	`electricity`	`traffic`	`wiki`	`dominick`
# time series	370	963	9013	100014
time granularity	hourly	hourly	daily	weekly
domain	$\mathbb{R}_+$	$[0,1]$	$\mathbb{N}$	$(-100,100)$
encoder length	96	96	60	16
decoder length	24	24	60	12
batch size	256	256	256	256
learning rate	5e-3	5e-3	5e-3	5e-3
learning rate decay factor	0.4	0.4	0.4	0.4
# learning rate max decays	4	4	4	4
early stopping patience	20	20	20	20
# LSTM layers	2	2	2	2
# LSTM nodes	80	80	80	80

Table 3: Dataset details and RNN parameters.

A Details on Accuracy Scores

The summations of all the metrics are over all time series, i.e., $i = 1, \dots N$, and over the whole prediction range, i.e., $t = T - t_0 + 1, \dots, T$, unless the range is explicitly provided. The value m in the MASE and MSIS metrics is the seasonal frequency which is set to 1 for yearly , weekly and daily data, 4 for quarterly data, 12 for monthly data, and 24 for hourly data, following the M4 competition definitions. Finally, ρ defines the prediction interval, i.e., $\rho = 0.05$ for a 95% interval, and $\hat{u}$, $\hat{l}$ are the $1 - \rho/2$, $\rho/2$ quantiles of the predictive distribution.

$$\text{QL\#} = \frac{\sum_{i,t} \Lambda(\hat{q}(\#), z_{i,t})}{\sum_{i,t} |z_{i,t}|}, \tag{18}$$

$$\text{QLm} = \frac{1}{99}(\text{QL1} + \text{QL2} + \cdots + \text{QL99}), \tag{19}$$

$$\text{ND} = \frac{\sum_{i,t} |z_{i,t} - \hat{z}_{i,t}|}{\sum_{i,t} |z_{i,t}|}, \tag{20}$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N(T-t_0)} \sum_{i,t} (z_{i,t} - \hat{z}_{i,t})^2}}{\frac{1}{N(T-t_0)} \sum_{i,t} |z_{i,t}|}, \tag{21}$$

$$\text{sMAPE} = \frac{1}{N(T-t_0)} \sum_{i,t} \frac{2|z_{i,t} - \hat{z}_{i,t}|}{|z_{i,t}| + |\hat{z}_{i,t}|} \tag{22}$$

$$\text{MASE} = \frac{1}{N(T-t_0)} \sum_{i} \frac{\sum_t |z_{i,t} - \hat{z}_t|}{\frac{1}{(T-m)} \sum_{t=m+1}^{T} |z_{i,t} - z_{i,t-m}|}, \tag{23}$$

$$\text{MSIS} = \frac{1}{N(T-t_0)} \sum_{i} \frac{\sum_t \hat{u}_{i,t} - \hat{l}_{i,t} + \frac{2}{\rho}(\hat{l}_{i,t} - z_{i,t})\mathcal{I}_{[z_{i,t} < \hat{l}_{i,t}]} + \frac{2}{\rho}(z_{i,t} - \hat{u}_{i,t})\mathcal{I}_{[z_{i,t} > \hat{u}_{i,t}]}}{\sum_{t=m+1}^{T} |z_{i,t} - z_{i,t-m}|}. \tag{24}$$

B Details on Data sets and Hyperparameters

All the hyperparameters (shown in Table 3) were selected by performing a grid search only on the `electricity` dataset, and were used as default values on all the other datasets, i.e., `traffic`, `wiki`, `dominick`, as well as the dataset of the M4 competition.

Evaluating the CRPS Integral for Linear Splines

In Section 3.3 we gave an analytic expression for the CRPS integral when the quantile function is a linear spline. Here we will provide the basic steps of the solution. Consider a value $\tilde{a}$ such that $q(\tilde{a}) = z$. Then, the CRPS integral can be written as follows:

$$\int_0^1 2\Lambda_\alpha(q(\alpha), z)d\alpha = \int_0^1 2(\alpha - \mathcal{I}_{[z<q(\alpha)]})(z - q(\alpha)) \tag{25}$$

$$= \int_0^{\tilde{a}} 2\alpha(z - q(\alpha)) + \int_{\tilde{a}}^1 2(\alpha - 1)(z - q(\alpha)), \tag{26}$$

since $q(\alpha)$ is non-decreasing.

The value of $\tilde{a}$ can be found by solving the equation $q(\tilde{a}) = z$:

$$\gamma + \sum_{l=0}^{L} b_l(\tilde{a} - d_l)_+ = z \Longleftrightarrow \tilde{a} = \frac{z - \gamma + \sum_{l=0}^{l_0} b_l d_l}{\sum_{l=0}^{l_0} b_l}, \tag{27}$$

where the index l_0 of the summation, with $0 \leq l_0 \leq L$ and $d_0 = 0$, is such that $d_{l_0} \leq \tilde{a} \leq d_{l_0+1}$, since the terms for $l = l_0 + 1, \ldots, L$ become zero due to the $(\cdot)_+$ function.

The index l_0 is straightforward to compute: we know that $q(d_{l_0}) \leq q(\tilde{a}) \leq q(d_{l_0+1})$, therefore we can sequentially evaluate the spline at the knot points d_l, $\forall l$ and find the largest knot d_l such that $q(d_l) \leq q(\tilde{a})$. The index of this knot is the index l_0 and it can be found in $O(L)$ time.