevity.

For this work, we consider two classes of sample-based generative models: implicit generative models, which map noise directly to samples, and distributional diffusion models, which generate samples via an iterative denoising process. Both represent $\hat{F}_x$ implicitly via samples $\{\hat{y}_m\}_{m=1}^M$. *Implicit generative models* (Mohamed and Lakshminarayanan, 2016) map a conditioning input x and a noise vector $\varepsilon_m \sim \mathcal{N}(0, I)$ directly to a sample $\hat{y}_m \sim \hat{F}_x$ via $\hat{y}_m = f_\theta(x, \varepsilon_m)$; drawing M independent noise vectors in parallel yields M samples in a single forward pass. *Diffusion models* (Ho et al., 2020; Song et al., 2020) gradually add noise to the clean target y_0 and learn a reverse denoising process conditioned on x.[1] At each reverse step, a denoiser estimates the clean target y_0 from a noisy observation y_t and the conditioning input x; in the standard formulation this prediction is a point estimate. Since y_0 is uncertain given (y_t, x), we consider the special class of *distributional diffusion models* (De Bortoli et al., 2025; Kneissl et al., 2025), whose denoiser models a full distribution over y_0. At inference, samples from $\hat{F}$ are obtained by iterating the full reverse chain.

A scoring rule $S(\hat{F}, y)$ assigns a scalar score to a predictive distribution $\hat{F}$ given the realised outcome y. It is *proper* if the true conditional F minimises the expected score, $\mathbb{E}_{Y\sim F}[S(F, Y)] \leq \mathbb{E}_{Y\sim F}[S(\hat{F}, Y)]$ for all $\hat{F}$, and *strictly proper* if F is the unique minimiser. The *energy score* (ES) (Gneiting and Raftery, 2007) is a popular example of a strictly proper scoring rule that admits a sample-based estimator (see Section A.1). For univariate targets the energy score reduces to the *continuous ranked probability score* (CRPS), widely used in meteorological forecast verification (Gneiting et al., 2007) and as a training objective for state-of-the-art ML weather prediction models (Lang et al., 2024b; Price et al., 2023). Both model classes introduced above are commonly trained by minimising strictly proper scoring rules, and in particular the energy score.

[1] We use y_0, y_t for the clean and noisy target following our regression notation; the standard diffusion literature uses x_0, x_t.

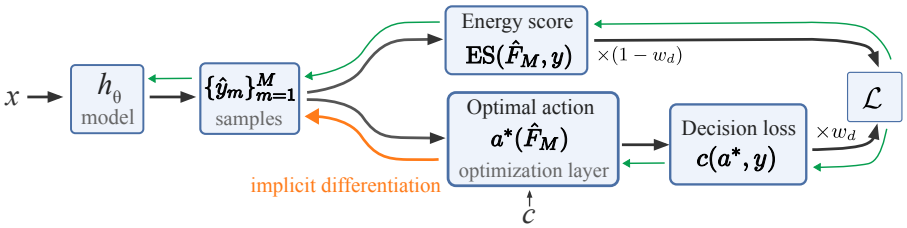


Figure 1: **Method overview.** In the forward pass the model h_θ produces samples $\{\hat{y}_m\}$ given input x. These feed two branches: the *Energy Score* (CRPS) computes the standard scoring rule loss given the observed target y; the *Optimization layer* solves Eq. (1) to obtain a^*, and the *decision loss* evaluates $c(a^*, y)$. Gradients flow back from energy score via standard autodiff (green arrows) and from decision loss through the optimization layer via implicit differentiation (orange arrow).

3 Decision-Making Under Uncertainty

Given a probabilistic forecast $\hat{F}$ over $\mathcal{Y}$, a decision maker selects an action $a \in \mathcal{A}$ to minimise expected cost under the forecast. This is modelled via a cost function $c : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$ that encodes the preferences of a particular decision maker; the Bayes-optimal action is

$$a^*(\hat{F}) \coloneqq \operatorname*{argmin}_{a \in \mathcal{A}} \mathbb{E}_{Y \sim \hat{F}}[c(a, Y)] \tag{1}$$

$$\approx \operatorname*{argmin}_{a \in \mathcal{A}} \frac{1}{M} \sum_{m=1}^{M} c(a, \hat{y}_m) = a^*(\hat{F}_M). \tag{2}$$

The expectation ensures that rare but sufficiently costly events still influence the optimal action. For sample-based models, the expectation is replaced by its Monte Carlo estimate over samples $\hat{y}_m \overset{\text{iid}}{\sim} \hat{F}$, $m = 1, ..., M$, and $\hat{F}_M$ denotes the empirical distribution[2] over $\{\hat{y}_m\}_{m=1}^M$.

The realised cost when acting on forecast $\hat{F}$ and observing the true outcome y is $c(a^*(\hat{F}), y)$. A forecast is *decision-calibrated* (Zhao et al., 2021) if its predicted expected cost matches the realised cost in expectation over inputs:

$$\mathbb{E}_X \, \mathbb{E}_{Y \sim \hat{F}_X} \big[c(a^*(\hat{F}_X), Y) \big] = \mathbb{E}_{(X,Y)} \big[c(a^*(\hat{F}_X), Y) \big]. \tag{3}$$

Intuitively, that means a decision-calibrated model correctly anticipates the cost of its own optimal action. We measure decision miscalibration as the mean absolute difference between predicted cost and realised cost across test points $\frac{1}{N} \sum_{i=1}^{N} \Big| \mathbb{E}_{Y \sim \hat{F}_{x_i}} \big[c(a^*(\hat{F}_{x_i}), Y) \big] - c(a^*(\hat{F}_{x_i}), y_i) \Big|$.

In principle, minimising a strictly proper scoring rule leads to optimal decisions: F is the unique minimizer of the expected score (Gneiting and Raftery, 2007), and the true distribution yields optimal decisions by construction.[3] In practice, however, limited data, model capacity, and stochastic optimisation yield only approximate local minima (Blasiok et al., 2023), with no awareness of where c is sensitive (Donti et al., 2017), so a model minimising the energy score may be decision-miscalibrated and incur high decision cost.

4 Decision-Aware Training

We address the gap between scoring-rule training and downstream decision quality by augmenting the scoring rule objective with a differentiable decision loss that directly penalises suboptimal decisions (Figure 1). We first establish the theoretical grounding and introduce the combined objective (Section 4.1), then analyse the gradient chain through the optimal action (Section 4.2), and describe the model architectures and training procedure (Section 4.3).

[2] $\hat{F}_M \coloneqq \frac{1}{M} \sum_{m=1}^{M} \delta_{\hat{y}_m}$.

[3] In fact, distribution calibration (Song et al., 2019) would suffice for optimal decisions, but even this weaker condition is not attainable nor verifiable in practice, motivating a focus on decision calibration directly (Derr et al., 2025; Sahoo et al., 2021).

4.1 Loss formulation and theoretical grounding

The decision problem of Section 3 suggests a natural training signal: penalise the model's optimal action by the cost it incurs at the realised outcome. We define the *decision loss*

$$S_c(\hat{F}, y) = c\Big(a^*(\hat{F}),\, y\Big)\,, \tag{4}$$

where $a^*(\hat{F})$ is the Bayes-optimal action in Eq. (1). The key observation is that S_c is itself a proper scoring rule.

Proposition 1 (Gneiting and Raftery 2007; Hartline et al. 2025)**.** *The decision loss S_c is a proper scoring rule: $\mathbb{E}_{Y\sim F}[S_c(F,Y)] \le \mathbb{E}_{Y\sim F}[S_c(\hat{F},Y)]$ for all forecasts $\hat{F}$, which follows from the definition of $a^*(F)$ as the Bayes action under F. However, it is not strictly proper: there can be $\hat{F} \neq F$ with $a^*(\hat{F}) = a^*(F)$ that achieve the same minimum.*

This means S_c is part of the same loss family as the energy score.[4] However, since the minimum is not unique, optimising S_c alone admits degenerate solutions. The energy score provides the missing uniqueness: as a strictly proper scoring rule, it admits F as its unique minimiser, and a positive linear combination of a strictly proper and a proper scoring rule remains strictly proper (Lemma 1, Section A.2). This result directly applies to the training of sample-based generative models, offering a natural extension to the existing scoring rule training objective.

Combined loss. Hence, we define our loss as convex combination of energy score and decision loss

$$\mathcal{L} = (1 - w_d)\,\mathrm{ES} + w_d\, S_c\,, \qquad w_d \in [0, 1), \tag{5}$$

with a unique minimum at $\hat{F} = F$. The decision weight w_d trades off distributional accuracy against decision awareness.[5] The energy score serves as an anchor, eliminating degenerate solutions admitted by S_c alone, while S_c concentrates the gradient on the cost-relevant region. This combination yields a theoretically grounded objective that is both decision-aware and distribution preserving.

4.2 Optimizing the decision loss

Training with the decision loss S_c requires computing a^* from samples via Eq. (2); the backward pass must then differentiate a^* with respect to those samples. For general cost functions, a^* has no closed-form expression in the samples $\{\hat{y}_m\}$, so we cannot differentiate it with respect to $\{\hat{y}_m\}$ directly; instead we differentiate the optimality conditions that define a^* treating it as an implicit function of $\{\hat{y}_m\}$ (implicit differentiation). We compute a^* within a *differentiable optimization layer* (Amos and Kolter, 2017; Donti et al., 2017; Blondel et al., 2022) over the feasible set $\mathcal{A} = [a_{\min}, a_{\max}]$, treating model samples as fixed. This requires c to be differentiable in a and the expected cost in Eq. (1) to have positive curvature at a^*, allowing implicit differentiation through a^*.

The decision loss $S_c = c(a^*, y_{\mathrm{obs}})$ depends on model parameters only through the samples $\{\hat{y}_m\}$ and hence through a^*. The gradient with respect to each sample is

$$\frac{\partial S_c}{\partial \hat{y}_m} = \left.\frac{\partial c}{\partial a}\right|_{a^*} \cdot \frac{\partial a^*}{\partial \hat{y}_m}, \tag{6}$$

where $\partial a^*/\partial \hat{y}_m$ is obtained via implicit differentiation, whereas the remaining upstream gradient with respect to model parameters is computed via standard automatic differentiation.

Per-sample gradient. When $a^* \in (a_{\min}, a_{\max})$, implicit differentiation gives

$$\frac{\partial a^*}{\partial \hat{y}_m} = -\left[\frac{1}{M}\sum_{j=1}^{M} \frac{\partial^2 c}{\partial a^2}(a^*, \hat{y}_j)\right]^{-1} \frac{1}{M}\,\frac{\partial^2 c}{\partial a\,\partial \hat{y}_m}(a^*, \hat{y}_m). \tag{7}$$

[4] In fact, Hartline et al. (2025) build on Kleinberg et al. (2023) to establish the stronger result that the space of decision losses induced by bounded cost functions is *equivalent* to the space of proper scoring rules: every proper scoring rule arises as the optimal action cost of some cost function. Our argument only requires the direction stated in Proposition 1 (S_c is proper).

[5] We discuss heuristics for choosing w_d in practice in Appendix E.

The gradient signal is concentrated on samples $\hat{y}_m$ where the cost has non-zero cross-curvature, $\partial^2 c/\partial a \partial \hat{y}_m \neq 0$, i.e., where small perturbations of $\hat{y}_m$ change how sensitive the cost is to a; samples in flat or saturated cost regions contribute almost nothing (full derivation in Section A.3). In sample-based models, the noise input ε enables flexible distributional shapes, and gradients can update specific parts of the predictive distribution. However, this flexibility can also lead to distributional degeneration at large w_d in practice.

Constraint boundary blocking. Eq. (7) applies when a^* is interior, where the first-order condition $\nabla_a \frac{1}{M} \sum_{m=1}^{M} c(a^*, \hat{y}_m) = 0$ holds. When a^* reaches a constraint boundary ($a^* \in \{a_{\min}, a_{\max}\}$), Eq. (7) does not apply. In this case the gradient is simply zero: geometrically, small perturbations of the samples do not move a^* off the boundary, so $\partial a^*/\partial \hat{y}_m = 0$ (formal derivation via KKT conditions in Section A.3), blocking the decision loss gradient chain in Eq. (6) entirely.[6] The energy score is then the only active training signal.

4.3 Model architectures and training procedure

For this work, we consider two sample-based generative architectures representative of those used in operational forecasting systems (eg., Lang et al., 2024a; Couairon et al., 2024). We extend both architectures with a differentiable optimization layer that computes a^* from the model samples.

Implicit generative model. An MLP maps input x and M independent noise draws to M samples in a single forward pass (details in Section B).

Distributional diffusion model. An MLP denoiser produces M samples of y_0 at each denoising step. The decision loss acts at the denoiser level, while the predictive distribution $\hat{F}_x$ at inference is shaped by the full reverse chain, an additional sampling layer absent in the implicit generative model.

Training procedure. At each training step, the model is trained by optimising the combined loss (Eq. (5)). In practice, both loss terms are evaluated from a finite set of samples $\{\hat{y}_m\}_{m=1}^{M}$ produced by the model for each input x in the batch (Figure 1). First, the energy score is evaluated using the unbiased fair estimator (see Appendix Section A.1), with gradients computed via standard autodiff. Second, a^* is computed by minimising the expected cost over the same samples (Eq. (2)) via the optimisation layer forward pass, treating the model samples as fixed. The decision loss S_c is then evaluated and its gradient flows back through a^* via implicit differentiation (Section 4.2). Computing a^* from M samples is the only source of a $O(1/M)$ finite-sample bias in the objective that vanishes as $M \to \infty$; we show that the objective nevertheless remains approximately strictly proper and that the bias is small in practice (Section A.4). Both losses are normalised to unit scale before training, so that w_d directly controls the relative weighting between the two terms independently of the cost function's magnitude; further implementation details are given in Section B.

5 Experiments

In the following we evaluate our method on one synthetic and two real-world decision tasks.

5.1 Synthetic Decision Task

Setup. To study the mechanism underlying decision-aware training in a controlled setting where the ground-truth conditional distribution is known, we construct a synthetic dataset with the following structure. The target $Y \mid X$ follows a bimodal Gaussian mixture distribution with fixed mode positions ($\mu_1 = 0$, $\mu_2 = 2.5$), fixed standard deviations ($\sigma_1 = 0.2$, $\sigma_2 = 0.4$), but a sinusoidally varying mixture weight $w(x)$, with mode 2 (the right mode at $y = 2.5$ in Figure 2 left) having lower marginal weight 0.4. The decision maker must choose a protection level $a \in [0, 1]$ against a hazard concentrated above a threshold at $y = 1.0$ (between the two modes): higher a reduces exposure to the hazard but incurs a growing protection cost. The threshold is set so that the mixture weight $w(x)$ is the key determinant of the optimal action and mode-weight tracking, thus the key diagnostic. We train a distributional diffusion model on this task across multiple data and training seeds; details are in Section C.1. We also use this synthetic setup to investigate the effect of training data size on the benefit of decision-aware training in Appendix Section D.

[6] The measure-zero case where the unconstrained minimiser coincides exactly with the boundary is excluded.

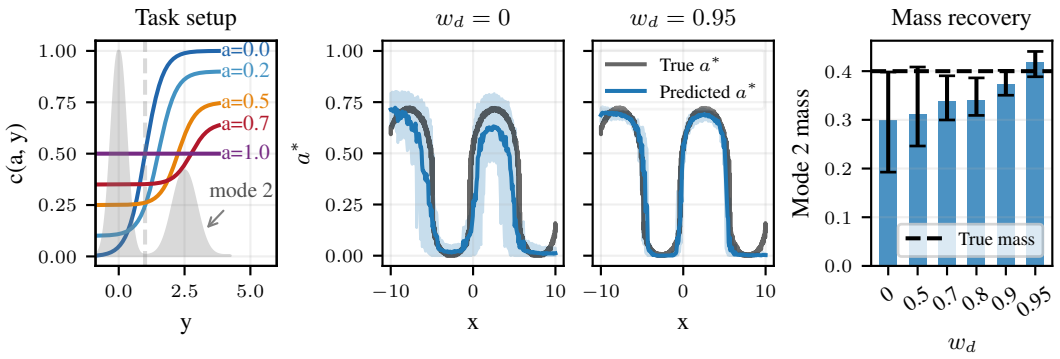


Figure 2: **Synthetic decision task.** *Left:* Cost function $c(a, y)$ for five protection levels a, overlaid with the marginal $p(y)$ (grey). The threshold at $y = 1.0$ separates the two modes; mode positions are fixed across x, only the mixture weight varies. *Center (two panels):* Predicted $a^*(x)$ for $w_d = 0$ and $w_d = 0.95$. Pure energy score ($w_d = 0$, left) fails to track the ground truth at the transitions; decision-aware training (ours) recovers the correct shape ($w_d = 0.95$, right). *Right:* Marginal mode 2 mass vs. w_d; dashed line marks the true mass (0.4). Average mass approaches the true value for larger decision weights w_d. Statistics show mean and 10th–90th percentiles across seeds.

Results. With this setup in place, we first ask whether a standard proper scoring rule suffices. Under pure energy score training ($w_d = 0$), the diffusion model roughly captures the extreme values of $a^*(x)$ but fails to reproduce the sinusoidal transitions of a^* between extremes, reflecting inaccurate mode weight tracking across x (Figure 2). Inspecting the mode weights of the marginal predictive distributions shows that the costly mode (mode 2 in Figure 2) is systematically underweighted when trained with energy score alone ($w_d = 0$).

Adding the decision loss directly addresses this: $a^*(x)$ tracks the ground truth more closely as w_d grows (Figure 2). The marginal mode weights are progressively corrected as w_d increases, with probability mass shifting toward mode 2, improving decision skill. At the extreme $w_d = 1$ (no energy score anchor), a^* tracking survives (Figure C.7) but the predictive distribution degenerates, confirming empirically that the energy score is necessary as a distributional anchor (Figure C.8). Results for all data seeds and aggregate metrics are provided in Section C.1.

5.2 Wind Power Dispatch

Setup. Wind power output is inherently uncertain, thus wind power operators must commit to a power production level (dispatch) hours before realised output is known; over-commitment incurs shortfall penalties while under-commitment forgoes revenue (Bruninx et al., 2025; Botterud et al., 2011; Bourry et al., 2008). The power curve P maps wind speeds to power output and introduces a non-trivial complication: at high wind speeds, turbines are shut down for safety (cut-off event), so production drops from peak output to zero; this cut-off regime is the most consequential for the dispatch decision since a missed cut-off event incurs the largest shortfall penalty. Accurate probabilistic modelling of this regime is therefore crucial for optimal decision-making. We study wind power dispatch at an offshore wind farm (North Sea, Norway), where the decision maker must commit to a normalized power dispatch level $a \in [0, 1]$ based on a forecast. We model this task via a cost function with quadratic shortfall penalty (Bruninx et al., 2025)

$$c(a, y) = -a + \lambda \operatorname{relu}(a - P(y))^2, \tag{8}$$

where $-a$ reflects the revenue from committing to dispatch level a (higher commitment earns more), $P(\cdot)$ is a differentiable power curve (Figure C.10), and λ controls the shortfall penalty. The modelled power curve has three regions: a *ramp* region where output rises with wind speed (3–15 m/s at turbine hub height), a *rated* region of flat maximum output (15–22 m/s), and a *cut-off* region where output drops sharply to zero above ≈22 m/s (≈5% of observations). We forecast wind speed (at 10 m) 6 h ahead from ERA5 reanalysis data (Hersbach et al., 2020), using lagged wind speed observations as inputs. Forecast samples are extrapolated to turbine hub height speed, passed through $P(\cdot)$, and

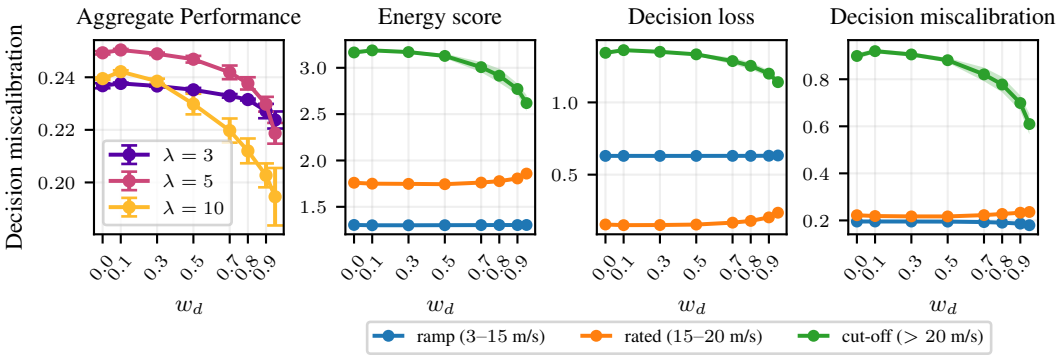


Figure 3: **Wind power dispatch results** (diffusion model). *Left:* aggregate decision miscalibration vs w_d for all λ (mean $\pm$ 1 std across seeds). Increasing w_d improves decision miscalibration. This effect scales with shortfall penalty λ. *Right (three panels):* conditional metrics by power-curve region of the observed wind speed vs. w_d (mean $\pm$ 1 std across seeds) for $\lambda = 5$. Improvements concentrate in the cut-off region; the rated region degrades for large w_d; the ramp region is mostly unaffected.

a^* is solved via the optimisation layer; the observed cost is then computed using the true observed power (see Section C.2.1 for details). We train a distributional diffusion model and an implicit generative model across $w_d \in \{0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95\}$ with three seeds (training details in Section C.2.1). We sweep $\lambda \in \{3, 5, 10\}$ and report $\lambda = 5$ as the primary setting.

Results. At the aggregate level, decision-aware training improves decision miscalibration monotonically for $w_d > 0.5$ in the diffusion model (Figure 3, left), with larger λ yielding greater improvements. The energy score also improves with w_d for $\lambda \leq 5$, but degrades at large w_d for $\lambda = 10$, while aggregate decision loss remains relatively flat throughout (Figure C.11). The implicit generative model shows behaves similarly (Figure C.15), except for the energy score, which degrades monotonically for all λ, indicating a clearer tradeoff between distributional accuracy and decision quality.

These aggregate metrics, however, obscure where the decision loss acts. Disaggregating performance by power-curve region of the observed wind speed (Figure 3) shows that the effect concentrates in the cut-off region[7], where forecast errors are most costly. At $w_d = 0.9$, the energy score improves by 12%, decision loss by 11%, and decision calibration by 22% on average compared to $w_d = 0$. This gain comes at the cost of a slight monotonic degradation in the rated region, with no effect in the ramp region. For the implicit generative model, the regional picture is less clean. The baseline energy score model is already well calibrated in the cut-off region, leaving little room for the decision loss to exploit; only $\lambda = 10$ yields a consistent regional improvement (see Figure C.16).

Two further observations are notable for the correction mechanism. First, the cut-off improvements in the diffusion model scale monotonically with λ (Figure C.13), as a higher penalty produces stronger corrections in the tail. However, at very high w_d, the energy score degrades again, suggesting the decision loss overshoots the optimal distributional correction. Second, the correction is directly visible as a median shift $\Delta Q_{0.5}(v)$ in the predictive distributions of both models (Figures C.14 and C.18). Corrections are negative in the ramp region, corresponding to more conservative dispatch, and positive beyond the cut-off threshold, placing more mass in the tail where power drops to zero.

Mechanism. The regional pattern of improvements and degradations follows from how the decision-loss gradient propagates through the power curve. In the *ramp* region ($\partial P/\partial y > 0$, a^* interior), gradients propagate and produce a shift to more conservative values in the predictive distribution, but aggregate metrics remain unaffected. In the *rated* region, $\partial P/\partial y \approx 0$ suppresses the decision loss gradient, leaving the energy score at reduced weight $(1 - w_d)$, which explains the degradation. In the *cut-off* region, $\partial P/\partial y \ll 0$ and a^* is interior; the gradient is strong, providing explicit supervision on the rare, high-cost tail events that energy score under-weights due to their low frequency.

[7]We use 20 m/s as the analysis threshold rather than the physical cut-off speed of 22 m/s to include events near the cut-off threshold where power already falls sharply.

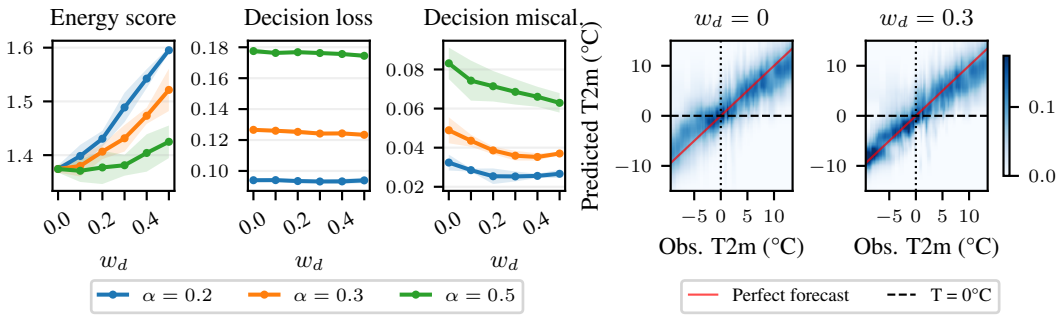


Figure 4: **Frost protection results** (implicit generative model). *Left (three panels):* Aggregate energy score, decision loss, and decision miscalibration vs. w_d for $\alpha \in \{0.2, 0.3, 0.5\}$ (mean $\pm$ std over seeds). Aggregate decision loss remains mostly flat with small improvement for $\alpha \geq 0.3$; energy score degrades with w_d; decision miscalibration improves most visibly for $\alpha = 0.5$. *Right (two panels):* Conditional predictive density for $w_d = 0$ and $w_d = 0.3$ ($\alpha = 0.3$) conditioned on observed temperature. A warm bias present at $w_d = 0$ ($T < 0$ °C) is corrected at $w_d = 0.3$, with the predictive distribution concentrating closer to the diagonal.

5.3 Frost Protection

Setup. Frost is among the most economically damaging meteorological hazards (Juurakko et al., 2021; White and Haas, 1975). Decision makers must commit to protection measures before observing temperatures, but protection comes at a cost and should be applied only when required. We study this frost protection task in Berlin using ERA5 reanalysis data (Hersbach et al., 2020), forecasting temperature (at 2 m, T2m) 24 h ahead from lagged ERA5 temperature observations over a spatial patch, restricted to winter months. The decision maker chooses a frost protection level $a \in [0, 1]$ before observing the realised temperature y. The cost function has the same sigmoid threshold structure as in Section 5.1, but mirrored: the hazard lies below the threshold. It penalises both unprotected frost (false negatives, FN) and unnecessary protection (false positives, FP), with α setting the FP/FN cost ratio; lower α penalises unprotected frost more strongly. Unlike the rare cut-off event in Section 5.2, frost occurs in about 25% of observations, making this a common event, where the energy score is less likely to under-weight the relevant area. We train an implicit generative model and a distributional diffusion model across $w_d \in \{0.0, 0.1, \ldots, 0.9\}$ and $\alpha \in \{0.2, 0.3, 0.5\}$ with three seeds (see details in Section C.3.1); diffusion model results are in Section C.3.3.

Results. At the aggregate level, the energy score degrades with w_d, more steeply for lower α, while aggregate decision loss is mostly flat (Figure 4). For $w_d > 0.5$, some forecasts exhibit spread collapse around the freezing point, which we attribute to excessive decision loss correction. Decision miscalibration nevertheless improves for all α, most visibly for $\alpha = 0.5$ with an improvement of approximately 15% at $w_d = 0.3$–0.5. Consistent with this, forecasts of sub-zero temperatures align more closely with observations when trained with the decision loss (Figure 4, right).

The combination of improved decision calibration and nearly flat aggregate cost is explained by the higher base rate of frost. Because frost events are common, the energy score already allocates substantial gradient signal to the threshold region and likely yields a near-optimal a^* without the decision loss. The decision loss still sharpens cost estimation, as the calibration gains show, but its impact on aggregate cost is marginal. A conditional breakdown (Figure C.24) shows that the no-frost regime drives the improvement in decision calibration, while the frost regime degrades slightly.

Mechanism. Due to the sharp sigmoid cost function, the decision loss gradient is concentrated near the 0 °C threshold. Therefore, quantiles closer to the threshold (cost-sensitive region) shift more than those further away, as shown by the quantile shift plot in Figure C.25, aligning with the gradient analysis in Section 4.2. Unlike in the wind power cut-off region, many forecasts here straddle the threshold, making the decision loss gradient signal dense and its effects visible at $w_d = 0.1$–0.2, whereas wind power cut-off improvements require $w_d \approx 0.5$–0.7 because cut-off events are rare.

6 Related Work

The relevant prior work spans statistics, decision theory, and machine learning. These strands have not previously been connected; we combine them to enable decision-aware training.

Decision-focused learning. Decision-focused learning integrates the downstream decision task into the training objective, most commonly in predict-then-optimise pipelines where a point prediction is followed by a combinatorial or linear programme. Elmachtoub and Grigas (2022) and Wilder et al. (2019) replace the prediction loss with a surrogate of the task loss. Donti et al. (2017) embeds a differentiable optimisation layer and trains end-to-end. All target point predictions rather than distributions, and evaluate only task loss; we additionally measure decision calibration (Zhao et al., 2021), which assesses whether the model correctly prices the cost of its own decisions. Our setting differs fundamentally: we work with probabilistic generative models that must produce forecast distributions, and the decision maker evaluates expected cost under the full distribution. To our knowledge, no existing decision-focused learning method operates in the sample-based setting.

Weighted proper scoring rules. Weighted proper scoring rules (Gneiting and Ranjan, 2011; Lerch et al., 2017) emphasize specific regions of the output space via a fixed, pre-specified weight function $w(y)$ that depends only on the observation y. In a decision problem, however, the cost function $c(a, y)$ depends jointly on the action and the observation. As a result, the cost-relevant region depends on $a^*(\hat{F}_x)$, and thus on the forecast itself which varies across inputs: low-wind and high-wind forecasts require precision in entirely different parts of the output space. No static reweighting can represent this input-dependent shift. Our method requires no pre-specified weight: the decision loss gradient concentrates automatically on the cost-relevant region, which changes for different a^* and therefore adapts to the predicted distribution. Static reweighting approaches therefore improve calibration in a fixed region, while our method identifies which region matters for the decision at hand, and calibrates there. Thus, they address a different problem and are not comparable as baselines.

Differentiable optimisation. Differentiable optimisation layers (Amos and Kolter, 2017; Donti et al., 2017; Blondel et al., 2022) enable gradients to flow through the solution of an optimisation problem embedded in a neural network layer. We use this machinery to differentiate through the optimal action a^*, connecting the decision loss to model parameters via implicit differentiation.

7 Conclusion

Decision-aware training for generative models is theoretically grounded and empirically effective, but its impact is selective. Whether a region of the output space receives a corrective signal depends on the optimal action: interior optima steer approximation errors toward less costly regions; boundary optima leave the original scoring-rule objective in full control. We demonstrated the mechanism on synthetic data, where decision-aware training corrects mode weight misattribution and improves downstream decisions; on wind power dispatch, where it concentrates improvements in the rare but costly cut-off region; and on frost protection, where it redistributes probability mass around the freezing threshold to improve decision calibration. The benefit is most pronounced when data or capacity limits leave energy score with suboptimal approximation errors in cost-relevant regions, a regime characteristic of operational forecasting (Section D).

A practical limitation is that the optimal decision weight w_d is task-dependent and requires tuning: at moderate w_d, a sweet spot typically exists where decision loss improves with little degradation in distributional accuracy, but the right balance depends on the cost function and data regime. The method also requires a dedicated model per decision-maker, since the cost function is embedded into the training objective. The optimisation layer adds computational overhead at each training step; further details are provided in Section B. For distributional diffusion models, our gradient analysis (Section 4.2) characterises the decision loss training signal per denoising step, but does not account for inference-time reverse chain dynamics; a score-function formulation for the distributional denoiser could clarify when and why the effect carries through, and remains an open theoretical question.

Extending decision-aware training to large-scale probabilistic weather prediction models (Lang et al., 2024b; Bonev et al., 2025) is an interesting direction for future work. These systems share the same sample-based scoring rule training setup and are deployed in operational contexts where the gap between forecast skill and decision skill is operationally consequential.

REPRODUCIBILITY STATEMENT

All details required to reproduce our results, including code availability, data sources, model architectures, hyperparameters, and training procedures, are provided in Section B.

ETHICS STATEMENT

Decision-aware training produces forecasts that are more actionable for specific downstream tasks, with direct applications in safety-critical domains such as weather hazard prediction, where better forecasts can directly reduce societal harm. A potential risk is that if the cost function encodes biased priorities, the model will optimise for them, potentially reinforcing skewed downstream decisions.

AI USE STATEMENT

We used generative AI tools (Claude, Anthropic, including the Claude Code coding assistant) in the following ways. *Theory:* for the analysis in Section A.4, to identify key ingredients of the argument and to draft the proof; for the gradient derivation in Section 4.2, to refine and verify it. *Research:* to refine hypotheses about the mechanisms behind our results, to provide feedback on the design of ablation experiments, and to interpret results. *Data and code:* to refine and extend the training and evaluation code and to build the ERA5 preprocessing pipeline; the core method and training code were implemented by the authors. *Writing:* to create and modify figures, draft and edit parts of the paper for readability, and identify relevant literature.

We did not use generative AI tools to generate the synthetic data set or to formulate the proposed method and its main claims; translation and qualitative data analysis are not applicable to this work.

We have reviewed all AI-assisted work: all derivations were worked through and checked by the authors, AI-assisted code was verified, and all references were checked against the original sources.

REFERENCES

M. Abbes and J. Belhadj. Wind resource estimation and wind park design in el-kef region, tunisia. *Energy*, 40(1):348–357, 2012.

B. Amos and J. Z. Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *International conference on machine learning*, pages 136–145. PMLR, 2017.

J. Blasiok, P. Gopalan, L. Hu, and P. Nakkiran. When does optimizing a proper loss yield calibration? In *Advances in Neural Information Processing Systems*, volume 36, pages 72071–72095, 2023.

M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert. Efficient and modular implicit differentiation. In *Advances in neural information processing systems*, volume 35, pages 5230–5242, 2022.

B. Bonev, T. Kurth, A. Mahesh, M. Bisson, J. Kossaifi, K. Kashinath, A. Anandkumar, W. D. Collins, M. S. Pritchard, and A. Keller. Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale. *arXiv preprint arXiv:2507.12144*, 2025.

A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Sumaili, and V. Miranda. Wind power trading under uncertainty in lmp markets. *IEEE Transactions on power systems*, 27(2):894–903, 2011.

F. Bourry, J. Juban, L. Costa, and G. Kariniotakis. Advanced strategies for wind power trading in short-term electricity markets. In *European Wind Energy Conference & Exhibition EWEC 2008*, pages 8–pages. EWEC, 2008.

M. Bruninx, T. Verstraeten, J. Kazempour, and J. Helsen. Day-ahead bidding strategies for wind farm operators under a one-price balancing scheme. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, pages 719–726, 2025.

G. Couairon, R. Singh, A. Charantonis, C. Lessig, and C. Monteleoni. Archesweather & archesweathergen: a deterministic and generative model for efficient ml weather forecasting. *arXiv preprint arXiv:2412.12971*, 2024.

V. De Bortoli, A. Galashov, J. S. Guntupatti, G. Zhou, K. Murphy, A. Gretton, and A. Doucet. Distributional diffusion models with scoring rules. *arXiv preprint arXiv:2502.02483*, 2025.

R. Derr, J. Finocchiaro, and R. C. Williamson. Three types of calibration with properties and their semantic and formal relationships. *arXiv preprint arXiv:2504.18395*, 2025.

P. Donti, B. Amos, and J. Z. Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in neural information processing systems*, volume 30, 2017.

A. N. Elmachtoub and P. Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022.

C. A. T. Ferro. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923, 2014. doi: 10.1002/qj.2270.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

T. Gneiting and R. Ranjan. Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422, 2011.

T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, 2007.

J. Hartline, Y. Wu, and Y. Yang. Smooth calibration and decision making. *arXiv preprint arXiv:2504.15582*, 2025.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in neural information processing systems*, volume 33, pages 6840–6851, 2020.

C. L. Juurakko, V. K. Walker, et al. Cold acclimation and prospects for cold-resilient crops. *Plant Stress*, 2:100028, 2021.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

B. Kleinberg, R. P. Leme, J. Schneider, and Y. Teng. U-calibration: Forecasting for an unknown agent. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5143–5145. PMLR, 2023.

C. Kneissl, C. Bulte, P. Scholl, and G. Kutyniok. Improved probabilistic regression using diffusion models. *arXiv preprint arXiv:2510.04583*, 2025.

S. G. Krantz and H. R. Parks. *The implicit function theorem: history, theory, and applications*, volume 202. Springer, 2002.

S. Lang, M. Alexe, M. Chantry, J. Dramsch, F. Pinault, B. Raoult, M. C. Clare, C. Lessig, M. Maier-Gerber, L. Magnusson, et al. Aifs–ecmwf's data-driven forecasting system. *arXiv preprint arXiv:2406.01465*, 2024a.

S. Lang, M. Alexe, M. C. Clare, C. Roberts, R. Adewoyin, Z. B. Bouallègue, M. Chantry, J. Dramsch, P. D. Dueben, S. Hahner, et al. Aifs-crps: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. *arXiv preprint arXiv:2412.15832*, 2024b.

S. Lerch, T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting. Forecaster's dilemma: extreme events and forecast evaluation. *Statistical Science*, pages 106–127, 2017.

S. Mohamed and B. Lakshminarayanan. Learning in implicit generative models. *arXiv preprint arXiv:1610.03483*, 2016.

I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, et al. Gencast: Diffusion-based ensemble forecasting for medium-range weather. *arXiv preprint arXiv:2312.15796*, 2023.

P. Rilstone, V. K. Srivastava, and A. Ullah. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics*, 75(2):369–395, 1996.

R. Sahoo, S. Zhao, A. Chen, and S. Ermon. Reliable decisions with threshold calibration. In *Advances in Neural Information Processing Systems*, volume 34, pages 1831–1844, 2021.

H. Song, T. Diethe, M. Kull, and P. Flach. Distribution calibration for regression. In *International Conference on Machine Learning*, pages 5897–5906. PMLR, 2019.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

S. Thrun and A. Schwartz. Issues in using function approximation for reinforcement learning. In *Connectionist Models Summer School*, Hillsdale, 1993.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

H. van Hasselt. Double Q-learning. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

Y. Wang, Q. Hu, L. Li, A. M. Foley, and D. Srinivasan. Approaches to wind power curve modeling: A review and discussion. *Renewable and Sustainable Energy Reviews*, 116:109422, 2019.

G. F. White and J. E. Haas. *Assessment of Research on Natural Hazards*. MIT Press, Cambridge, MA, 1975.

B. Wilder, B. Dilkina, and M. Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 1658–1665, 2019.

S. Zhao, M. Kim, R. Sahoo, T. Ma, and S. Ermon. Calibrating predictions to decisions: A novel approach to multi-class calibration. In *Advances in Neural Information Processing Systems*, volume 34, pages 22313–22324, 2021.

A THEORY

This section provides a definition of the energy score, a proof of Lemma 1 and a detailed derivation of the decision loss gradient in Eq. (7).

A.1 ENERGY SCORE

The energy score (Gneiting and Raftery, 2007) evaluates a predictive distribution $\hat{F}$ against a realised outcome y via

$$\mathrm{ES}(\hat{F}, y) = \mathbb{E}_{Y\sim\hat{F}}\|Y - y\| - \frac{1}{2}\mathbb{E}_{Y,Y'\sim\hat{F}}\|Y - Y'\| \tag{9}$$

For sample-based models, both expectations are approximated by Monte Carlo estimates over the same M samples $\{\hat{y}_m\}_{m=1}^{M}$ forming the empirical distribution $\hat{F}_M$. We use the fair (unbiased) estimator (Ferro, 2014),

$$\mathrm{ES}_{\mathrm{fair}}(\hat{F}_M, y) = \frac{1}{M}\sum_{m=1}^{M}\|\hat{y}_m - y\| - \frac{1}{2M(M-1)}\sum_{m=1}^{M}\sum_{j=1}^{M}\|\hat{y}_m - \hat{y}_j\|. \tag{10}$$

For univariate targets the energy score reduces to the continuous ranked probability score (CRPS).

A.2 PROOF OF LEMMA 1

Lemma 1. *Let S_1 be a strictly proper scoring rule and S_2 be a proper scoring rule. Then $S = (1 - w_d)\, S_1 + w_d\, S_2$ is strictly proper for any $w_d \in [0, 1)$.*

Proof. The result follows immediately from the definitions. Let F be the true distribution and $\hat{F} \neq F$ any forecast. Then

$$\begin{aligned}\mathbb{E}_{Y\sim F}[S(\hat{F}, Y)] &= (1 - w_d)\,\mathbb{E}[S_1(\hat{F}, Y)] + w_d\,\mathbb{E}[S_2(\hat{F}, Y)] \\ &> (1 - w_d)\,\mathbb{E}[S_1(F, Y)] + w_d\,\mathbb{E}[S_2(F, Y)] \\ &= \mathbb{E}_{Y\sim F}[S(F, Y)],\end{aligned}$$

where the strict inequality uses $\mathbb{E}[S_1(\hat{F}, Y)] > \mathbb{E}[S_1(F, Y)]$ (strict properness of S_1) with weight $(1 - w_d) > 0$, and $\mathbb{E}[S_2(\hat{F}, Y)] \geq \mathbb{E}[S_2(F, Y)]$ (properness of S_2) with weight $w_d \geq 0$. □

A.3 DERIVATION OF PER-SAMPLE GRADIENT (EQ. (7))

Problem setup. Define $f(a; \{\hat{y}_j\}_{j=1}^{M}) := \frac{1}{M}\sum_{j=1}^{M} c(a, \hat{y}_j)$; we write $f(a)$ for brevity, making the sample dependence explicit only where the argument requires it. The optimal action solves

$$\min_{a}\ f(a) \quad \text{s.t.} \quad a \geq a_{\min}, \quad a \leq a_{\max}. \tag{11}$$

Since the feasible action set $[a_{\min}, a_{\max}]$ satisfies Slater's condition, the KKT conditions are necessary at the optimal action a^*. Rewriting the constraints in standard form $g_1(a) = a_{\min} - a \leq 0$, $g_2(a) = a - a_{\max} \leq 0$, the Lagrangian is

$$\mathcal{L}(a, \mu_1, \mu_2) = f(a) + \mu_1(a_{\min} - a) + \mu_2(a - a_{\max}), \tag{12}$$

with KKT conditions at a^*:

$$\begin{aligned}&\text{(stationarity)} && f'(a^*) - \mu_1 + \mu_2 = 0, && (13)\\ &\text{(dual feasibility)} && \mu_1,\, \mu_2 \geq 0, && (14)\\ &\text{(complementary slackness)} && \mu_1(a^* - a_{\min}) = 0, \quad \mu_2(a_{\max} - a^*) = 0, && (15)\end{aligned}$$

where $f'(a^*) := \partial f / \partial a\,(a^*)$.

Interior solution. When a^* is interior, both constraints are inactive, so complementary slackness forces $\mu_1 = \mu_2 = 0$ and all KKT conditions reduce to the first-order condition:

$$f'(a^*) = \frac{1}{M}\sum_{j=1}^{M} \frac{\partial c}{\partial a}(a^*, \hat{y}_j) = 0. \tag{16}$$

To obtain $\partial a^*/\partial \hat{y}_m$ we differentiate the optimality condition Eq. (16) implicitly with respect to $\hat{y}_m$.

> *Implicit function theorem (Blondel et al., 2022; Krantz and Parks, 2002) (IFT).* If $F\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable with $F(x_0, \theta_0) = 0$ and $\partial F/\partial x \neq 0$ at (x_0, θ_0), then there exists a unique smooth function $x^*(\theta)$ near θ_0 satisfying $x^*(\theta_0) = x_0$ and $F(x^*(\theta), \theta) = 0$.

Applying this with $x = a$, $\theta = \{\hat{y}_j\}_{j=1}^M$, and $F(a; \{\hat{y}_j\}_{j=1}^M) := f'(a; \{\hat{y}_j\}_{j=1}^M)$, the IFT requires $\partial F/\partial a \neq 0$ at a^*:

$$\left.\frac{\partial F}{\partial a}\right|_{a^*} = \frac{1}{M}\sum_{j=1}^{M} \frac{\partial^2 c}{\partial a^2}(a^*, \hat{y}_j) =: H > 0, \tag{17}$$

where $H > 0$ follows from positive curvature of f at a^*; the IFT therefore guarantees that a^* is locally a smooth function of $\{\hat{y}_j\}_{j=1}^M$. In our setting, a^* is already defined as a function of $\{\hat{y}_j\}$; the role of the IFT is therefore to establish smoothness of this function, licensing the implicit differentiation below — not to establish the functional dependence itself. Since $F(a^*(\{\hat{y}_j\}); \{\hat{y}_j\}_{j=1}^M) = 0$ holds for all $\{\hat{y}_j\}$, its total derivative with respect to any $\hat{y}_m$ is zero; applying the chain rule gives

$$\left.\frac{\partial F}{\partial a}\right|_{a^*} \cdot \frac{\partial a^*}{\partial \hat{y}_m} + \left.\frac{\partial F}{\partial \hat{y}_m}\right|_{a^*} = 0, \qquad \left.\frac{\partial F}{\partial \hat{y}_m}\right|_{a^*} = \frac{1}{M}\frac{\partial^2 c}{\partial a\, \partial \hat{y}_m}(a^*, \hat{y}_m), \tag{18}$$

where only the $j = m$ term in the sum contributes to $\partial F/\partial \hat{y}_m$. Solving for $\partial a^*/\partial \hat{y}_m$ gives Eq. (7) in Section 4.2.

Boundary solution. Consider $a^* = a_{\max}$ (the case $a^* = a_{\min}$ is symmetric). Since g_1 is inactive, complementary slackness gives $\mu_1 = 0$, and Eq. (13) becomes

$$f'(a_{\max}; \{\hat{y}_j\}_{j=1}^M) = -\mu_2 \leq 0, \tag{19}$$

where the inequality follows from dual feasibility $\mu_2 \geq 0$. Excluding the measure-zero case $f'(a_{\max}) = 0$ (the unconstrained minimiser coincides exactly with $a_{\max}$), we have $f'(a_{\max}) < 0$ strictly.

Since f depends on $\hat{y}_m$ through the m-th summand, $f'(a_{\max}; \{\hat{y}_j\}_{j=1}^M)$ is continuous in $\hat{y}_m$ and therefore remains strictly negative under small perturbations $\hat{y}_m \to \hat{y}_m + \delta$. Since $f'(a_{\max})$ remains strictly negative (the objective still pushes beyond the boundary), the constraint g_2 stays active and a^* stays at $a_{\max}$; the multiplier μ_2 adjusts to maintain Eq. (13) without a^* moving. Consequently,

$$\frac{\partial a^*}{\partial \hat{y}_m} = 0. \tag{20}$$

A.4 Finite-sample approximation of the combined loss

Proposition 1 and the strict properness of Eq. (5) are population-level results, stated for exact expectations. In training, both loss terms are evaluated on a finite number M of model samples. This section shows that the resulting finite-M objective remains approximately strictly proper.

Setup. We drop the input x for brevity. Let F be the true conditional distribution, $\hat{F}$ the model's predictive distribution, and $\hat{F}_M = \frac{1}{M}\sum_{m=1}^{M} \delta_{\hat{y}_m}$ the empirical distribution of M i.i.d. samples $\hat{y}_m \sim \hat{F}$, as drawn in one training step. Let $a^*(\hat{F})$ denote the Bayes action (Eq. (1)) and $a^*(\hat{F}_M)$ its sample-based counterpart computed by the optimisation layer (Eq. (2)). Let

$$J(a) = \mathbb{E}_{Y\sim F}[c(a, Y)], \qquad a_F^* = a^*(F) = \arg\min_a J(a)$$

denote the true expected cost of action a and the true optimal action. Taking expectations over the observation $Y \sim F$, the population objective of Eq. (5) is

$$L_\infty(\hat{F}) = (1 - w_d)\, \mathbb{E}_{Y\sim F}\big[\mathrm{ES}(\hat{F}, Y)\big] + w_d\, J\big(a^*(\hat{F})\big),$$

since $\mathbb{E}_{Y\sim F}[S_c(\hat{F}, Y)] = J(a^*(\hat{F}))$. It is strictly proper, i.e. uniquely minimised at $\hat{F} = F$. Because the samples are redrawn at every step, training instead minimises the finite-M objective

$$L_M(\hat{F}) = \mathbb{E}_{\hat{F}_M}\Big[(1 - w_d)\, \mathbb{E}_{Y\sim F}\big[\mathrm{ES}_{\mathrm{fair}}(\hat{F}_M, Y)\big] + w_d\, J\big(a^*(\hat{F}_M)\big)\Big],$$

where $\mathbb{E}_{\hat{F}_M}$ averages over sample draws and $\mathrm{ES}_{\mathrm{fair}}$ is the fair energy score estimator (Section A.1).

Decomposition. The fair estimator is unbiased, $\mathbb{E}_{\hat{F}_M}[\mathrm{ES}_{\mathrm{fair}}(\hat{F}_M, y)] = \mathrm{ES}(\hat{F}, y)$ for all $\hat{F}$ and y (Ferro, 2014). The finite-M effect therefore enters only through the decision term:

$$L_M(\hat{F}) = L_\infty(\hat{F}) + w_d\, b_M(\hat{F}), \qquad b_M(\hat{F}) = \mathbb{E}_{\hat{F}_M}\big[J(a^*(\hat{F}_M))\big] - J\big(a^*(\hat{F})\big). \tag{21}$$

The bias b_M is the expected difference in true cost between acting on M samples and acting on the full predictive distribution.

Size of the bias. Because it is computed from finitely many samples, the sample-based action $a^*(\hat{F}_M)$ scatters around $a^*(\hat{F})$ and systematically incurs excess cost (analogous to the overestimation bias of Q-learning (Thrun and Schwartz, 1993; van Hasselt, 2010)). Write $\bar{a} = a^*(\hat{F})$ and $\Delta_M = a^*(\hat{F}_M) - \bar{a}$. Expanding the true expected cost around $\bar{a}$,

$$J\big(a^*(\hat{F}_M)\big) = J(\bar{a}) + J'(\bar{a})\, \Delta_M + \tfrac{1}{2}\, J''(\bar{a})\, \Delta_M^2 + o(\Delta_M^2),$$

and taking the expectation over sample draws gives

$$b_M(\hat{F}) = J'(\bar{a})\, \mathbb{E}_{\hat{F}_M}[\Delta_M] + \tfrac{1}{2}\, J''(\bar{a})\, \mathbb{E}_{\hat{F}_M}[\Delta_M^2] + o(1/M). \tag{22}$$

Minimisers of empirical averages such as $a^*(\hat{F}_M)$ are called M-estimators; standard asymptotics give a scatter $\mathbb{E}[\Delta_M^2] = O(1/M)$ (van der Vaart, 1998, Ch. 5) and a mean shift $\mathbb{E}[\Delta_M] = O(1/M)$ (Rilstone et al., 1996), so both terms are $O(1/M)$. We assume that there exists a constant $C > 0$, independent of M and $\hat{F}$, such that

$$|b_M(\hat{F})| \le C/M \quad \text{for all } M \text{ and all forecasts considered}, \tag{23}$$

which requires well-determined optimal actions, i.e. unique minima with positive curvature.

Steepness away from the truth. The finite-M term $w_d\, b_M$ is of order $1/M$, but how far it can move the minimiser depends on how steeply the population loss L_∞ rises away from F: if the rise is steep, a small perturbation barely shifts the minimum; if it is flat, even a small one can shift it far. Strict properness guarantees that F is the unique minimiser of L_∞, but not how steep this rise is. For the energy score, the rise is known exactly: with $X, X' \sim \hat{F}$ and $Y, Y' \sim F$ independent, the expected score exceeds that of the truth by half the energy distance (Székely and Rizzo, 2013),

$$\mathbb{E}_{Y\sim F}\big[\mathrm{ES}(\hat{F}, Y)\big] - \mathbb{E}_{Y\sim F}\big[\mathrm{ES}(F, Y)\big] = \tfrac{1}{2}\, \mathrm{ED}(\hat{F}, F), \qquad \mathrm{ED}(\hat{F}, F) = 2\,\mathbb{E}\|X - Y\| - \mathbb{E}\|X - X'\| - \mathbb{E}\|Y - Y'\|,$$

a squared distance between distributions that is zero if and only if $\hat{F} = F$. The decision term can only add to this, since $J(a^*(\hat{F})) \ge J(a_F^*)$, so

$$L_\infty(\hat{F}) - L_\infty(F) \ge \tfrac{1 - w_d}{2}\, \mathrm{ED}(\hat{F}, F). \tag{24}$$

The energy score thus guarantees that L_∞ rises at least quadratically in the distance from F, at a rate set by $1 - w_d$ and independent of the cost function.

Result. Two effects compete: (i) the growth of L_∞ away from F and (ii) the finite-M term. A forecast $\hat{F}$ can only outperform F under the training objective if the gain from (ii) exceeds the loss from (i). By Eq. (21),

$$L_M(\hat{F}) - L_M(F) = \big[L_\infty(\hat{F}) - L_\infty(F)\big] + w_d\, \big[b_M(\hat{F}) - b_M(F)\big] \;\ge\; \underbrace{\tfrac{1 - w_d}{2}\, \mathrm{ED}(\hat{F}, F)}_{\text{(i) min. loss for leaving } F} \;-\; \underbrace{\frac{2\, w_d\, C}{M}}_{\text{(ii) max. finite-}M\text{ gain}},$$

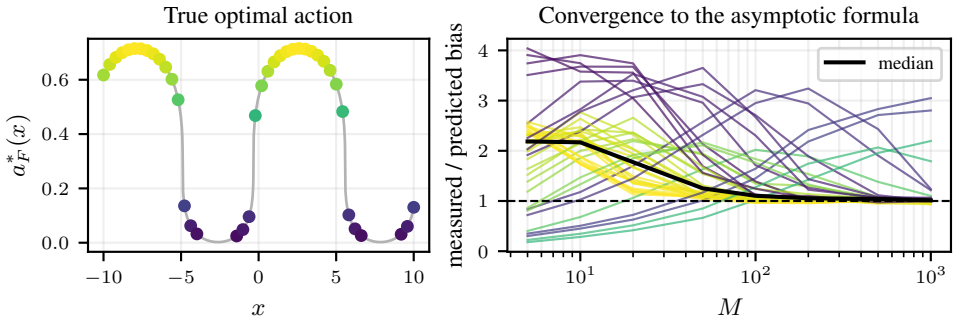


Figure A.1: Finite-M bias at the true distribution, per input (synthetic task). *Left:* true optimal action $a_F^*(x)$; dots mark the inputs x analysed on the right, at their optimal action (interior optima, $0.02 < a_F^* < 0.98$). *Right:* ratio of measured bias (Eq. (21)) to its asymptotic prediction (Eq. (26)), colours as in the left panel. The median (black) converges to 1. Inputs on the plateaus of a_F^* (yellow) converge quickly, as their optimal action is insensitive to the sampled proportion of the two modes; where a_F^* changes steeply or lies near the action bound (green, purple), convergence is slower.

using Eq. (24) and Eq. (23) (each of $b_M(\hat{F})$ and $b_M(F)$ is at most C/M in magnitude). Any forecast far enough from F that (i) exceeds (ii) is thus beaten by the truth. If F is among the forecasts considered, the minimiser $\hat{F}^*$ of L_M satisfies $L_M(\hat{F}^*) \le L_M(F)$ and must therefore lie where (i) does not exceed (ii), i.e.

$$\mathrm{ED}(\hat{F}^*, F) \;\le\; \frac{4\,w_d}{1-w_d} \cdot \frac{C}{M}. \tag{25}$$

The minimiser of the finite-M objective thus lies within $O(1/M)$ of the truth in energy distance. The bound vanishes at $w_d = 0$, where L_M reduces to the fair energy score and is exactly strictly proper for every M, and diverges as $w_d \to 1$, where the energy-score anchor disappears.

The result concerns the objective itself: computing the optimal action from finitely many samples moves its minimiser by at most $O(1/M)$ in energy distance. It does not guarantee that training reaches this target: neural networks parameterise only a restricted family of distributions, training data are finite, and stochastic optimisation may stop at a local optimum. The resulting errors are precisely what decision-aware training targets: where they affect the decisions, the decision loss directs the model's limited capacity towards reducing them. The method thus trades a small, controlled distortion of the target for a correction of errors that can be considerably larger.

Empirical check. The only approximate step above is the size of the bias b_M in Eq. (22). We measure it on the synthetic task, where the true distribution is known, first at the truth and then for trained models.

At the truth, the bias can be predicted: the first-order term vanishes since a_F^* minimises J, and M-estimator asymptotics give

$$b_M(F) \approx \frac{\mathrm{Var}_{Y\sim F}\big[\partial_a c(a_F^*, Y)\big]}{2\,J''(a_F^*)\,M}, \tag{26}$$

i.e. the bias is larger when the cost's sensitivity to the action varies strongly with the outcome, and smaller when the expected cost is sharply curved at its minimum. The measured bias converges to this prediction and hence decays as $1/M$ (Figure A.1): the median ratio falls from 2.2 at $M = 5$ to 1.01 at $M = 1000$. For small M, the prediction underestimates the bias, most strongly for inputs where the optimal action changes steeply with the proportion of the two modes.

For trained models (Figure A.2), the true expected cost of acting on M samples follows a straight line in $1/M$ for $M \ge 10$, so the bias decays as $1/M$ also away from the truth, as assumed in Eq. (23). At the training value $M = 10$, it raises the expected cost by only 2.4% for $w_d = 0$ and 1.3% for $w_d = 0.95$ (medians over 3×3 seeds).

Effect on training. Finally, we vary the ensemble size used during training, $M \in \{5, 10, 20, 50\}$, on the synthetic task. Decision-aware training improves all three metrics consistently across ensemble sizes (Figure A.3), so the choice of M does not affect our conclusions. We use $M = 10$ throughout.

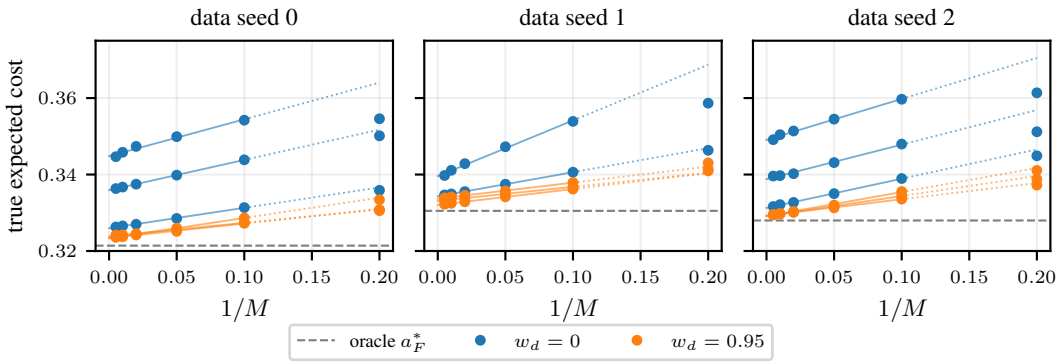


Figure A.2: Finite-M bias for trained models (synthetic task; one panel per data seed, one line per trained model, coloured by w_d). Markers: true expected cost of acting on M model samples, $\mathbb{E}_{\hat{F}_M}[J(a^*(\hat{F}_M))]$, averaged over test inputs. Lines: linear fits in $1/M$ on $M \in \{10, \ldots, 200\}$; the dotted continuation to $M = 5$ (not used in the fit) shows the deviation at small M. The intercept estimates the cost of acting on the full predictive distribution, $J(a^*(\hat{F}))$, so the vertical distance of a marker to the intercept is the finite-M bias b_M (Eq. (21)). Dashed: cost of the true optimal action, $J(a_F^*)$; the gap to the intercept is the remaining model error. One $w_d = 0$ run of data seed 1, which did not recover the second mode, lies outside the plotted range.

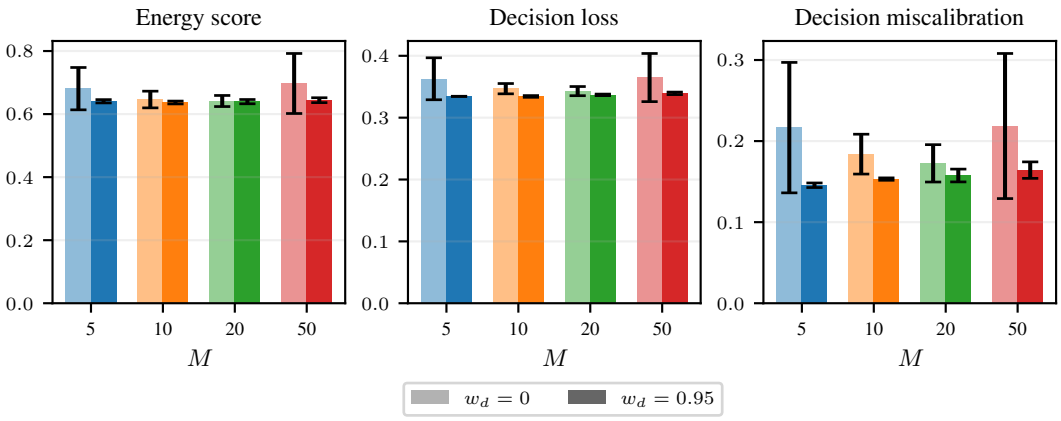


Figure A.3: Sample size sensitivity of decision-aware training benefit (synthetic task). For each $M \in \{5, 10, 20, 50\}$ (number of samples drawn per input during training) we compare the pure energy-score baseline ($w_d = 0$, transparent bars) against a decision-aware run ($w_d = 0.9$, solid bars) on the synthetic task ($N_{train} = 1600$, mean $\pm$ std over training seeds). Decision-aware training improves all three metrics consistently across M, showing robustness to the sample size.

B Implementation Details

Parts of our implementation build on code from Kneissl et al. (2025), released under the MIT licence. Code for reproducing our experiments can be found at `https://anonymous.4open.science/r/decision-aware-training-2114`.

Training. We use the Adam optimizer(Kingma and Ba, 2014) with learning rate 10^{-3} to train the models over 2000 epochs. Loss scales of energy score and decision loss are estimated once before training and fixed throughout; optionally, a short energy score pre-training phase precedes scale estimation to ensure realistic samples and cost values. We perform 200 epochs of energy score pre-training for the wind power dispatch task, since we observe that untrained predictions yield

outliers with non-realistic costs. We validate performance every 10 epochs on a validation set and apply early stopping with a patience of 600 epochs.

Distributional diffusion model. Distributional diffusion models (De Bortoli et al., 2025; Kneissl et al., 2025) keep the forward noising process and reverse chain of standard diffusion models (Ho et al., 2020) but change the denoiser. Instead of predicting a single estimate of the clean target y_0 from (y_t, x), trained with a mean-squared error, the denoiser receives additional latent noise and outputs samples of y_0, trained with a strictly proper scoring rule such as the energy score so that they follow the full conditional distribution of y_0 given (y_t, x). At inference, each reverse step uses a draw from this distribution instead of a point estimate. The denoiser is an MLP with DDPM-style linear noise schedule over $T = 100$ steps, with 5 hidden layers of dimension 50 and ReLU activations. $T = 100$ is selected based on De Bortoli et al. (2025); Kneissl et al. (2025), where improvements saturate for larger T. We also ran pilot experiments with larger T but found no improvements. At each step the denoiser predicts a distribution over ϵ (equivalently y_0) given noisy y_t, drawing M samples per step. Conditional input x and a per-sample noise vector are injected via FiLM (Feature-wise Linear Modulation) conditioning at every hidden layer, giving each of the M samples a unique modulation.

Implicit generative model. The implicit generative model is an MLP with 5 hidden layers of dimension 50 and ReLU activations. It maps the concatenation of conditioning input x and noise vector $\varepsilon \sim \mathcal{N}(0, I)$ directly to a sample $\hat{y}_m$; M independent noise draws in parallel yield M samples in a single forward pass.

Optimization layer. The optimal action a^* is obtained by projected gradient descent on $\frac{1}{M}\sum_m c(a, \hat{y}_m)$ over $\mathcal{A}$, implemented within the optimization layer forward pass. We use 50 optimization steps during training, which small pilot experiments confirmed to be sufficient for convergence; at inference we use 200 steps for higher precision. Second-order derivatives required for computing the decision loss gradient (Eq. (7)) are computed via PyTorch `autograd`.

Decision-aware training adds computational overhead relative to standard energy score training through the optimisation layer, which is invoked at every forward pass. The overhead scales with the number of samples M (the optimisation objective is evaluated over all M samples per batch element) and the number of iterations required to solve Eq. (1) to convergence.

Computational resources We use an internal compute cluster with NVIDIA A100 GPUs. Training of all models required approximately 50 GPU hours.

Data The ERA5 dataset used in the real-world experiments s provided by the Copernicus Climate Change Service under a licence permitting research use (Hersbach et al., 2020).

C Additional Results

This section presents additional experimental details and results for all three experiments in the main paper.

C.1 Synthetic Decision Task

We use a 1600/200/400 train/val/test split across 3 data seeds and 3 training seeds, training a distributional diffusion model. The denoiser uses $M = 10$ samples during training (Section A.4 analyses and ablates this choice). At inference, we draw 200 samples from the learned predictive distributions via independent reverse chain runs. The cost function belongs to the same sigmoid threshold family as the frost hazard cost; positive curvature of the expected cost at interior solutions is verified in Section C.3.1.

Figure C.4 shows the data generating process (bimodal conditional GMM with input-dependent mode weights and ground-truth optimal action $a^*(x)$) alongside the cost function. Figure C.5 shows aggregate energy score, decision loss, and decision calibration vs. w_d, pooled across all seeds. Figure C.6 shows the predictive marginal distributions across seeds and w_d values: mode 1 is overrepresented at $w_d = 0$ (pure energy score), with mode 2 mass progressively recovering as

w_d increases. Figure C.7 shows the predicted $a^*(x)$ across data seeds and w_d: tracking improves consistently, though some of the troughs (boundary $a^* = 0$, no gradient) remain a blind spot. The optimal action is tracked correctly even at $w_d = 1.0$ (no energy score anchor), but in this setting the underlying distribution degenerates to extreme values (Figure C.8). This provides empirical evidence that the energy score is necessary as a distributional anchor. Figure C.9 shows the conditional mode 2 mass as a function of input x: dec loss partially recovers the true $w(x)$ around the peaks but overshoots at high w_d.

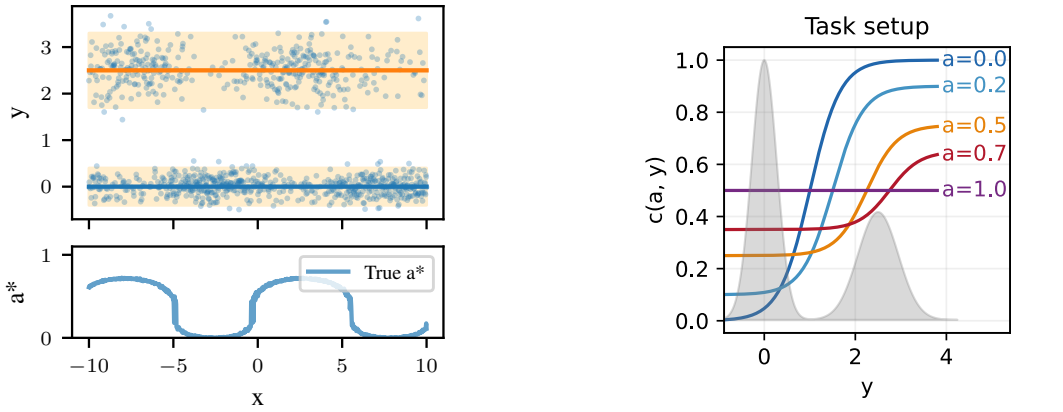


Figure C.4: *Left:* bimodal conditional GMM (top) and ground-truth optimal action $a^*(x)$ (bottom). *Right:* cost function $c(a, y)$ for five protection levels a, overlaid with the marginal $p(y)$.

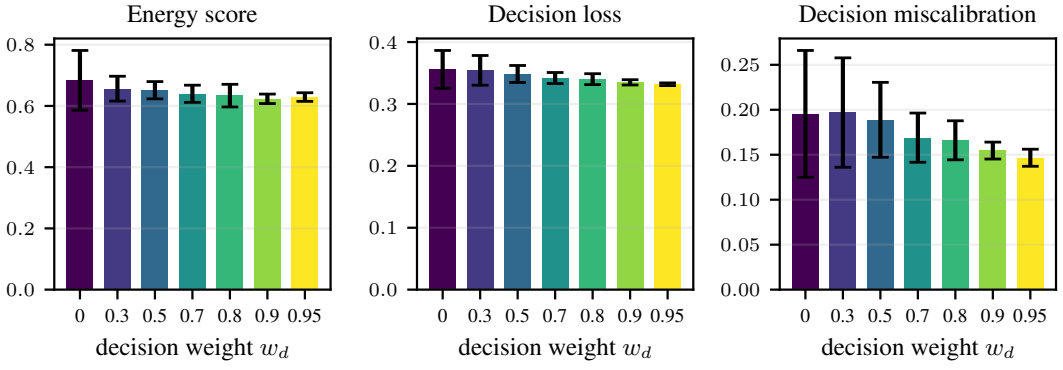


Figure C.5: Aggregate energy score, decision loss, and decision miscalibration vs. w_d (mean $\pm$ std, pooled across training and data seeds). energy score and decision loss improve slightly with w_d; average decision calibration improves more consistently throughout.

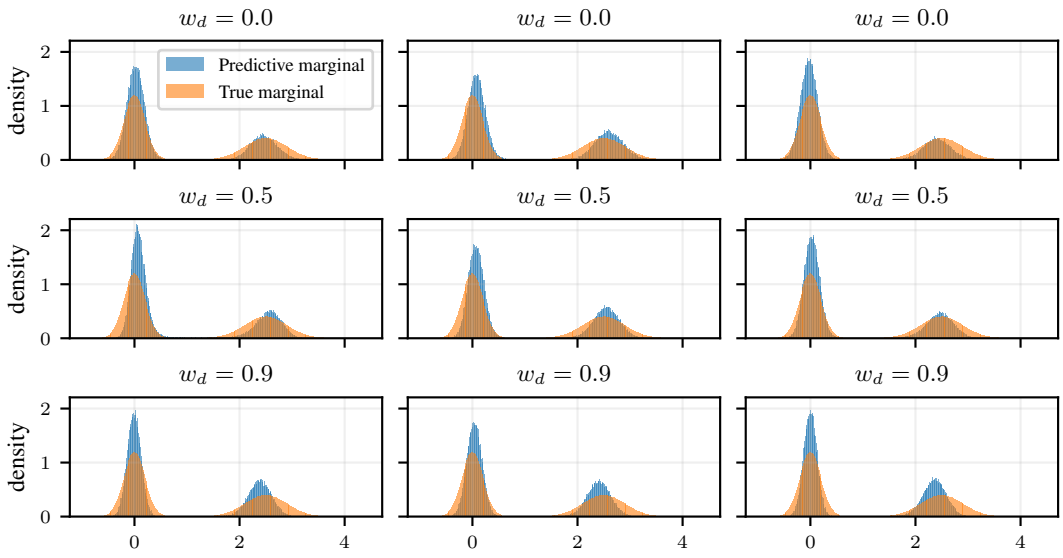


Figure C.6: Predictive marginal distributions for $w_d \in \{0.0, 0.5, 0.9\}$ across three training seeds (one seed per column). Mode 1 is overrepresented at $w_d = 0$; mode 2 mass recovers with increasing w_d.

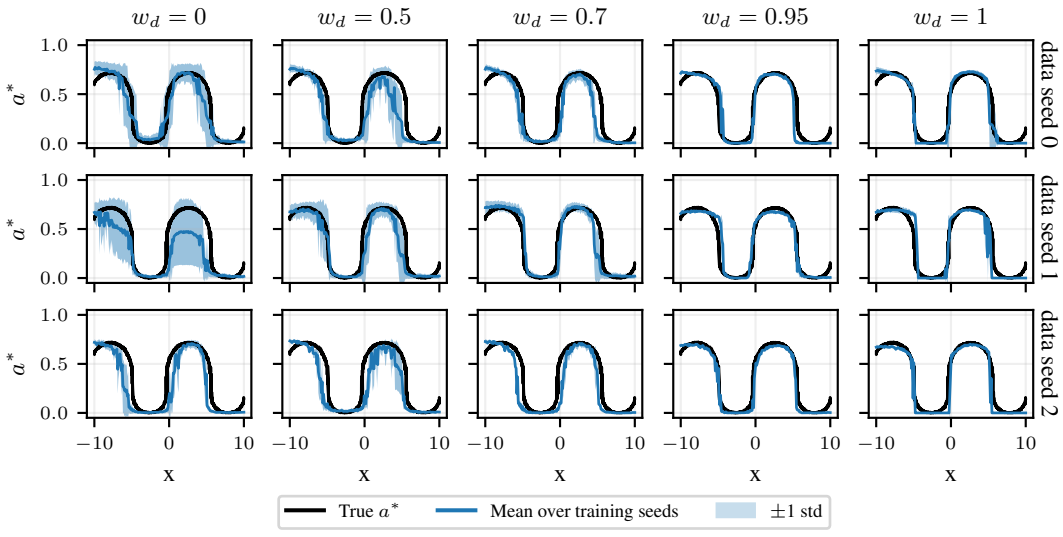


Figure C.7: Predicted $a^*(x)$ (mean $\pm$ std over training seeds) vs ground truth across w_d and data seeds (one seed per row). Pure energy score training ($w_d = 0$) fails to track the ground truth action at the transitions between the extremes; decision-aware training ($w_d > 0.0$) improves the tracking of $a*$ and gets closer to the ground truth for large w_d. Some of the troughs ($a^* = 0$, boundary case) that exist for $w_d = 0$ remain a blind spot throughout. Notably, a^* tracking remains accurate at $w_d = 1.0$ (pure decision loss training) despite the underlying predictive distribution degenerating (Figure C.8).

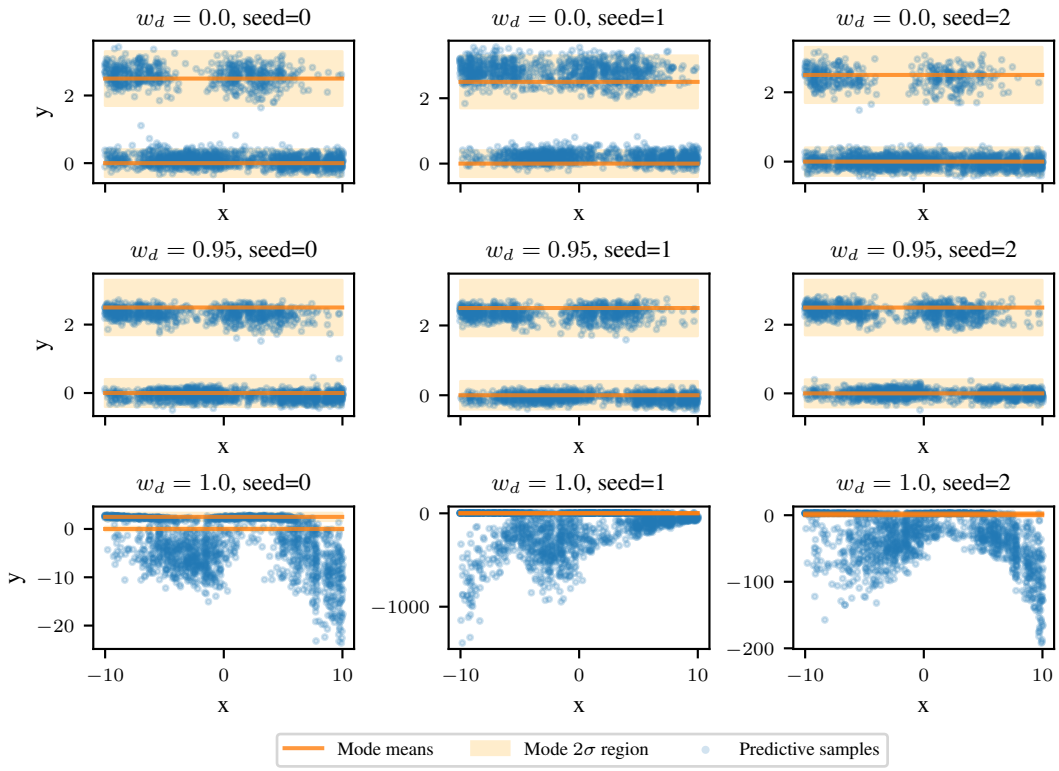


Figure C.8: Predictive samples (blue) vs GMM mode means and $\pm 2\sigma$ regions (orange) across decision loss weights $w_d \in 0.0, 0.9, 1.0$ and three data seeds. For $w_d \in 0.0, 0.9$ samples cluster around both true modes. At $w_d = 1.0$ the energy score anchor is removed and the predictive distribution degenerates to extreme values (observe the different y axes). Combined with the preserved $a^*(x)$ tracking for $w_d = 1.0$ (Figure C.7), this confirms empirically that the energy score is necessary as a distributional anchor.

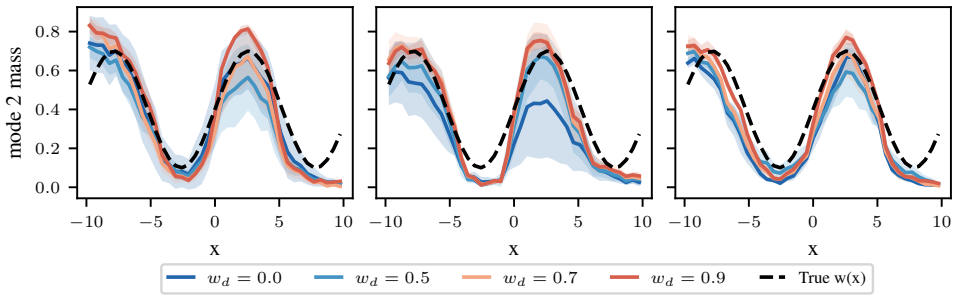


Figure C.9: Conditional mode 2 mass estimated within 40 x-bins for different w_d, per data seed (one seed per panel). Lines show mean over training seeds; shaded band is ± 1 std. Decision loss often improves the conditional mode mass estimate (moving towards $w(x)$) around the peaks and the transitions but can overshoot at high w_d.

C.2 Wind Power Dispatch

In this section we report experimental details of the wind power dispatch experiment followed by additional results for the implicit generative model as well as results for the distributional diffusion model on this task.

C.2.1 EXPERIMENTAL SETUP

The power curve maps 10m wind speed to normalised power output via a differentiable sigmoid approximation (Wang et al., 2019); hub-height wind speeds are recovered via a power-law profile (Abbes and Belhadj, 2012). The dispatch cost is $c(a, y) = -a + \lambda \operatorname{relu}(a - P(y))^2$, where $a \in [0, 1]$ is the committed power fraction, $P(y)$ is the normalised power output, and λ controls the shortfall penalty. We do not apply a penalty for power overproduction since this case can be avoided by windpower providers by dynamically reducing power output (curtailment) to match the promised amount (Bruninx et al., 2025). In this case, only opportunity costs remain ($c = -a$), since a larger a could have been chosen to increase revenue.

ERA5 inputs are observed wind speed at t, $t-6$ h, and $t-12$ h. We use 2005–2019 for training, 2020 for validation, and 2021 for testing. The test set contains 105 cut-off events ($v > 20$ m/s at hub height), 230 rated, and 1074 ramp observations. The implicit generative model uses $M = 10$ samples during training and $M = 500$ at inference. The distributional diffusion model uses $M = 10$ samples for the probabilistic denoiser during training; at inference, 500 samples are drawn via independent reverse chain runs.

Positive curvature of $f(a) = \frac{1}{M}\sum_j c(a, \hat{y}_j)$ (Section A.3) at interior solutions holds analytically: since $\frac{\partial^2}{\partial a^2} c(a, \hat{y}_j) = 2\lambda$ when $P(\hat{y}_j) < a$ and 0 otherwise, we have $H = (2\lambda/M)\,\#\{j : P(\hat{y}_j) < a^*\}$. At any interior optimum a^*, at least one sample must satisfy $P(\hat{y}_j) < a^*$, otherwise $f'(a^*) = -1 < 0$, contradicting optimality, so $H > 0$.

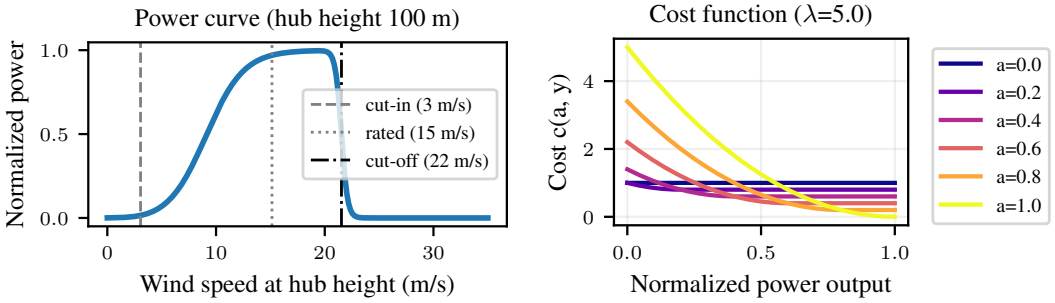


Figure C.10: *Left:* differentiable power curve used for converting wind speed at turbine hub height to normalized power output. *Right:* dispatch cost $c(a, y)$ for $\lambda = 5$ as a function of committed power a and observed power output y.

C.2.2 ADDITIONAL RESULTS: DISTRIBUTIONAL DIFFUSION MODEL

Figure C.11 shows aggregate metrics vs w_d: energy score improves then overshoots at high w_d; aggregate decision loss is flat; decision calibration improves consistently for all λ. Figure C.12 shows the full regional breakdown for each λ. Figure C.13 shows the cut-off region ablation across λ. Figure C.14 shows the rolling-window median shift.

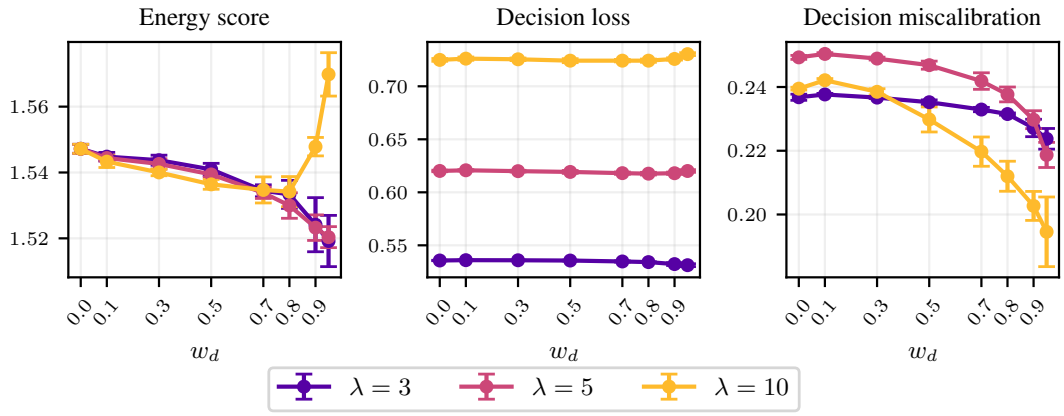


Figure C.11: Aggregate metrics vs w_d for $\lambda \in \{3, 5, 10\}$, distributional diffusion model. Energy score improves with w_d, but degrades at high w_d for $\lambda = 10$. Average decision loss stays relatively across w_d. Decision miscalibration improves with w_d, with stronger improvements at higher shortfall penalty λ.

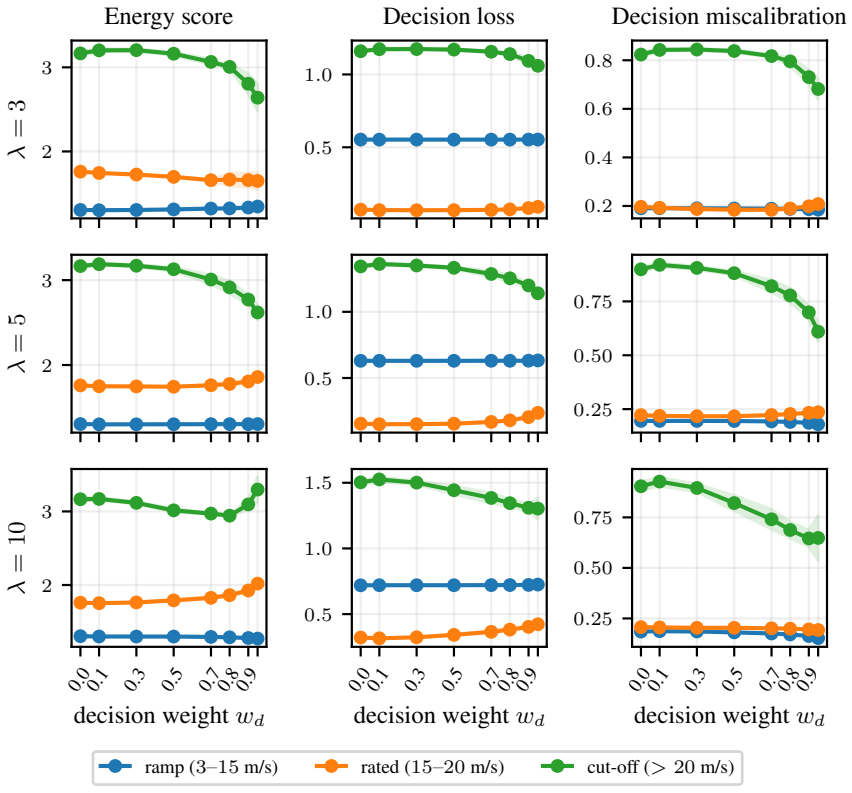


Figure C.12: Regional metrics (cut-off, ramp, rated) vs w_d for $\lambda = 3$ (top), $\lambda = 5$ (middle), and $\lambda = 10$ (bottom), distributional diffusion model. Performance in the cut-off region improves with w_d, with stronger improvements at higher shortfall penalty λ; most rated region metrics degrade with w_d across all λ.

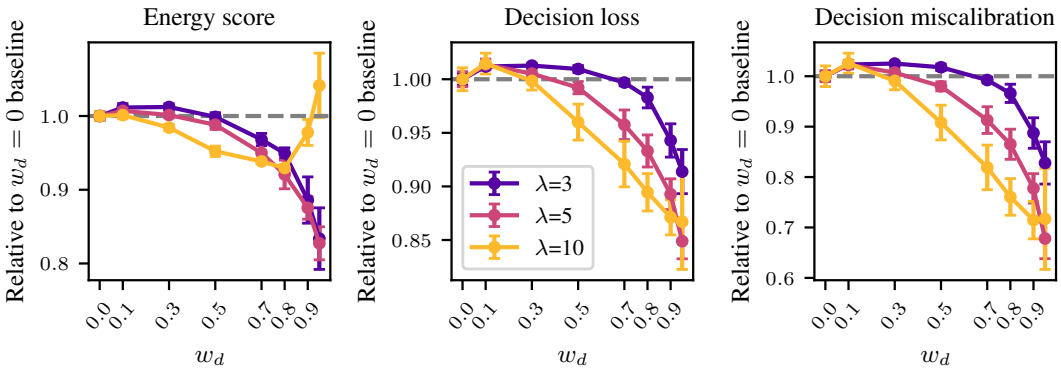


Figure C.13: Cut-off region metrics vs w_d across $\lambda \in \{3, 5, 10\}$, distributional diffusion model. We observe a monotonic improvement pattern across λ for $w_d \geq 0.5$. For larger λ, the decision loss gradient provides a stronger learning signal in the cut-off region leading to stronger improvements. Energy score degrades for large λ and w_d, incidating overshoot (distributional fit degrades).

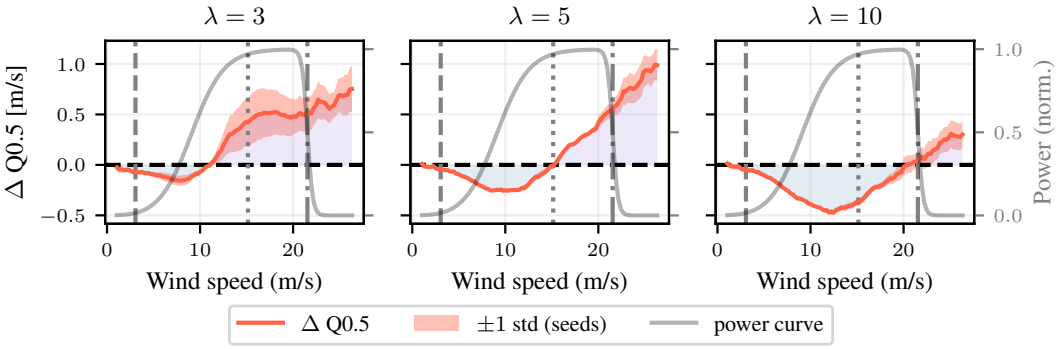


Figure C.14: Rolling-window median shift $\Delta Q_{0.5}(v)$ conditioned on observed wind speed for $\lambda \in \{3, 5, 10\}$, distributional diffusion model. Similar pattern as for the implicit generative model: negative corrections in the ramp region and positive corrections above the cut-off threshold.

C.2.3 Implicit generative model results

Figure C.15 shows the aggregate performance of the implicit generative model with respect to energy score, decision loss and decision miscalibration across seeds and decision weights w_d. Figure C.16 shows the regional breakdown for $\lambda \in \{3, 5, 10\}$: in the cut-off region, decision loss and decision miscalibration improve consistently only for $\lambda = 10$. With the biased uncorrected energy score estimator, which leaves larger tail errors, improvements appear for all λ (Section C.2.4). Figure C.17 illustrates the effect of decision-aware training on the predictive distributions: the model shifts mass to more conservative values in the ramp region and sharpens forecasts near the cut-off threshold. Figure C.18 shows the rolling-window median shift $\Delta Q_{0.5}(v)$: the correction is negative in the ramp region and turns positive beyond cut-off.

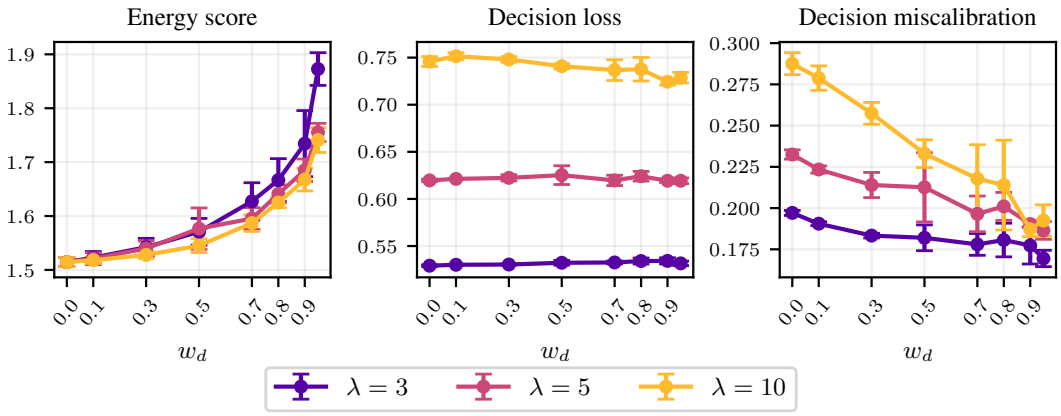


Figure C.15: Aggregate metrics vs w_d for shortfall penalty $\lambda \in \{3, 5, 10\}$, implicit generative model. Compared to the diffusion model (Figure C.11), we see a clearer trade-off between distributional accuracy (energy score) and decision quality across w_d. Aggregate decision loss remains relatively flat, whereas decision miscalibration improves with w_d, with stronger improvements at higher λ, mirroring the pattern observed for the diffusion model.

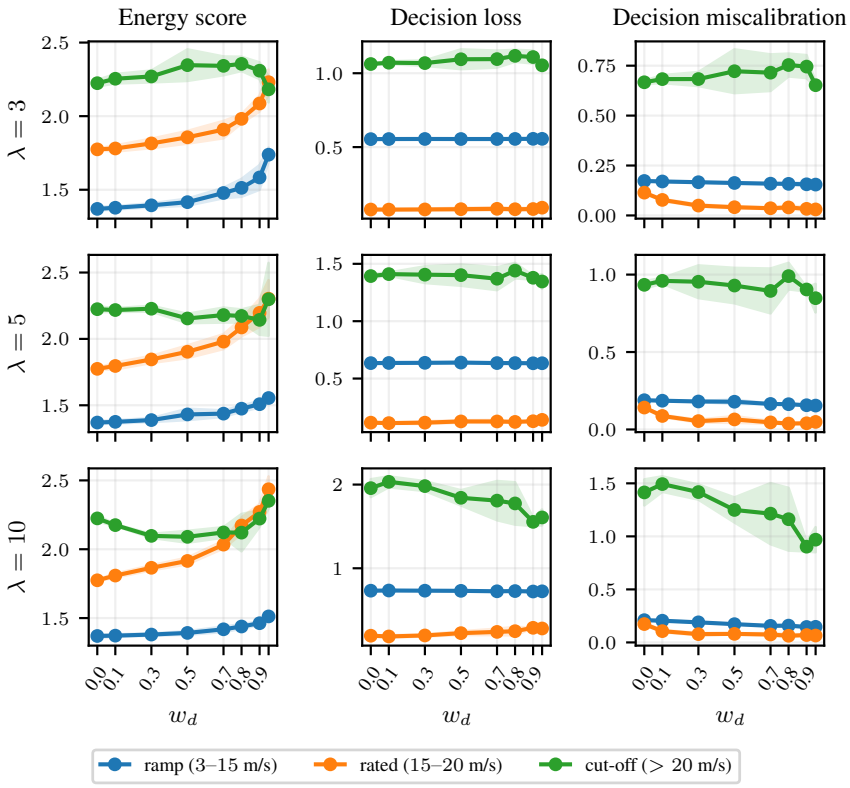


Figure C.16: Regional metrics (cut-off, ramp, rated) vs w_d for $\lambda = 3$ (top row), $\lambda = 5$ (middle row) and $\lambda = 10$ (bottom row), implicit generative model. For $\lambda \in \{3, 5\}$ the baseline ($w_d = 0$) is already well-calibrated in the cut-off region, leaving little room for improvement; for $\lambda = 10$ a consistent improvement emerges through decision-aware training ($w_d > 0$), as the larger shortfall penalty amplifies the cost of even small forecast errors in the cut-off region.

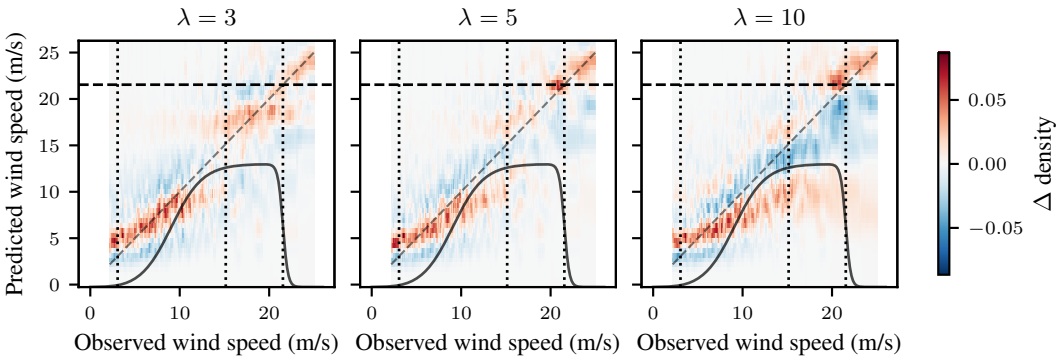


Figure C.17: Difference in conditional forecast density (conditioned on observed wind speed) between $w_d = 0.7$ and $w_d = 0.0$ (red: density gain, blue: density loss), for $\lambda \in \{3, 5, 10\}$ (left to right). The gray curve shows the power curve (observed wind speed axis). In the ramp region, decision loss shifts predicted mass below the diagonal, reflecting more conservative dispatch. Near cut-off, the model moves mass into the tail to anticipate cut-off costs, with the correction growing with λ: at $\lambda = 5$ the gained mass aligns with the diagonal, indicating that the model produces more confident forecasts near cut-off that are also well-aligned with observed wind speeds. At $\lambda = 10$ the gained mass around cut-off shifts above the diagonal, indicating that the model exaggerates predicted wind speeds to hedge against the larger shortfall penalty.

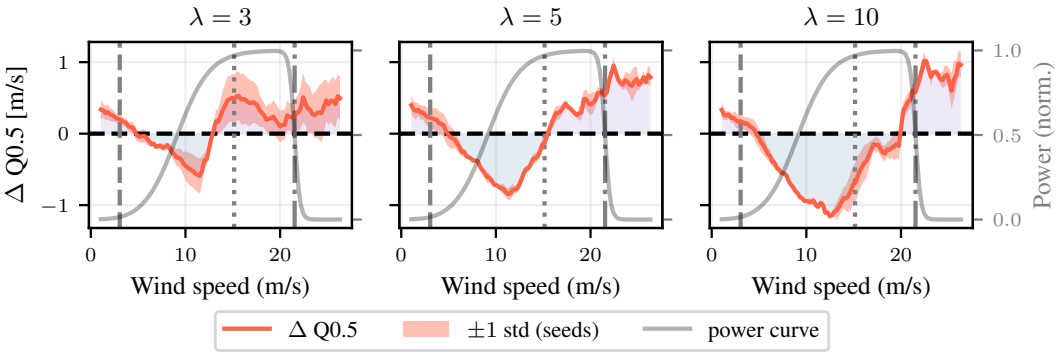


Figure C.18: Rolling-window median shift $\Delta Q_{0.5}(v)$ conditioned on observed wind speed for $\lambda \in \{3, 5, 10\}$, implicit generative model. Correction is negative in the ramp (more conservative dispatch) and positive in the cut-off region (more mass in costly tail where power drops to zero). This effect strengthens with λ. The model learns different correction behavior across outputs that reflects the cost function structure. Window size ± 1.5 m/s.

C.2.4 Decision-aware training with the uncorrected energy score

The main experiments use the unbiased fair energy score estimator (Section A.1). A common alternative, and the default in widely used scoring-rule libraries, is the uncorrected estimator, which normalises the spread term by M^2 rather than $M(M-1)$:

$$\mathrm{ES}_{\mathrm{uncorr}}(\hat{F}_M, y) = \frac{1}{M}\sum_{m=1}^{M}\|\hat{y}_m - y\| - \frac{1}{2M^2}\sum_{m=1}^{M}\sum_{j=1}^{M}\|\hat{y}_m - \hat{y}_j\|.$$

Since the M diagonal terms vanish, the spread term is underestimated by a factor $(1 - 1/M)$ relative to the fair estimator. Spread is thus under-rewarded relative to proximity to the observation, and minimisation favours underdispersed forecasts (Ferro, 2014), with the tails most affected. This provides a controlled setting to increase the residual error that energy score training leaves in the

tail, and thus to test whether the benefit of decision-aware training grows with it. We repeat the wind power dispatch experiment with the implicit generative model using the uncorrected estimator, keeping all other settings fixed ($\lambda \in \{3, 5, 10\}$, same w_d grid and seeds).

Aggregate results. The uncorrected baseline ($w_d = 0$) starts with higher decision loss and decision miscalibration than its fair counterpart (Figure C.19; cf. Figure C.15). Increasing w_d reduces both, reaching approximately the level obtained under the fair estimator; the effect grows with λ. The energy score remains flat up to moderate w_d and increases at large w_d.

Regional results. The improvement concentrates in the cut-off region (Figure C.20). Under the uncorrected estimator, decision loss and decision miscalibration in this region decrease with w_d for all λ, with larger λ yielding stronger improvements; under the fair estimator, a comparable improvement appears only for the largest penalty, $\lambda = 10$ (Figure C.16). The energy score in the cut-off region also improves under the uncorrected estimator, approaching the fair baseline: decision-aware training corrects the underdispersed tail itself, not only the resulting decisions. The rated region degrades as in the fair case, since $\partial P/\partial y \approx 0$ and $a^* \approx 1$ leave the energy score as the only active signal; the ramp region is largely unaffected.

Interpretation. The results confirm that the benefit of decision-aware training grows with the error the scoring rule leaves in cost-relevant regions (Section D): with the larger tail error of the uncorrected estimator, improvements appear for all λ, and at large w_d the two estimators attain nearly identical decision quality. The correction is local, however: in the ramp and rated regions the uncorrected estimator remains worse. The fair estimator and the decision loss are therefore complementary: the former reduces distributional error everywhere, the latter corrects the residual error that matters for the decision. The conditional density shift induced by the decision loss (not shown) is qualitatively identical to the fair case (Figure C.17), with approximately twice the magnitude. This is consistent with two sources of tail underdispersion acting in the same direction: the estimator bias, and the density-proportional allocation of the energy score gradient. The fair estimator removes only the former, leaving a smaller residual that decision-aware training corrects when the penalty is large enough. We report this comparison for the implicit generative model only; for the distributional diffusion model, the two estimators yield similar results.

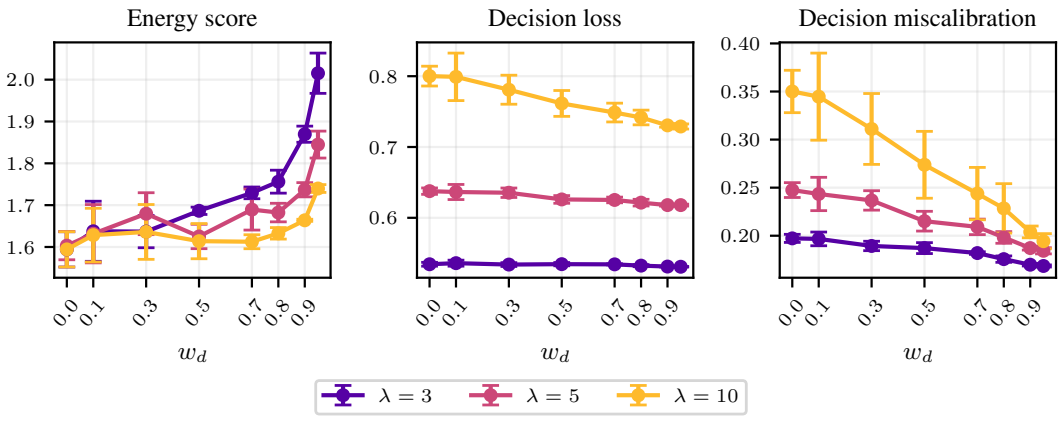


Figure C.19: Wind power dispatch, implicit generative model, uncorrected energy score estimator: aggregate metrics vs. w_d for $\lambda \in \{3, 5, 10\}$ (mean $\pm$ 1 std over seeds). Compare Figure C.15 (fair estimator).

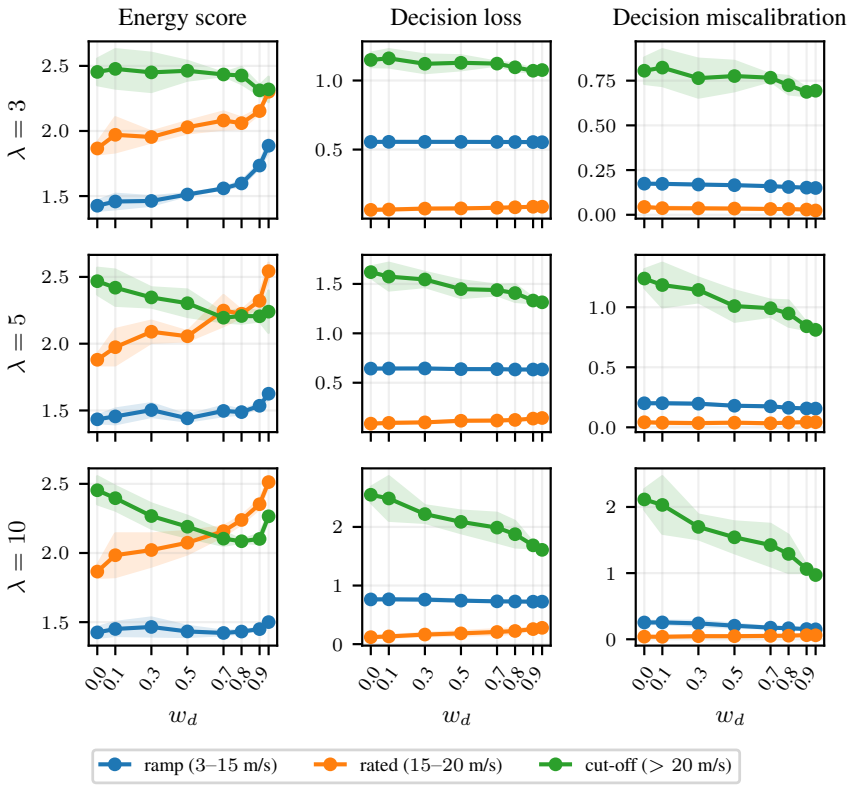


Figure C.20: As Figure C.19, conditional on the power curve region of the observed wind speed. Compare Figure C.16 (fair estimator).

C.3 Frost Protection

In this section we report experimental details of the frost protection experiment followed by additional results for the implicit generative model as well as results for the distributional diffusion model on this task.

C.3.1 Experimental setup

ERA5 inputs are 2 m temperature from a 3×3 spatial patch (centered around the target grid point) at three time steps (27 features), restricted to November–March. The implicit generative model uses $M = 10$ samples during training and $M = 500$ at inference. The distributional diffusion model uses $M = 10$ samples from the probabilistic denoiser during training; at inference, 500 samples are drawn via independent reverse chain runs. The cost function is a sigmoid threshold at 0 °C; $\alpha = \frac{FP}{FN}$ controls the asymmetry between false positive and false negative costs (Figure C.21). The gradient derivation (Section A.3) requires strict convexity of $f(a) = \frac{1}{M}\sum_j c(a, \hat{y}_j)$ at interior solutions a^*. This was verified empirically across 100,000 randomly drawn sample sets of size $M = 50$ from the marginal training distribution, with $H > 0$ (strict convexity) in all cases.

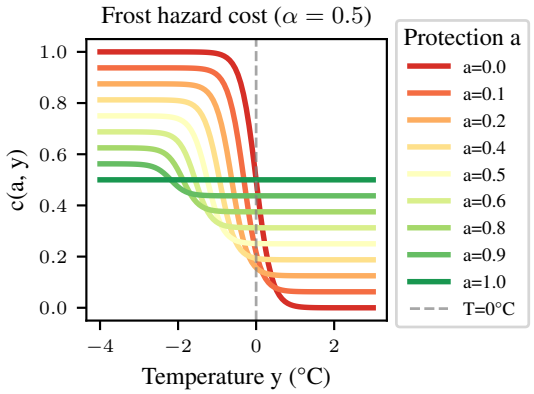


Figure C.21: Frost protection cost function for $\alpha = 0.5$.

C.3.2 Additional results: implicit generative model

The main text shows aggregate metrics for all α values and the conditional KDE for $\alpha = 0.3$. Figure C.22 shows the conditional KDE for all three α values at $w_d = 0.3$, alongside the shared $w_d = 0$ baseline. Figure C.23 shows the marginal 2m temperature (T2m) distributions for $\alpha = 0.5$: the marginal gets fitted better around the decision-critical freezing point. Figure C.24 shows the conditional breakdown by frost and no-frost regime for $\alpha \in \{0.2, 0.3, 0.5\}$: the no-frost regime improves in terms of decision loss and decision calibration while the frost regime degrades slightly or stays constant across w_d. Figure C.25 shows the quantile shift in the predictive distributions introduced through decision-aware training: quantiles closer to the cost-sensitive threshold region get shifted more strongly, supporting the claim of probability mass being redistributed around the decision threshold.

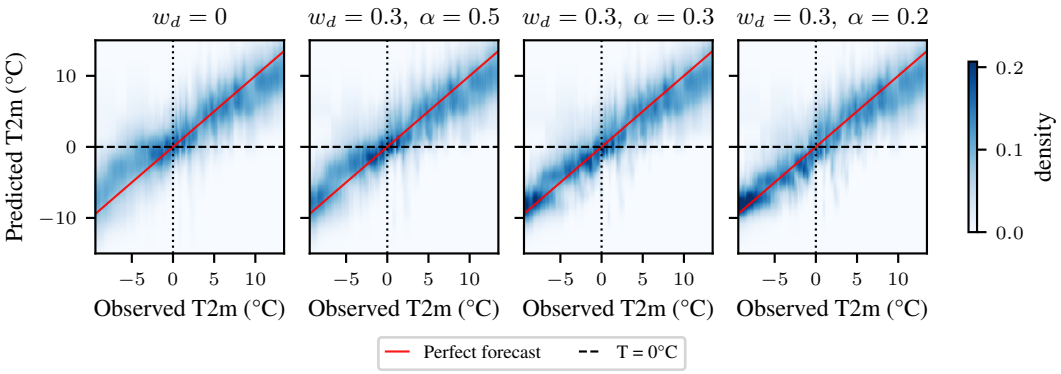


Figure C.22: Rolling-window conditional predictive density of temperature forecasts for $w_d = 0$ (leftmost, shared baseline) and $w_d = 0.3$ for $\alpha = 0.2$, $\alpha = 0.3$, and $\alpha = 0.5$ (left to right), implicit generative model. A warm bias for sub-zero temperatures visible at $w_d = 0$ is corrected at $w_d = 0.3$ progressively with α, with the predictive distribution concentrating closer to the diagonal in the frost region. Window size ± 1°C.

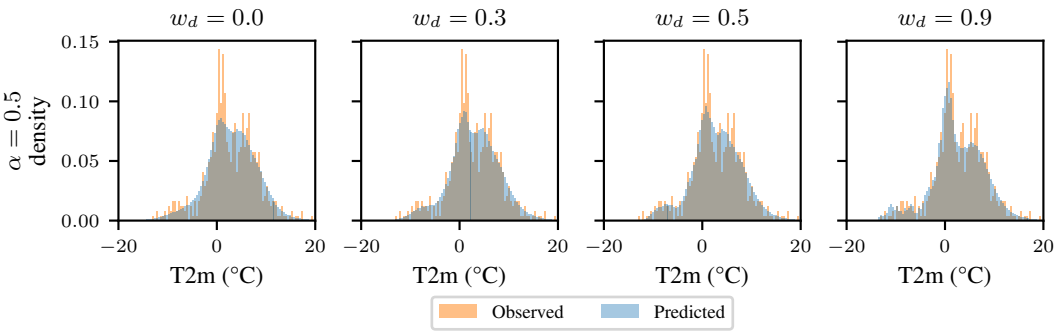


Figure C.23: Marginal T2m distribution (observed vs predicted) for $\alpha = 0.5$ at $w_d \in \{0.0, 0.3, 0.5, 0.9\}$, implicit generative model. As w_d grows, the model fits the marginal distribution better around the (decision-critical) freezing point.

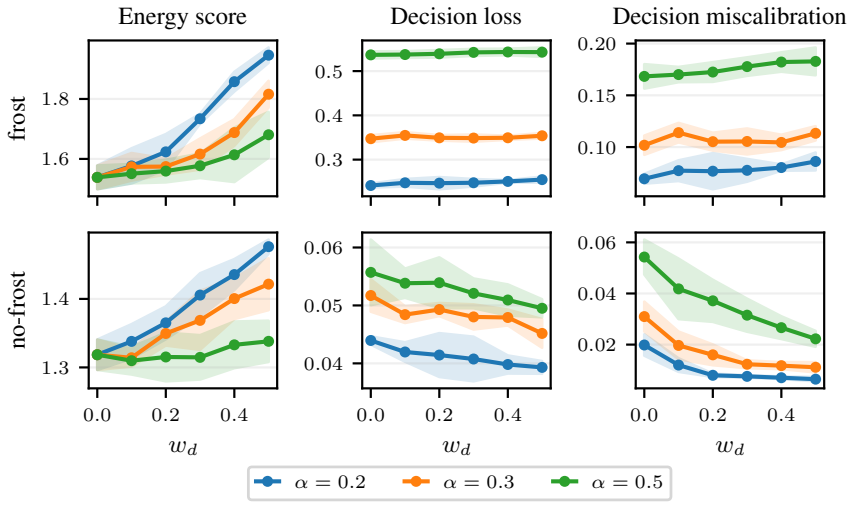


Figure C.24: Conditional metrics vs w_d for $\alpha \in \{0.2, 0.3, 0.5\}$ conditioned on observed temperatures (below or above 0 °C), implicit generative model: frost region (top row) and no-frost region (bottom row). At $\alpha = 0.5$, the no-frost region improves while the frost region degrades slightly (decision loss and miscalibration). Lowering α (increasing cost for missed frost) degrades energy score more quickly.

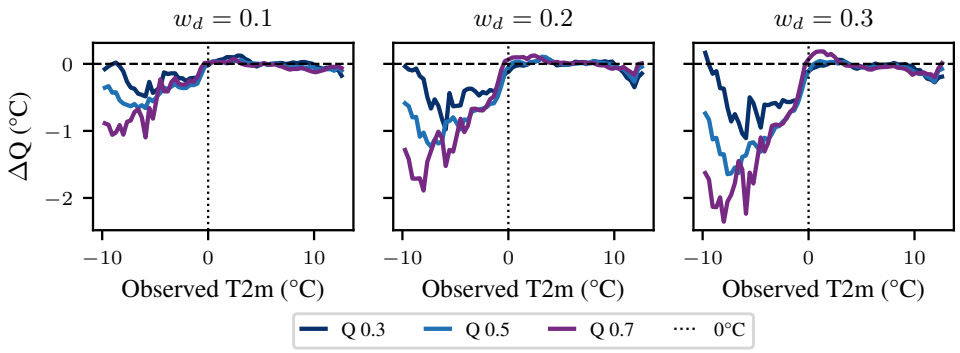


Figure C.25: Rolling quantile shift (relative to $w_d = 0$) for $w_d \in \{0.1, 0.2, 0.3\}$, implicit generative model ($\alpha = 0.3$, window ± 1°C). For observations near the threshold (obs T2m ≈ 0°C), all quantiles shift more uniformly as the whole ensemble straddles the cost-active region. For deep-cold observations (obs T2m < -5°C), the upper quantiles (Q 0.7) are closest to the 0°C threshold and therefore sit closest to the cost-active region, shifting substantially more than the lower quantiles (Q 0.3), which are further from the threshold and receive a weaker gradient. This quantile ordering directly illustrates the sample-level redistribution effect of the decision loss gradient predicted by the gradient analysis in Section 4.2. The effect scales consistently with w_d across all three panels.

C.3.3 Distributional diffusion model results

For the distributional diffusion model, the decision loss does not improve the forecast distributions as expected. The baseline ($w_d = 0$) already exhibits a cold bias: for sub-zero observations, the conditional forecast distribution is shifted toward colder temperatures relative to the observed values, as visible in the KDE plots (Figure C.26). Decision-aware training ($w_d > 0$) does not correct this artifact. energy score degrades monotonically with w_d in all α settings, yet decision miscalibration improves for $\alpha < 0.5$. Notably, the same sigmoid cost function with a distributional diffusion model yields clear improvements on the synthetic decision task (Section 5), so the failure is not inherent to the cost structure or model class. We do not have a complete explanation for this failure mode, one hypothesis is that samples from the baseline model ($w_d = 0$) are already outside of the strong gradient region of the cost function (too cold), thus the decision loss has no handle to correct them. We also note that the reverse chain dynamics at inference time are not captured by the per-step training objective (Section 7), potentially introducing compounding errors at inference.

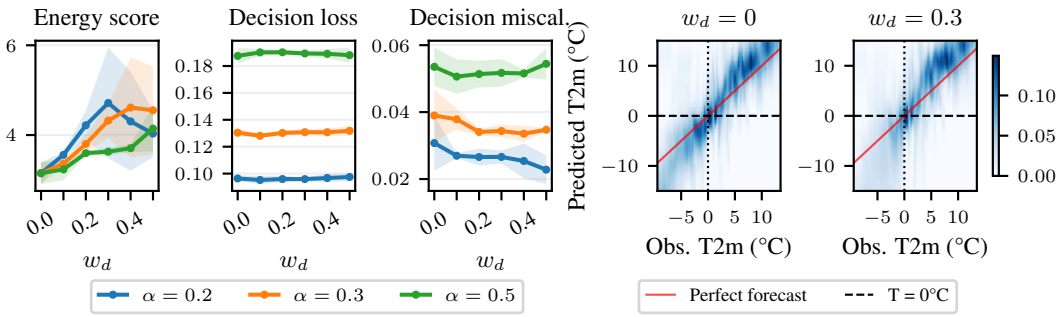


Figure C.26: Frost protection results, distributional diffusion model. Same plot as for implicit generative model in Figure 4. *Left (three panels):* Aggregate energy score, decision loss, and decision miscalibration vs w_d for $\alpha \in \{0.2, 0.3, 0.5\}$ (mean $\pm$ 1 std over seeds). *Right (two panels):* Conditional predictive density for $w_d = 0$ and $w_d = 0.3$ ($\alpha = 0.3$) conditioned on observed temperature. The baseline ($w_d = 0$) already exhibits a cold bias for sub-zero observations visible in the conditional KDE. Decision-aware training ($w_d = 0.3$) does not correct this bias. Energy score degrades monotonically with larger w_d for all α while decision loss stays flat throughout. For $\alpha < 0.5$, decision calibration nonetheless improves with w_d, suggesting the expected cost of the optimal action becomes better aligned with the observed cost even as the overall distributional quality degrades.

D Applicability and Scale

Decision-aware training is beneficial whenever energy score training yields suboptimal approximation errors in cost-relevant regions. This regime arises broadly in practice: approximation errors are unavoidable for any model trained with finite data and limited capacity, and energy score training provides no guarantee that these errors avoid cost-relevant regions.

To characterise how the benefit of decision-aware training scales with training data size, we vary the training set size N_{train} on the synthetic task while holding the model architecture and task fixed. The main-text experiments use $N_{\text{train}} = 1600$; here we sweep $N_{\text{train}} \in \{400, 800, 1600, 3200, 6400, 12800\}$ with three training seeds per size.

Figures D.27 and D.28 summarise the result. The relative improvements in decision loss and decision miscalibration shrink as N_{train} grows and vanish at $N_{\text{train}} = 12800$. The improvement in optimal action (a^*) MAE (mean absolute error of $a^*(x)$ against the ground-truth $a^*(x)$ across test inputs) also shrinks but remains visible across the full range studied. The $a^*(x)$ recovery panels (Figure D.28) show the same story qualitatively: the baseline ($w_d = 0$) fails to track the true $a^*(x)$ at small N_{train} and only converges to it for the largest sizes, whereas decision-aware training tracks $a^*(x)$ accurately throughout.

Operational forecasting tasks typically fall into the regime where decision-aware training helps: meteorological time series exhibit strong temporal and spatial autocorrelation that reduces the effective sample size below the nominal dataset size. Additionally, rare high-impact events are often the most costly, but by definition less frequent, and thus tend to be underweighted by energy score training.

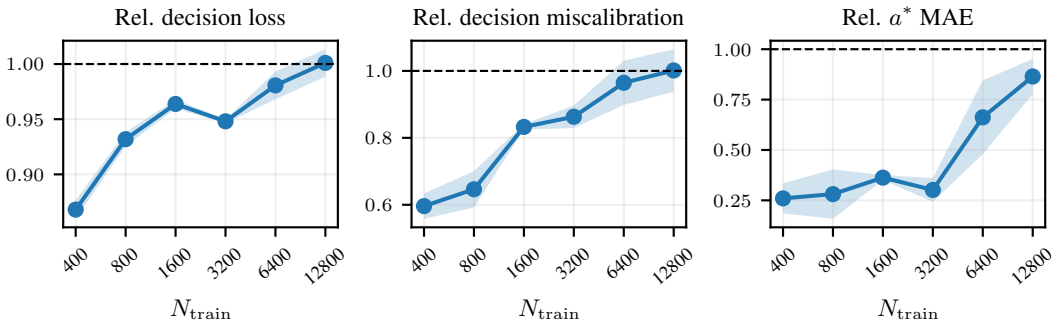


Figure D.27: Relative performance of decision-aware training ($w_d = 0.95$) against the energy-score baseline ($w_d = 0$) on the synthetic decision task (single data seed) for different training set sizes $N_{\text{train}} \in \{400, 800, 1600, 3200, 6400, 12800\}$. Values below 1.0 (dashed line) indicate improvement over baseline. Decision loss and decision miscalibration improvements shrink as N_{train} grows and vanish at $N_{\text{train}} = 12800$. Optimal action a^* MAE improvement also shrinks for larger N_{train} but persists across the range studied. Shaded bands show ± 1 std over training seeds.

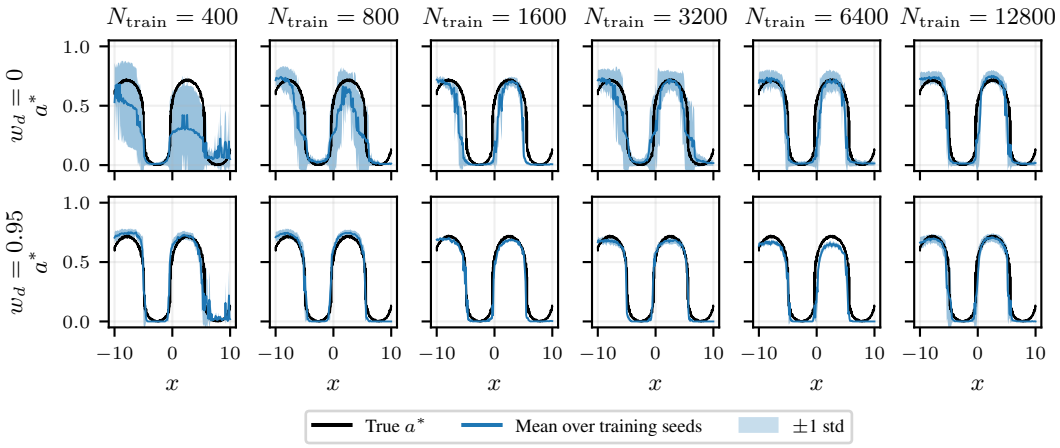


Figure D.28: Optimal action $a^*(x)$ recovery on the synthetic decision task (single data seed) across training set sizes $N_{\text{train}} \in 400, 800, 1600, 3200, 6400, 12800$ for the energy-score baseline ($w_d = 0$, top row) and decision-aware model ($w_d = 0.95$, bottom row). Blue lines show the mean over training seeds, shaded bands show ± 1 std. Decision-aware training recovers the true a^* (black) more accurately. Improvements are largest at small-to-medium N_{train} where the baseline fails to track a^*; at larger N_{train} the baseline already tracks a^* well, leaving less room for improvement. Nonetheless, a^* recovery gains remain visible across the full range studied.

E Choosing the decision weight w_d in practice

The decision weight w_d is a user-preference knob rather than a hyperparameter with a single optimum. It encodes a trade-off between distributional accuracy (small w_d) and decision specialization (large w_d), and the appropriate value depends on how the user weights these two goals and not on the task alone.

Nevertheless, the loss construction makes w_d interpretable and comparable across tasks. The loss-scale estimation step (Section 4.3) normalizes the energy score and the decision loss to comparable magnitude before training. This removes the cost-magnitude degree of freedom, so that $w_d \in [0, 1]$ carries a consistent meaning: the fraction of training signal allocated to the decision objective. Without this normalization, the useful range of w_d would depend on the cost scale: close to zero

(e.g. $[0, 0.1]$) for large-magnitude costs and squeezed toward one (e.g. $[0.9, 0.99]$) for small ones; no shared grid or cross-task comparison would be possible.

In practice, a coarse grid over w_d (e.g. $\{0.3, 0.5, 0.7\}$) is sufficient to identify the optimal range. Each candidate is evaluated on both validation energy score and validation decision cost (or decision miscalibration), and the user selects the largest decision improvement that remains within an acceptable degradation of the energy score. Task structure informs the starting point: when high-cost events are rare (e.g. wind-power cut-off), a larger w_d is needed for the decision term to contribute sufficient gradient signal, whereas for frequent events (e.g. frost) a smaller w_d suffices.