WaveBound: Dynamic Error Bounds for Stable Time Series Forecasting

Youngin Cho*, Daejin Kim*, Dongmin Kim, Mohammad Azam Khan, Jaegul Choo (*: equal contribution)

TL;DR

- Recent time series forecasting models suffer from fitting to irreducible noise in time series data.
- In image classification, flooding regularization[1] addresses overfitting by preventing the training loss from falling below a certain level.
- However, in time series forecasting, the difficulties of predictions are varied greatly for each feature and time step, even at different iterations.
- In this work, we propose dynamic error bounds for preventing the overfitting issue in time series forecasting.

Overfitting & Zero Training Loss

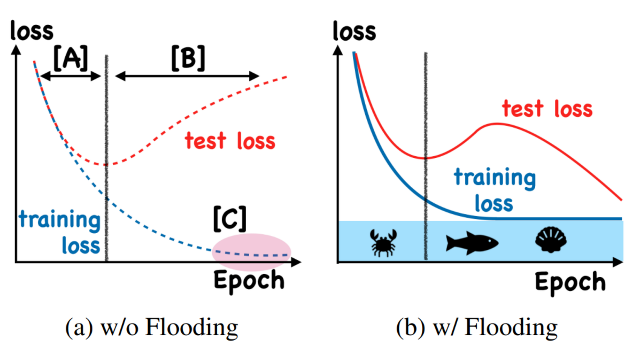
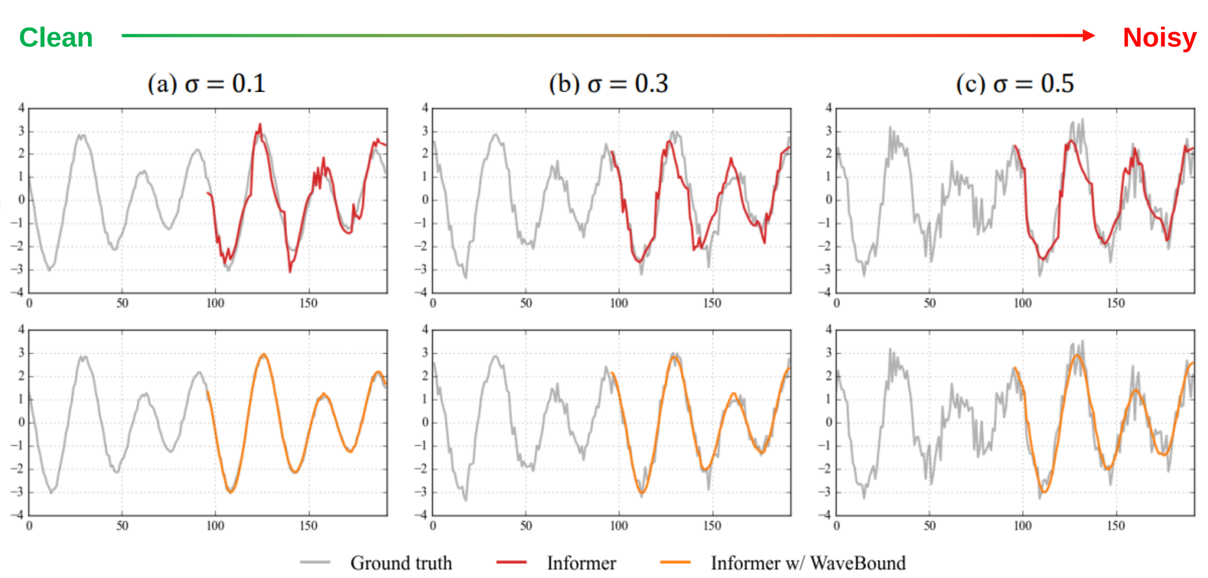


Image from Ishida *et al.* [1]

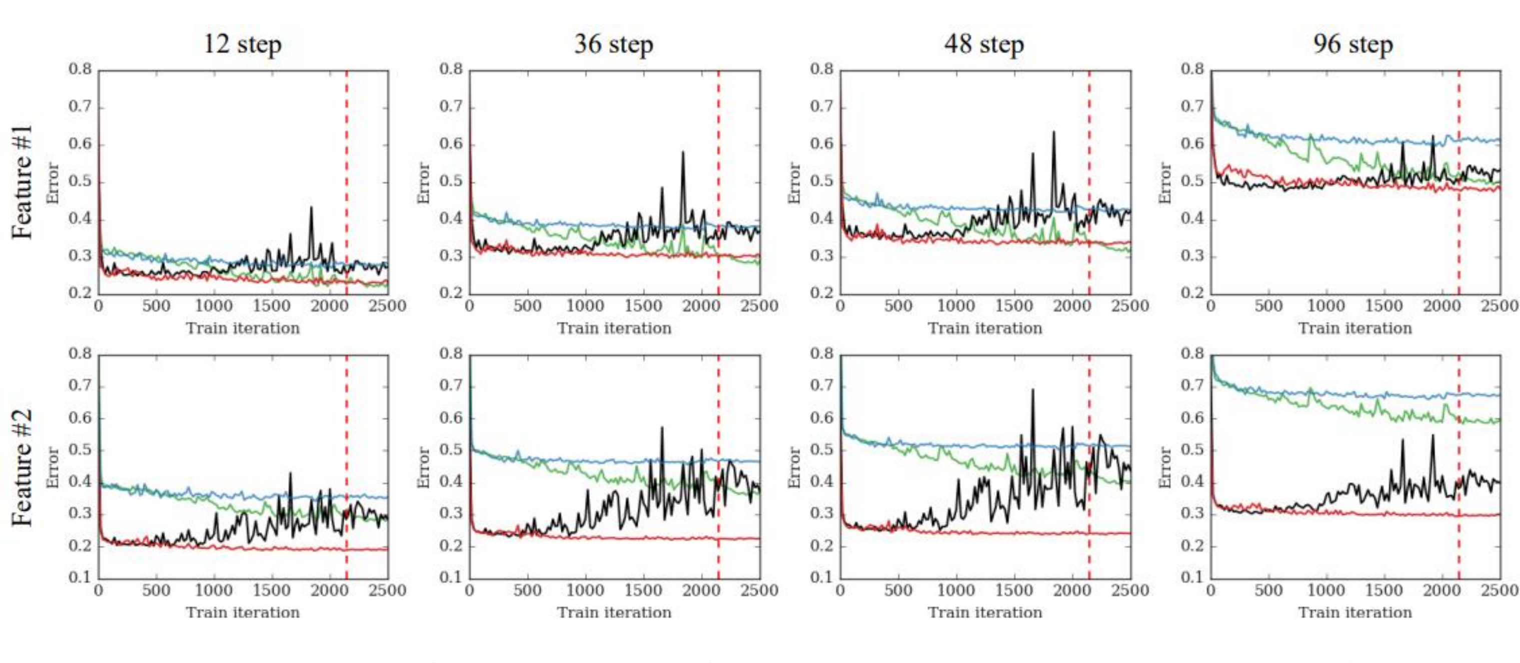
- Zero training loss implies that the model memorizes all training data.
- Flooding regularization [1] is proposed for addressing the overfitting issue by preventing the training loss from falling below a certain level.

Overfitting in Time Series Forecasting

- Recent forecasting models cannot properly address the overfitting issue.



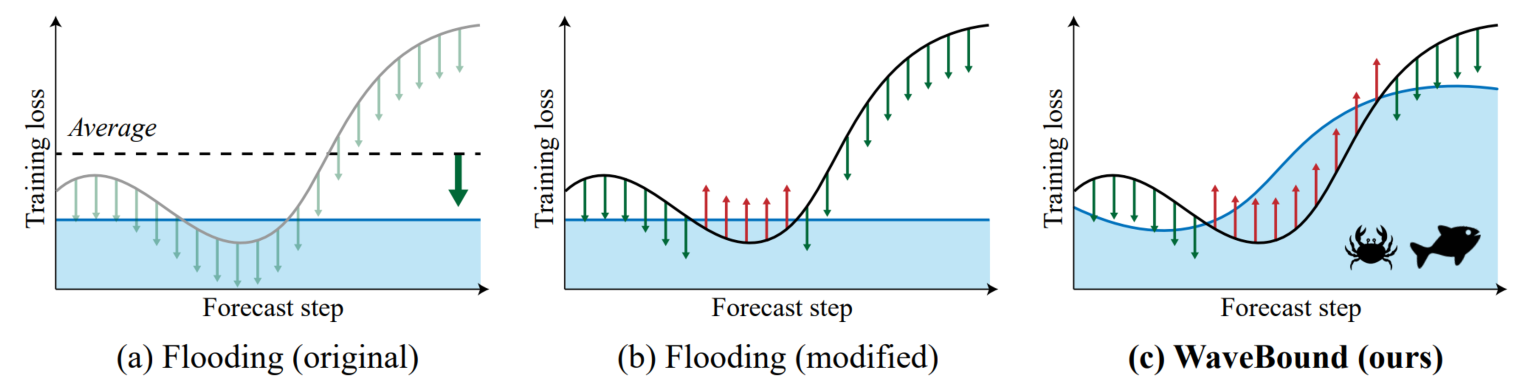
- In time series forecasting, the difficulty of prediction varies with each feature and time step. → Existing methods, such as early stopping and the original flooding regularization, cannot work effectively.



Proposed Method: WaveBound

To address the overfitting issue in time series forecasting, we design a novel regularization called **WaveBound** by dynamically finding the error bounds for each feature and time step at each train iteration.

Q. How to design proper error bounds?

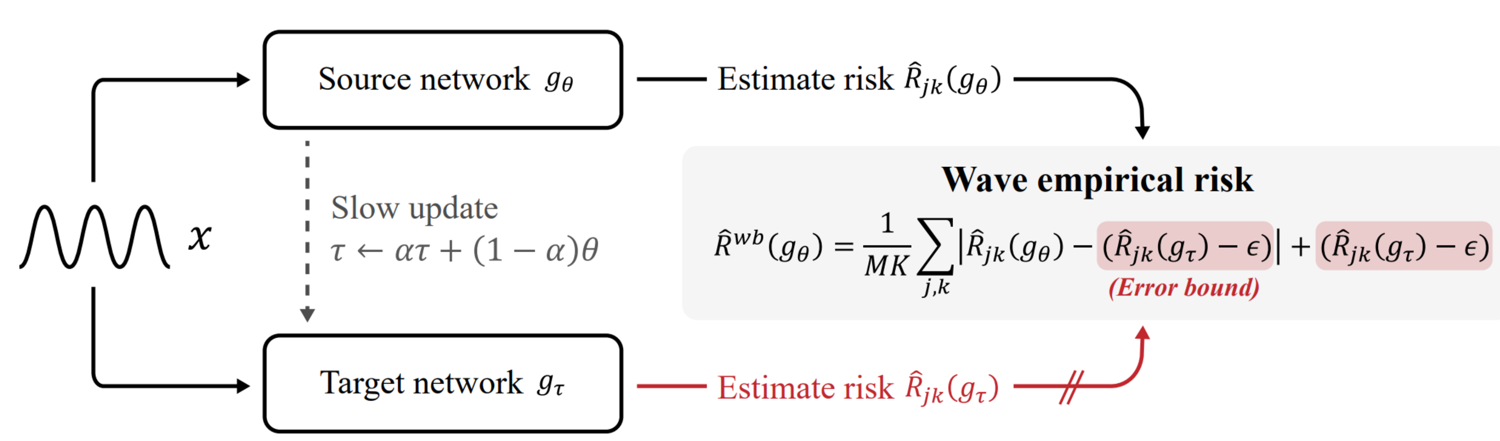


$|\hat{R}(g) - b|$ $\quad$ $\frac{1}{MK}\sum_{j,k}|\hat{R}_{jk}(g) - b|$ $\quad$ $\frac{1}{MK}\sum_{j,k}|\hat{R}_{jk}(g) - \hat{R}_{jk}(g^*) + \epsilon|$

- **(a)** The original flooding only bounds the 'mean' of the training loss, which is not feasible for time series forecasting.
- **(b)** Even if we consider each feature and time step, the fixed level of the error bound cannot consider the different difficulties of predictions for each feature and time step.
- **(c)** So, for each train iteration, we have to **estimate** the proper error bound for each feature and time step individually.

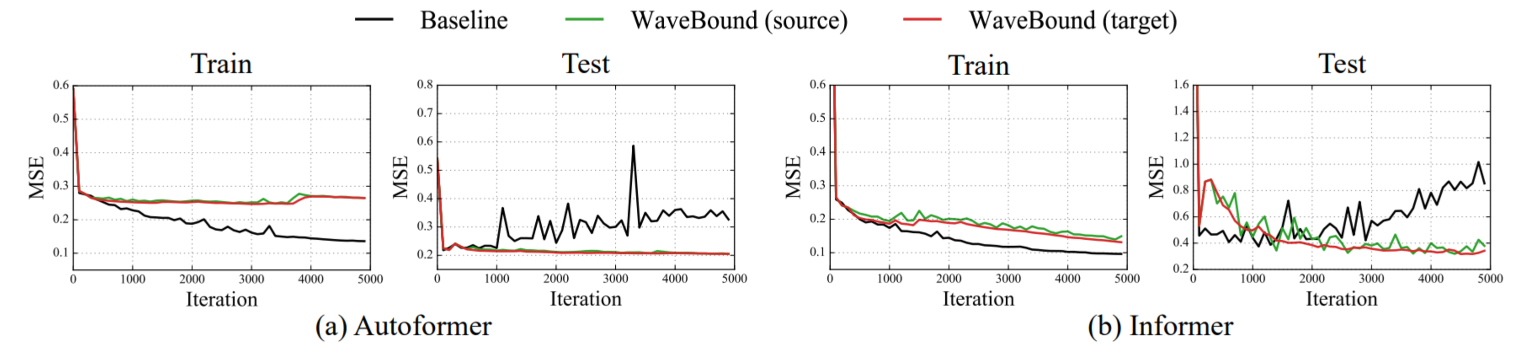
Method	Estimated risk (w/o constant)		Autoformer 96	Autoformer 336	Pyraformer 96	Pyraformer 336	Informer 96	Informer 336	LSTNet 96	LSTNet 336
Base model	$\hat{R}(g)$	MSE	0.202	0.247	0.256	0.278	0.335	0.369	0.268	0.284
		MAE	0.317	0.351	0.360	0.383	0.417	0.448	0.366	0.382
Flooding [10]	$\lvert\hat{R}(g) - b\rvert$	MSE	0.194	0.247	0.257	0.277	0.335	0.368	0.268	0.284
		MAE	0.309	0.351	0.360	0.382	0.416	0.447	0.366	0.381
Constant flooding	$\frac{1}{MK}\sum_{j,k}\lvert\hat{R}_{jk}(g) - b\rvert$	MSE	0.198	0.247	0.257	0.277	0.333	0.369	0.268	0.284
		MAE	0.314	0.351	0.360	0.382	0.415	0.448	0.366	0.382
WaveBound (Avg.)	$\lvert\hat{R}(g) - \hat{R}(g^*) + \epsilon\rvert$	MSE	0.194	0.221	0.248	0.288	0.302	0.322	0.208	0.246
		MAE	0.309	0.331	0.352	0.388	0.388	0.407	0.314	0.356
WaveBound (Indiv.)	$\frac{1}{MK}\sum_{j,k}\lvert\hat{R}_{jk}(g) - \hat{R}_{jk}(g^*) + \epsilon\rvert$	MSE	**0.176**	**0.217**	**0.241**	**0.269**	**0.289**	**0.305**	**0.185**	**0.217**
		MAE	**0.288**	**0.327**	**0.345**	**0.371**	**0.378**	**0.394**	**0.291**	**0.326**

Q. How to estimate error bounds?



To provide the dynamic error bounds, we employ the model updated with an exponential moving average (EMA). EMA model is known to be robust to noise and to memorize previous training data (in time series forecasting, memorizing the generalized pattern is important for handling noise!)

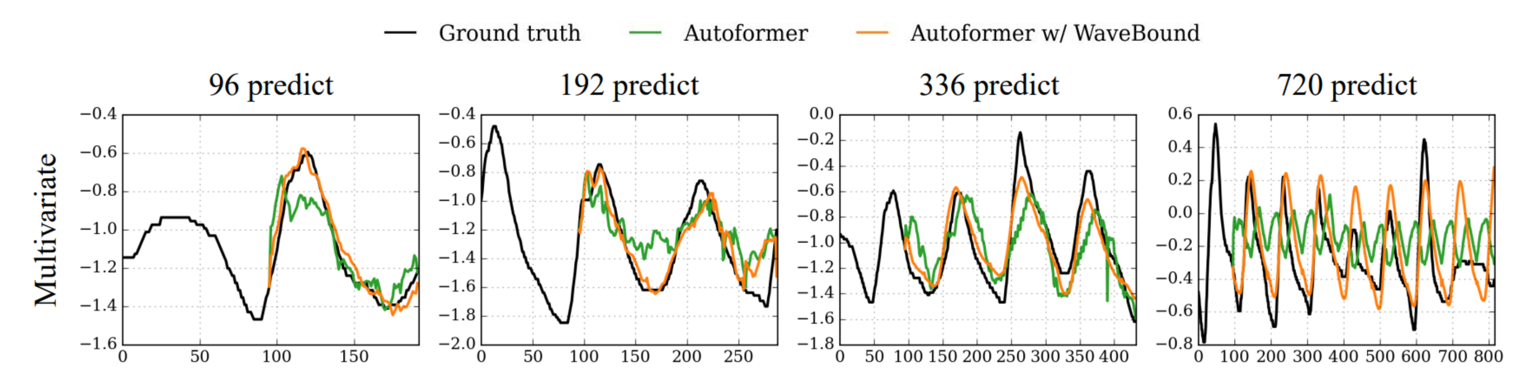
Generalization Gaps



Without WaveBound, the test loss of both Autoformer and Informer increases after a certain iteration, while the training loss of both models continues to decrease toward zero.

Forecasting Performance

Regardless of the model and dataset, the forecasting performance of the existing models is significantly improved when we use WaveBound.



Models		Autoformer [5] Origin MSE	Autoformer [5] Origin MAE	Autoformer [5] w/ Ours MSE	Autoformer [5] w/ Ours MAE	Pyraformer [6] Origin MSE	Pyraformer [6] Origin MAE	Pyraformer [6] w/ Ours MSE	Pyraformer [6] w/ Ours MAE	Informer [7] Origin MSE	Informer [7] Origin MAE	Informer [7] w/ Ours MSE	Informer [7] w/ Ours MAE	LSTNet [14] Origin MSE	LSTNet [14] Origin MAE	LSTNet [14] w/ Ours MSE	LSTNet [14] w/ Ours MAE
ETTm2	96	0.262	0.326	**0.204**	**0.285**	0.363	0.451	**0.281**	**0.386**	0.376	0.477	**0.334**	**0.429**	0.455	0.511	**0.268**	**0.368**
	192	0.284	0.342	**0.265**	**0.322**	0.708	0.648	**0.624**	**0.599**	0.751	0.672	**0.698**	**0.631**	0.706	0.660	**0.464**	**0.508**
	336	0.338	0.374	**0.320**	**0.356**	1.130	0.846	**1.072**	**0.829**	1.440	0.917	**1.087**	**0.845**	1.161	0.868	**0.781**	**0.695**
	720	0.446	0.435	**0.413**	**0.408**	2.995	1.386	**1.917**	**1.119**	3.897	1.498	**2.984**	**1.411**	3.288	1.494	**2.312**	**1.239**
ECL	96	0.202	0.317	**0.176**	**0.288**	0.256	0.360	**0.241**	**0.345**	0.335	0.417	**0.289**	**0.378**	0.268	0.366	**0.185**	**0.291**
	192	0.235	0.340	**0.205**	**0.317**	0.272	0.378	**0.256**	**0.360**	0.341	0.426	**0.298**	**0.388**	0.277	0.375	**0.197**	**0.304**
	336	0.247	0.351	**0.217**	**0.327**	0.278	0.383	**0.269**	**0.371**	0.369	0.448	**0.305**	**0.394**	0.284	0.382	**0.217**	**0.326**
	720	0.270	0.371	**0.260**	**0.359**	0.291	0.385	**0.283**	**0.377**	0.396	0.457	**0.311**	**0.398**	0.316	0.404	**0.250**	**0.350**
Exchange	96	0.153	0.285	**0.146**	**0.274**	**0.604**	**0.624**	0.615	0.627	0.979	0.791	**0.878**	**0.765**	0.483	0.518	**0.357**	**0.432**
	192	0.297	0.397	**0.262**	**0.373**	0.982	0.806	**0.953**	**0.797**	1.147	**0.854**	**1.136**	0.859	0.706	0.646	**0.621**	**0.593**
	336	0.438	0.490	**0.425**	**0.483**	1.264	**0.934**	**1.263**	0.944	1.592	1.014	**1.461**	**0.992**	1.055	0.800	**0.837**	**0.691**
	720	1.207	0.860	**1.088**	**0.810**	1.663	1.051	**1.562**	**1.016**	2.540	1.306	**2.496**	**1.294**	2.198	1.127	**1.374**	**0.894**
Traffic	96	0.645	0.399	**0.596**	**0.352**	0.635	0.364	**0.622**	**0.341**	0.731	0.412	**0.671**	**0.364**	0.735	0.446	**0.587**	**0.356**
	192	0.644	0.407	**0.607**	**0.370**	0.658	0.376	**0.646**	**0.355**	0.751	0.422	**0.666**	**0.360**	0.750	0.446	**0.595**	**0.365**
	336	0.625	0.390	**0.603**	**0.361**	0.668	0.377	**0.653**	**0.355**	0.822	0.465	**0.709**	**0.387**	0.778	0.455	**0.623**	**0.378**
	720	0.650	0.398	**0.642**	**0.383**	0.698	0.390	**0.672**	**0.364**	0.957	0.539	**0.764**	**0.421**	0.815	0.470	**0.648**	**0.383**
Weather	96	0.294	0.355	**0.227**	**0.296**	0.235	0.321	**0.193**	**0.272**	0.378	0.428	**0.355**	**0.415**	0.237	0.310	**0.202**	**0.275**
	192	0.308	0.368	**0.283**	**0.340**	0.340	0.415	**0.306**	**0.372**	0.462	0.467	**0.424**	**0.448**	0.277	0.343	**0.254**	**0.316**
	336	0.364	0.396	**0.335**	**0.370**	0.453	0.484	**0.403**	**0.441**	0.575	0.535	**0.506**	**0.484**	0.326	0.378	**0.309**	**0.358**
	720	0.426	0.433	**0.401**	**0.411**	0.599	0.563	**0.535**	**0.519**	1.024	0.751	**0.972**	**0.712**	0.412	0.431	**0.398**	**0.415**
ILI	24	3.468	1.299	**3.118**	**1.200**	4.822	1.489	**4.679**	**1.459**	5.356	1.590	**4.947**	**1.494**	7.934	2.091	**6.331**	**1.816**
	36	3.441	1.273	**3.310**	**1.240**	4.831	**1.479**	**4.763**	1.483	5.131	1.569	**5.027**	**1.537**	8.793	2.214	**6.560**	**1.848**
	48	3.086	1.184	**2.927**	**1.128**	4.789	1.465	**4.524**	**1.439**	5.150	1.564	**4.920**	**1.514**	7.968	2.068	**6.154**	**1.779**
	60	2.843	1.136	**2.785**	**1.116**	4.876	1.495	**4.573**	**1.465**	5.407	1.604	**5.013**	**1.528**	7.387	1.984	**6.119**	**1.758**

... see our paper for more results

Find more details in our paper! (arxiv.org)

References

[1] Takashi Ishida, Ikko Yamane, Tomoya Sakai, Gang Niu, and Masashi Sugiyama. Do we need zero training loss after achieving zero training error? In Proc. the International Conference on Machine Learning (ICML), 2020.