# Stability of loss functions for solar wind forecasting using Deep Learning


Mario Cobos Maestre   Manuel Flores-Soriano   David F. Barrero

Universidad de Alcalá


Universidad de Alcalá   ISG Intelligent Systems Group   SWE Space Weather Group

## Introduction

- Loss functions are essential for neural network training
- Lack of stability might lead to an inability to converge

* For simplicity, we assume operation on a normalized space with values in [-1, 1]

## Definitions

- **Stability**: resilience in the presence of frequent outliers in the distribution
- $\mathcal{X}$: any given dataset
- $\mathcal{Y}$: labels in a dataset, such that $\mathcal{Y} \subseteq \mathcal{X}$
- $\hat{\mathcal{Y}}$: set of outputs of deep neural network model $f$ for all $x \in \mathcal{X}$, such that $\hat{\mathcal{Y}} \subseteq \mathcal{X}$
- $N = \text{Size}(\mathcal{Y}) = \text{Size}(\hat{\mathcal{Y}})$. For plotting purposes, we fix $N = 300$
- $\mathcal{G}(\mu, \sigma)$ denotes a Gaussian distribution with mean $\mu$ and standard deviation $\sigma$

## MSE and MAE

- Both functions have a lower bound of zero
- Both functions are even
- Mean Squared Error:

$$\text{MSE}(\mathcal{Y}, \hat{\mathcal{Y}}) = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N}, \; y_i \in \mathcal{Y}, \; \hat{y}_i \in \hat{\mathcal{Y}}$$

- Mean Absolute Error:

$$\text{MAE}(\mathcal{Y}, \hat{\mathcal{Y}}) = \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{N}, \; y_i \in \mathcal{Y}, \; \hat{y}_i \in \hat{\mathcal{Y}})$$

## Individual contribution to output

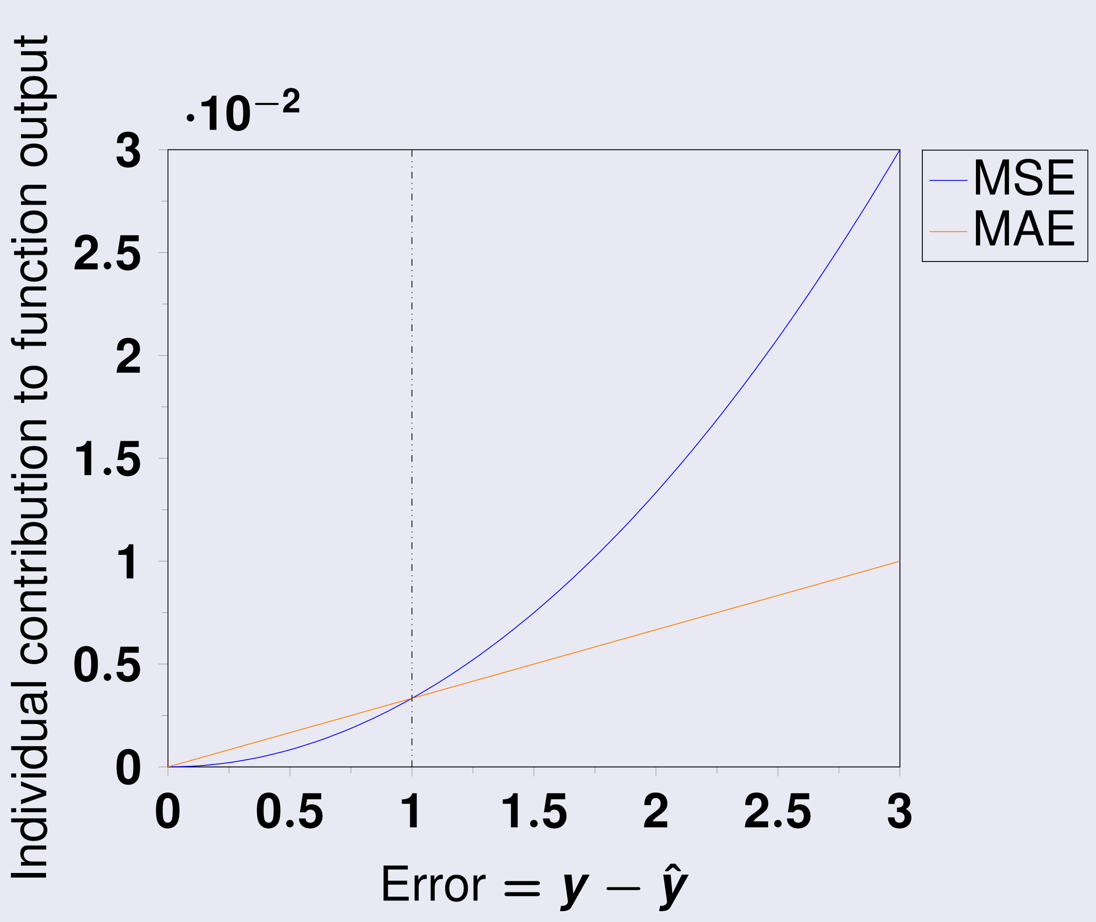
- MSE:

$$\text{MSE}_i(y_i, \hat{y}_i) = \frac{(y_i - \hat{y}_i)^2}{N}, \; y_i \in \mathcal{Y}, \; \hat{y}_i \in \hat{\mathcal{Y}}$$

- MAE:

$$\text{MAE}_i(y_i, \hat{y}_i) = \frac{|y_i - \hat{y}_i|}{N}, \; y_i \in \mathcal{Y}, \; \hat{y}_i \in \hat{\mathcal{Y}}$$

**Compared individual error contribution to function output**




## Derivability

- Second derivative provides information about the curvature of the loss function
- Let $E_i = y_i - \hat{y}_i$:

$$\frac{d\text{MSE}_i}{dE_i} = \frac{2E_i}{N} \rightarrow \frac{d^2\text{MSE}_i}{dE_i} = \frac{2}{N}$$

$$\frac{d\text{MAE}_i}{dE_i} = \frac{\text{Sign}(E_i)}{N} \rightarrow \frac{d^2\text{MAE}_i}{dE_i} = 0$$

- MSE codifies a more curved space than MAE

## Acknowledgements


This work was supported by project PID2020-119407GB-I00/AEI/10.13039/501100011033


## Considerations

- MSE weights higher errors more heavily
- MSE is less resilient to outliers than MAE $\rightarrow$ MSE is less stable than MAE
- MSE's approach can accelerate convergence in favorable scenarios $\rightarrow$ What conditions need to be fulfilled?

## Impact of data distribution

Consider a random distribution $\mathcal{D}$, with mean $\mu$ and standard deviation $\sigma$. If we take a random sample $s$, and the model currently assumes $\mathcal{D}$ to be $\mathcal{G}(\mu, \sigma)$:

- $s < \mu - \sigma$ or $s > \mu + \sigma \rightarrow E > 1.0$
- $\mu - \sigma \leq s \leq \mu + \sigma \rightarrow 0.0 \leq E \leq 1.0$

We consider $s < \mu - \sigma$ and $> \mu + \sigma$ as outliers. Depending on $\mathcal{D}$:

- $\mathcal{D} \sim \mathcal{G}(\mu, \sigma) \rightarrow P(s < \mu - \sigma) \sim P(s > \mu + \sigma) \ll P(\mu - \sigma \leq s \leq \mu + \sigma)$
- $\mathcal{D} \nsim \mathcal{G}(\mu, \sigma) \rightarrow$ the distribution needs to be examined to determine probabilities

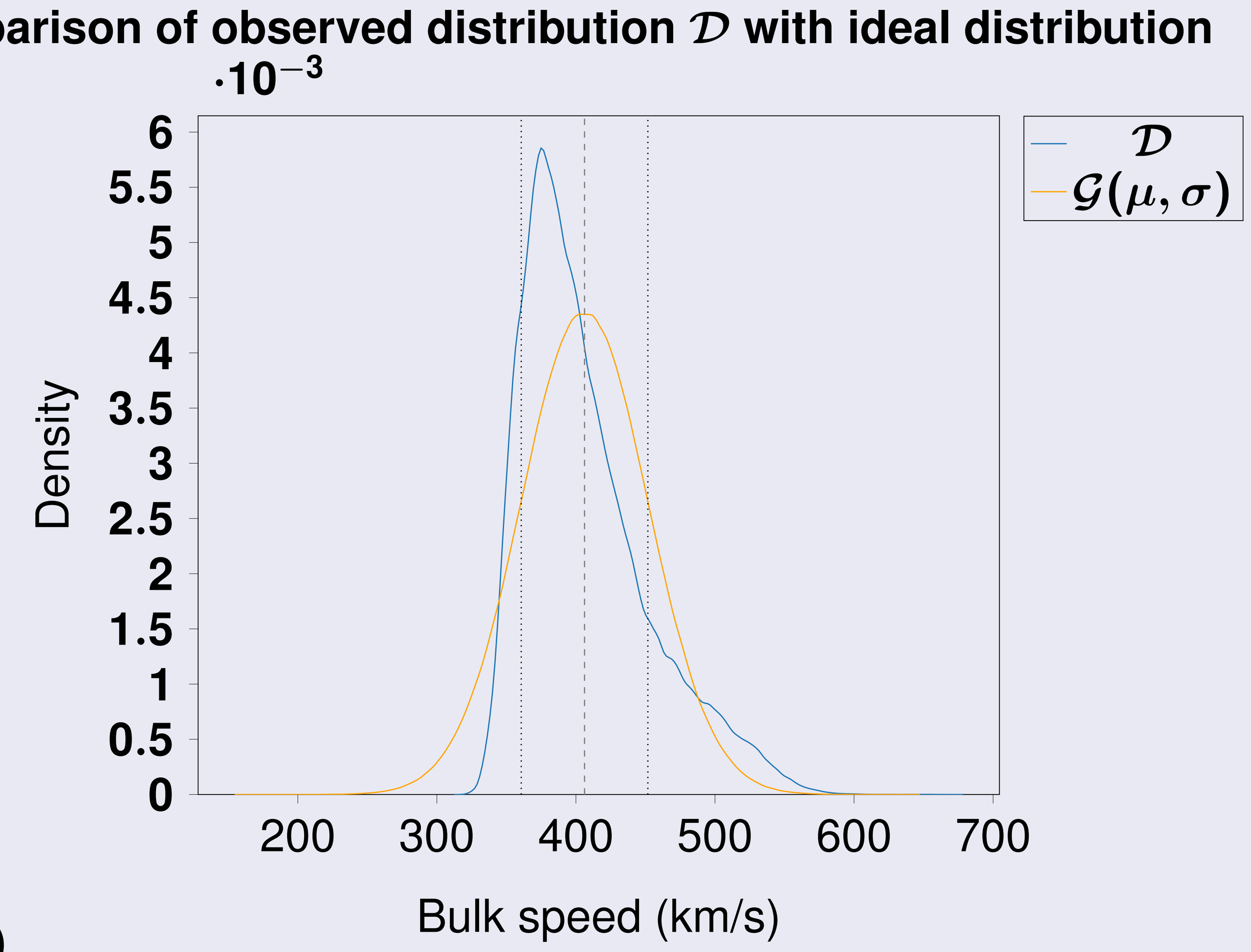
For $\mathcal{D} \nsim \mathcal{G}(\mu, \sigma)$:

- $P(s < \mu - \sigma) > P(s > \mu + \sigma) \rightarrow$ MSE will lead to the resulting model preferring lower values
- $P(s < \mu - \sigma) < P(s > \mu + \sigma) \rightarrow$ MSE will lead to the resulting model preferring higher values
- $P(s < \mu - \sigma) \sim P(s > \mu + \sigma) \sim P(\mu - \sigma \leq s \leq \mu + \sigma) \rightarrow$ MSE might not be able to converge

## Solar wind data distribution

- Dataset ranges from 1 January 2011 to 1 September 2021

**Comparison of observed distribution $\mathcal{D}$ with ideal distribution**




$\mathcal{G}(\mu, \sigma)$

- $\mathcal{D} \nsim \mathcal{G}(\mu, \sigma)$
- $P(s < \mu - \sigma) \sim P(s > \mu + \sigma) \sim P(\mu - \sigma \leq s \leq \mu + \sigma)$ observed
- MSE can lead to marked bias
  - The constant change in bias can lead to an inability to converge

## Conclusions

- MAE is a more stable loss function than MSE
- MSE is applicable to most scenarios
- MSE should converge in heavy tail distributions, at the expense of biasing the resulting model
- Solar wind bulk speed data is not normally distributed
- MAE is preferred as a loss function for solar wind bulk speed regression using deep neural networks

## References


1. Bar-Lev, S. K., Boukai, B. & Enis, P. On the mean squared error, the mean absolute error and the like. *Communications in Statistics - Theory and Methods* **28,** 1813–1822 (1999).
2. Willmott, C. J. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30,** 79–82 (2005).
3. Qi, J., Du, J., Siniscalchi, S. M., Ma, X. & Lee, C.-H. On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Processing Letters* **27,** 1485–1489 (2020).
4. Larrodera, C. & Cid, C. Bimodal distribution of the solar wind at 1 AU. *A&A* **635,** A44 (2020).