A Symmetric Relative-Error Loss Function for Intermittent Multiscale Signal Modelling

Sergio M. Vanegas Arias[1] , **Lasse Lensu**[1] and **Fredy Ruiz Palacios**[2]
[1]Department of Computational Engineering, LUT University, Lappeenranta, Finland
[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
sergio.vanegas.arias@lut.fi

Abstract

Multiscale signals represent a formidable modelling challenge in Machine Learning as the ubiquitous Mean Squared Error loss function neglects signal behaviour at smaller values. Several scale-equalizing error metrics have been devised to tackle this problem, amongst which the Mean Absolute Percentage Error (MAPE) remains the most widely used due to its simplicity and interpretability. However, by its very definition, MAPE introduces three major issues: asymptotic behaviour at zero-target values, asymptotic gradient behaviour at zero error, and accuracy loss for large signal scales. We address these limitations by proposing the Symmetric Mean Arctangent Squared Percentage Error (SMASPE), which builds up from the Mean Arctangent Absolute Percentage Error (MAAPE) and leverages a mathematically smoother definition along with user-provided signal bounds to extend its functionality. The numerical properties of SMASPE are explored, and its performance is tested in two real-life cases for deterministic and stochastic optimization. The experiments show a clear advantage of the proposed loss function, with an improvement of up to 42% with respect to MAAPE in terms of Mean Absolute Error for deep learning models when appropriate bounds are selected.

1 Introduction

When performing regression tasks, regardless of context or modelling technique, the most common way to determine the fitness of a model is to calculate the Mean Squared Error (MSE). First described in [Legendre, 1806] as an algebraic procedure for fitting linear equations to data and later proven to be optimal for the lowest-sampling-variance estimator under unbiased assumptions [Henderson, 1975], it has persisted as the go-to loss function in the field of Machine Learning (ML) due to its context-free nature and, more importantly in the context of back-propagation, function smoothness. Unfortunately, this generality renders the MSE unsuitable for measuring model fitness when the sample space is not normally distributed or signal outliers are significant [Liano, 1996]. Such is the case of multiscale modelling problems, amongst which physical models such as the Navier-Stokes equation [Steinhauser, 2016] or intermittent signal forecasts like that of Photo-Voltaic (PV) energy generation can be mentioned.

To illustrate the problems associated with minimizing the MSE for multiscale spaces, a synthetic example is presented: let $\mathbf{y}$ be the target signal defined by

$$\begin{aligned} y_i &= 1 + \sin^2\left(\frac{2\pi i}{20}\right)(1 + 9\beta_i) \\ \beta_i &\sim \mathcal{B}(1, 0.15), \quad i \in \{1, 2, \ldots, 1000\}, \end{aligned} \tag{1}$$

where each β_i is a sample drawn from a single-trial Binomial distribution with a 0.15 probability of success. Two dynamical system estimators, namely $\hat{\mathbf{y}}^{(\mathbf{1})}$ and $\hat{\mathbf{y}}^{(\mathbf{2})}$, are defined as

$$\hat{y}_i^{(1)} = y_i \cdot |\eta_i|, \quad \eta_i \sim \mathcal{N}(1.0, 0.15), \tag{2}$$

$$\hat{y}_i^{(2)} = \begin{cases} \hat{\theta} & \text{if } \beta = 0 \\ y_i & \text{otherwise,} \end{cases} \tag{3}$$

$$\hat{\theta} \coloneqq \arg\min_\theta \int_0^\pi \left[1 + sin^2 x - \theta\right]^2 dx.$$

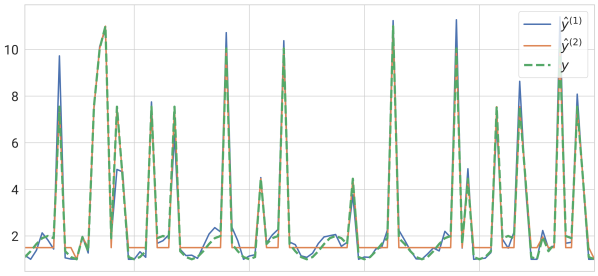
The first estimator introduces multiplicative noise to the reference signal. In contrast, the second one perfectly captures the peaks introduced by the intermittent signal β while holding the constant that minimizes the squared error in its absence. Simulating these conditions yields the curves shown in Figure 1a and MSEs of 0.215 and 0.107 for estimators 1 and 2 respectively, which empirically evidences how this metric mathematically neglects small-scale dynamics.

In ML regression, the traditional way to cope with these variations in scale is to fit the models to minimize the Mean Absolute Percentage Error (MAPE) [De Myttenaere *et al.*, 2016], defined as

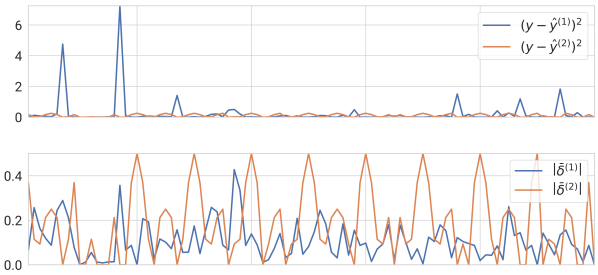
$$\mathrm{MAPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| = \frac{1}{N}\sum_{i=1}^{N}|\bar{\delta}_i|, \tag{4}$$

where $\bar{\delta}_i \coloneqq \frac{y_i - \hat{y}_i}{y_i}$ is a term henceforth referred to as relative sample error.

The MAPE divides each sample's error by its target value, making errors at lower scales numerically comparable to those of larger reference magnitude. Figure 1b illustrates the *equalization* effect of this approach, with the MAPE of the



(a) Reference signal and Estimators as defined in Equations (2) and (3).



(b) Squared and Relative Errors as *perceived* by the loss functions.

Figure 1: Synthetic multiscale example (100-sample window).

noisy signal now halving that of the small-scale-smoothing one (0.113 and 0.199 respectively). This effectively means a model trained to minimize the MAPE would better capture the behaviour of the signal. Unfortunately, the instability of this approach quickly becomes self-evident, since $|\bar{\delta}| \to \infty$ as $|y| \to 0$ save for a perfect prediction, and even then stability cannot be guaranteed due to quantization.

Division by 0 is not uncommon in informatics and mathematical analysis, and some standard approaches have been proposed to handle this exception for regression applications [Hyndman and Koehler, 2006; Hodson *et al.*, 2021]. However, [Kim and Kim, 2016] took an unconventional approach by interpreting $\bar{\delta}$ as the hypotenuse slope of the "right triangle" defined by the perpendicular sides $|y - \hat{y}|$ and $|y|$ to introduce Mean Arctangent Absolute Percentage Error (MAAPE), which measures the average angle of this slope as

$$\mathrm{MAAPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N} \arctan\left|\bar{\delta}_i\right|. \tag{5}$$

Albeit a clever solution to the asymptotic behaviour at $|y_i| \to 0^+$, choosing to preserve the absolute value of this ratio as a loss function introduces unwanted gradient behaviour at $\hat{y}_i \to y_i$. A smoother gradient can thus be attained by relaxing this one-to-one correspondence with the MAPE slope. Furthermore, by exploiting enforceable bounds, we can extend the definition to better capture the peak values of the signal without losing the small-scale dynamics.

This work modifies MAAPE and uses this extension as a functional unit to formulate the Symmetric Mean Average Squared Percentage Error (SMASPE). The specific contributions presented in this paper are the following:

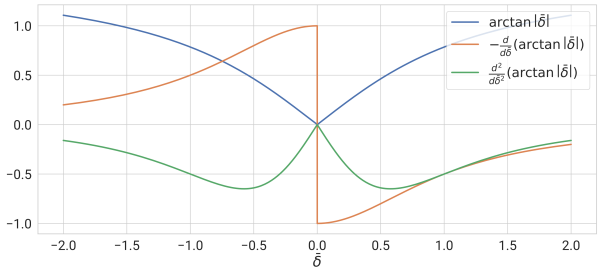


Figure 2: Sample-wise behaviour of the MAAPE and its first two derivatives as a function of the error ratio.

1. The formulation of Mean Arctangent Squared Percentage Error (MASPE), a quadratic-relative-error variation of the MAAPE, and an analytical study of the mathematical properties that render it a more suitable loss function for gradient-based optimization.
2. SMASPE, a symmetrical extension of the MASPE that exploits heuristic signal bounds to allow models to capture multiscale dynamics, and a graphical study of its behaviour as a function of the predicted and reference signal values.
3. A numerical validation of SMASPE's adequacy for deterministic and stochastic optimization algorithms.

The remainder of this paper is organized as follows. In Section 2, the benefits and drawbacks introduced using the above ratio's square instead of its absolute value are explored as a function of the auxiliary variable $\bar{\delta}$. Section 3 presents the symmetric bound-enforcement and studies its implications as a function of the reference and predicted values y and $\hat{y}$. In Section 4, two real-world examples are used to showcase the capabilities of SMASPE. Finally, in Section 5, a summary of the obtained experimental results and possible extensions to the proposed methodology are laid out.

The code and complementary material is made available as a GitHub repository at [Vanegas Arias, 2025].

2 Unidimensional Analysis: Improvements over MAAPE

Figure 2 illustrates the MAAPE and its derivatives as functions of $\bar{\delta}$, with the first derivative being multiplied by -1 to highlight the direction of minimization. As mentioned in Section 1, optimizing the absolute value of this ratio makes this function $\mathcal{C}^0$ (not continuously differentiable) over the real domain, which means it presents asymptotic behaviour at the origin and thus renders the optimization problem numerically unstable. This is the same phenomenon encountered when using the Mean Absolute Error (MAE) as a loss function, which makes the MSE better suited as a minimization target even if the MAE is being used as the main performance metric.

By loosening the trigonometrical correspondence with the MAPE and focusing on the utility of MAAPE as a minimiza-

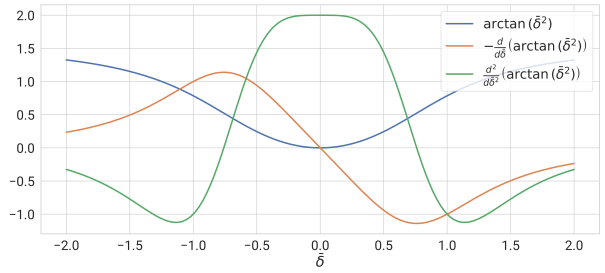


Figure 3: Sample-wise behaviour of the MASPE and its first two derivatives as a function of the error ratio.

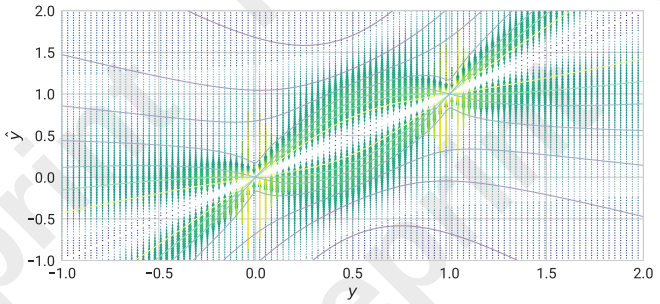


Figure 4: 2D behaviour of the SMASPE (isocontours) and its negative gradient w.r.t. $\hat{y}$. The colour scale is a logarithmic function of the magnitudes, where the yellow regions surrounding the user-set bounds ($\{y^-, y^+\} = \{0, 1\}$) indicate a larger gradient magnitude and the blue ones denote zones with a vanishing gradient.

tion target, the equivalent MASPE can be derived from Equation (5), yielding

$$\text{MASPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N} \arctan \bar{\delta}_i^2. \tag{6}$$

Since the arctangent is monotonically increasing, the predictor $\hat{\mathbf{y}}$ that minimizes MASPE for a given reference $\mathbf{y}$ also minimizes MAAPE. However, unlike its absolute-valued counterpart, Equation (6) belongs to $\mathcal{C}^\infty$ (proof in complementary material) over the real domain. Thus, its gradient lacks MAAPE's discontinuity at the origin and, consequently, its second derivative is no longer asymptotic at 0 error, as shown in Figure 3.

It is worth noting that the price the MAAPE (and by extension the MASPE) pays for its boundedness is the vanishing gradient at infinity. Indeed, in both Figures 2 and 3 it is evident how the loss functions converge to $\pi/2$ as $\bar{\delta} \to \pm\infty$, making its derivative converge to 0 by consequence. However, as shown in Section 3, this issue can be mitigated by enforcing loss-function symmetry w.r.t. known signal bounds.

3 Bidimensional Analysis: Symmetric Bounds

If by minimizing the expected squared error the MSE neglects small-scale system dynamics, then the relative-error approaches reviewed so far suffer an equivalent loss of accuracy at larger scales. Referring back to the synthetic example presented in Equation (1), the estimator defined by Equation (2) would be perceived to be twice as accurate as the one detailed in Equation (3) by the $\bar{\delta}$-based loss functions (MASPE included) despite consistently failing to capture the peak values of the intermittent signal. Therefore, an extension of the metrics explored so far that considers the behaviour of the loss as a function of both y and $\hat{y}$ is required.

Even if a system's dynamics are unknown, its domain can often be either characterized or empirically determined from data and thus be exploited to equalize measurable outputs reliably. Looking at the MASPE as a bidimensional function of y and $\hat{y}$, let $[y^-, y^+] \subset \mathbb{R}$ be a real, finite interval bounding the target signal; i.e., $-\infty < y^- \leq y_i \leq y^+ < \infty \quad \forall i$. Then, loss symmetry is enforced w.r.t. these bounds as

$$\begin{aligned}\text{SMASPE}\left(\mathbf{y}, \hat{\mathbf{y}}\right) =& \text{MASPE}\left(\mathbf{y} - y^-, \hat{\mathbf{y}} - y^-\right) \\ &+ \text{MASPE}\left(y^+ - \mathbf{y}, y^+ - \hat{\mathbf{y}}\right),\end{aligned} \tag{7}$$

thus defining the SMASPE. Since this modification is merely the sum of two smooth functions evaluated on shifted (inverted) versions of the original signals y and $\hat{y}$, the differentiability is preserved.

The second term of the above equation inverts the signal w.r.t. y^+ so that large-scale dynamics are perceived as small-scale ones and vice versa. This effectively yields a numerically stable loss function within the user-defined interval, as illustrated by Figure 4, preventing the vanishing error gradients towards the upper bound of the modelled signal.

The $[y^-, y^+]$ interval can be loosened as

$$[y^-_{\text{loose}}, y^+_{\text{loose}}] = [y^- - \gamma\Delta_y, y^+ + \gamma\Delta_y], \tag{8}$$

where $\Delta_y := y^+ - y^-$ and γ is a relaxation coefficient. This mitigates SMASPE's asymptotic behaviour near the symmetry bounds, further increasing the robustness of the approach. However, if overdone, the function may lose track of the difference in scale over the signal range, forfeiting the advantages of the relative-error approach. As evidenced by the vector field intensity in Figure 4, the trade-off SMASPE makes for its stability and domain coverage is its reliance on carefully defined bounds, which should follow a user-defined heuristic.

4 Applications

This section showcases the performance and versatility of the SMASPE with two real-life datasets. In Section 4.1, a comparative study for the real-life application presented in [Kim and Kim, 2016], Stock-Keeping Unit (SKU)-demand forecasting, is performed to illustrate the characteristics of SMASPE w.r.t. the State-of-the-Art (SotA). Then, in Section 4.2, a Deep Neural Network (DNN) is trained to forecast PV-generation using the SMASPE to highlight the advantages of the proposed loss function in backpropagation. In both experiments, the SMASPE is tested using tight and loose bounds ($\gamma = 0.1$, sensitivity analysis in complementary material).

The experiments were implemented in Python using the *Jax* [Bradbury *et al.*, 2018] and *Keras 3* [Chollet and others, 2015] stacks for ML model construction and training.

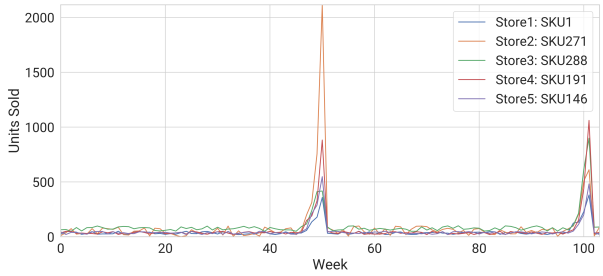
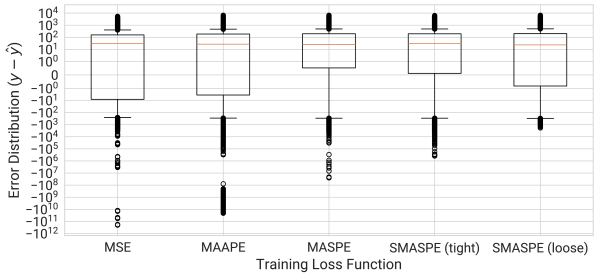
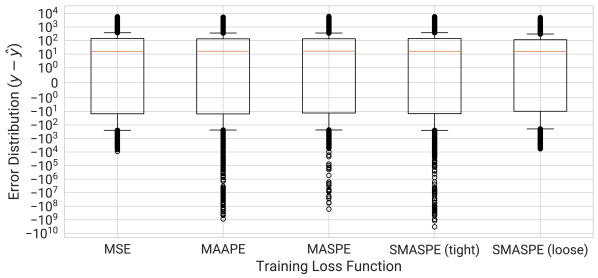


Figure 5: Intermittent demand patterns of four SKUs.



(a) ARIMA model.



(b) HW model.

Figure 6: SKU demand model error distribution boxplots.

4.1 BFGS Optimization: Stock-Level Next-Sample Forecasting

As the first experiment, the performance of the loss functions proposed in this paper was tested for the same use-case studied in [Kim and Kim, 2016]. The experiments in this section were carried out using a Kaggle dataset [Veera, 2020] which, to the authors' knowledge, is the only open-access dataset sampled with the same frequency (i.e., weekly) and containing approximately the same number of samples per sequence as the one used for the original MAAPE paper.

The dataset comprises 6000 time series, each containing 103 samples, representing the weekly sales volume per product across five retail stores, with one example from each of them being shown in Figure 5. Although not confirmed by the dataset description, it is reasonable to infer that the two sales peaks, approximately one year apart from each other, correspond to the sales near the holiday season, which explains the signal synchronization. Following the methodology of the MAAPE proposers [Kim and Kim, 2016], the first 95 samples were used to fit the models and the remaining 8 to evaluate their out-of-sample performance.

The considered models were the worst and best performers in [Kim and Kim, 2016]; namely, non-seasonal Auto-Regressive Integrated Moving Average (ARIMA) and Holt-Winters (HW) additive method (Triple Exponential Smoothing), respectively. Considering the length of the dataset, and since the authors did not provide the model parameters, a $(4, 1, 4)$ order ARIMA and a HW with seasonality 4 were fitted separately to each time series, initializing the ARIMA model weights to 0 and the convex sum coefficients of the HW method to 0.5. These weights were adjusted by minimizing the MSE, MAAPE, MASPE, and SMASPE[1]. The performance of the fitted models was evaluated using all loss functions tested, the MAE, and their median convergence time (measured in iterations) in all sequences.

Table 1 contains the numerical results of the described experimental pipeline evaluated over the **saturated** model outputs, since negative demand values are not valid. From these results, the following can be stated:

- Despite the HW model converging generally faster than the ARIMA due to its lower parameter count (3 for HW vs. 10 for ARIMA), the previously worst-performing architecture ends up being the most accurate according to all error metrics when fitted using the SMASPE with moderately loose bounds. This, however, comes at an evident cost in convergence speed, requiring the most iterations out of all evaluated model-loss configurations.
- Figure 6 provides some insight on MAAPE's poor performance for training the HW models. Indeed, despite all loss functions producing a similar distribution box, the MAAPE yields significantly more outliers of larger magnitude than the MSE or the loose SMASPE, explaining the difference in order of magnitude for the metrics in Table 1. This phenomenon could remain hidden for a small sample size, which might explain why the results in this paper contradict those of [Kim and Kim, 2016].
- Finally, even when performing numerically similar (on average) to its loose counterpart while converging significantly faster, the worst-case scenarios in Figure 7 illustrate how tight bounds for the SMASPE often lead to unstable model weights. This pathological behaviour makes itself numerically evident through the MAE and MSE metrics for the HW model.

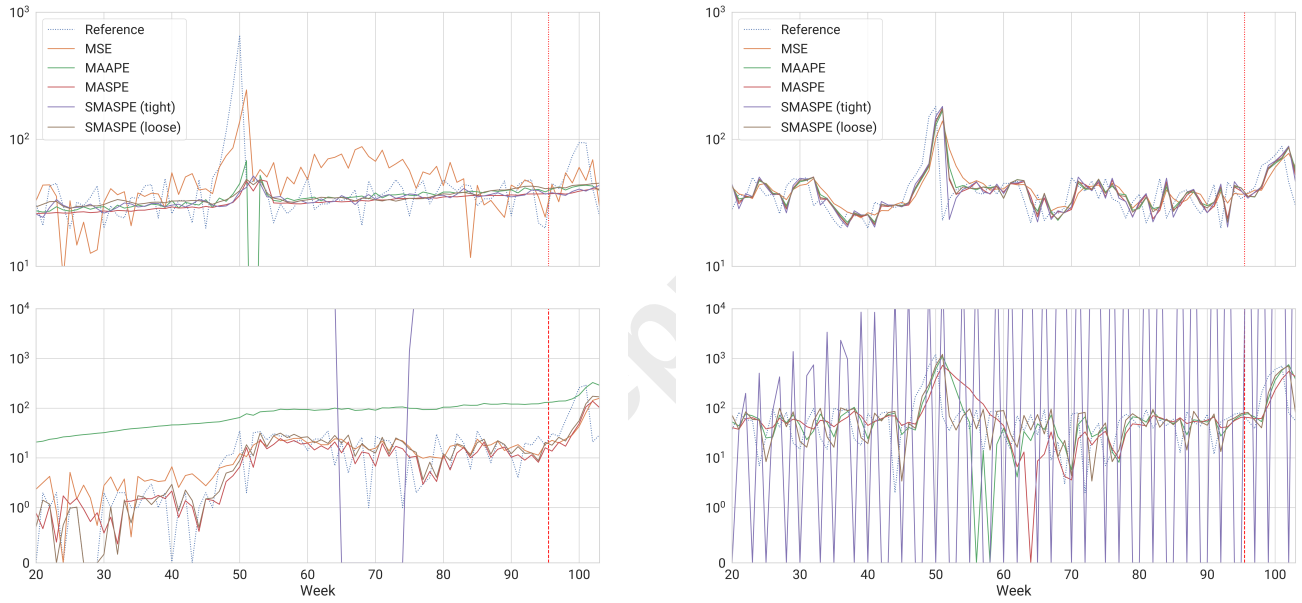
4.2 Backpropagation: PV Generation 3-Day-Ahead Forecasting

The second dataset is the *PV energy generation*. It comprises 821 full days (after clean-up) of hourly-sampled floating-point data including timestamps, weather measurements, and average generated power in kilowatts. Furthermore, the output signal was normalized from 0 to ~ 1 by dividing each daily curve over the median peak of the 15 preceding days to avoid

[1]Tight bounds set to each sequence's extrema.

Model	Loss Function	Median Iteration Number	Error Metrics (Out-of-Sample)					
			MAE	MSE	MAAPE	MASPE	SMASPE (tight)	SMASPE (loose)
ARIMA	MSE	23	4.32×10^{7}	6.70×10^{18}	0.648	0.554	1.07	0.653
	MAAPE	5	3.19×10^{8}	2.79×10^{18}	0.638	0.562	1.13	0.700
	MASPE	28	4.66×10^{3}	9.43×10^{10}	0.575	0.486	1.03	0.630
	SMASPE (tight)	9	3.09×10^{2}	1.62×10^{7}	0.608	0.527	1.07	0.657
	SMASPE (loose)	43	$\mathbf{2.08 \times 10^{2}}$	$\mathbf{2.25 \times 10^{5}}$	**0.528**	**0.429**	0.934	**0.553**
HW	MSE	**2**	2.31×10^{2}	2.35×10^{5}	0.614	0.522	1.00	0.636
	MAAPE	3	8.12×10^{4}	2.97×10^{13}	0.638	0.543	1.03	0.712
	MASPE	5	8.78×10^{3}	6.93×10^{11}	0.620	0.519	1.00	0.688
	SMASPE (tight)	4	1.79×10^{5}	2.68×10^{14}	0.637	0.543	1.03	0.716
	SMASPE (loose)	4	2.32×10^{2}	2.53×10^{5}	0.558	0.441	**0.912**	0.614

Table 1: SKU forecasting numerical results (the best performer per metric is highlighted).



(a) ARIMA model.

(b) HW model.

Figure 7: SKU forecasts for the best (top) and worst (bottom) case scenarios for both architectures according to the out-of-sample Integral Absolute Error (IAE). The vertical red line divides the in-sample and out-of-sample regions of the time series.

numerically neglecting generation values during low-sunlight seasons such as winter and prolonged storms, thus setting a heuristic for the SMASPE bounds.

Missing samples in the dataset impose a natural division that was used to determine the training, validation, and test partitions, resulting in a $59.1/29.7/11.2\%$ scheme. In addition to the historical PV generation signal, global radiation and air temperature measurements were used as exogenous features, the forecasts of which can be retrieved during deployment and used as additional inputs. Since only the historical measurements of weather variables were logged, forecast uncertainty was simulated by introducing additive noise with

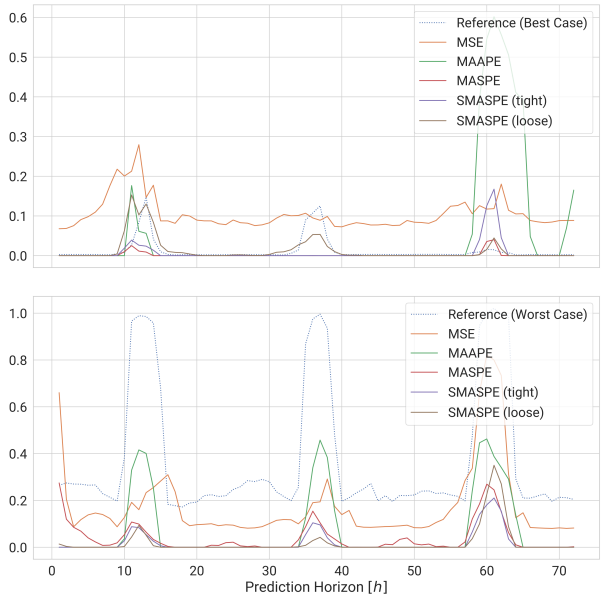


Figure 8: Best (top) and Worst (bottom) Case PV-generation forecasts.

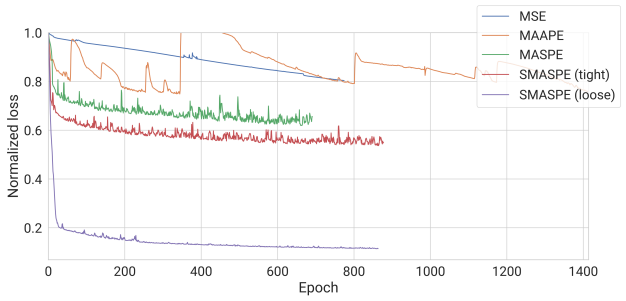
exponentially increasing variance; namely,

$$\begin{aligned} \tilde{x}[n] &= x[n] + \epsilon_x[n], \\ \epsilon_x &\sim \mathcal{N}(0, \varsigma^2 e^{2\beta n} \sigma_x^2), \\ n &\in \mathbb{N}_0, \end{aligned} \tag{9}$$

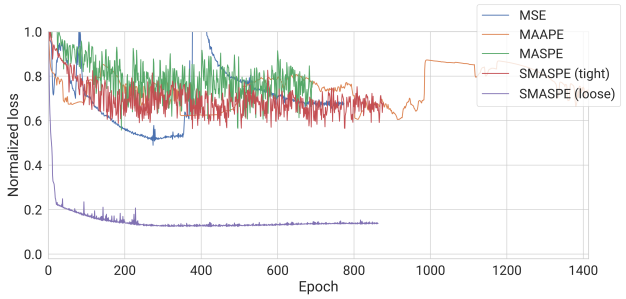
where σ_x denotes the variance of the original variable x, ς represents the standard deviation of the white noise for the first sample over the prediction horizon, and β sets the exponential growth factor of the noise variance as a function of the sample index over the prediction horizon. These parameters were calibrated to ensure that the noise had a standard deviation of 0.1% of each feature's standard deviation at the beginning of the forecast horizon and increased tenfold by the end of the forecast window (i.e. $\beta = \frac{\ln 10}{71}$).

The chosen architecture was an encoder/decoder Recursive Neural Network (RNN) with Gated-Recurrent Unit (GRU) layers similar to the scheme proposed by [Cho, 2014] with its hyperparameters fixed (1737 weights in total) for 72-hour context and forecast windows. The weights were randomly initialized five times and used as common starting points for all loss functions considered in Section 4.1, keeping the weights that achieved the median validation loss. These were minimized using the *Adam* optimizer (Learning Rate of 5×10^{-4}, 64 sequences per batch) for a maximum of 2000 epochs (500 patience epochs). Finally, rather than the actual extrema of the training partition, the expected $[0, 1]$ interval was used to define the tight bounds.

Table 2 contains the numerical results of the described experimental pipeline evaluated on the **saturated** model outputs, since negative generation values are not possible in this scenario. Figure 8 showcases the best and worst-case PV-generation forecasts based on the IAE for the displayed prediction horizon. From these, the following can be stated:



(a) Loss history for the training partition (unsaturated output signal to avoid gradient clipping).



(b) Loss history for the validation partition (saturated output signal).

Figure 9: PV-generation model loss values during training normalized w.r.t. their a-priori value for the best trials.

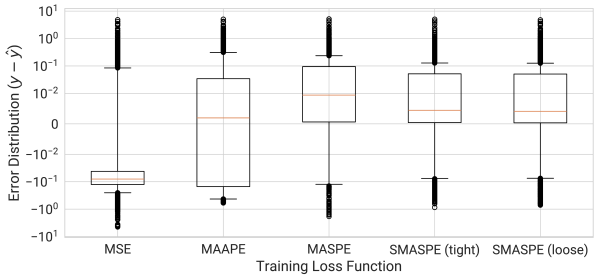


Figure 10: PV-generation model error distribution boxplots.

- Despite the MSE and MAAPE converging the fastest out of all considered loss functions, Figure 9b reveals that this is due to the divergence of the validation loss. Since the MASPE history does not exhibit the same behaviour, it can be inferred that this issue stems from model overfitting and gradients around zero-relative-error respectively. This makes the models trained using the fast-converging loss functions perform significantly worse than the proposed ones for all error metrics evaluated over the test partition, as illustrated by Figure 8.
- On the opposite end of the spectrum, and as shown in Figure 9, MASPE models take the longest to converge due to the vanishing gradients brought up in Section 2.

Loss Function	Median Epoch Number	Error Metrics (Testing Partition)					
		MAE	MSE	MAAPE	MASPE	SMASPE (tight)	SMASPE (loose)
MSE	**774***	0.179	0.143	1.13	1.15	1.28	0.649
MAAPE	855	0.185	0.132	0.901	0.866	1.02	0.550
MASPE	1050	0.123	0.121	0.726	0.655	0.786	0.243
SMASPE (tight)	974	0.109	9.79×10^{-2}	0.729	0.661	0.779	0.223
SMASPE (loose)	904	**0.107**	$\mathbf{9.25 \times 10^{-2}}$	**0.715**	**0.641**	**0.756**	**0.215**

Table 2: PV Generation Forecast numerical results (the best performer per metric is highlighted).

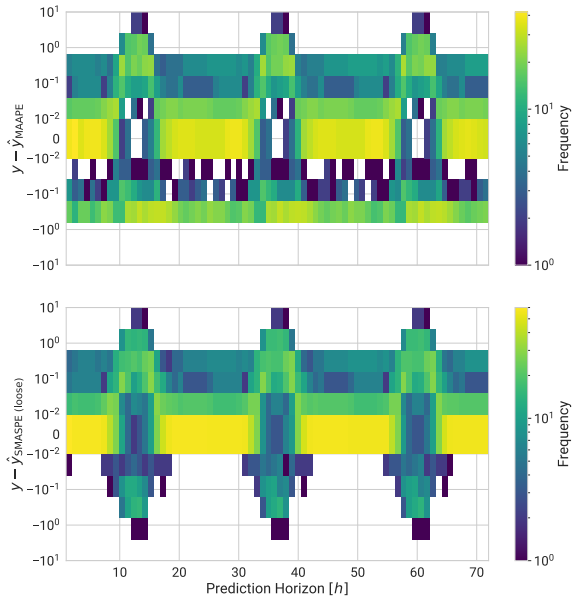


Figure 11: PV-generation signed-error heatmaps for the MAAPE (top) and the loose SMASPE (bottom) as a function of their position within the prediction horizon.

Furthermore, Figure 10 reveals all proposed loss functions converge towards similar error distributions, yet the loose SMASPE remains the clear winner in terms of convergence speed and stability.

- The numerical results in Table 2 validate the increase in accuracy achieved by SMASPE, yielding relative improvements of 42% and 13% in terms of MAE w.r.t. MAAPE and MASPE-trained models respectively. Moreover, Figure 11 provides further insight into the signed-error behaviour: while the MAAPE-trained model fails at capturing PV-generation peaks, the relaxed SMASPE-trained one achieves a near-symmetrical distribution at these critical points over the prediction horizon.

5 Conclusion

We have proposed the Mean Arctangent Squared Percentage Error (MASPE) as an improvement over MAAPE to prevent the asymptotic behaviour of the gradient at zero relative error. Moreover, we have proposed a novel loss function, the Symmetric Mean Average Squared Percentage Error (SMASPE), using the MASPE as a building block to exploit a priori knowledge on the bounds of the target signal and capture its dynamics across all numerical scales. An analysis of the mathematical properties that render these metrics more suitable as loss functions for gradient-based optimization was laid out, followed by a numerical validation of the theorized advantages for both deterministic and stochastic optimization applications. The results not only show that a SMASPE with properly calibrated bounds surpasses both the ubiquitous MSE and the SotA MAAPE for deterministic optimization in terms of accuracy, but it also significantly outperforms the latter when targetted by stochastic optimization algorithms (e.g., ADAM). As possible extensions, we propose multiple symmetry intervals for scalar output models and the generalization of the SMASPE to vectorial outputs, either using hypercubes for independent dimensions or hyperspheres for vectors with known magnitude bounds.

Acknowledgements

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

This programme has received funding from the European Union through Marie Skłodowska-Curie actions under project number 101081466: SEED - Systems and Engineering Science Doctorate.

Special thanks to Dr. Jose Rodrigo Rojo García, who selflessly lent his expertise during the formalization of the proof for the smoothness of the SMASPE.

References

[Bradbury *et al.*, 2018] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas,

Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018.

[Cho, 2014] Kyunghyun Cho. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[Chollet and others, 2015] François Chollet et al. Keras. https://keras.io, 2015.

[De Myttenaere *et al.*, 2016] Arnaud De Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi. Mean absolute percentage error for regression models. *Neurocomputing*, 192:38–48, 2016.

[Henderson, 1975] Charles R Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447, 1975.

[Hodson *et al.*, 2021] Timothy O Hodson, Thomas M Over, and Sydney S Foks. Mean squared error, deconstructed. *Journal of Advances in Modeling Earth Systems*, 13(12):e2021MS002681, 2021.

[Hyndman and Koehler, 2006] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[Kim and Kim, 2016] Sungil Kim and Heeyoung Kim. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679, 2016.

[Legendre, 1806] Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes: avec un supplément contenant divers perfectionnemens de ces méthodes et leur application aux deux comètes de 1805*. Courcier, 1806.

[Liano, 1996] Kadir Liano. Robust error measure for supervised neural network learning with outliers. *IEEE Transactions on Neural Networks*, 7(1):246–250, 1996.

[Steinhauser, 2016] Martin Steinhauser. *Computational multiscale modeling of fluids and solids*. Springer, Berlin, Germany, 2 edition, December 2016.

[Vanegas Arias, 2025] Sergio Mauricio Vanegas Arias. A symmetric relative-error loss function for intermittent multiscale signal modelling (github repository). https://github.com/sergiovaneg/SMASPE, 2025. Accessed: 2025-05-29.

[Veera, 2020] Pearl Veera. Weekly sku level product sales transactions. https://www.kaggle.com/datasets/pearlveera/weekly-sku-level-product-sales-transactions, 2020.