We thank the reviewers for their meaningful and valuable comments, which help to improve the quality of our work.

R1: • **Training details:** your first interpretation is correct: as stated in paper l.149, we independently train STRIPE-shape (resp. STRIPE-time) by optimizing Eq. (1) with $\mathcal{L}_{quality} + \lambda \mathcal{L}_{diversity-shape}$ (resp. $\mathcal{L}_{quality} + \lambda \mathcal{L}_{diversity-time}$) with the same optimization setting. Both models have their own encoder, decoder and diversification mechanism. A variant consists in jointly learning the STRIPE-shape and STRIPE-time modules with the sequential sampling scheme in Algo. 1, which we experimentally found to be inferior, probably due to the increased optimization difficulty of the overall sequential model. Therefore, the sequential shape and time scheme in section 3.2 is only used at inference time. **We will be glad to clarify these details in the training scheme and diversity sampling earlier in section 3, to add pseudo-code to the final version if accepted. We will also publish our code on Github with pretrained models for reproducibility.** In the sequential sampling scheme, the ordering shape+time is actually important since the notion of time diversity between two time series is only meaningful if they have a similar shape (so that computing the DTW optimal path for has a sense). We do not alternate the objectives since this would result in a combinatorial increase in the number of trajectories, and we have a fixed budget of N trajectories during inference.

• **MSE-loss with S+T diversity:** we show below the requested results. It follows the general trends observed in the submission: STRIPE S+T leads to a large diversity gain compared to the baselines without deteriorating quality. We will be glad to add this result in Table 1 of the final paper if accepted.

	$H_{quality}(MSE)$	$H_{quality}(DILATE)$	$H_{diversity}(MSE)$	$H_{diversity}(DILATE)$	CRPS
STRIPE S+T	12.4 ± 1.0	48.7 ± 0.7	18.1 ± 1.6	62.0 ± 5.4	72.2 ± 3.1

• **Eq.(4):** Thank you for your thorough reading, we made a typo ($\mathcal{K}$ instead of $\mathbf{I}$ before the minus sign), which will be corrected. It is equivalent to your proposal since: $\mathcal{L}_{diversity}(\mathcal{K}) = -Tr(\mathbf{I} - (\mathcal{K} + \mathbf{I})^{-1}) = -Tr(\mathcal{K}(\mathcal{K} + \mathbf{I})^{-1})$.

Paper novelty and differences to ref [27] (R2 & R3): our submission addresses the problem of probabilistic time series forecasting by looking for a **set of sharp and diverse predictions**. It is in sharp contrast with the deterministic context of DILATE [27]. Our main differences with [27] are as follows:

1. We propose a diversification mechanism based on DPPs. Although our shape and time diversity criteria are inspired from [27], **we design DPP kernels that we prove to be PSD**, an important requirement for the DPP framework. Note that this property is non-trivial since DTW is not a proper distance (triangle inequality not satisfied).

2. **R3:** We introduce a sequential sampling scheme to disentangle the shape and time diversity features. Ablation studies validate the relevance of this diversification procedure, and **bring a crucial take-home message on the importance of using different losses for prediction quality and diversity**.

Note that our DPP diversification method also paves the way to other PSD kernels based on shape and time (e.g. edit distance, signal correlation) but also to different diversity criteria, e.g. the autoregressive kernel [Cuturi & Doucet'11].

R3 on experiments: We conduct extensive experiments on synthetic and real-world datasets and detailed ablation studies that validate our claims, as highlighted by Reviewers 1,2,4,5. Note that Fig 1 in submission represents a comparative visualization obtained by running source code of deterministic [27] or probabilistic [54] baselines for the synthetic dataset. We will add comparative visualizations for the other datasets in supplementary if accepted.

Non-stationarity and seasonality (R4 and R5): Thank you for your relevant comments, Traffic and Electricity datasets indeed show daily, weakly, yearly periodic patterns. In this work, we are more interested in finer intraday temporal scales, where these signals can present sharp fluctuations that are challenging for many applications (e.g. short-term renewable energy forecasts for load adjustment in a smart-grid). We have not leveraged seasonality-based algorithms to keep the method generic. We can however note that we obtain better results (in paper Table 3) than N-Beats [33], a recent deep method which incorporates seasonality and trend models. Introducing seasonality and extrinsic features (such as special events) is an appealing future perspective to help models focus on the relevant non-stationary parts of these signals, as well as quantifying their impact on diversity.

R5: • **Choice of number N of trajectories:** our motivation is to summarize the predictive distribution by providing a small set ($N = 10$) of diverse and probable scenarii to a decision-maker, especially in the context where sharp fluctuations can occur. Larger values of N would be indeed useful to estimate the full predictive distribution and computing quantiles (which is not our main purpose here). However, note that in paper Fig 5, we show results of our approach STRIPE S+T compared to the strong DeepAR baseline [43] when N increases from 10 to 100. In this case, STRIPE S+T still has a better quality and diversity than DeepAR: at $N = 100$, $H_{quality} = 30.3$, $H_{diversity} = 42.2$ for STRIPE S+T *vs.* 67.7 and 58.9 for DeepAR (lower is better) ; results for $N \in [10, 100]$ will be added in supplementary.

• **DeepAR implementation:** we use the PytorchTS code (which wraps the GluonTS implementation in Pytorch).

Formatting : Thank you, we will correct the acronym typos (R2), correct the DeepAR reference (R5) and check all other references.