A3I Project

Regional Frequency Analysis (RFA) for rainfall extremes with probabilistic neural networks

Chiara Angileri
Niccolò Marzi
Shola Oshodi

Rainfall Extreme Predictions

Why is it important?

Estimating rainfall extremes is fundamental in hydrological analysis and bears significant implications for various real-world applications.

Understanding, modeling and predicting weather and climate extremes is identified as a major area necessitating further progress in climate research.

Reliable predictions of extremes are needed on short and long time scales to reduce potential risks and damages that result from weather and climate extremes.

Project Description

GOAL

Estimate the **probability distribution** of **Annual Maximum** rainfall data for various **durations** (1, 3, 6, 12, 24 hours) using information from diverse measurement stations, in order to analyse extreme precipitation events.

CHALLENGES

Observations regarding precipitation extremes are often **too sparse** to ensure accurate estimation due to the **rarity of extreme events** themselves and the limited spatial distribution of monitoring stations.

SOLUTION

Regional frequency analysis (RFA) stands as a **robust methodology for deriving estimates of flood** frequency curves in locations **lacking sufficient gauge data,** trasfering observed data from other gauged location to the target site.

Dataset

Overview

The dataset comprises **annual maximum rainfall depths** recorded for various durations (1, 3, 6, 12, 24 hours) from the years 1916 to 2019. The recordings are obtained from **2338** gauged locations situated in the **northern and central** regions of Italy.

Main data information

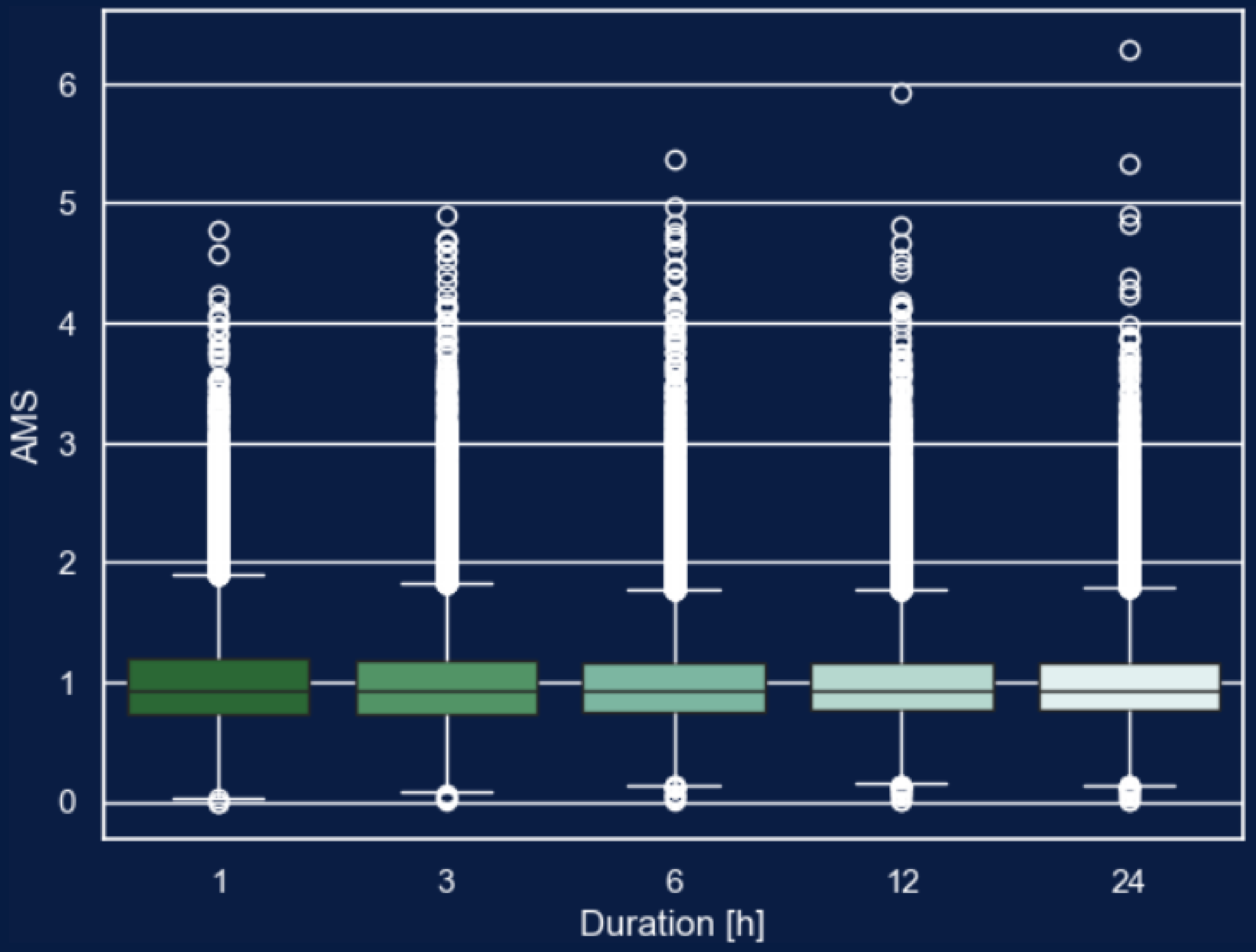
AMS

MORPHOCLIMATIC DESCRIPTORS
[morphological, climatic, geographic]

DURATIONS

AMS Inspection

Boxplot illustrating the AMS values calculated across different time durations, considering the entire dataset.

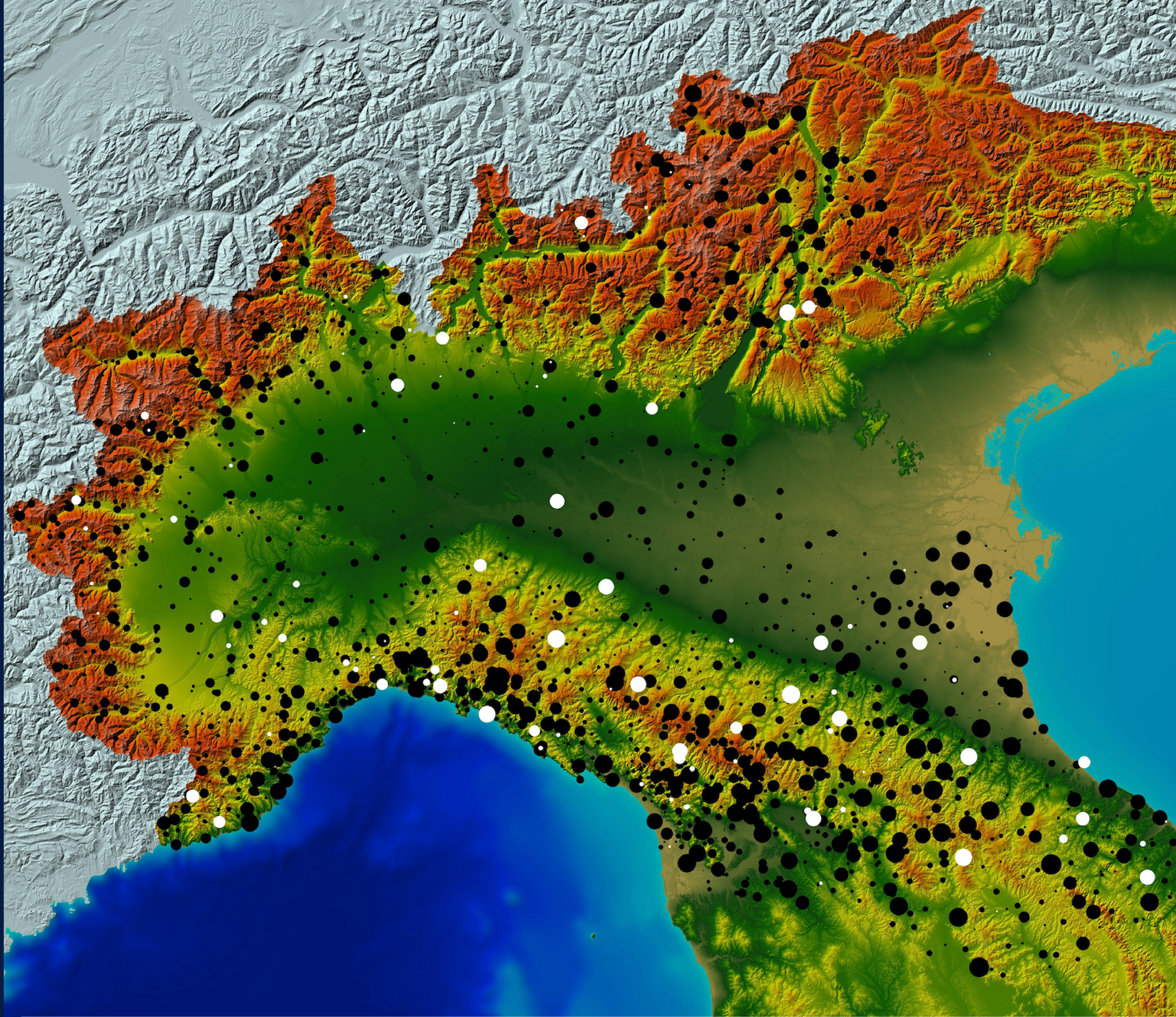


Split

TRAIN **233218** observations

VALIDATION **25914** observations

TEST **15415** observations



Black - Train stations
White - Test stations
The dimension of the points is proportional to their num of observations

Preprocessing

1 **Standardization**

Because of different nature and magnitude, data need to be standardized.

2 **OHE**

The duration values are transformed into one-hot encoded vectors.

3 **AMS scaling**

To use the Beta distribution, AMS values must be scaled to the range]0, 1[.

GUMBEL DISTRIBUTION

- **location (μ)**
- **scale (β): [β > 0]**

Support Range

The Gumbel distribution is defined for real values across the entire x-axis, without constraints on the range of values it can generate.

- The Gumbel distribution exhibits **heavy-tailed behavior**, assigning relatively high probabilities to extreme events occurring far from the mean, making it suitable for modeling rare but significant events.

- Widely studied, with a long history of application in extreme **value analysis** and hydrological risk assessment.

BETA DISTRIBUTION

- **alpha (α) : [α > 0]**
- **beta (β): [β > 0]**

Support Range

The Beta distribution is bounded between 0 and 1 .

- Our decision to experiment with the beta distribution was motivated by its remarkable **flexibility**, with the hope of achieving a customized and adaptable approach to modeling precipitation data.

- Commonly applied in: Bayesian statistics, reliability engineering, quality control and modeling success rates or proportions.

Neuro-Probabilistic Models

1. Probabilistic model of our phenomenon.

$$\text{AMS} \sim \text{Gumbel}(\mu(x,\theta), \beta(x,\theta))$$

$$\text{AMS} \sim \text{Beta}(\alpha(x,\theta), \beta(x,\theta))$$

2. Train the model for maximum log likelihood.

$$\arg\min_{\theta} -\sum_{i=1}^{m} \log f(\text{AMS}_i, (\mu(x_i,\theta), \beta(x_i,\theta)))$$

$$\arg\min_{\theta} -\sum_{i=1}^{m} \log f(\text{AMS}_i, (\alpha(x_i,\theta), \beta(x_i,\theta)))$$

Distribution parameters guesses

The original dataset already contains parameters estimates for Gumbel distribution obtained from statistical fitting.

Gumbel

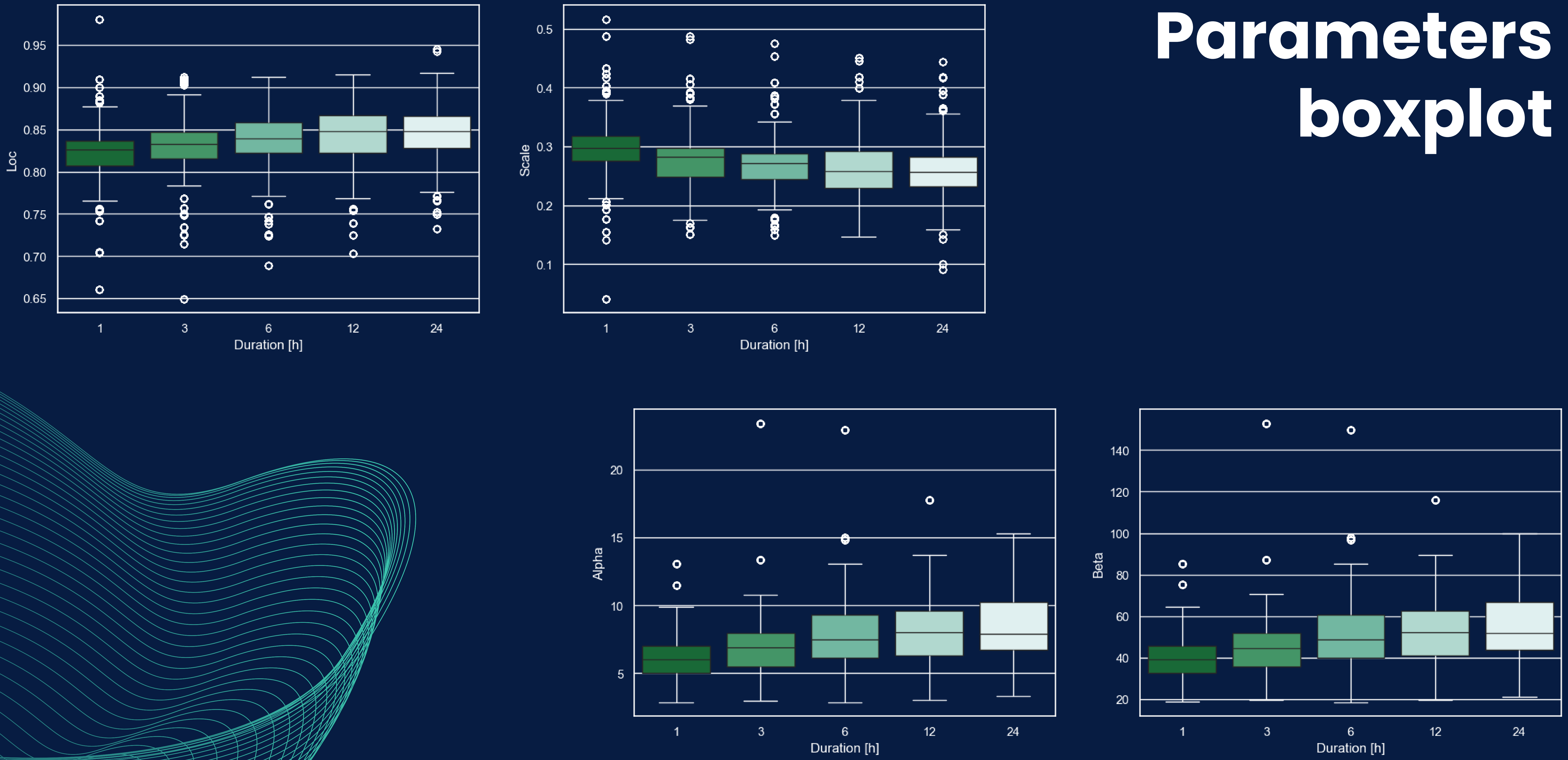
Gumbel parameters estimates are in range [0, 1].

Beta

The Beta distribution is highly flexible, and thus, a few examples may not provide a reliable estimation of its parameters through statistical fitting.

Beta parameters estimates can vary by orders of magnitude.

Parameters boxplot



Models

For each of the two distributions mentioned, three distinct models are constructed.

MAP-model

As input, it exclusively utilizes the Mean Annual Precipitation and durations, with MAP being the most important and significant morphoclimatic indicator for modeling the Annual Maximum Series (AMS).

PCA-model

As input, it uses the duration and the first five principal components resulting from the application of PCA on all morphoclimatic indicators.

Full-model

As input, it utilizes all the morphoclimatic indicators and durations.

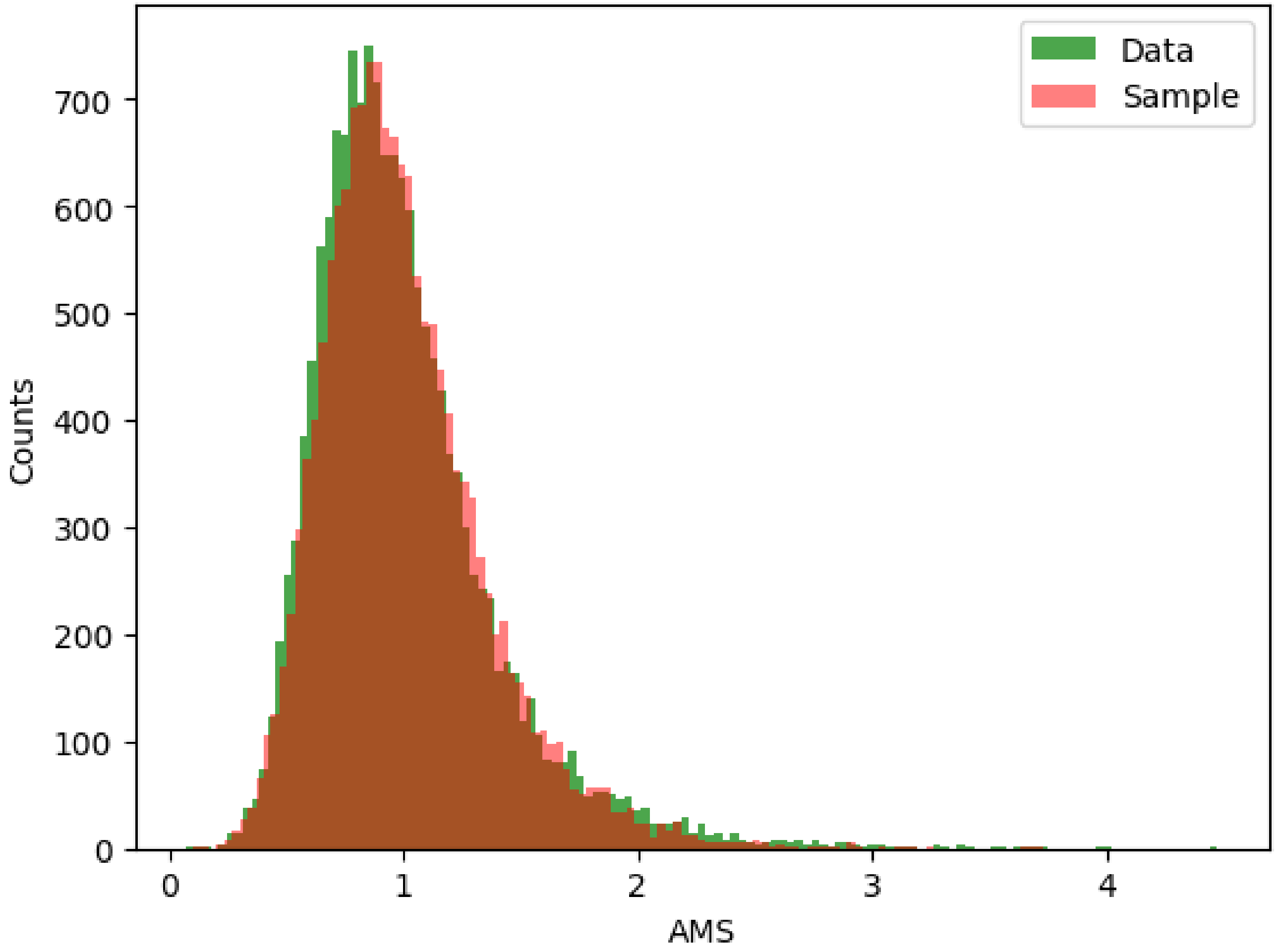
Training Recipe

Both models used to predict the Gumbel and Beta distributions are trained under the same conditions.

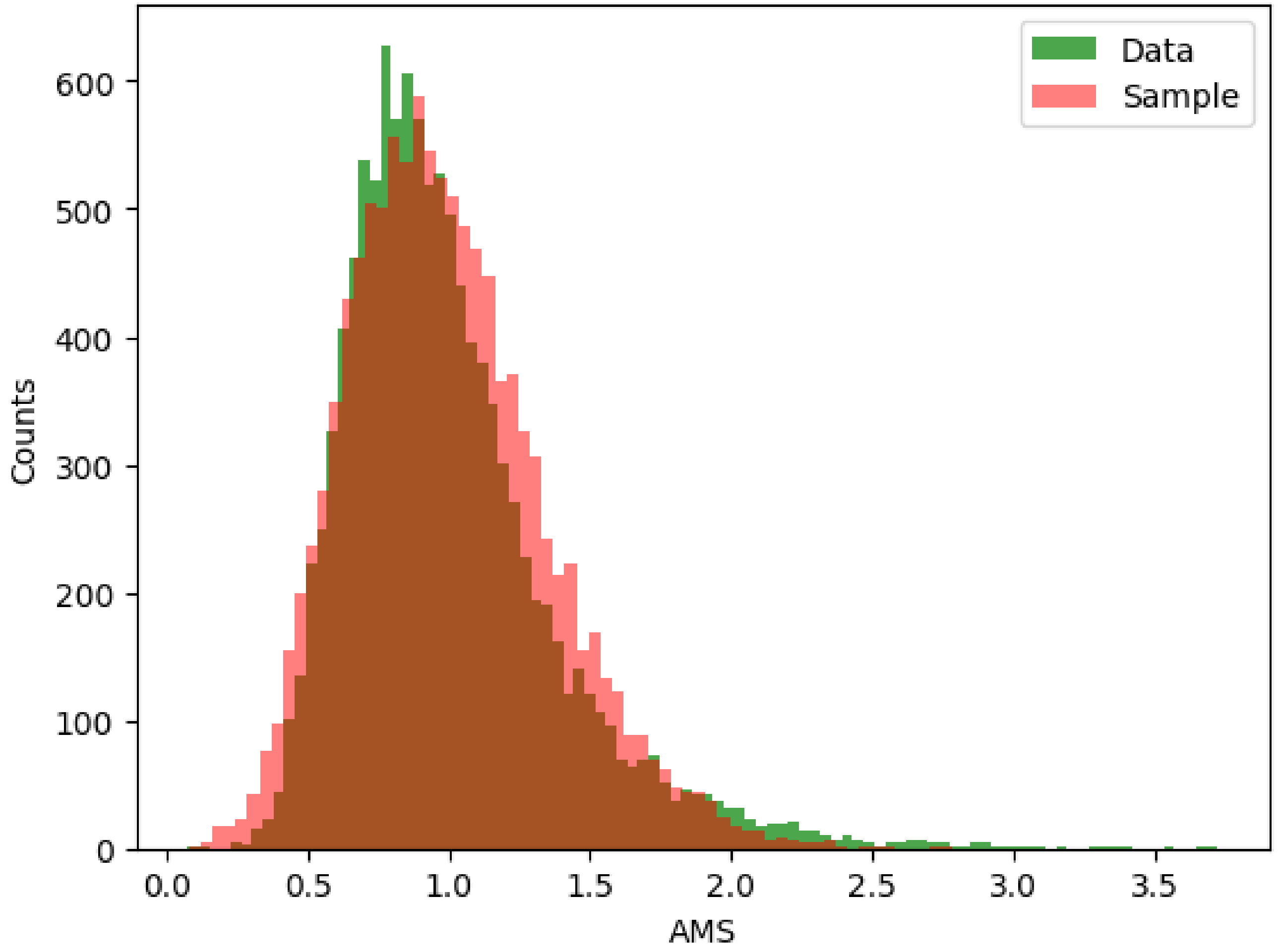
Optimizer	Learning Rate	Loss	Batch Size
Adam	10e-3	NegLogLikelihood	128
Hidden layers	**Epochs**	**Patience**	**Min delta**
[256, 128]	90	30	10e-3

Evaluation – First insight

1. Sampling from predicted distributions and overlaying results with original samples (AMS).
2. **Visual comparison** provides initial insights into the relationship between actual and predicted distributions.
3. **Precision limited** due to the necessity for multiple observations.



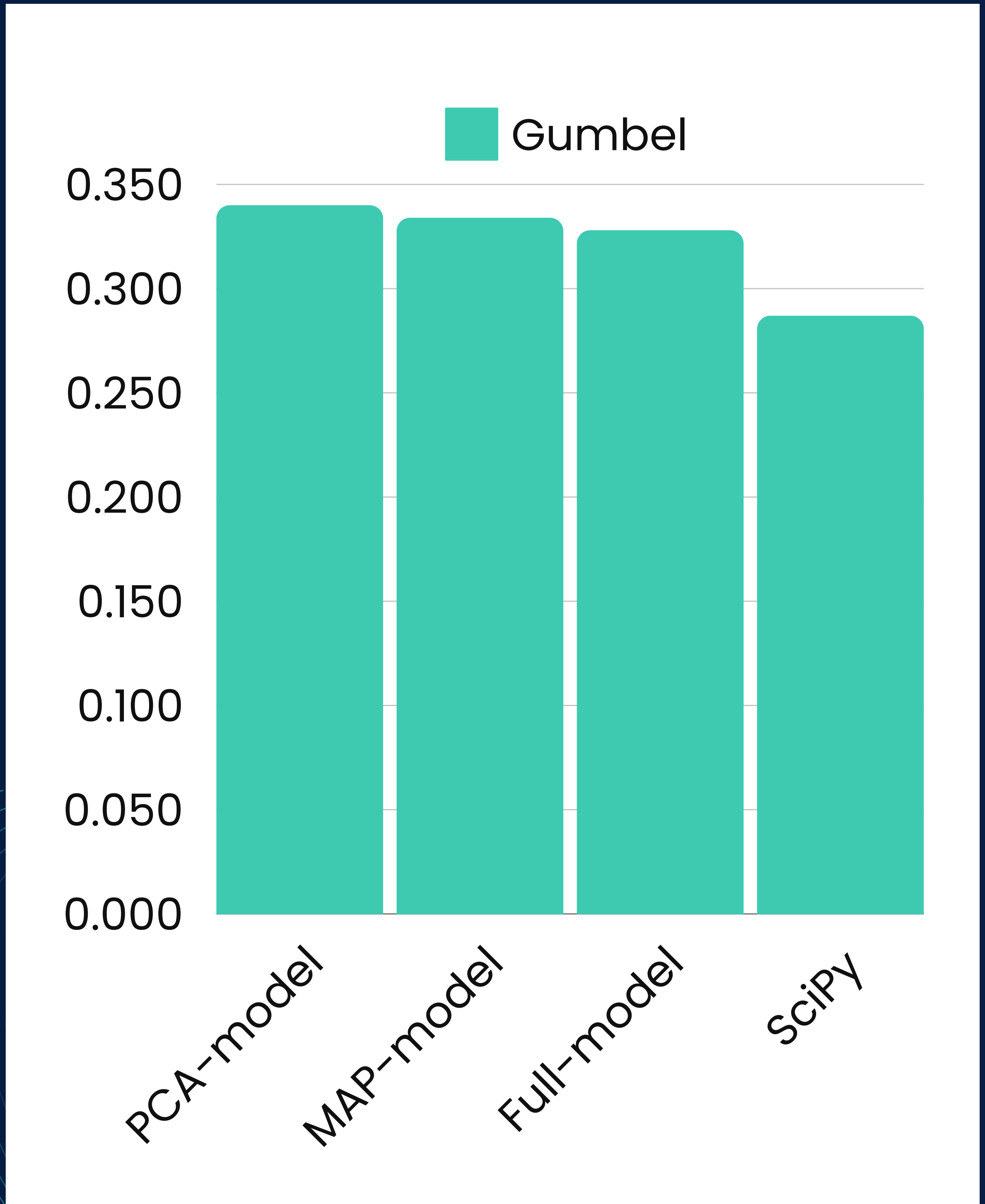
Sample comparison – Full Gumbel model

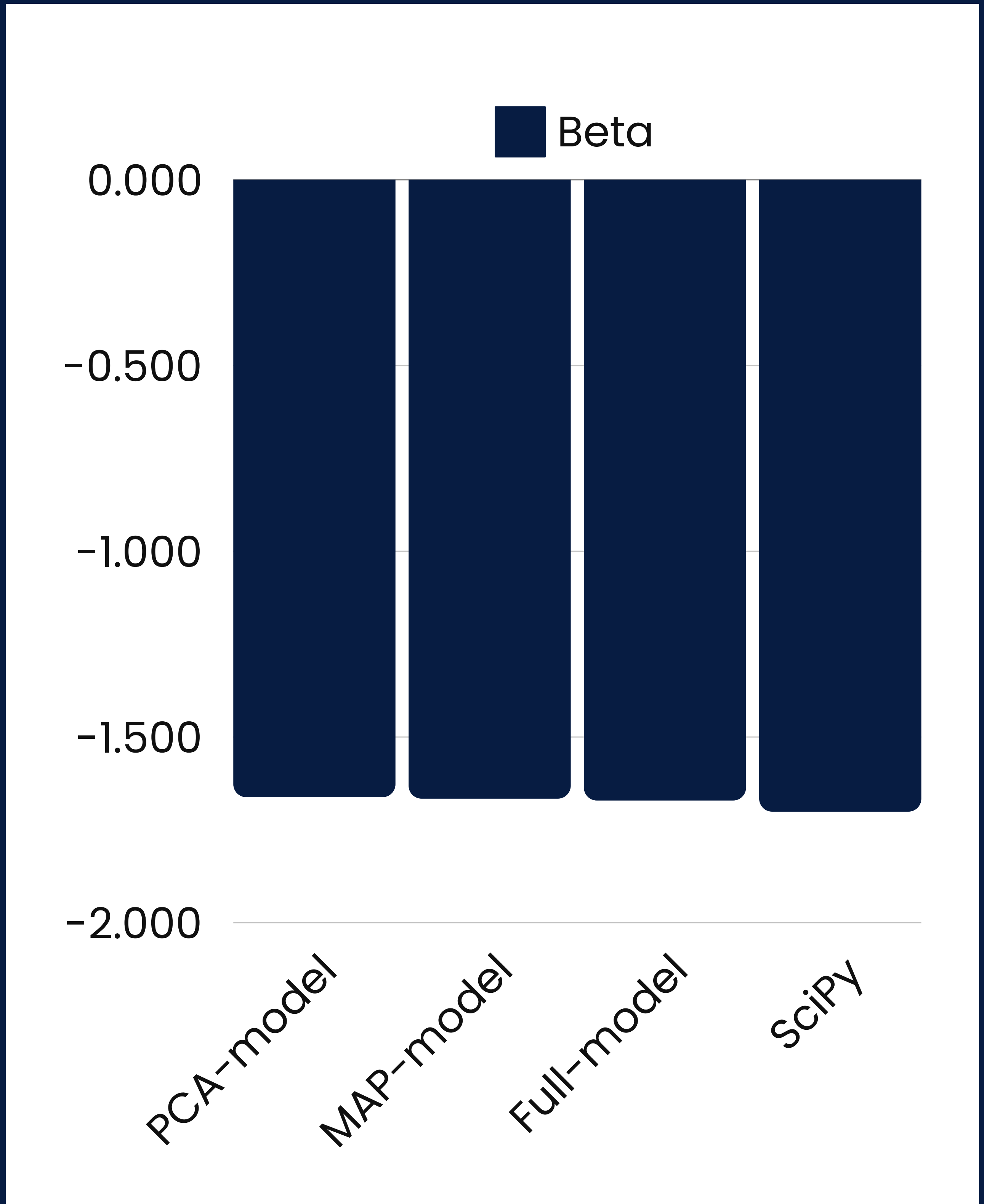


Sample comparison – Full Beta model

Loss

Compare the loss on the test set among the three models for both distributions.





Metrics

Three metrics employed for evaluating predicted parameters of Gumbel and Beta distributions w.r.t. the ones derived from statistical fitting.

Relavitve BIAS (BIASr)

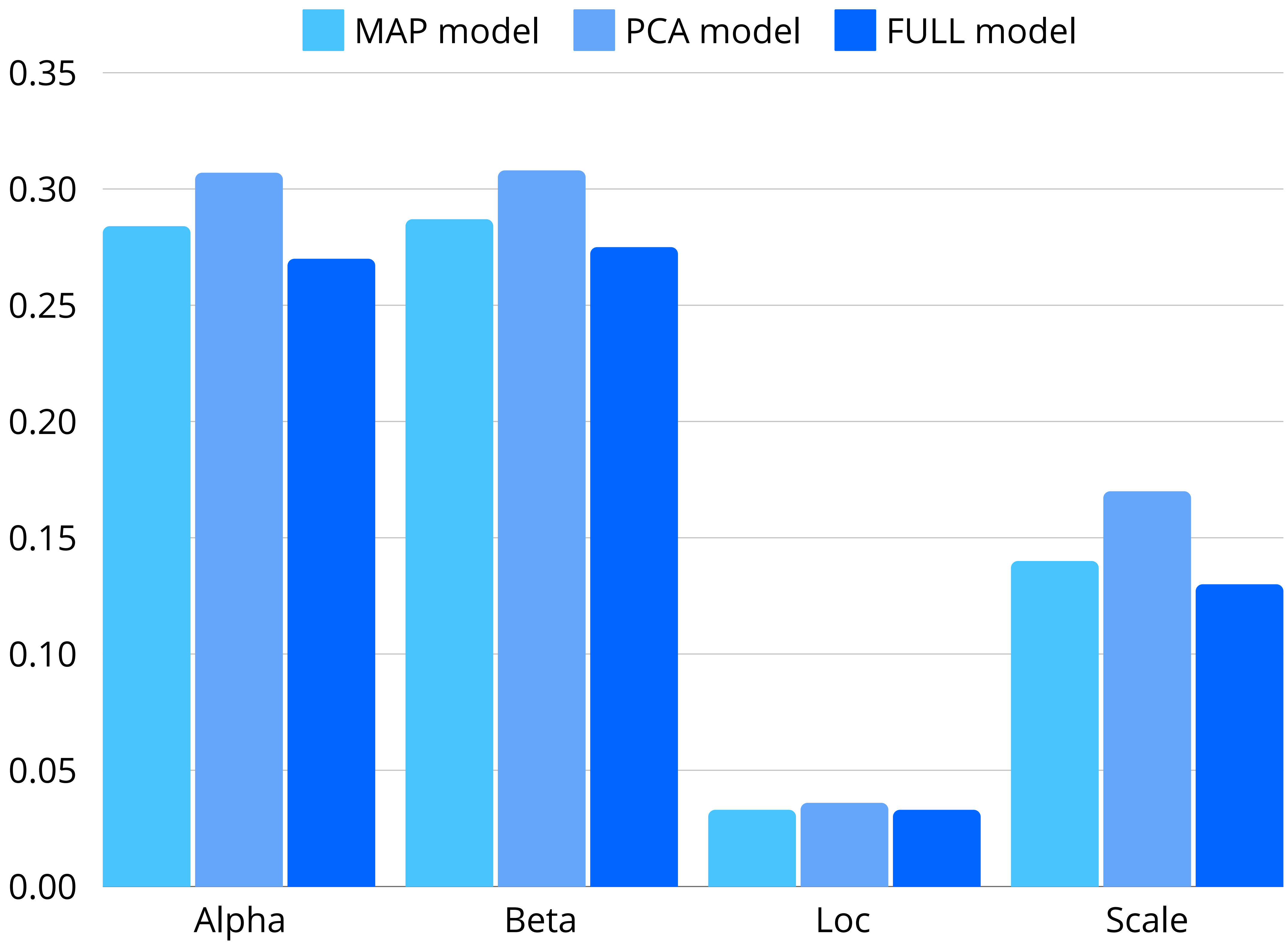
1. Quantifies systematic error in models.
2. Highlights bias in model predictions.
3. Consistent negative/positive bias implies systematic over/underestimation.

Mean Absolute Percentage Error (MAPE)

1. Measures error as a percentage of actual values.
2. Lower MAPE indicates that the estimated parameters are similar to the fitted ones.

Pearson's correlation coefficient (PCC)

1. Measures linearity in the relationship between reference values and model predictions.
2. Indicates the strength and direction of the relationship.
3. A high positive PCC suggests a strong positive linear relationship.

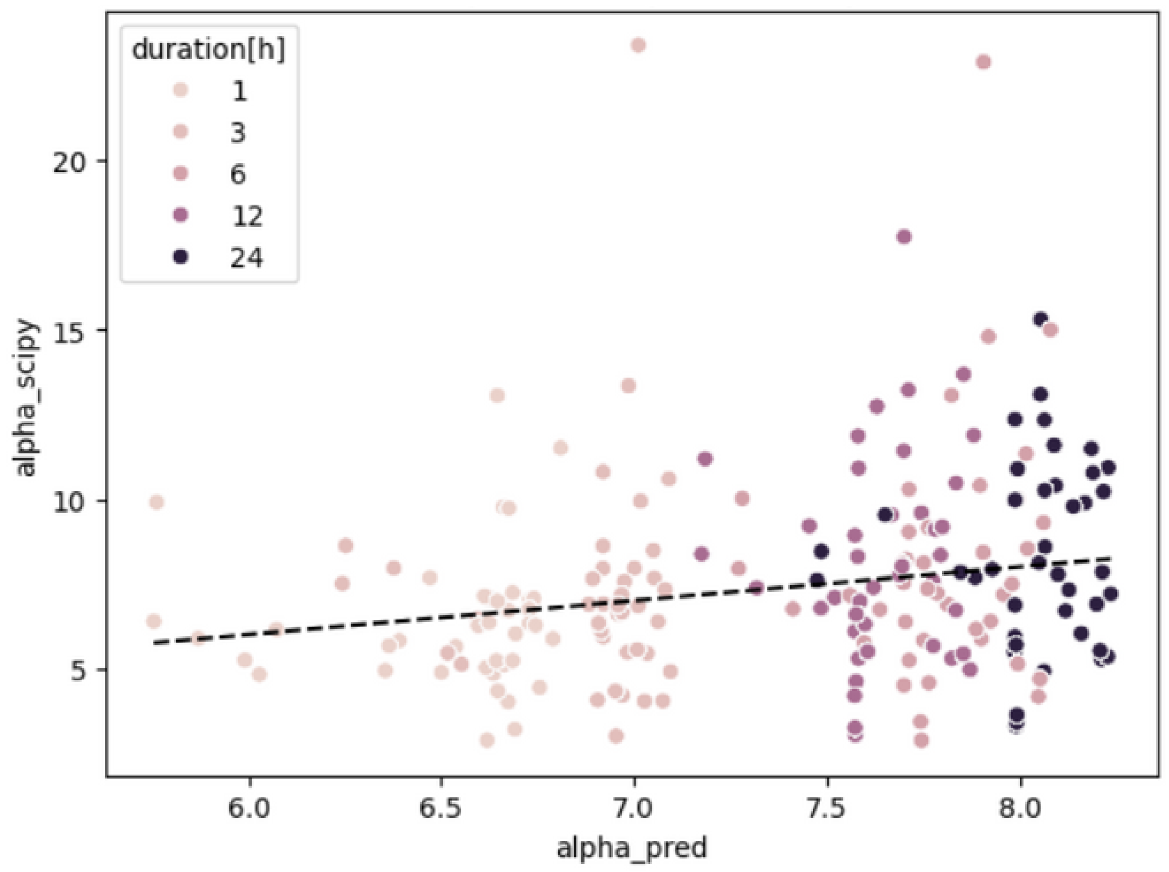


MAPE
Gumbel vs. Beta Comparison

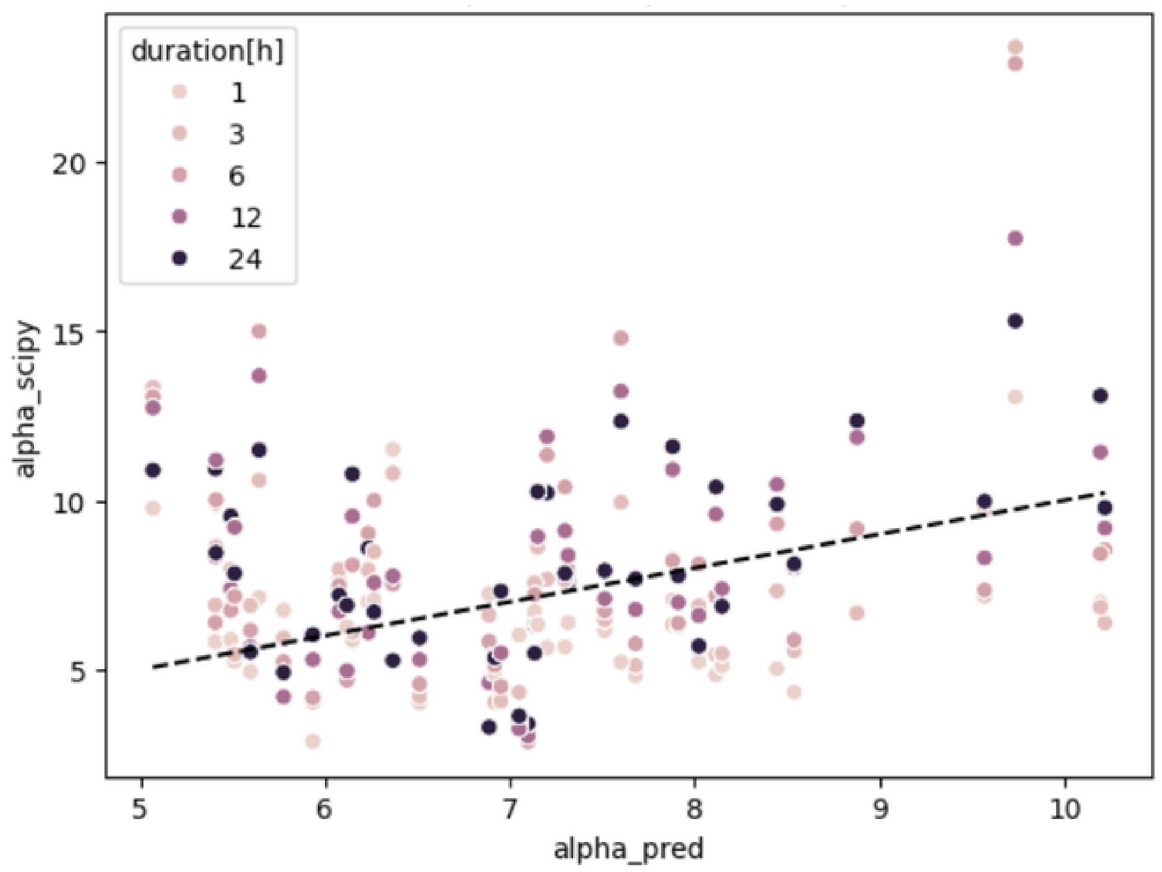
1. The adjacent chart displays the overall MAPE results for the parameters of the examined distributions.

2. Notably, both Beta and Gumbel distributions exhibit lower errors when incorporating morphoclimatic aspects as model inputs.

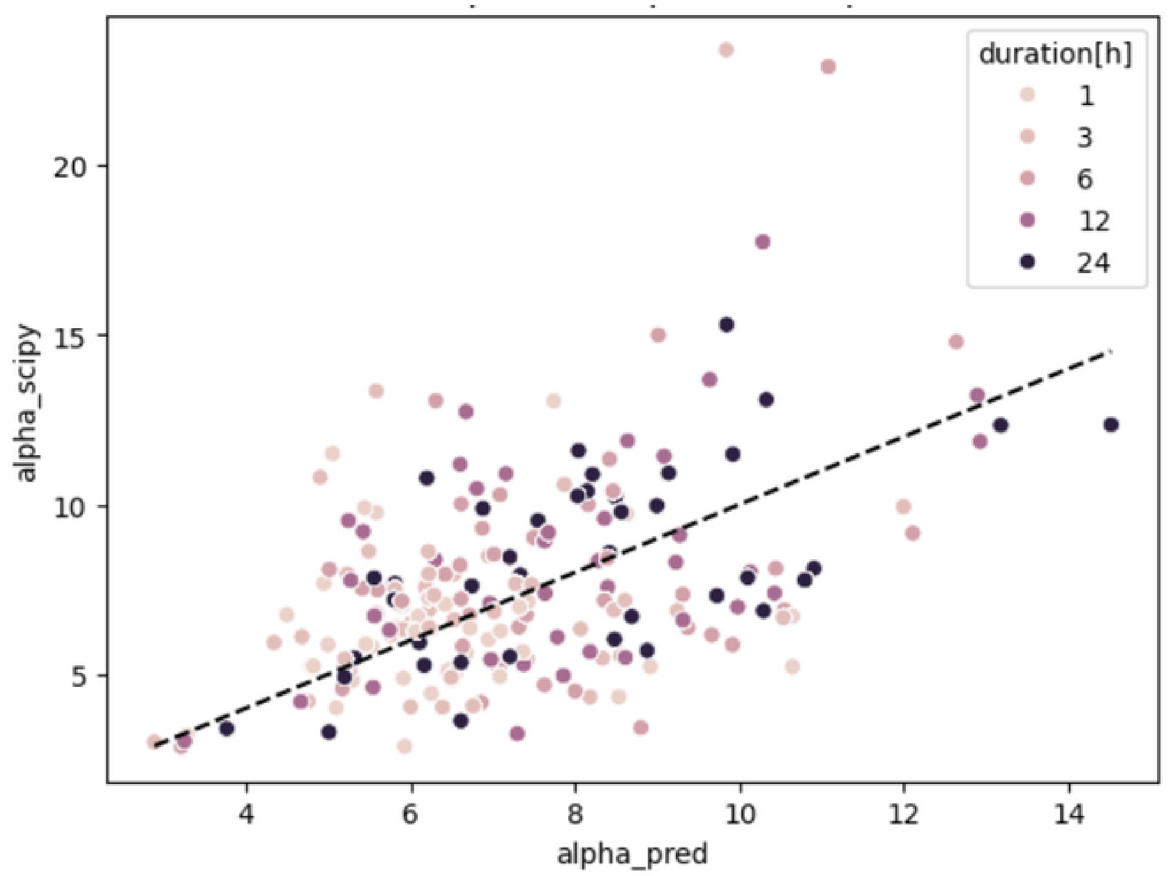
Comparative Scatter Plots of Alpha Parameters



MAP model

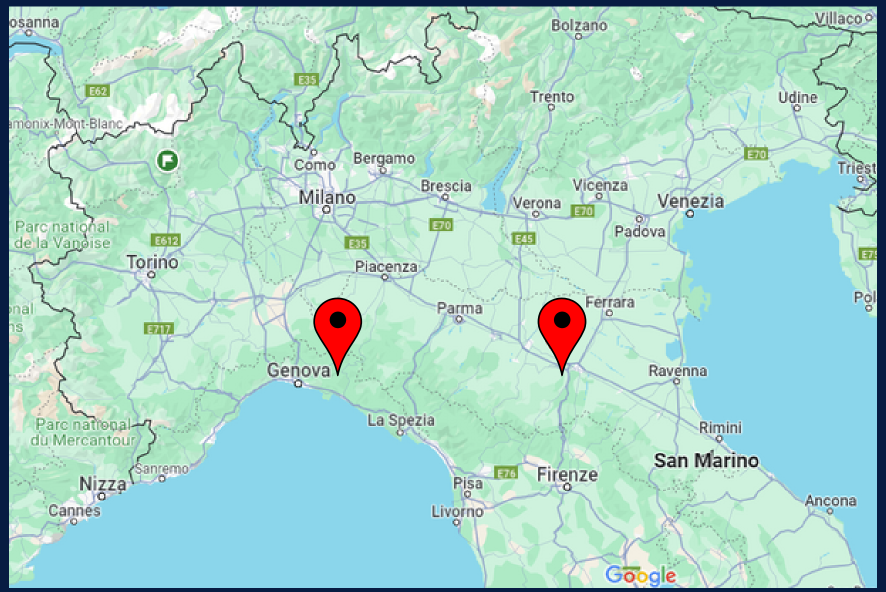


PCA model

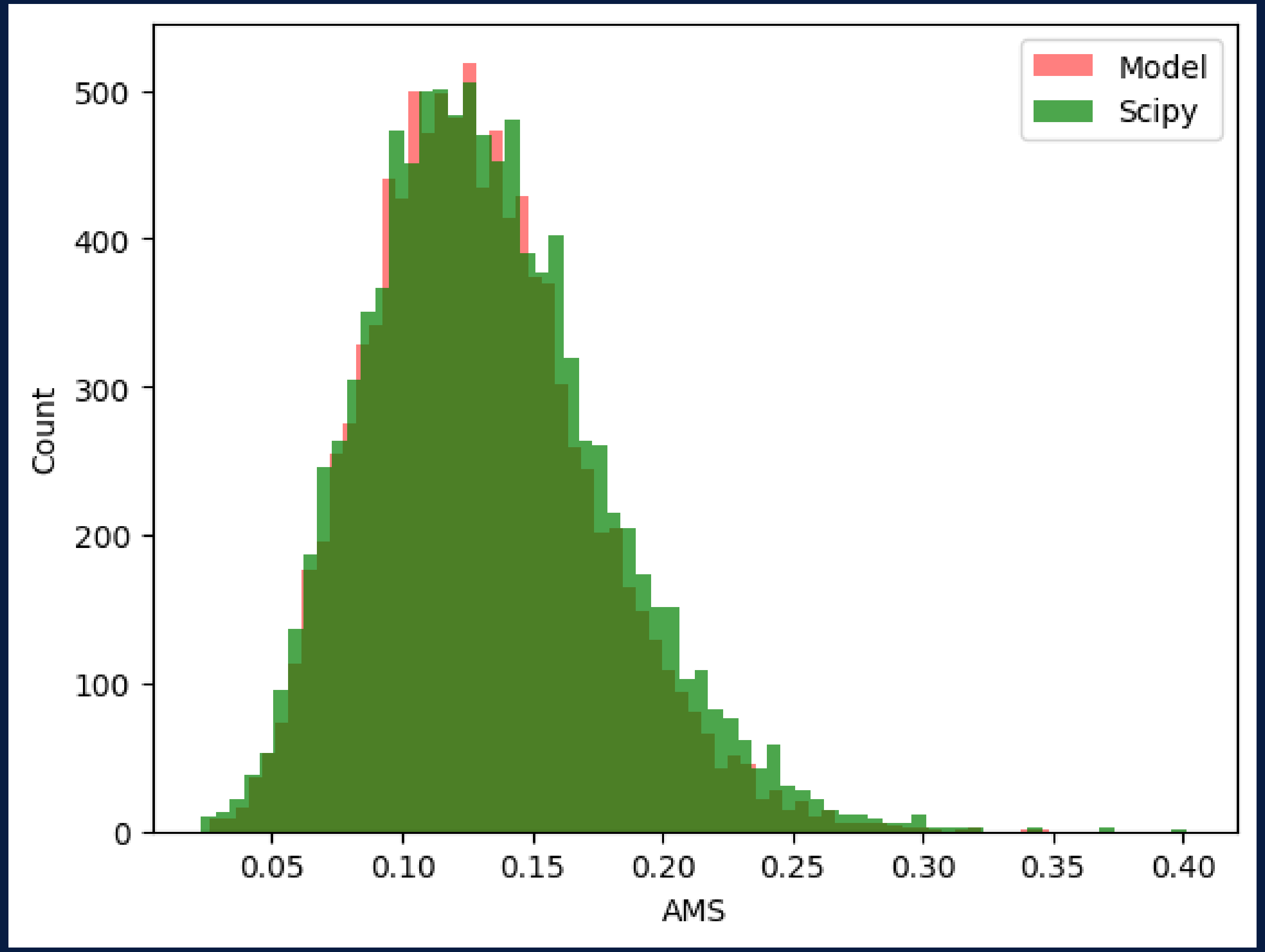


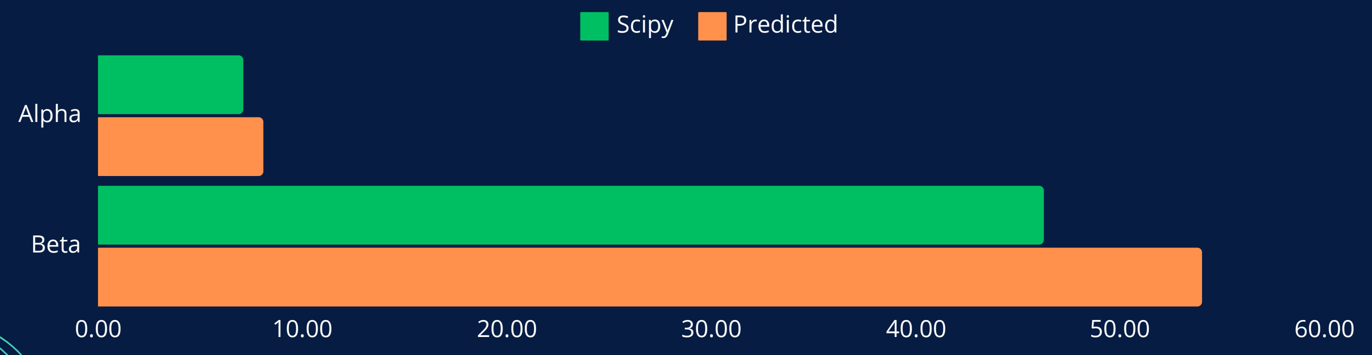
FULL model

Stations example

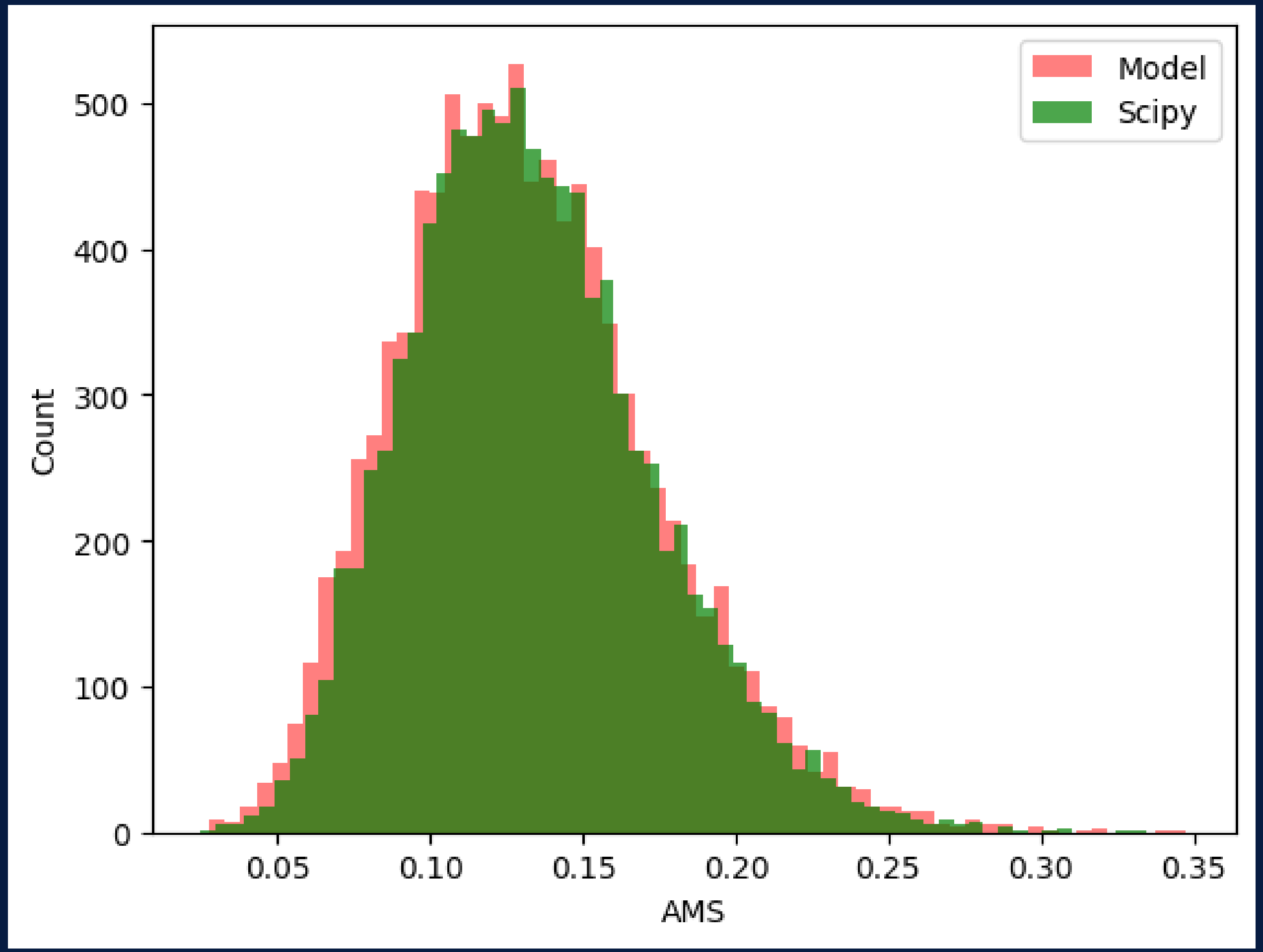


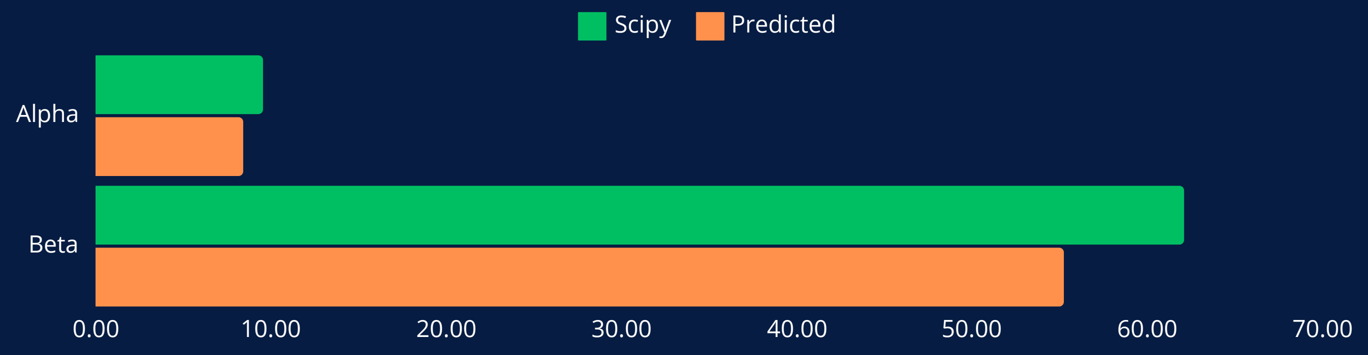
Anzola Dell'Emilia (BO)





Viganego (GE)





Distribution

Gumbel distribution confirms to be the most suitable distribution to analyse extreme precipitation events.

Model

Using all morphoclimatic indicators allows a better estimate of the AMS distribution.

Parameters

Parameters used for comparison obtained from statistical fitting must not be considered as gold standards.

Conclusions

Thanks for your attention!

A3I Project

Project Work

A3I

KS statistics

1. It's a nonparametric method for comparing one-dimensional continuous or discontinuous probability distributions.

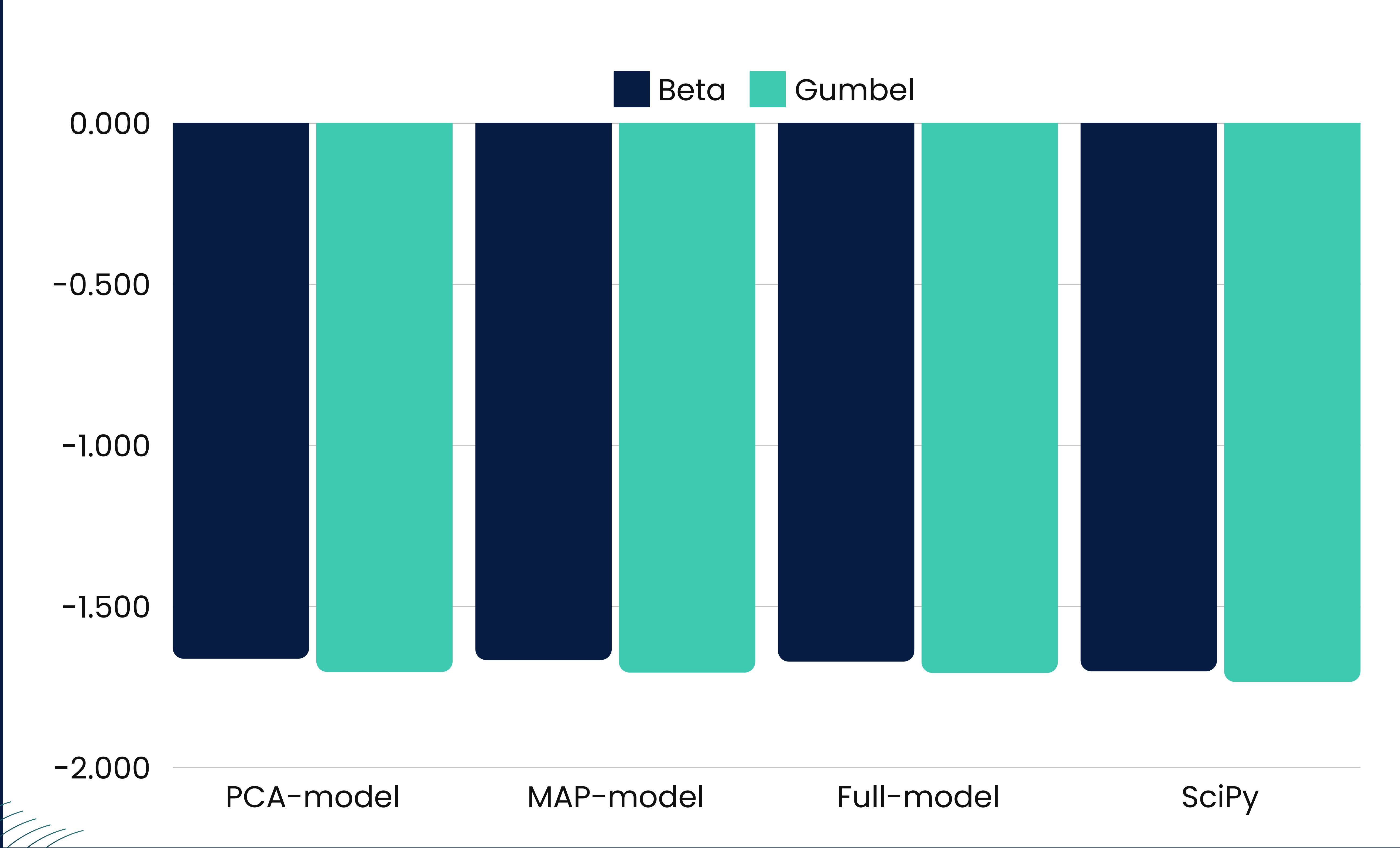
2. It assesses whether a sample comes from a given reference distribution (one-sample KS test) or if two samples come from the same.

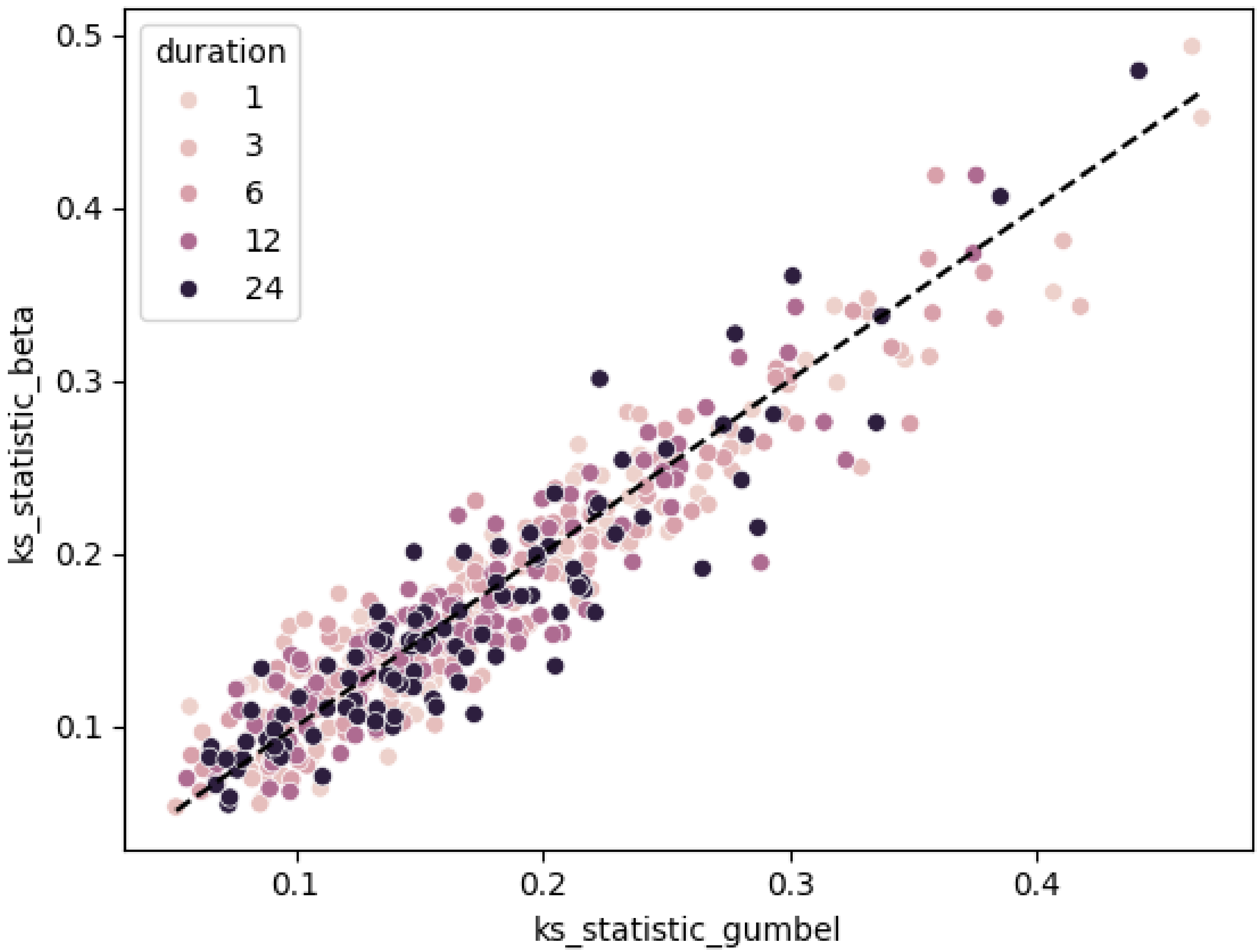
3. **In the context of predicting AMS, the Kolmogorov-Smirnov test allows us to determine which distribution better approximates the data.**

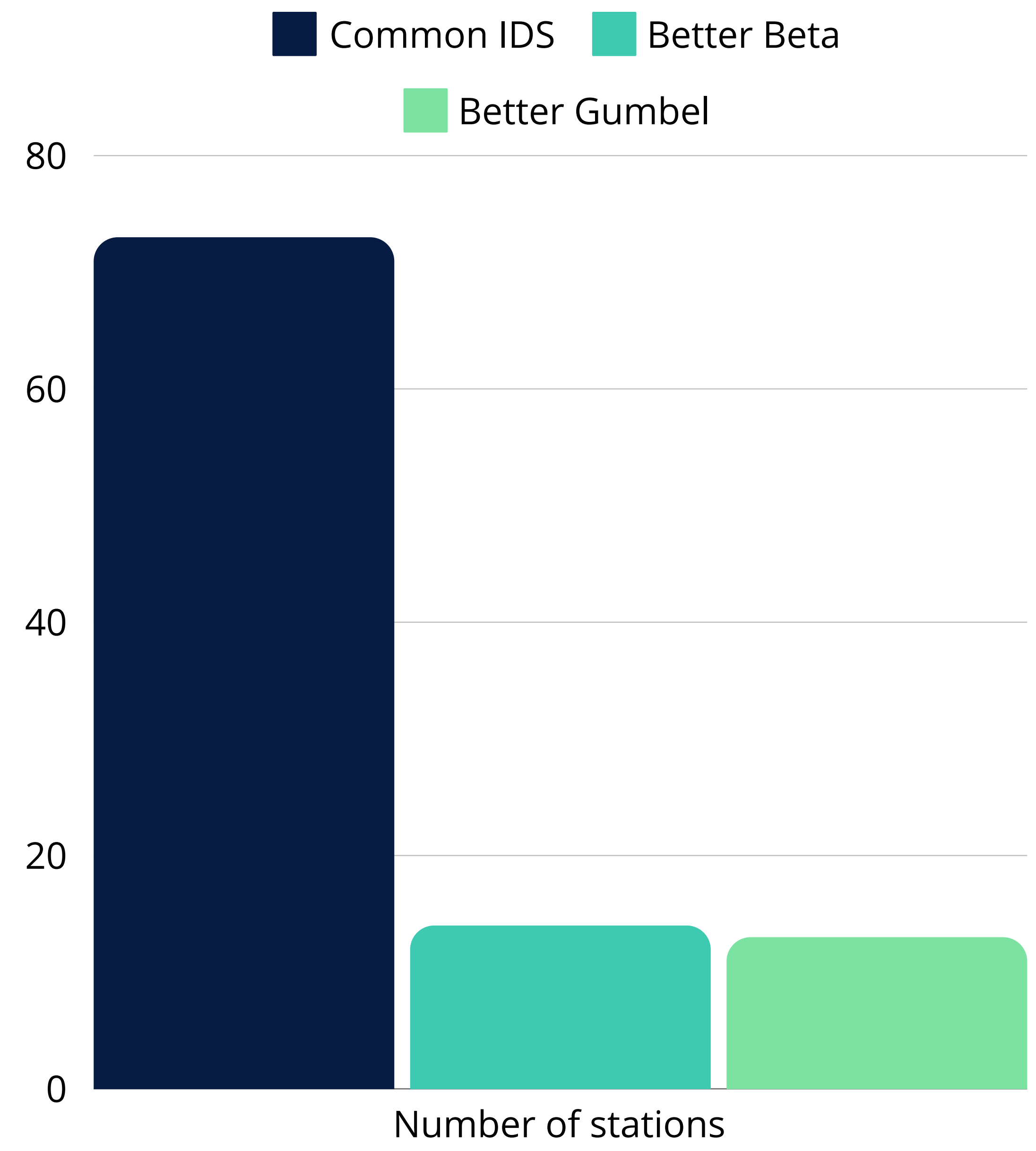
Since it is a scale-sensitive measure, it was necessary to rescale the AMS of the Gumbel between 0 and 1, like the Beta distribution.

Loss

Having the **same scale** allows us to compare the losses of Beta and Gumbel.







FULL model

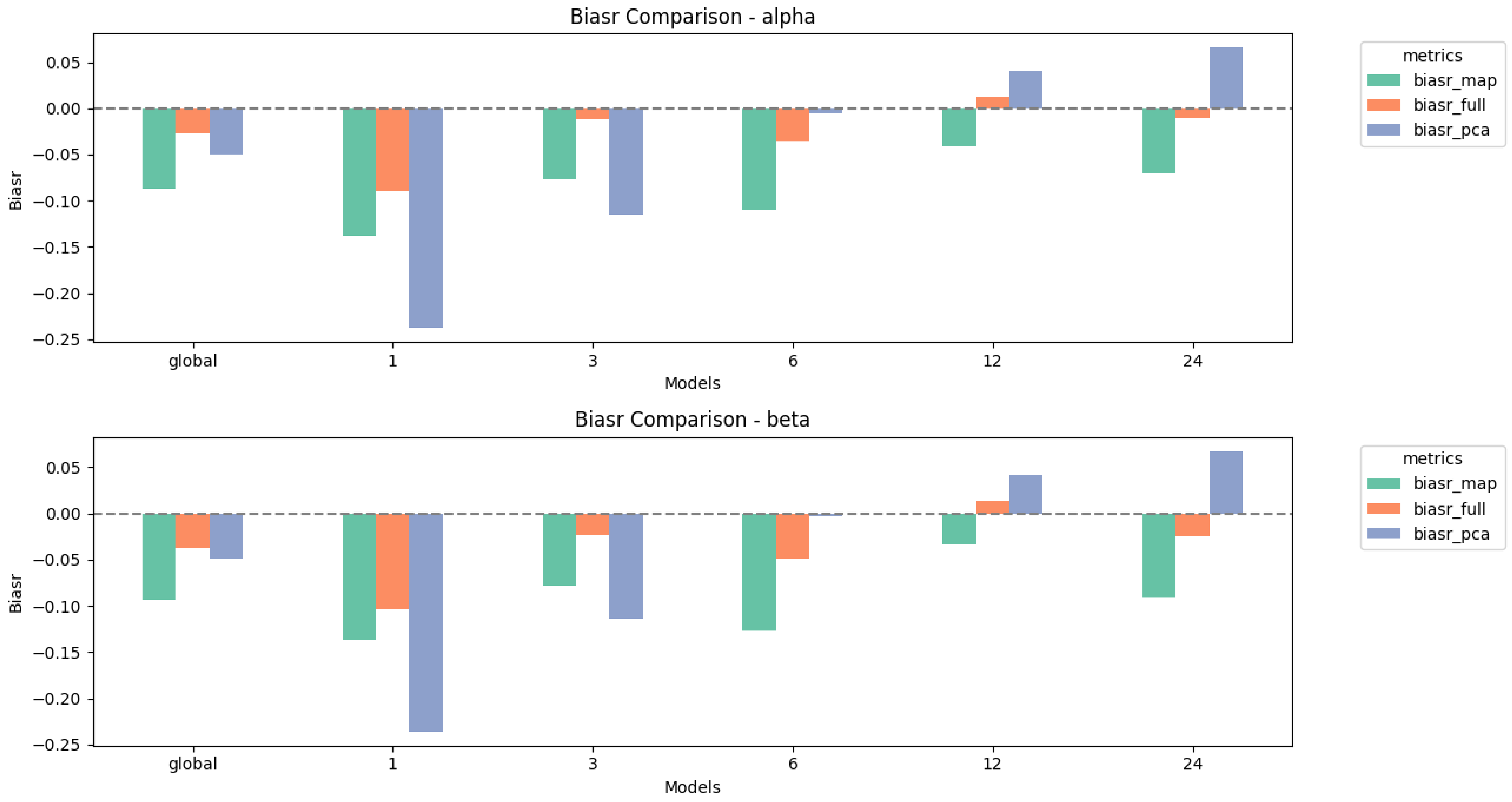
Comparative KS Statistics

Conclusion

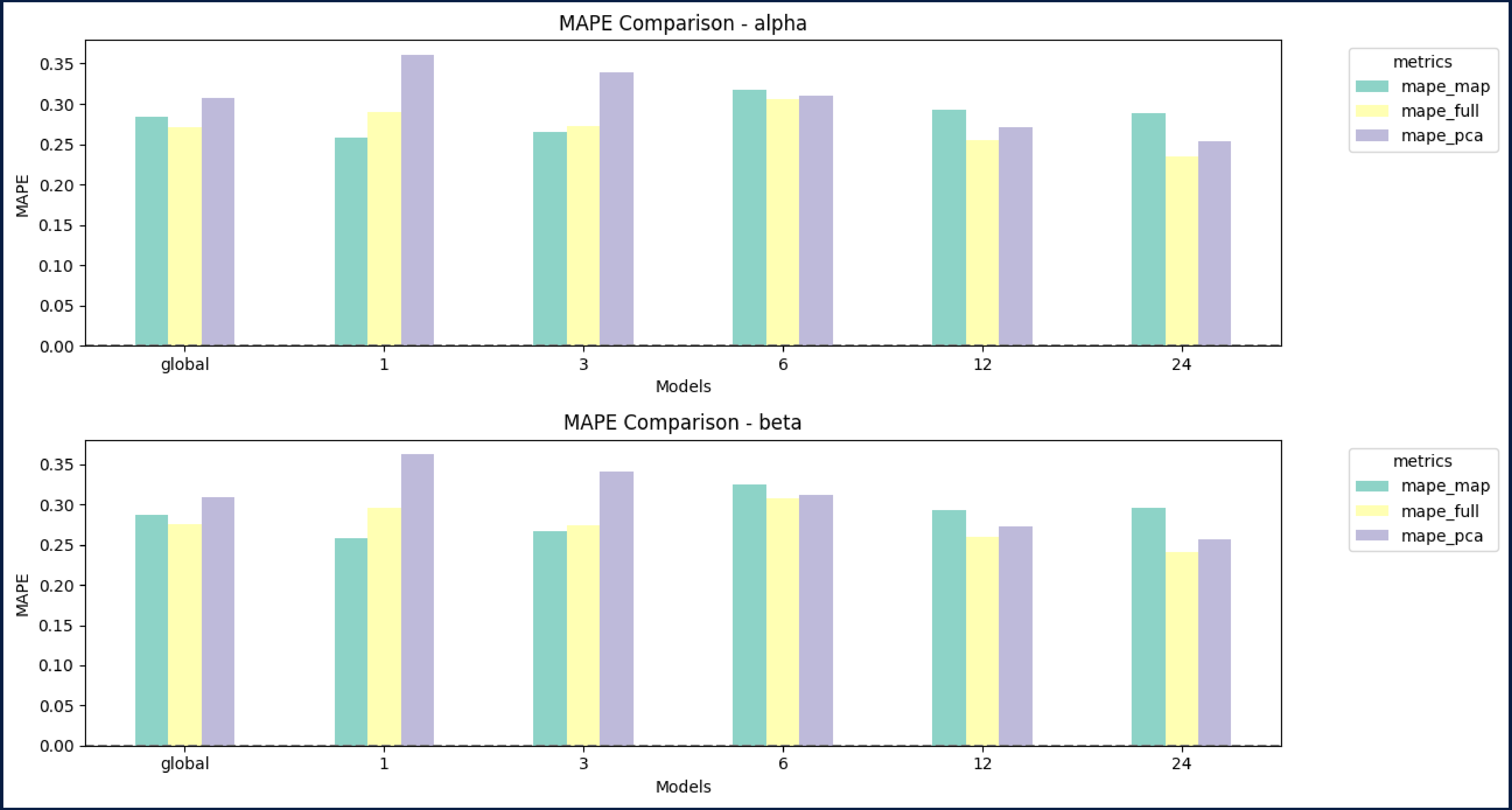
By calculating the KS statistic for all the meteorological stations in the test, we are able to conclude that:

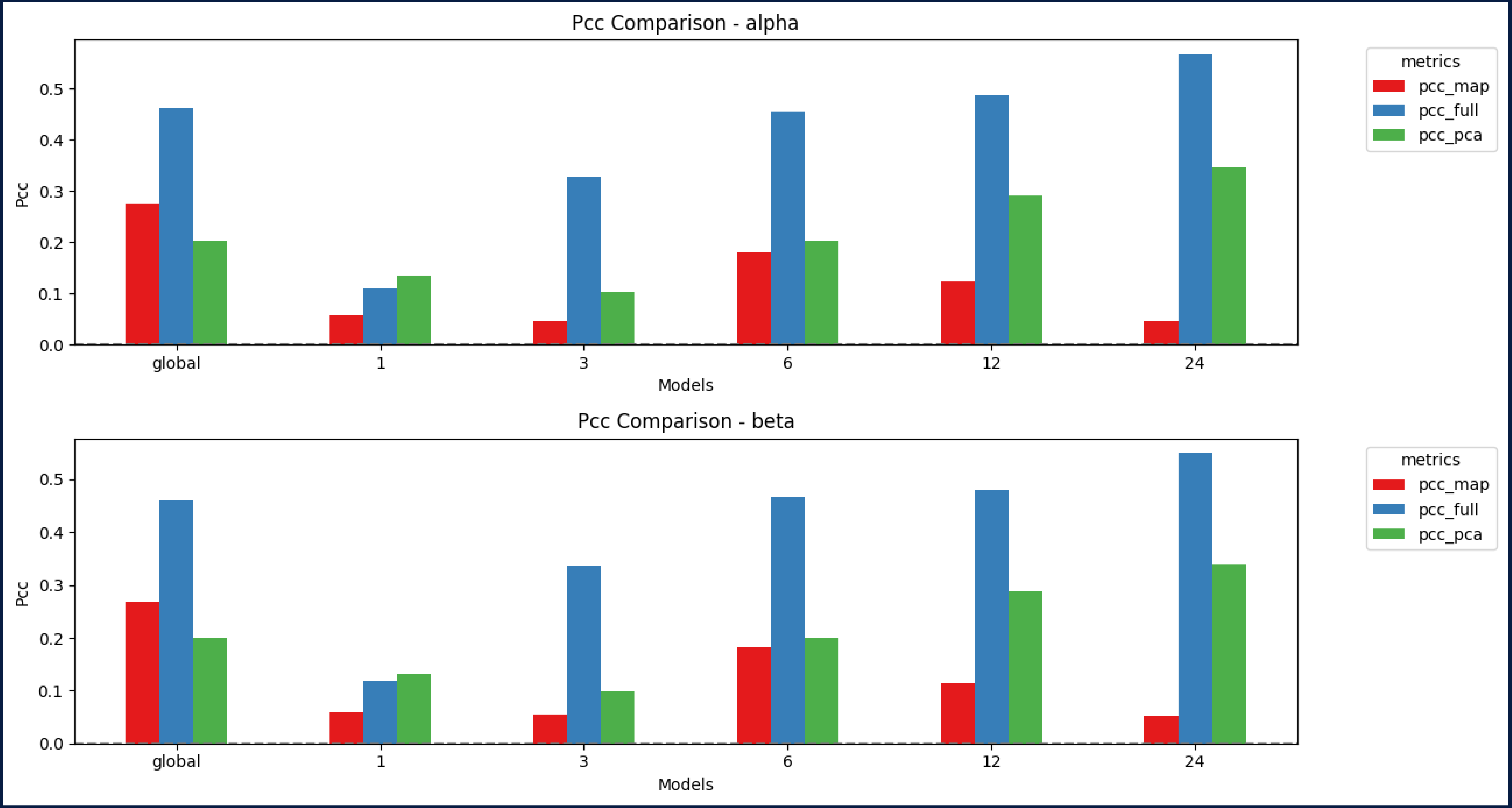
1. Unexpectedly, the **beta distribution** shows itself capable of obtaining good results.
2. Statistically there isn't a distribution that significantly **prevails** over the other.
3. Various factors have been investigated to identify any **patterns**, yet none have been detected, neither at the morphological level nor within the considered time interval.

THE END

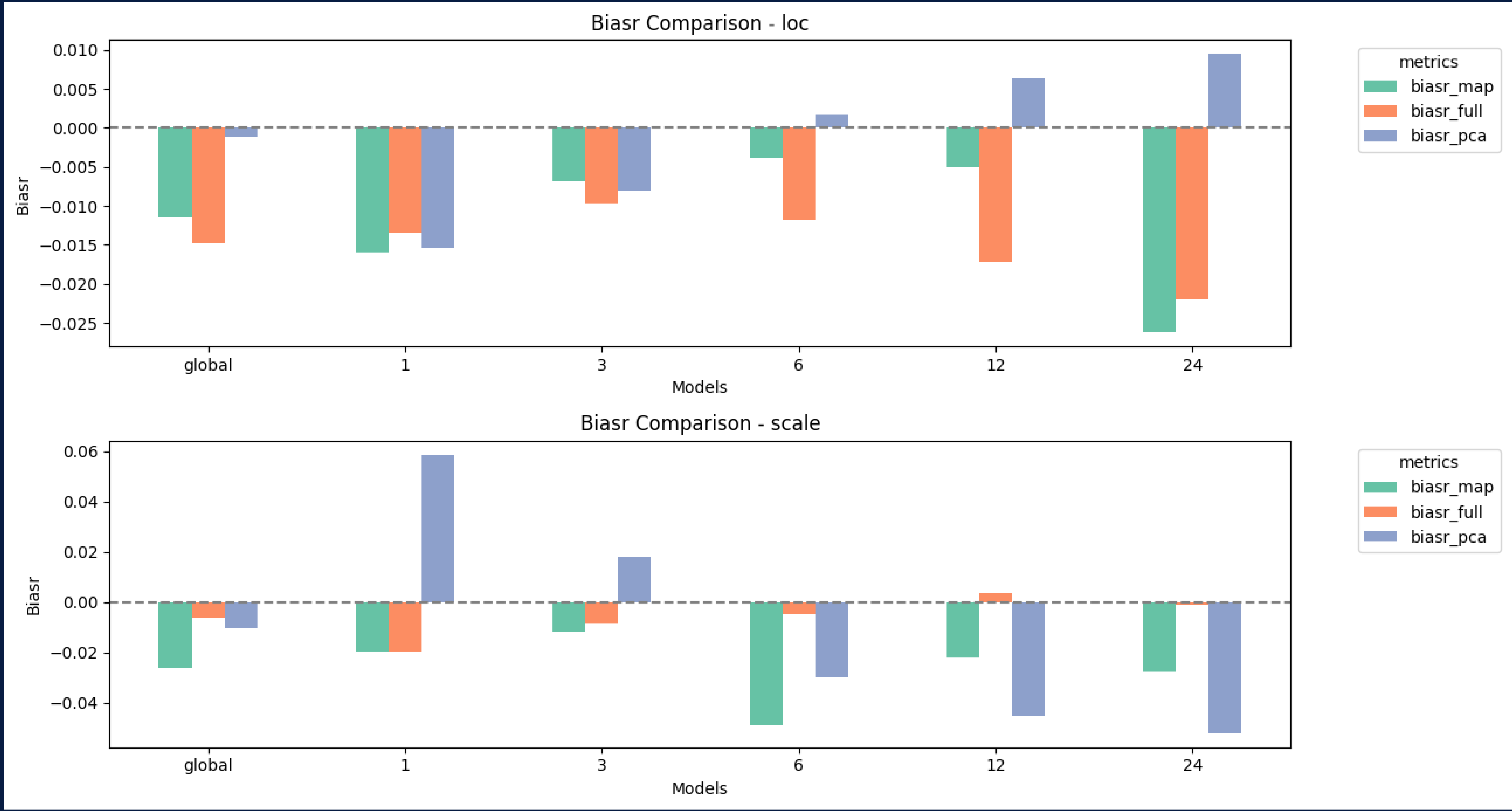


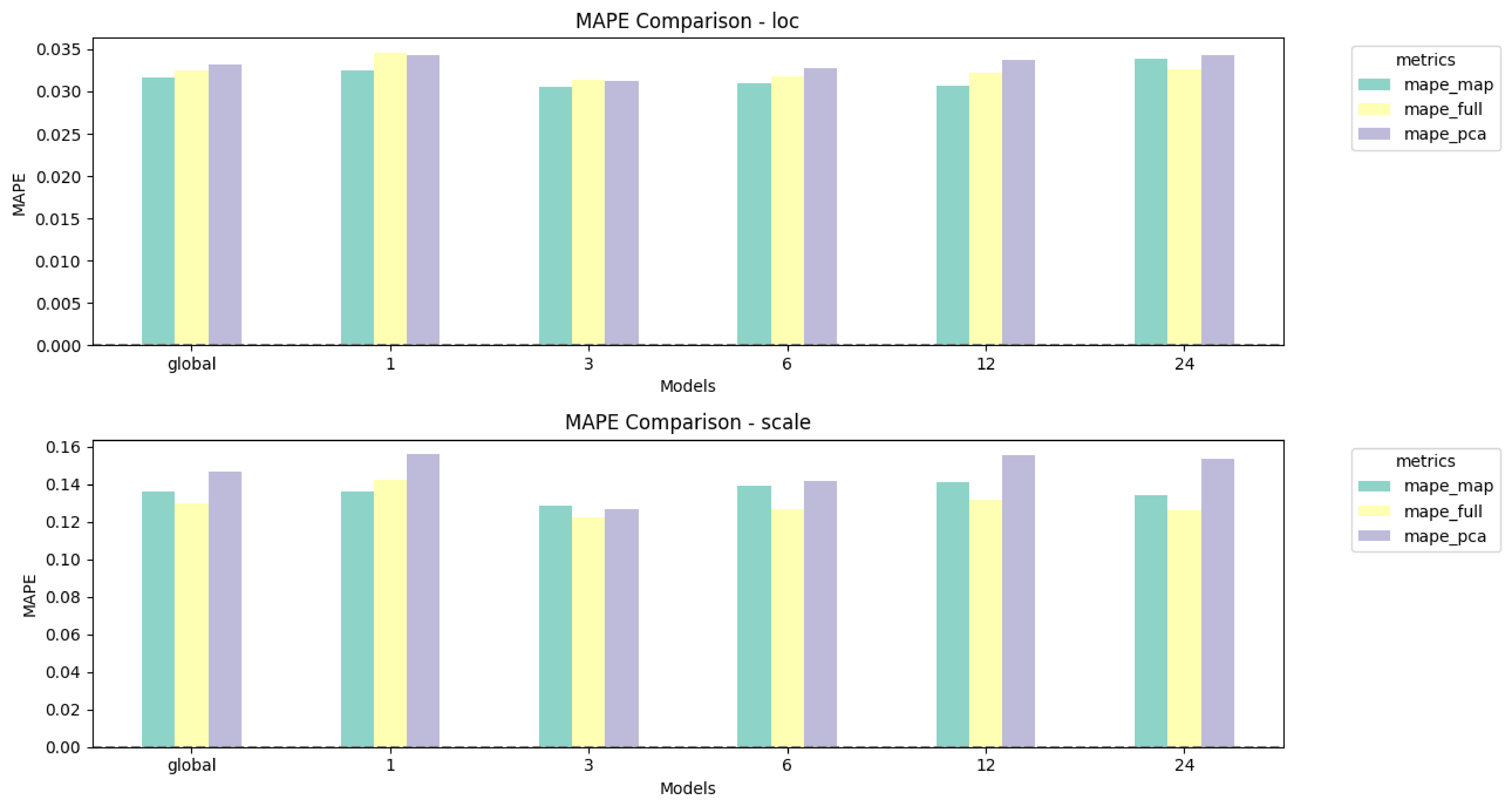
Biasr Comparison - alpha
Biasr
Models
global
1
3
6
12
24
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
metrics
biasr_map
biasr_full
biasr_pca
Biasr Comparison - beta
Biasr
Models

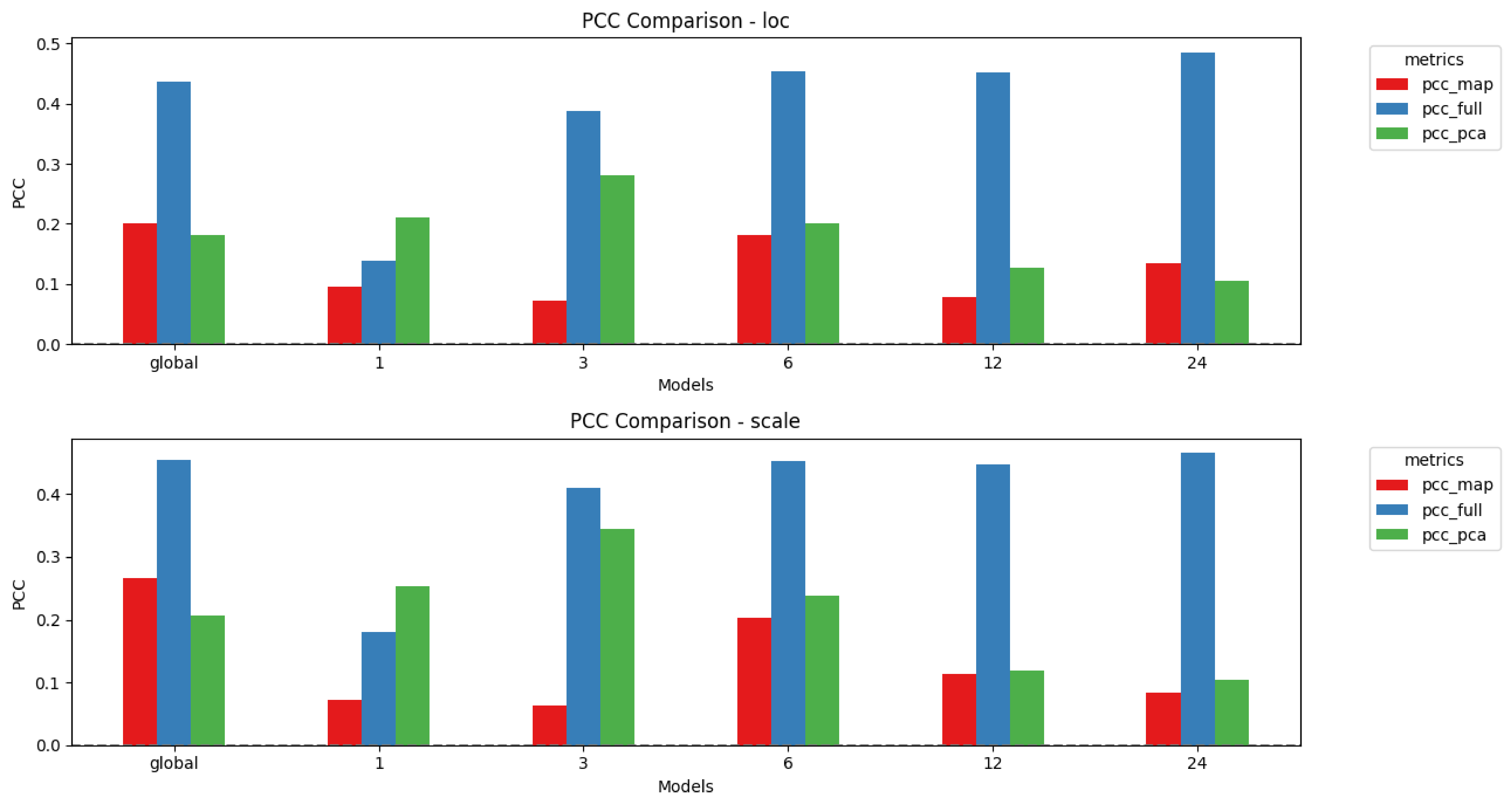

MAPE Comparison - alpha
metrics
mape_map
mape_full
mape_pca
MAPE
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
global
1
3
6
12
24
Models
MAPE Comparison - beta
metrics
mape_map
mape_full
mape_pca
MAPE
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
global
1
3
6
12
24
Models



Pcc Comparison - alpha
Pcc
Models
global
1
3
6
12
24
0.0
0.1
0.2
0.3
0.4
0.5
metrics
pcc_map
pcc_full
pcc_pca
Pcc Comparison - beta
Pcc
Models
global
1
3
6
12
24
0.0
0.1
0.2
0.3
0.4
0.5
metrics
pcc_map
pcc_full
pcc_pca


Biasr Comparison - loc
0.010
0.005
0.000
−0.005
−0.010
−0.015
−0.020
−0.025
Biasr
global
1
3
6
12
24
Models
metrics
biasr_map
biasr_full
biasr_pca
Biasr Comparison - scale
0.06
0.04
0.02
0.00
−0.02
−0.04
Biasr
global
1
3
6
12
24
Models
metrics
biasr_map
biasr_full
biasr_pca


MAPE Comparison - loc
MAPE
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
global
1
3
6
12
24
Models
metrics
mape_map
mape_full
mape_pca
MAPE Comparison - scale
MAPE
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
global
1
3
6
12
24
Models
metrics
mape_map
mape_full
mape_pca


PCC Comparison - loc
PCC
Models
global
1
3
6
12
24
metrics
pcc_map
pcc_full
pcc_pca
PCC Comparison - scale
PCC
Models
global
1
3
6
12
24
metrics
pcc_map
pcc_full
pcc_pca