

**Table A1.** Hyperparameter description of the developed models.

Model type	Model hyperparameters	Value range of model hyperparameters
Statistical time series	$\delta_1$ : Number of autoregressive terms.	No less than 0.
	$\delta_2$ : Number of nonseasonal difference.	No less than 0.
	$\delta_3$ : Number of lagged forecast errors.	No less than 0.
Shallow learning	$\alpha_1$ : Number of neighbors to use by default for $k$ neighbors queries.	Start: 1, end: 9, step: 2.
	$\alpha_2$ : Power parameter for the Minkowski metric.	Start: 1, end: 4, step: 1.
	$\alpha_3$ : The regularization parameter.	Choose from 0.015, 0.01, 0.03, 0.15, 0.2, 0.5, 1.0, 1.5, 50, 70, or 90.
	$\alpha_4$ : Maximum number of iterations taken for the solvers to converge.	Choose from 50, 70, 90, 1500, 2000, or 2500.
	$\alpha_5$ : The maximum depth of the tree.	Start: 4, end: 8, step: 1.
Deep learning	$\beta_1$ : Number of channels for first layer Conv1D.	Choose from 1, 2, 4, or 8.
	$\beta_2$ : Kernel size for first layer Conv1D.	Choose from 1 or 3.
	$\beta_3$ : Type of the RNN cell.	Choose from “GRU” or “LSTM”.
	$\beta_4$ : Number of RNN cells for each layer.	Choose from 1, 2, 4, 8, 16, 32, or 64.
	$\beta_5$ : Type of the RNN cell for the skip layer.	Choose from “GRU” or “LSTM”.
	$\beta_6$ : Number of RNN cells for each layer for skip part.	Choose from 1, 2, 4, 8, 16, 32, or 64.
	$\beta_7$ : Dropout regularization parameter.	Follow a quantized uniform distribution with a minimum of 0, a maximum of 0.5, and a step size of 0.05.
	$\beta_8$ : Number of samples per batch.	Follow a quantized random integer distribution between 16 and 128, with a step size of 16.
	$\beta_9$ : Max epochs during training.	Follow a quantized random integer distribution between 50 and 300, with a step size of 16.
	$\beta_{10}$ : A tuning parameter in an optimization algorithm.	Uniformly distributed between 1e-4 and 1e-2.
	$\beta_{11}$ : Number of epochs to wait for improvement before terminating.	Follow a quantized random integer distribution between 5 and 20, with a step size of 5.
	$\beta_{12}$ : The type of the specific paddle RNN module.	Choose from “LSTM” or “GRU”.
	$\beta_{13}$ : A list containing the dimensions of the hidden layers of the fully connected NN.	Choose randomly from a set of options, including single-layer configurations with 16, 32, 64, or 128 units, as well as two-layer and three-layer configurations with the same number of units per layer.
	$\beta_{14}$ : The number of features in the hidden state $h$ of the RNN module.	Follow a quantized random integer distribution between 32 and 256, with a step size of 32.
	$\beta_{15}$ : The number of recurrent layers.	Randomly generated integer between 1 and 3.
	$\beta_{16}$ : The fraction of neurons that are dropped in all-but-last RNN layers.	Draw from a quantized uniform distribution ranging from 0 to 0.5 in steps of 0.05.
	$\beta_{17}$ : Boolean value indicating whether the generic architecture of N-BEATS is used.	Randomly choose from “True” or “False”.
	$\beta_{18}$ : The number of stacks that make up the whole model.	Randomly generated integer between 2 and 5.
	$\beta_{19}$ : The number of blocks making up each stack.	Randomly generated integer between 2 and 5.
	$\beta_{20}$ : The number of fully connected layers preceding the final forking layers in each block of every stack.	Randomly generated integer between 1 and 5.
	$\beta_{21}$ : Determines the number of neurons that make up each fully connected layer in each block of every stack.	Follow a quantized random integer distribution between 32 and 256, with a step size of 32.
	$\beta_{22}$ : The dimensionality of the waveform generator parameters.	Follow a quantized random integer distribution between 32 and 256, with a step size of 32.
	$\beta_{23}$ : The degree of the polynomial used as waveform generator in trend stacks.	Randomly generated integer between 2 and 5.
	$\beta_{24}$ : The number of heads in the multi-head attention mechanism.	Randomly choose from 1, 2, 4, or 8.
	$\beta_{25}$ : The number of encoder layers in the encoder.	Randomly generated integer between 1 and 10.
	$\beta_{26}$ : The number of decoder layers in the decoder.	Randomly generated integer between 1 and 10.
	$\beta_{27}$ : The dimension of the feedforward network model.	Follow a quantized random integer distribution between 32 and 256, with a step size of 32.
	$\beta_{28}$ : The activation function of encoder/decoder intermediate layer.	Randomly choose from “ReLU” or “GeLU”.
	$\beta_{29}$ : The expected feature size for the input/output of the transformer’s encoder/decoder.	Follow a quantized random integer distribution between 32 and 256, with a step size of 32.
Ensemble learning	$\theta_1$ : The maximum depth of the tree.	Start: 4, end: 8, step: 1.
	$\theta_2$ : The number of trees in the forest.	Start: 50, end: 100, step: 10.