


FULL PAPER

Open Access



BaHaMAs: a method for uncertainty quantification in geodetic time series and its application in short-term prediction of length of day

Mostafa Kiani Shahvandi^{1*} , Siddhartha Mishra² and Benedikt Soja¹

Abstract

Some of the important geodetic time series used in various Earth science disciplines are provided without uncertainty estimates. This can affect the validity of conclusions based on such data. However, an efficient uncertainty quantification algorithm to tackle this problem is currently not available. Here we present a methodology to approximate the aleatoric uncertainty in time series, called Bayesian Hamiltonian Monte Carlo Autoencoders (BaHaMAs). BaHaMAs is based on three elements: (1) self-supervised autoencoders that learn the underlying structure of the time series, (2) Bayesian machine learning that accurately quantifies the data uncertainty, and (3) Monte Carlo sampling that follows the Hamiltonian dynamics. The method can be applied in various fields in the Earth sciences. As an example, we focus on Atmospheric and Oceanic Angular Momentum time series (AAM and OAM, respectively), which are typically provided without uncertainty information. We apply our methodology to 3-hourly AAM and OAM time series and quantify the uncertainty in the data from 1976 up to the end of 2022. Furthermore, since Length of Day (LOD) is a geodetic time series that is closely connected to AAM and OAM and its short-term prediction is important for various space-geodetic applications, we show that the use of the derived uncertainties alongside the time series of AAM and OAM improves the prediction performance of LOD on average by 17% for different time spans. Finally, a comparison with alternative uncertainty quantification baseline methods, i.e., variational autoencoders and deep ensembles, reveals that BaHaMAs is more accurate in quantifying uncertainty.

Keywords Machine learning, Uncertainty quantification, Length of day

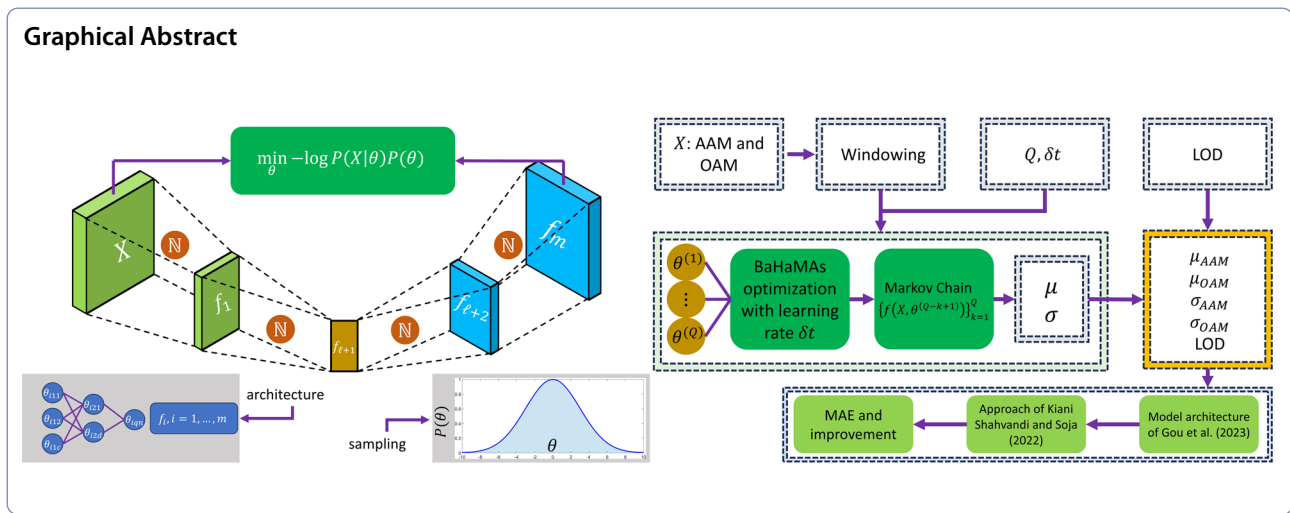
*Correspondence:

Mostafa Kiani Shahvandi
mkiani@ethz.ch

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



1 Introduction

Quantifying the uncertainty in geodetic time series is an essential task (Kiani Shahvandi and Soja 2022a), since it provides a measure of the reliability of the data (Hüllermeier and Waegeman 2021). This type of uncertainty is called aleatoric, which is related to the inherent characteristics of the phenomenon the time series represents. Geodetic time series and the associated uncertainty information can be subsequently used in various other applications that use these data as input. However, often in practice, information regarding the aleatoric uncertainty is not provided. For instance, climate variables such as zonal and meridional winds provided by European Centre for Medium-Range Weather Forecasts (see e.g., ECMWF; Uppala et al. Uppala et al. 2005) lack rigorous uncertainty estimates. As a result, products such as Atmospheric and Oceanic Angular Momentum functions (AAM and OAM, respectively) that rely on these data lack uncertainty information as well, such as those provided by the GFZ German Research Centre for Geosciences (GFZ; Dobslaw et al. 2010, Dobslaw and Dill 2018). When predicting these time series, it is possible to derive the so-called prediction uncertainty, which is a measure of prediction reliability. This can be achieved using methods such as deep ensembles (DEs; Lakshminarayanan et al. 2016), wherein various models with different parameters are fitted simultaneously to the same time series. Subsequently, prediction statistics are computed from the predictions of individual models. However, assigning uncertainty to the original time series requires a different approach, as no prediction is made. The so-called autoencoders (see e.g., Baldi 2012) can be considered as a viable approach for approximating the aleatoric uncertainty in the data by learning it from the time

series itself. In autoencoders, the input to the mapping pipeline and its output are the same, and the algorithm tries to discover the hidden structure in the data and reconstruct the input. However, to avoid a simple linear mapping that fully recovers the input (in other words avoiding input memorization) a so-called *bottleneck* is designed in the algorithm whereby the degree of predictability of the time series is analyzed. In other words, if not for the bottleneck, the parameters of the algorithm would be derived such that the input X is passed through the identity function $f(X) = X$, fully recovering the input. The usual bottlenecks used in autoencoders are batch normalization (Lofe and Szegedy 2015), layer normalization (Ba et al. 2016), or instant normalization (Ulyanov et al. 2016). These normalize the output of the neural network in each layer, preventing the autoencoder from converging to the identity function.

A variety of autoencoder approaches have been proposed. One of the most important classes of these approaches are the so-called Variational AutoEncoders (VAEs; Kingma and Welling Kingma and Welling 2014), which are based on Bayesian models. In VAEs, a model approximates the posterior distribution, typically Gaussian, with unknown mean and standard deviation. By random sampling from a normal distribution and iteratively minimizing the loss function (which is based on the negative logarithm of posterior and prior likelihoods, both having a normal distribution) the mean and standard deviation of the approximating posterior are derived. Although widely used for deriving a distribution over the output of the autoencoders, Yang et al. (2021) have shown that Hamiltonian Monte Carlo sampling (HMC; Neal 2011) is more accurate than variational sampling, partly due to the fact that the samples follow Hamiltonian dynamics (Lowenstein 2012). HMC

thereby accurately represents the geometry of the problem, for example by finding the best optimization path and thus facilitating the loss minimization. We therefore present a new autoencoder method that uses HMC sampling, named Bayesian HMC Autoencoders (BaHaMAs). BaHaMAs is a general method that can be used for various time series, especially in the field of Earth sciences, where knowledge of uncertainty plays a pivotal role in the analyses (Kiani Shahvandi and Soja 2022a). To present the applicability of the method to Earth science data, we focus on the atmosphere and ocean and their role in causing variations of the rotation of the Earth (Brzeziński et al. 2002). This is achieved by analyzing the net effect of global atmospheric winds and pressure variations, as well as oceanic currents and mass redistribution, resulting in the AAM and OAM time series (Gross 2015).

As mentioned, AAM and OAM time series are usually provided without uncertainty information. AAM and OAM are widely used in different studies for the prediction of Earth Orientation Parameters (see e.g., EOPs; Gross 2015), including Dill et al. (2019); Kiani Shahvandi et al. (2022b); Kiani Shahvandi and Soja (2022d); Gou et al. (2023). Notwithstanding the improvements in predicting EOPs when AAM and OAM are used, the prediction accuracy can reach a limit due to the unquantified uncertainties of AAM and OAM, obscuring the predictability of EOPs. In case these uncertainties are available, it would be possible to design and use algorithms that directly take into account the uncertainty in the data for prediction purposes (Kiani Shahvandi and Soja 2022a). Among EOPs, this is particularly important for the case of Length of Day (LOD, unit milliseconds (ms)), because it is directly proportional to the AAM and OAM variations from subdaily to interannual time scales (Gross 2015; Kiani Shahvandi et al. 2024d). Therefore, we primarily focus on improving the prediction of LOD by quantifying the uncertainties in AAM and OAM. Note that the net sum of Hydrological and Sea-Level Angular Momentum functions (HAM and SLAM, respectively) as provided by GFZ (see, for instance, Dill et al. 2019) is close to zero (due to mass conservation in the continent-ocean system), thus the contribution of HAM and SLAM to the LOD prediction are not analyzed. We mainly focus on the short-term prediction of LOD (up to 30 days into the future) using the data provided by the International Earth Rotation and Reference Systems Service (IERS). The reason is the 4-week latency between the rapid and final data, implying that this forecasting horizon is most widely demanded and after this time interval the final observations are made available, making predictions less useful (Kiani Shahvandi et al. 2023a).

It should be mentioned that there are some studies that quantify the uncertainty in the AAM and OAM data.

Lehmann and Névir (2012) quantify the uncertainties in the AAM data by comparing them with radiosonde observations and analyzing the agreement between these two approaches. Quinn et al. (2019) follow a similar path for quantifying the OAM uncertainties by comparing their OAM time series with those of other studies. Koot et al. (2006), however, follow a different approach more similar to ours by analyzing the noise level in the AAM data by mathematical means rather than by comparing it with other series. Similarly, Li et al. (2023) analyze the noise level in the angular momentum data for the purpose of EOP prediction but do not quantify the uncertainties in the data. Dill et al. (2023) quantify the uncertainties in AAM but rather for the forecasts, implying what they provide is, as mentioned before, the prediction uncertainty rather than the uncertainty in the data. However, currently no uncertainty information is provided for AAM and OAM data by operational data centers. We therefore present our method for uncertainty quantification in these time series. In summary, the goals of the study are as follows:

- presenting a novel and general methodology for uncertainty quantification in time series,
- demonstrating the usefulness of the uncertainties derived for AAM and OAM by incorporating them in the prediction of LOD.

The rest of this paper is organized as follows. In Section Methods, the methodology is described in detail. In Section Data description, the data used for the numerical studies are described. Section Results and discussions is used to present the numerical results and analyses. Finally, Section Conclusions is used for drawing conclusions.

2 Methods

2.1 Rationale behind BaHaMAs

As mentioned in the Introduction, BaHaMAs is based on the assumption of the predictability of time series (see e.g., Xu et al. 2019). In fact, BaHaMAs captures epistemic uncertainty due to the model deficiencies (Sullivan 2015). Therefore, the degree to which we can predict the time series is a representation of aleatoric uncertainty. BaHaMAs essentially approximates the aleatoric uncertainties through a loss function akin to traditional least squares. Least squares estimators are shown to be unbiased (Lukman et al. 2020); therefore, the uncertainties estimated by BaHaMAs are mathematically convergent to the 'true' uncertainties, provided that sufficient number of samples are used in the algorithm. Nevertheless, we note that these uncertainties are approximate values and do not necessarily correspond to the 'true' yet

unknown uncertainties. For instance, instrument measurement errors might not be captured if their effect in the time series is not evident. In the following sections, however, we derive a condition on the BaHaMAs that guarantees the convergence to the 'true' uncertainties. Hence, the uncertainties derived by BaHaMAs are useful to gain insight to the time series and its inherent uncertainty.

Considering that BaHaMAs is based on autoencoders and Bayesian machine learning with HMC sampling, in the following we present the methodology in two parts. We first discuss the autoencoder part.

2.2 Autoencoders

We denote the input to the autoencoder f as X , where X can be multidimensional (such as batch dimension, feature dimension, among others). The autoencoder f itself can be written as the composition of m neural networks f_1, \dots, f_m as in Eq. (1). It is important to mention that these neural networks are parametric functions, i.e., having parameters $\theta_1, \dots, \theta_m$ that must be optimized to best represent and reconstruct the input:

$$f = f_m(f_{m-1} \dots (f_1)) \quad (1)$$

As mentioned in the Introduction, to create a bottleneck that allows to quantify the aleatoric uncertainty, we apply batch, layer, or instance normalization to the output of each function (except for f_m , which is the final output that needs to be mapped to the observations). Therefore, the following operator in Eq. (2) is applied to f_1, \dots, f_{m-1} :

$$\mathbb{N}(f_i) = \frac{f_i - \mu_{f_i}}{\sqrt{\sigma_{f_i}^2 + \epsilon}}, \quad i = 1, \dots, m-1 \quad (2)$$

where μ and σ denote the mean and standard deviation of the output f_i , respectively, and ϵ is a numerical stabilizer that prevents division by zero in the optimization phase (in case $\sigma_{f_i}^2$ is too small, the numerical results would be unstable). Similar to the case of DEs, a value of $\epsilon = 10^{-8}$ has proven to be quite effective in deriving accurate and representative uncertainties (Kiani Shahvandi et al. 2023a, 2024a, b) and is therefore adopted here as well. Even though the form of the normalization operator \mathbb{N} is similar for batch, layer, and instance normalization, the computation is different. In batch normalization μ, σ are computed for each feature in X separately, whereas in layer normalization they are computed for each input across all features. In case of multidimensional time series input (such as sequence of images), μ, σ are computed for each batch separately and over spatial dimensions (such as rows and columns in an image). However, if the input is a one-dimensional time series the

instance normalization is identical to batch normalization. The suitability of each of these three different normalization schemes depends on the problem under consideration. In Section Results and discussions we analyze the impact of these different approaches on the prediction accuracy of LOD. Finally, we note that even though our approach is based on the application of normalization operators as bottleneck, autoencoders in general do not need normalization. Even the very presence of bottlenecks is questioned (Yong and Brintrup 2022), as the main purpose of autoencoders is dimensionality reduction. However, since the application of normalization in autoencoders improves their sensitivity to out of distribution samples (Zhou 2022), we have applied them here in our approach.

2.3 Bayesian HMC

With the design of the architecture of the autoencoder as above, we proceed to the optimization phase, where the difference between the output of the autoencoders and the input should be minimized. In this phase, Bayesian HMC is applied. In Bayesian HMC, the posterior distribution in the Bayes theorem (see, for instance, Jaynes 2012), denoted by $P(\theta|X)$ where θ represents the parameters of autoencoder f , is approximated as in Eq. (3):

$$P(\theta|X) \approx P(X|\theta)P(\theta) \quad (3)$$

where $P(\theta)$ is the prior distribution of parameters and $P(X|\theta)$ is the likelihood of observations given the parameters $\theta = \{\theta_i\}_{i=1}^m$. $P(\theta)$ is usually formulated as having a normal distribution with zero mean and variance as a parameter that needs to be optimized in the algorithm (Yang et al. 2021). On the other hand, $P(X|\theta)$ has a Gaussian distribution with mean f . The optimization is done by minimizing $-\log(P(X|\theta)P(\theta)) = -\log(P(X|\theta)) - \log(P(\theta))$, where \log denotes the natural logarithm function. Following Yang et al. (2021), we provide the explicit form of the optimization problem in Eq. (4). Note that constants such as $\frac{N}{2} \log(2\pi)$ that would appear in the loss function are omitted, since they neither participate in the optimization nor influence the results:

$$\begin{aligned} \min_{\theta, \sigma_\theta} -\log(P(\theta|X)) &= \min_{\theta, \sigma_\theta} (-\log(P(X|\theta)) - \log(P(\theta))) \\ &= \min_{\theta, \sigma_\theta} \left(\frac{1}{2} \sum_{i=1}^N (f(X_i, \theta) - X_i)^2 \right. \\ &\quad \left. + \frac{1}{2} \sum_{j=1}^{N_\theta} \log(\sigma_{\theta_j}^2) + \frac{|\theta_j|^2}{\sigma_{\theta_j}^2} \right) \end{aligned} \quad (4)$$

In this equation, N is the total number of samples in the time series and N_θ the total number of parameters. $\sigma_\theta = \{\sigma_{\theta_i}\}_{i=1}^m$ are the standard deviation of parameters considered as hyperparameters, which can be optimized together with the parameters θ by gradient descent and its variants (Bengio 2000; Bergstra and Bengio 2012). In this case, the uncertainty of parameters is part of the parameters denoted as θ , which are then jointly optimized (in other words $\theta \leftarrow [\theta, \sigma_\theta]$). To optimize Eq. (4), Bayesian HMC is applied, where Q samples $\theta^{(k)}$, $k = 1, \dots, Q$ are generated and used to compute the mean and standard deviation of the autoencoder output. First, $\theta^{(0)}$ is generated by random sampling from normal distribution $\mathcal{N}(0, 1)$. Then, for each sample $\theta^{(k)}$ we follow the optimization scheme as follows. First, an auxiliary variable r at step $k - 1$ (i.e., $r^{(k-1)}$) is sampled from multidimensional normal distribution $\mathcal{N}(0, I_{N_\theta})$, where I denotes the identity matrix. This is due to the fact that Hamiltonian dynamics (in differential form) is represented by two intertwined variables θ and r . $\theta^{(k-1)}$ and $r^{(k-1)}$ are subsequently used as initial values θ_0 and r_0 for the optimization step with learning rate δt and L epochs, as in Eq. (5):

$$U(\theta) = \frac{1}{2} \sum_{i=1}^N (f(X_i, \theta) - X_i)^2 + \frac{1}{2} \sum_{j=1}^{N_\theta} \log(\sigma_{\theta_j}^2) + \frac{|\theta_j|^2}{\sigma_{\theta_j}^2}$$

for $s = 0, \dots, L - 1$:

$$r_s \leftarrow r_s - \frac{1}{2} \delta t \nabla U(\theta_s)$$

$$\theta_{s+1} \leftarrow \theta_s + \delta t r_s$$

$$r_{s+1} \leftarrow r_s - \frac{1}{2} \delta t \nabla U(\theta_{s+1})$$
(5)

The final values θ_L and r_L from Eq. (5) are passed through the Metropolis-Hastings step (see, for instance, Robert 2015), whereby a random number p is generated from the uniform distribution $\mathcal{U}(0, 1)$ and used to ensure that the derived values sufficiently satisfy the Hamiltonian condition as follows:

$$H(\theta, r) = U(\theta) + \frac{1}{2} |r|^2$$

$$\alpha = \min \left(1, \exp(H(\theta_L, r_L)) - H(\theta^{(k-1)}, r^{(k-1)}) \right)$$

if $p \geq \alpha$:

$$\theta^{(k)} = \theta_L$$

if $p < \alpha$:

$$\theta^{(k)} = \theta^{(k-1)}$$
(6)

After performing the mentioned steps for all the samples $k = 1, \dots, Q$, a Markov chain is derived:

$\{f(X, \theta^{(Q-k+1)})\}_{k=1}^Q$. The mean of this chain, μ , is the reconstructed value. What we are interested in is the standard deviation of this chain, σ , which represents the aleatoric uncertainty in the data, computed as in Eq. (7):

$$\mu = \frac{1}{Q} \sum_{k=1}^Q f(X, \theta^{(Q-k+1)})$$

$$\sigma = \sqrt{\frac{1}{Q} \sum_{k=1}^Q (f(X, \theta^{(Q-k+1)}) - \mu)^2}$$
(7)

2.4 Convergence property of BaHaMAs

Here, we discuss whether the uncertainties derived by BaHaMAs correspond to the 'true' uncertainties. In other words, our goal is to derive certain conditions under which the uncertainties derived by BaHaMAs are convergent to the actual uncertainties in the data. For this purpose, we need to analyze the consistency (or unbiasedness) of BaHaMAs. The general approach for consistency theorems is based on theorems derived by Jennrich (1969) and Wu (1981). Following Wu (1981), we present the following proposition.

Proposition 1 *The uncertainties derived by BaHaMAs correspond to the 'true' uncertainties provided that for two different set of parameters θ and θ' the quantity $|f(X, \theta) - f(X, \theta')|$ scales proportional to N^{-p} , where $\frac{1}{2} < p \leq 1$.*

The proof of the proposition is given in the Appendix.

Proposition 1 provides important constraints on the convergence property of BaHaMAs: to derive accurate uncertainties, the number of samples Q from the parameters space should be appropriate. Furthermore, the role and choice of bottleneck is important in that the algorithm should not fully reconstruct the input, as the condition above will be violated. In Section Results and discussions we discuss the appropriate choices of Q and bottleneck more thoroughly.

2.5 BaHaMAs applied to sequences

The relations in Eq. (7) are valid if the output appears only once in the computations. In other words, if the inputs and outputs to the algorithm are taken from a moving window in the time series, each output would appear several times in the calculations. One of the reasons to use moving windows on the time series is to take into account the temporal dependency in the data. For instance, consider that for the first sample the first 30 values (i.e., 1,...,30) of the time series are taken as input (and because the architecture is an autoencoder they are

the outputs as well). Then, the second sample contains the next 30 values with one shift (i.e., 2,...,31). But these samples have 29 overlapping values, which must be taken into account when computing μ, σ from Eq. (7).

Let us assume samples are taken in a moving window of length M from $X_i, i = 1, \dots, N$. Then, the total number of samples is $N - M + 1$, each with length M . The maximum number of times a sample appears in the outputs is then equivalent to $z = \min(M, N - M + 1)$. Note, however, our goal is to assign uncertainty to each individual value in the time series and therefore, $X_i, i = 1, \dots, N$ should be considered. The first z values in $X_i, i = 1, \dots, N$ for each of the N outputs appear $h_i = i, i = 1, \dots, z$ times respectively. The next $N - 2z$ values all appear $h_i = z, i = z + 1, \dots, N - z$ times. Finally, the last z values appear $h_i = N - i + 1, i = N - z + 1, \dots, N$ times. Based on these, the relations in equations in (7) are modified as in Eq. (8):

$$\begin{aligned} \mu &= \frac{1}{Q} \sum_{k=1}^Q \sum_{j=1}^{h_k} \frac{1}{h_k} f^{(j)}(X, \theta^{(Q-k+1)}) \\ \sigma &= \sqrt{\frac{1}{Q} \sum_{k=1}^Q \sum_{j=1}^{h_k} \frac{1}{h_k} \left(f^{(j)}(X, \theta^{(Q-k+1)}) - \mu \right)^2} \end{aligned} \quad (8)$$

where $f^{(j)}$ represents the j -th time the value appears in the outputs. It is important to note that the definition of σ can be changed to represent only the mean of the individual standard deviations for each $f^{(j)}$ in Eq. (8). Even though simpler, we consider Eq. (8) to be more accurate, because it takes into account the contribution of each output to the mean and standard deviation by the number of instances they appear in the output. Finally, the same convergence property as in Proposition 1 applies here as well.

2.6 Architecture of BaHaMAs

It should be mentioned that since BaHaMAs is based on autoencoders, m (i.e., number of neural networks) is usually an even number in order for the algorithm to represent a symmetrical architecture together with the input. The hidden dimension of the encoder part of autoencoders, i.e., neural networks $f_1, \dots, f_\ell, \ell = \frac{m-2}{2}$, form a descending sequence. The lowest hidden dimension is that of the intersection between the encoder and decoder parts of the autoencoder, i.e., $f_{\ell+1}$ (in our numerical analyses it has dimension 1). On the other hand, the hidden dimension of neural networks $f_{\ell+2}, \dots, f_m$ form an ascending sequence and that of f_m matches the dimension of the input X . Note that in the general case each of $f_i, i = 1, \dots, m$ are neural networks with arbitrary number of layers. However, since we are using m neural

networks and to facilitate training and interpretation, we use neural networks with only one layer. Therefore, the autoencoder is a stack of m neural networks each with one layer. Finally, it should be mentioned that the form of $f_i, i = 1, \dots, m$ can be chosen from a vast number of possible architectures. We focus on Multi-Layer Perceptron (MLP; refer to Bishop 2006). Additionally, we use Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber 1997) to analyze the influence of model architecture on the robustness of the results. In summary, in Fig. 1, we have shown the schematic architecture of BaHaMAs in its general case.

2.7 Application of BaHaMAs in earth rotation analysis and its importance

To analyze the influence of the derived uncertainties as in Eq. (8), we follow the approach of Kiani Shahvandi and Soja (2022a), in which the loss function is modified according to the observational uncertainty (reciprocal of the variance) to pay more attention to the values that are more accurate. The usefulness of these uncertainties can be analyzed by computing the prediction performance improvement of an independent – yet related – time series in case of using these uncertainties against the case when they are not. As mentioned, the time series that we use for this purpose is that of LOD. For the prediction of LOD, we use the model proposed by Gou et al. (2023), which is based on LSTM neural networks. This model is composed of an Encoder-Decoder LSTM architecture which contains three consecutive LSTM layers with 10, 14, and 7 hidden neurons, respectively, followed by two time-distributed dense layers with 14 and 1 hidden neurons, respectively. Our analyses are mainly based on the application of BaHaMAs to the AAM and OAM time series. However, we also analyze the case where we also use AAM and OAM forecasts to improve the prediction accuracy.

We focus on this application, because prediction of LOD is an important task in spacecraft navigation and orientation of deep space telescopes (Gou et al. 2023). Thus, by improving the LOD predictions, we contribute to the improvement in important space geodetic applications. Furthermore, since AAM and OAM are quantities that can be used for the study of weather and climate, knowledge of the uncertainty of AAM and OAM plays an important role in understanding ways to improve the accuracy of AAM and OAM, thus benefiting the environmental studies.

2.8 Evaluation metrics

We note that our metric for prediction accuracy is Mean Absolute Error (MAE), as established in various EOP prediction studies (Kalarus et al. 2010; Kiani Shahvandi

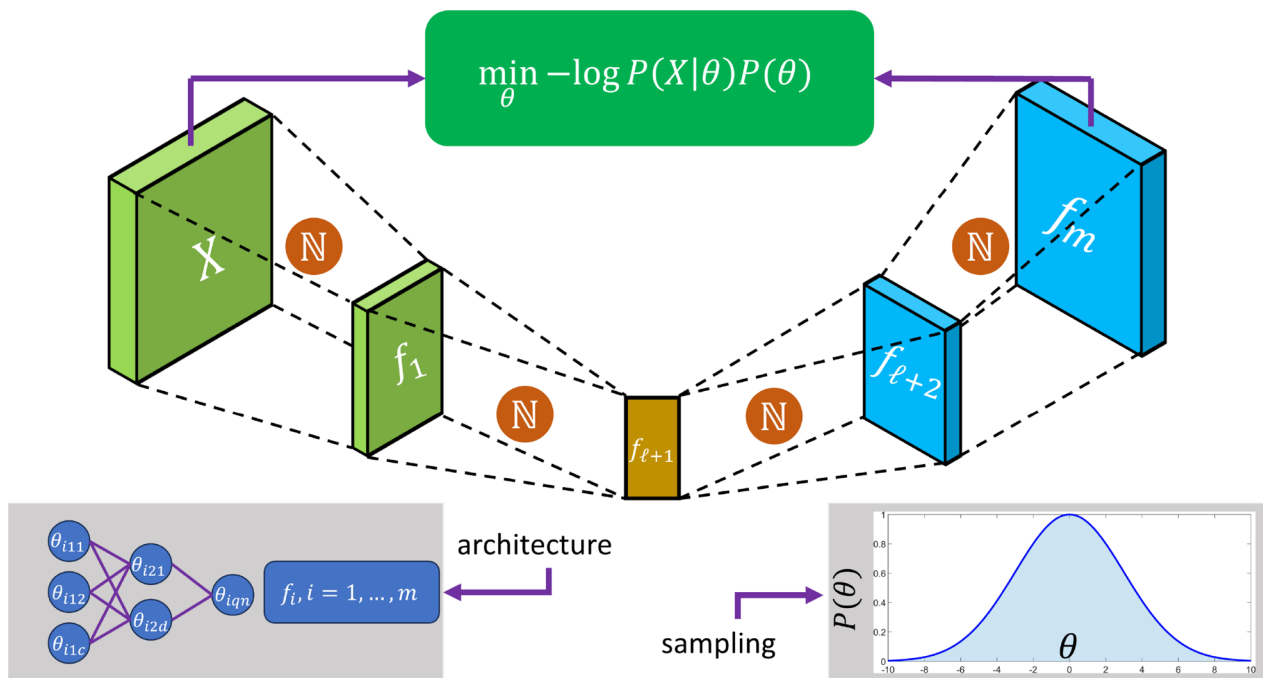


Fig. 1 Proposed architecture of BaHaMAs in its general case where the input X is multidimensional and each of the neural networks $f_i, i = 1, \dots, m$ has multiple layers with parameters $\theta_{i,l,n}$ (i : neural network number; l : layer number; n : hidden neuron number). Sampling of θ through HMC is performed from normal and Gaussian distributions. N represent the normalization in Eq. (2)

et al. 2022b). If the predicted and the final value to evaluate against are \hat{y} and y , respectively, MAE is computed as in Eq. (9) based on D prediction values (i.e., total number of predictions is D):

$$MAE = \frac{1}{D} \sum_{j=1}^D |y_j - \hat{y}_j| \tag{9}$$

The improvement (in percentage) is computed by comparing the MAE of the cases with and without the AAM and OAM uncertainties considered, denoted by MAE^w and MAE^{wo} respectively, as in Eq. (10):

$$\text{improvement} = 100\% \frac{MAE^{wo} - MAE^w}{MAE^{wo}} \tag{10}$$

Figure 2 summarizes the algorithm for the quantification of uncertainty in AAM and OAM, and their use in the prediction of LOD.

3 Data description

3.1 Data used in the analyses

For the numerical results presented in the paper, we use three different data sets. First, we use the AAM and OAM series provided by GFZ (Dobslaw et al. 2010). The temporal resolution of the data is 3 h and the data span

the range 1976 to the present. Since angular momentum is either caused by mass redistribution or motion relative to the reference frame (Gross 2015), the data are divided into mass and motion terms. GFZ uses numerical weather model data of ECMWF (Uppala et al. 2005) and the climate-ocean forcing models to compute the net effect of global pressure and zonal and meridional wind variations through integral formulas defining AAM and OAM. Even though for LOD variations the motion terms are by far the most important (Gross 2015), we also analyze the mass terms. We only analyze the axial component of these series, because they are directly proportional to LOD. The equatorial components could be analyzed for the case of polar motion prediction in future studies.

Regarding LOD, we use the recent IERS EOP series named IERS 20 C04 (Petit and Luzum 2010). The reason to use this series is that Kiani Shahvandi et al. (2023a) have shown this series is more accurate than its predecessor, i.e., IERS 14 C04 (Bizouard et al. 2019) for the application of EOP prediction, mainly due to the rigorous alignment to the International Terrestrial Reference Frame 2020 (Altamimi et al. 2023).

For the highest accuracy in the prediction of EOPs, it is important to use AAM and OAM forecasts. The

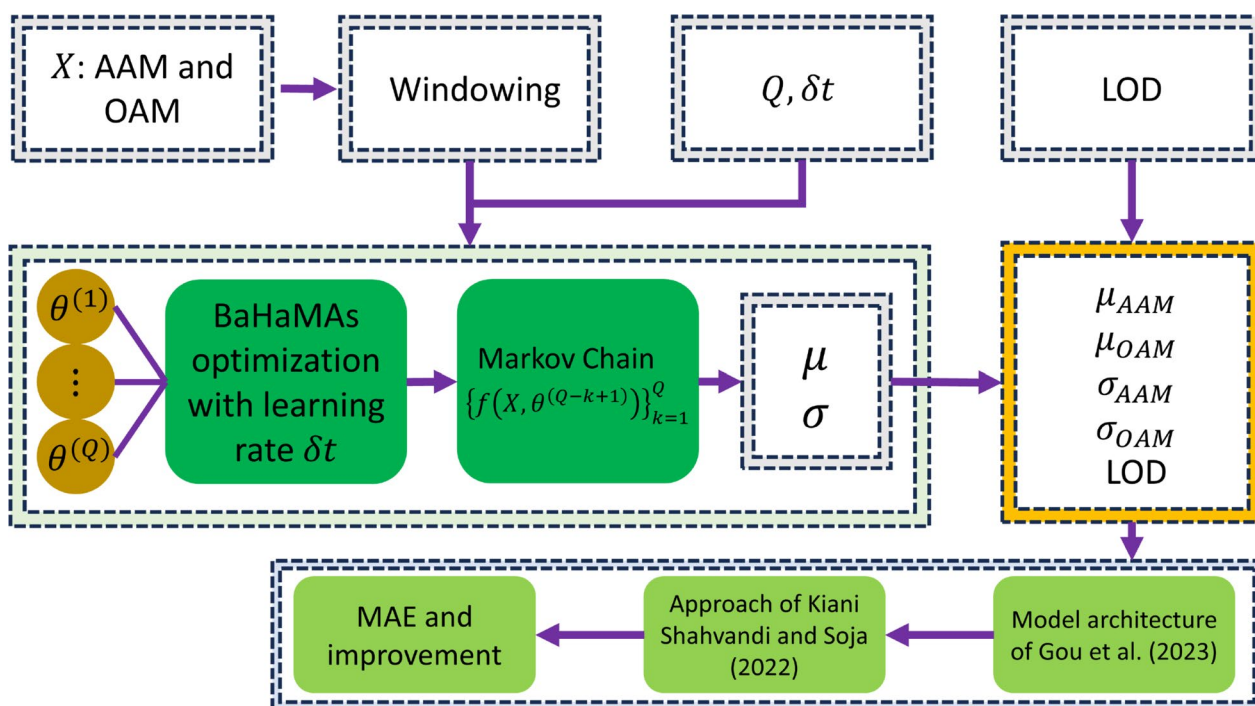


Fig. 2 Summary of the BaHaMAs algorithm for the quantification of AAM and OAM uncertainty and their use in the prediction of LOD

length of the forecasts is also important (Kur et al. 2022). For these reasons, we use the 14-day-ahead AAM and OAM forecasts provided by Geodetic Prediction Center (GPC; Soja et al. 2022) at ETH Zurich, which are shown to present state-of-the-art accuracy compared to the currently available 6-day-ahead forecasts (Kiani Shahvandi et al. 2022c). However, these series are provided without uncertainty information. We therefore apply our algorithm to these forecast data as well to first quantify the uncertainties and subsequently use them in the prediction of LOD.

Figure 3 shows the mass terms of axial AAM and OAM from 1976 to the end of 2022, i.e., our study period. Similarly, in Fig. 4 we display the motion terms of AAM and OAM in the study period. These are the input and output of the BaHaMAs algorithm. Each of the series is analyzed individually, because autoencoders take one time series at a time.

3.2 Studies and their evaluation intervals

The LOD observations since 1976 to the end of 2022 are shown in Fig. 5. Since the goal is to analyze the improvement in prediction performance once the uncertainties in the AAM and OAM data are derived and used in the prediction of LOD, we present two studies to rigorously analyze the results. In study 1, the LOD, AAM, and OAM data (with their uncertainties)

in the range 1976 to the end of 2000 are used for training, and LOD predictions are made in the range 2001 to the end of 2005. In study 2, on the other hand, the training and prediction intervals are 1976 to the end of 2017, and 2018 to the end of 2022, respectively. These separate intervals are long enough to give us robust results, thus adding confidence to the conclusions drawn.

4 Results and discussion

4.1 Configurations of BaHaMAs

In our default setting, we use MLP (with tangent hyperbolic activation function) as the base neural network. According to Gou et al. (2023), an input sequence length of 30 is effective for the prediction of LOD using AAM data. Considering this and that the temporal resolution of AAM and OAM data is 3 h (8 times per day), we chose our windowing size to be 240. Then, the first and second neural networks reduce the dimension from 240 to 8, and from 8 to 1. Subsequently, the decoder part increases the dimension from 1 to 8, and finally from 8 to 240 again. Furthermore, we utilize $Q = 50000$ samples, learning rate $\delta t = 5 \times 10^{-4}$, and number of epochs $L = 200$. However, we investigate the effect of sample size and model architecture on the prediction accuracy of LOD. The value chosen for the learning rate is confirmed and used in various studies (Gou et al. 2023; Kiani Shahvandi et al. 2023a,

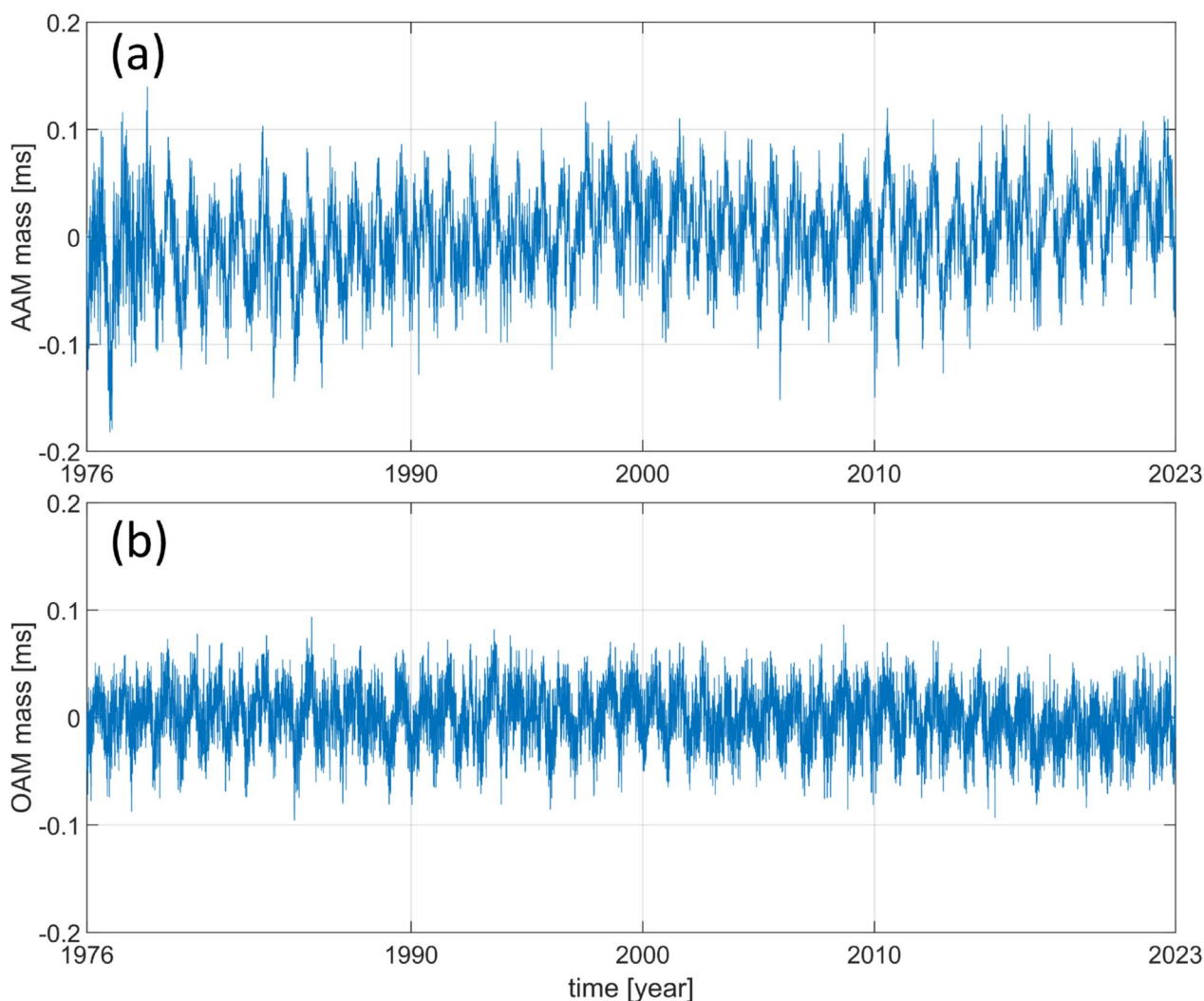


Fig. 3 Mass terms of **a** AAM and **b** OAM in the range 1976 to the end of 2022. The unit is ms

2024a). Furthermore, the number of epochs is quite robust as convergence is usually achieved in much fewer epochs than 200. The reason is the use of early stopping (Prechelt 2012) in the computations wherein if the loss function is not reduced after a certain number of epochs (in our case 30, from Gou et al. 2023), we assume that the minimum of the loss function has already been reached. Finally, the layer normalization is applied, but the impact of this choice and the comparison with other normalizations are presented in the following analyses.

4.2 Uncertainties derived by BaHaMAs

We show the results of applying BaHaMAs to the AAM and OAM in Fig. 6. The distribution of the uncertainties are also shown (at the level of one standard deviation).

From these figures, most of the uncertainties are below 0.06 ms. For AAM and OAM mass terms (as well as OAM motion term), the uncertainties are mostly around 0.01 ms. This is already close to the uncertainties of LOD time series as provided by IERS (around 0.01 ms), implying that AAM and OAM have the potential to improve the prediction performance of LOD prediction. For AAM motion term, the uncertainties in earlier times are larger and around 0.02 ms, but over time the uncertainties are reduced such that in the interval 2010–2023 they are close to 0.01 ms. By dividing the time series values with their associated uncertainty, overall the approximate signal-to-noise ratio is 90% or higher, implying that AAM and OAM time series are sufficiently accurate. The larger uncertainties around 1986 are due to the large and anomalous change of the AAM motion term in the time series

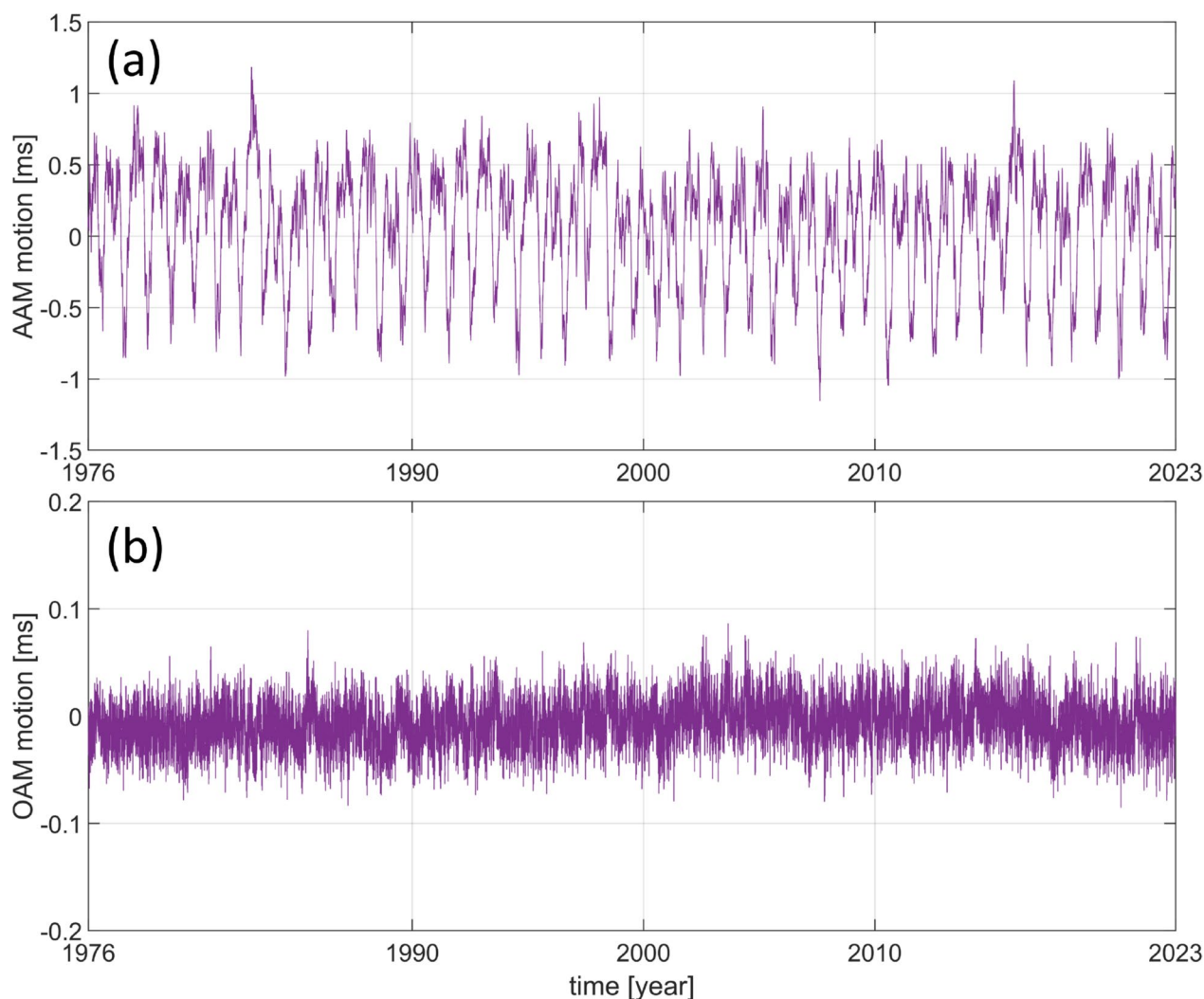


Fig. 4 Motions terms of **a** AAM and **b** OAM in the range 1976 to the end of 2022. The unit is ms

itself. This could be potentially attributed to the 1986 El Niño event which had a distinctive characteristic (Chen and Li 2018), causing the LOD to change rapidly.

4.3 Validating the uncertainties: comparison with differences between two independent series

We present the differences between AAM and OAM series of GFZ and another independent series, namely Systèmes de Référence Temps-Espace (SYRTE), in Figs. 7–8. Since these two data centers provide the time series independently, the mentioned differences might approximate the aleatoric uncertainty provided that the exact same mathematical approach is used to compute AAM and OAM. As shown from the figures, the differences for OAM mass and motion terms, as well as AAM mass term are in good agreement with the

uncertainty values we derived using BaHaMAs. However, for the AAM motion term, the differences are large and almost five times bigger than those derived by BaHaMAs. This might imply that the uncertainties derived by BaHaMAs for the AAM motion term are overoptimistic. Nevertheless, it is crucial to note that there are visibly pronounced signals in the AAM motion term differences that are probably related to the omission of some contributors to the AAM variations, such as the El Niño Southern Oscillation (ENSO) – represented as Multivariate ENSO Index (MEI; Wolter and Timlin 1993) – that exhibits similar amplitude and variations on interannual time scales (Zheng et al. 2003; Yu et al. 2021). By analyzing the differences between AAM and OAM series of GFZ and several other institutes, Dobslaw and Dill (2019) also conclude that there are

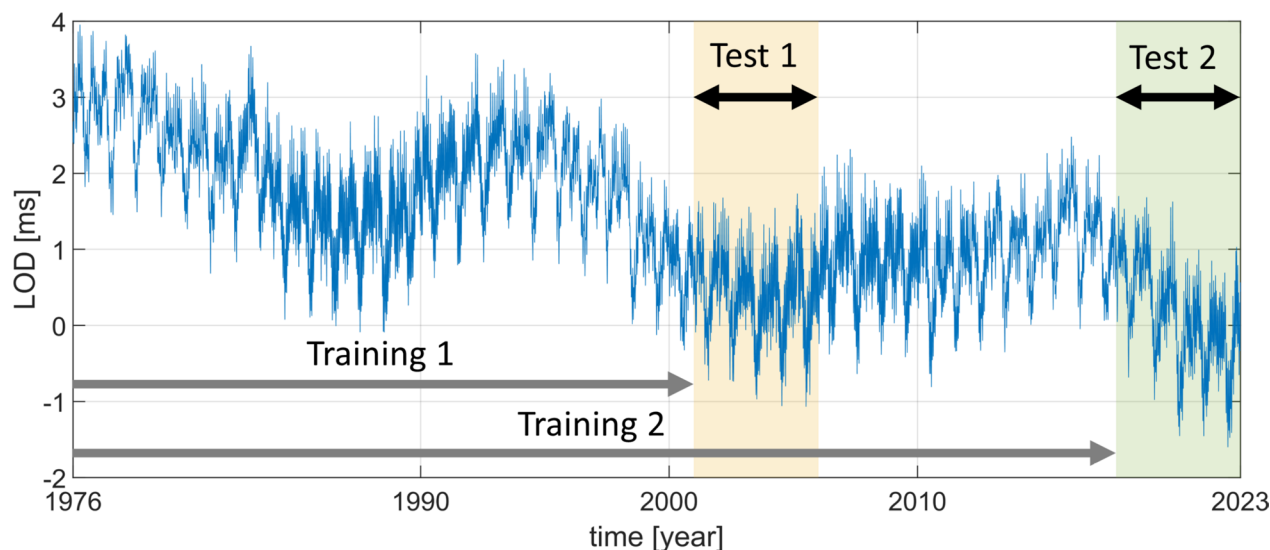


Fig. 5 LOD time series (in ms) from 1976 to the end of 2022. Two separate studies are presented. First: training in the range 1976–2001 and predicting in 2001–2006. Second: training in the range 1976–2018 and predicting in 2018–2023

large discrepancies that are most probably due to the computation of the AAM and OAM themselves (such as inaccuracies in the removal of tidal variations) and not so much related to the intrinsic uncertainty in the input data. Therefore, our derived uncertainty values might in fact more accurately represent the aleatoric uncertainties. Further evidence supporting our conclusion comes from the fact that the AAM motion term can be predicted to the future as accurately as 0.01 ms (Kiani Shahvandi et al. 2022c, 2023b, 2024c), suggesting that the aleatoric uncertainty in the data is similar to the value that we derive using BaHAMAs.

4.4 Prediction of LOD

Next, we use the uncertainties derived alongside the AAM and OAM time series in a machine learning framework in the loss function of which the uncertainties are considered (here only for the observations). This is inspired by the weighted least squares methodology and based on the approach of Kiani Shahvandi and Soja (2022a). The predictions made are then compared with the final IERS 20 C04 time series and the MAE is computed. We present the MAE for the case where uncertainties are used and compare it with the case when no uncertainty information is used in the LOD prediction algorithm, Fig. 9. As observed, in both studies mentioned in Fig. 5 the prediction accuracy is improved, confirming the usefulness of the uncertainties derived. The average improvement across all prediction horizons is 17% and 13% for study 1 and 2 respectively. We note that, longer predictions are less accurate and therefore, the potential

of improving predictions by using uncertainties is better realized (as much as 22%. The numerical values are summarized in the Appendix B).

4.5 Analyzing the impact of sample size

To test the impact of sample size (Q) on the robustness of the uncertainties derived, we compute the average MAE of LOD across all prediction horizons for Q in the range 1 to 50,000 (with increment 1). The results are presented in Fig. 10. As seen, the average MAE with $Q < 5000$ is almost twice as large as that of $Q = 50000$, implying that more than 5000 samples are needed for high accuracies. As suggested by this diagram, the relationship between sample size and the average MAE is exponential, resembling the traditional Monte Carlo sampling that scales with the reciprocal of the square root of sample size (Thomopoulos 2013). Even though as the sample size increases the average MAE decreases, no significant gain in accuracy is obtained after around $Q = 30000$. Nevertheless, we choose to use $Q = 50000$ samples to ensure that the highest accuracy is obtained in all other analyses as well, as presented in the following.

4.6 Analyzing the impact of model architecture

We analyze the influence of the model architecture on the robustness of the derived uncertainties by analyzing the corresponding LOD prediction accuracy. All the analyses so far have been based on the architecture of MLP. In Fig. 11, however, we compare the MAE of MLP with that of LSTM. While LSTM is less accurate than MLP in study 1 by as much as 0.5%, it presents a

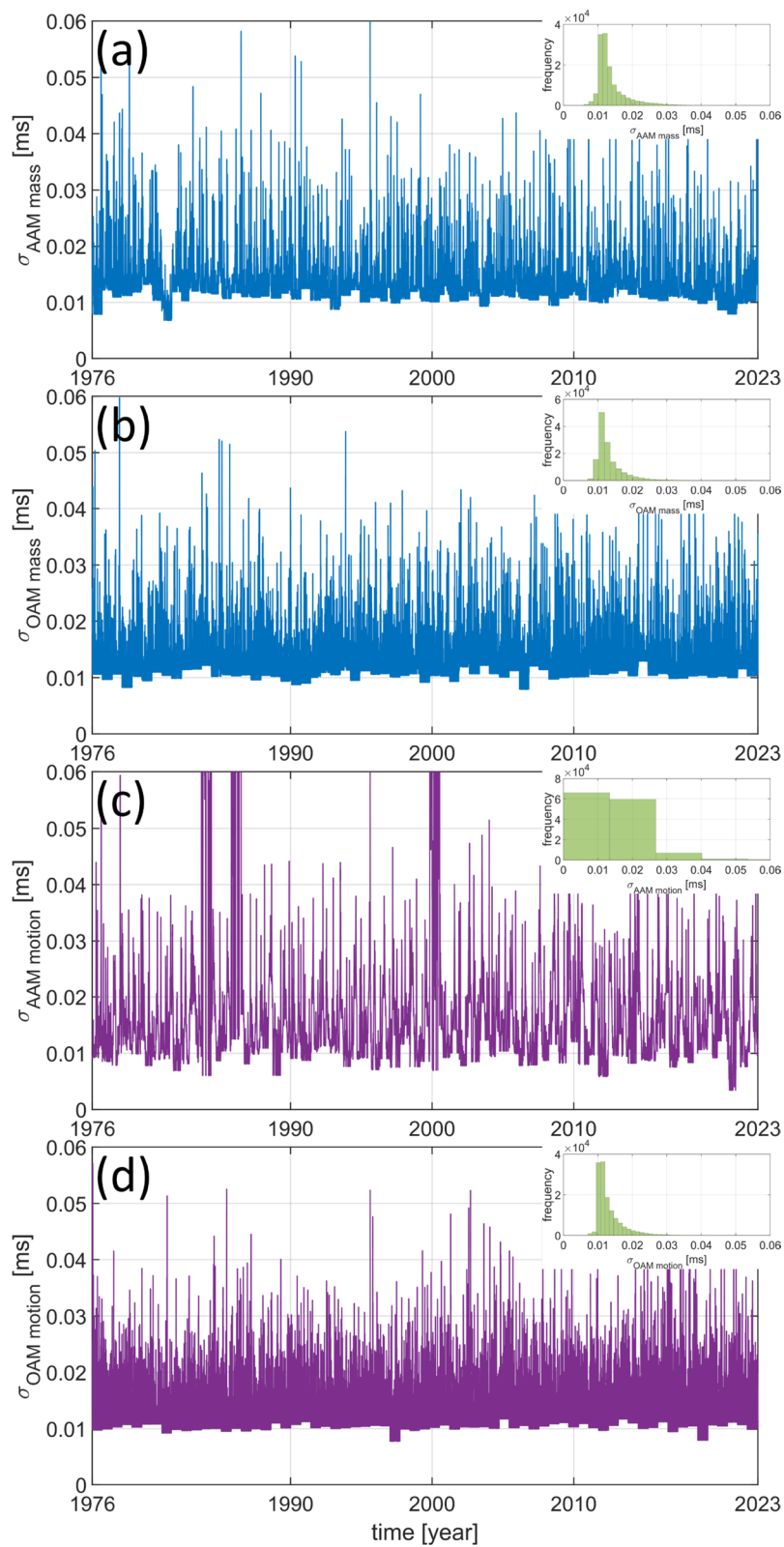


Fig. 6 Reconstructed aleatoric uncertainty by BaHaMAs for **a** AAM mass term, **b** OAM mass term, **c** AAM motion term, and **d** OAM motion term. The distribution of uncertainties (histogram) are also shown. The unit is ms

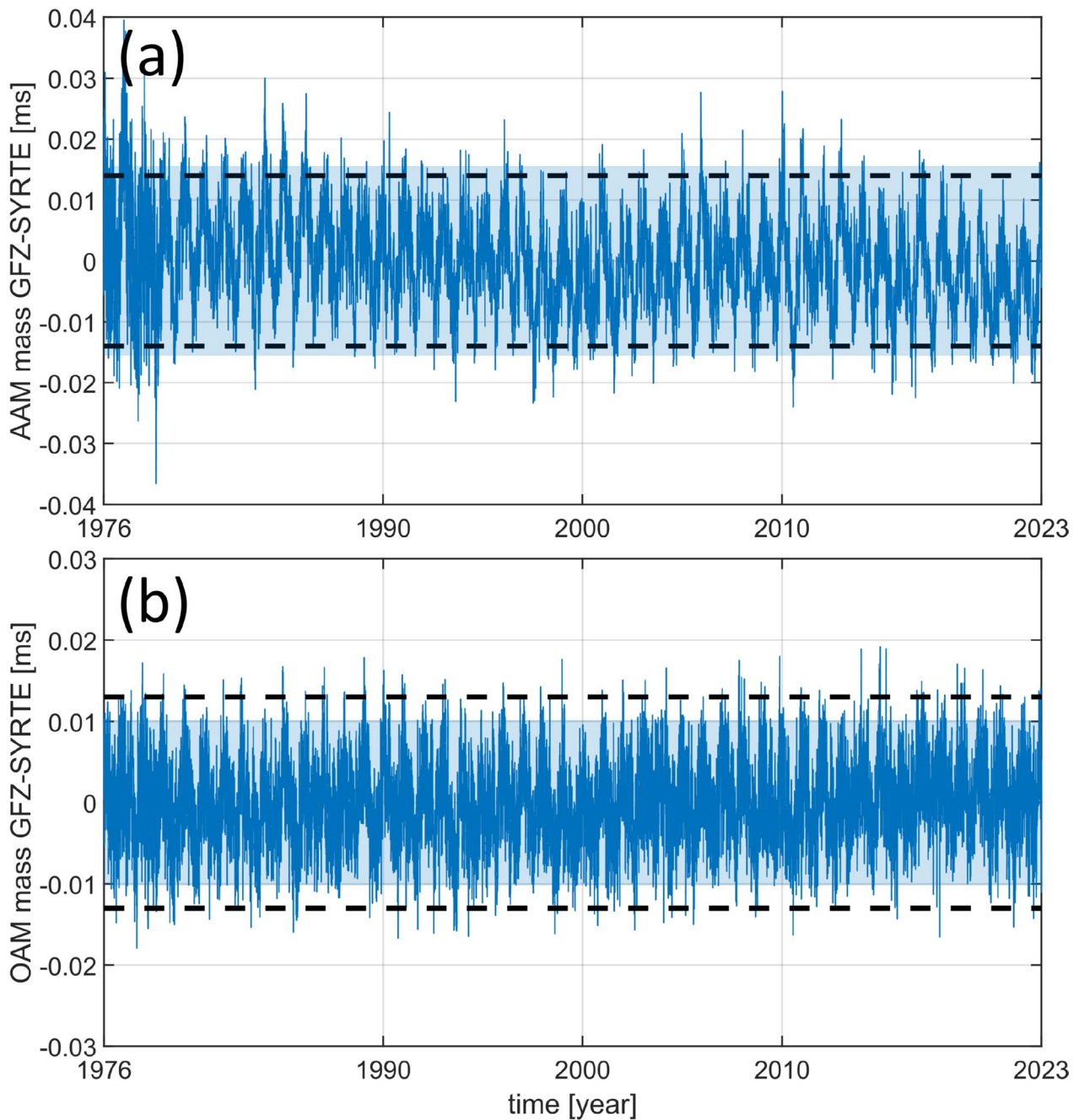


Fig. 7 Differences between mass terms of **a** AAM and **b** OAM of GFZ and SYRTE. The unit is ms. The shaded areas represent the region containing values in one standard deviation (computed across all these difference values). The horizontal black dashed lines show the region in one standard deviation derived from BaHaMAs

2% improvement on average compared to MLP in study 2. These values confirm that the role of model architecture on the uncertainties derived is marginal. This also implies that the epistemic uncertainties due to the model deficiencies are captured well enough by sampling a large

number of candidates for the prior such that changing the architecture from MLP to LSTM does not fundamentally alter the results. Therefore, we proceed to follow the MLP architecture in the remaining analyses.

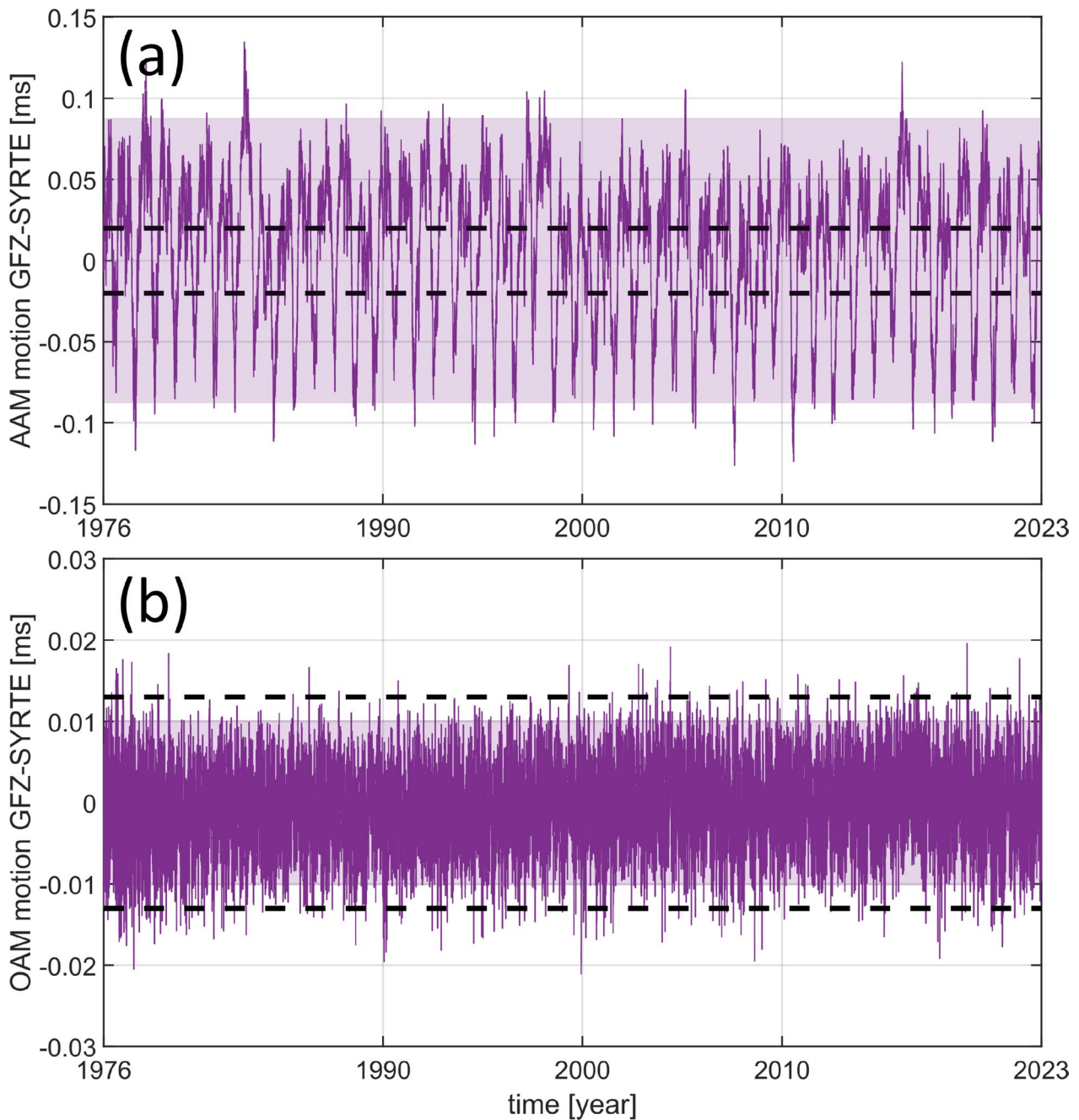


Fig. 8 Differences between motion terms of **a** AAM and **b** OAM of GFZ and SYRTE. The unit is ms. The shaded areas represent the region containing values in one standard deviation (computed across all these difference values). The horizontal black dashed lines show the region in one standard deviation derived from BaHaMAs

4.7 Analyzing the impact of normalization choice

To analyze the influence of the bottleneck (i.e., normalization type) in the architecture of BaHaMAs, we compare the batch and layer normalization approaches (note that instance normalization in our problem is equivalent to batch normalization). The results are presented in Fig. 12. Similar to the model architecture, it appears

that the role of normalization type on the uncertainties derived and the prediction accuracy of LOD is marginal. More precisely, however, layer normalization is on average 0.5% and 3% more accurate than the batch normalization, confirming the suitability of our choice of normalization type.

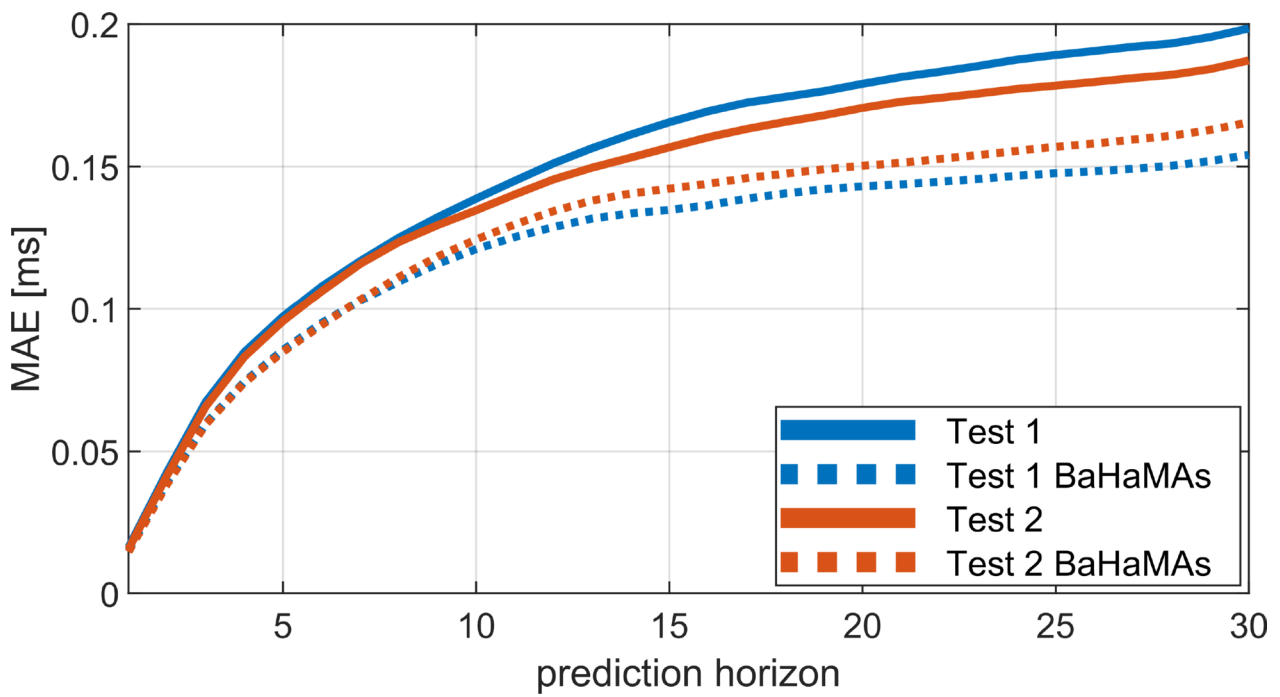


Fig. 9 Prediction accuracy of LOD in terms of MAE in ms for both studies 1 and 2. The results of two cases are compared: with and without the use of uncertainty values derived for AAM and OAM using BaHaMAs

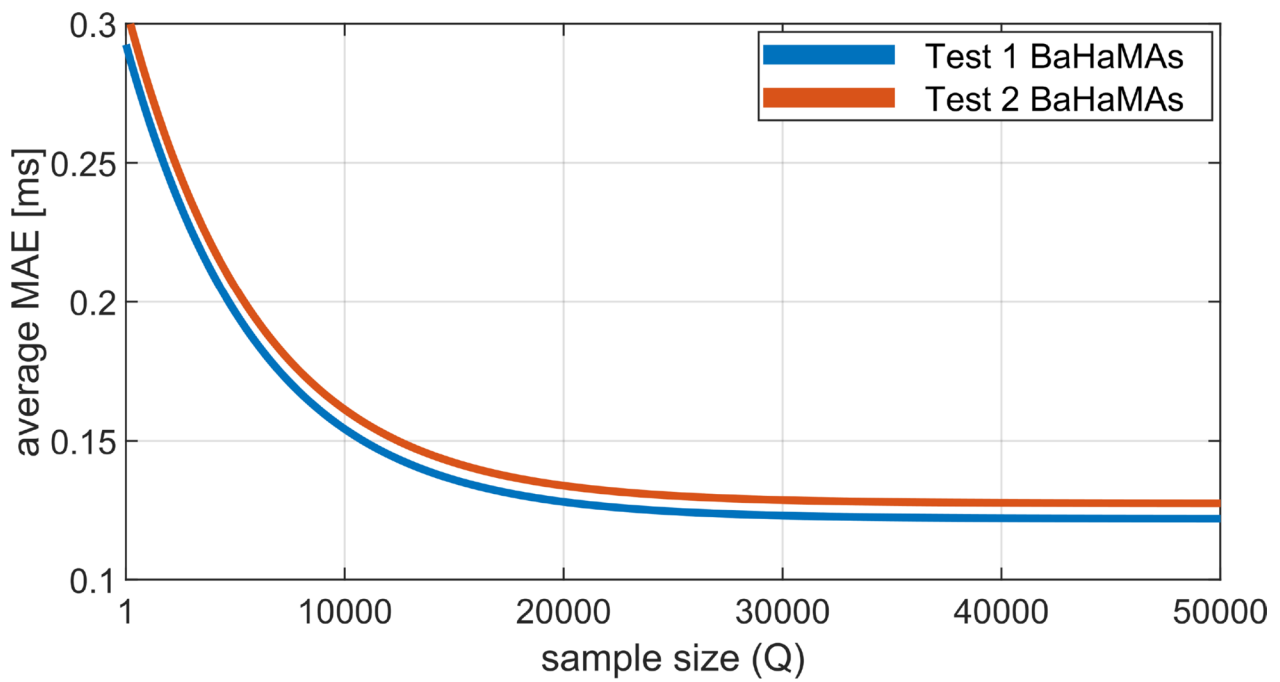


Fig. 10 Impact of sample size Q on the average prediction accuracy of LOD, for studies 1 and 2

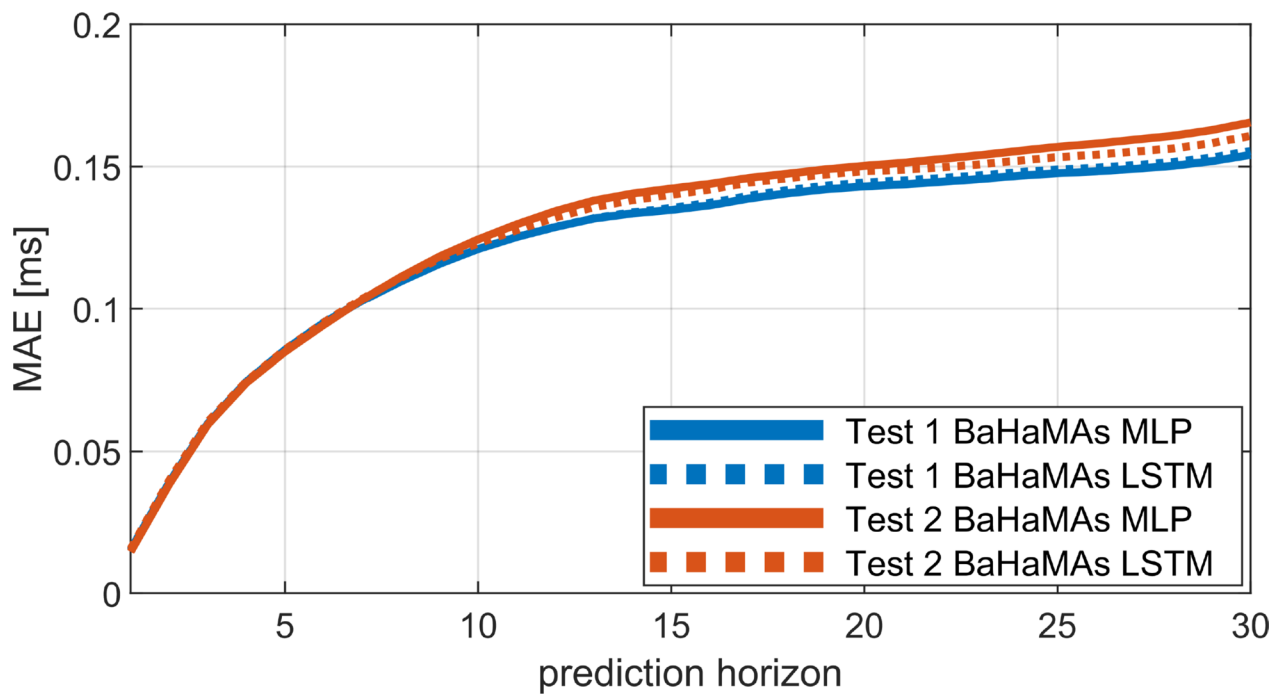


Fig. 11 Impact of the model architecture on the prediction accuracy of LOD using BaHaMAs algorithm for the determination of AAM and OAM uncertainties

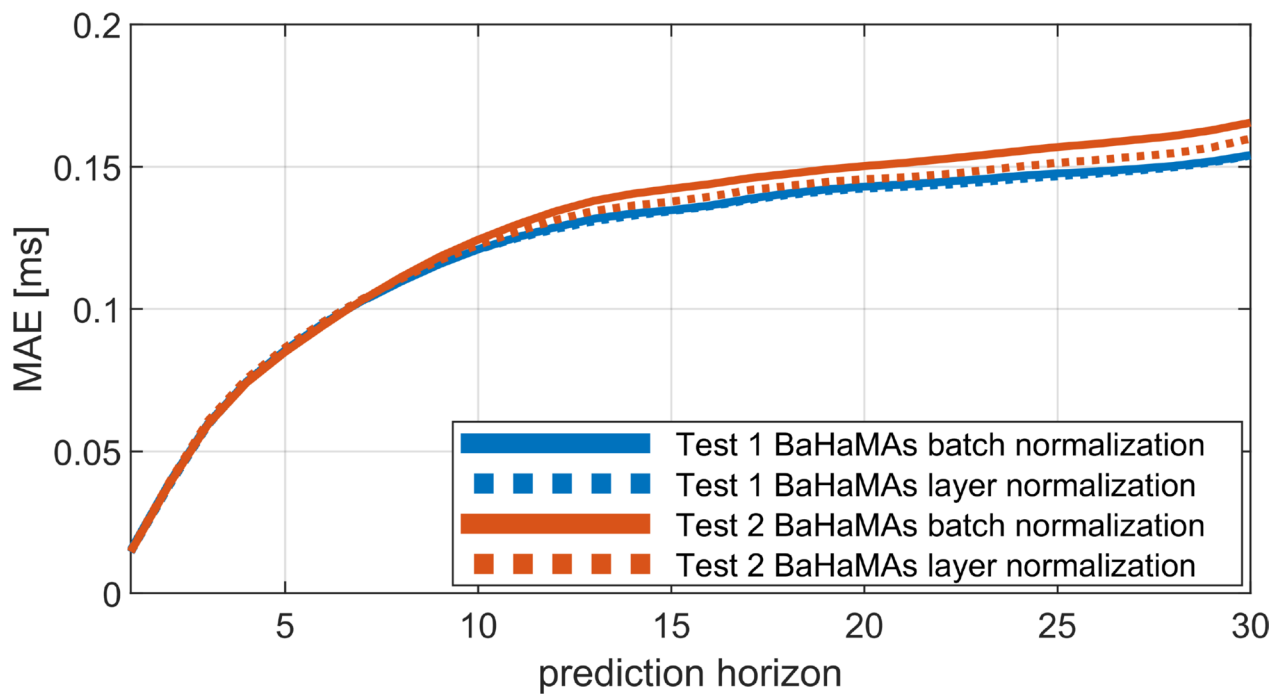


Fig. 12 Impact of the normalization type in BaHaMAs on the prediction accuracy of LOD

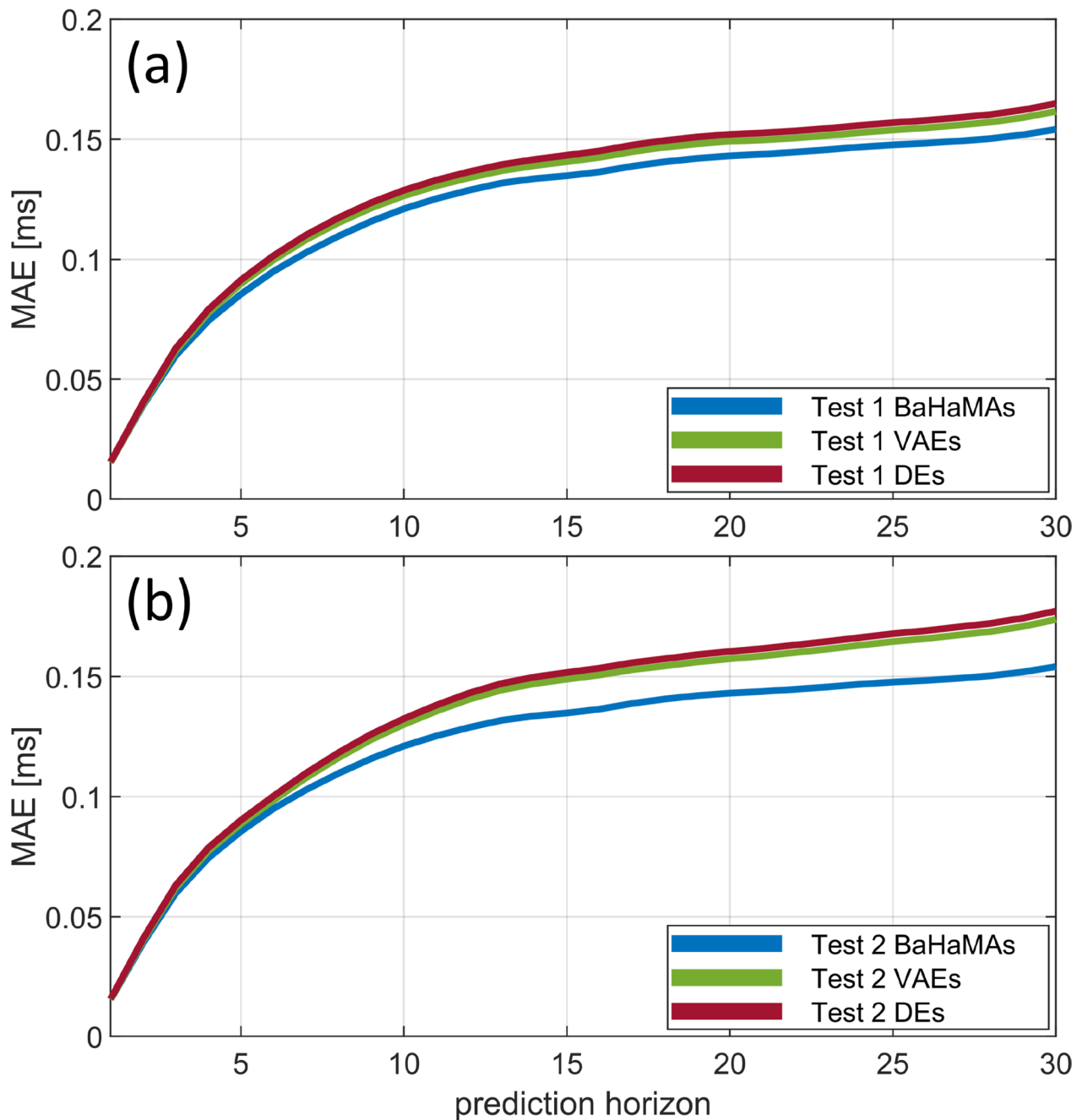


Fig. 13 Impact of the uncertainty algorithm on the prediction accuracy of LOD. The results of BaHaMAs, VAEs, and DEs are compared for study 1 and 2 in (a) and (b), respectively

4.8 Comparison of BaHaMAs with VAEs and DEs as baselines

To compare BaHaMAs with alternative uncertainty quantification algorithms, we present the results of VAEs and DEs. Similar to BaHaMAs, VAEs use $Q = 50000$ samples. For DE, 10 models are trained,

as suggested in several studies (Kiani Shahvandi et al. 2023a, 2024a) to present robust analyses for the prediction of EOPs (in fact we also tested the impact of the number of DE models, but after 10 models we did not see significant improvement in the results, implying that the results presented here are robust).

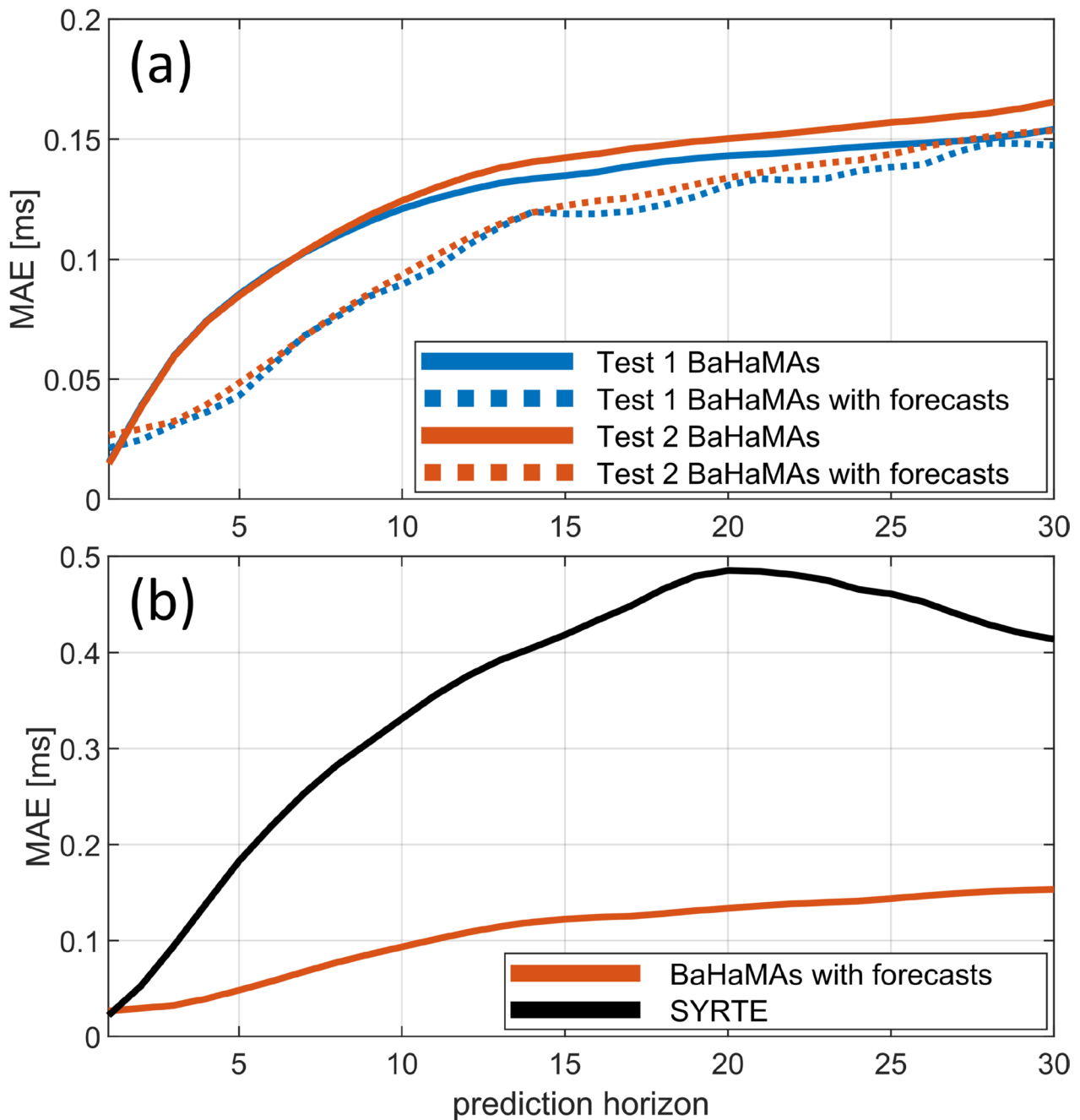


Fig. 14 **a** Impact of including AAM and OAM forecasts on the prediction accuracy of LOD, when the uncertainties are determined using BaHaMAs. **b** Comparison of the prediction accuracy of our algorithm with the independent predictions provided by SYRTE. Note here that the results are based on the predictions in the range May 20, 2021 to the end of 2022

The results are presented in Fig. 13, indicating a higher accuracy of BaHaMAs. In study 1, the average improvement of using BaHaMAs instead of VAEs and DEs is 5% and 7%, respectively. On the other hand, in study 2 the corresponding values are 9% and 10%.

4.9 Analyzing the impact of using AAM and OAM forecasts and comparison with independent predictions as baselines

Since it has been shown that AAM and OAM forecasts have significant impact on the prediction accuracy

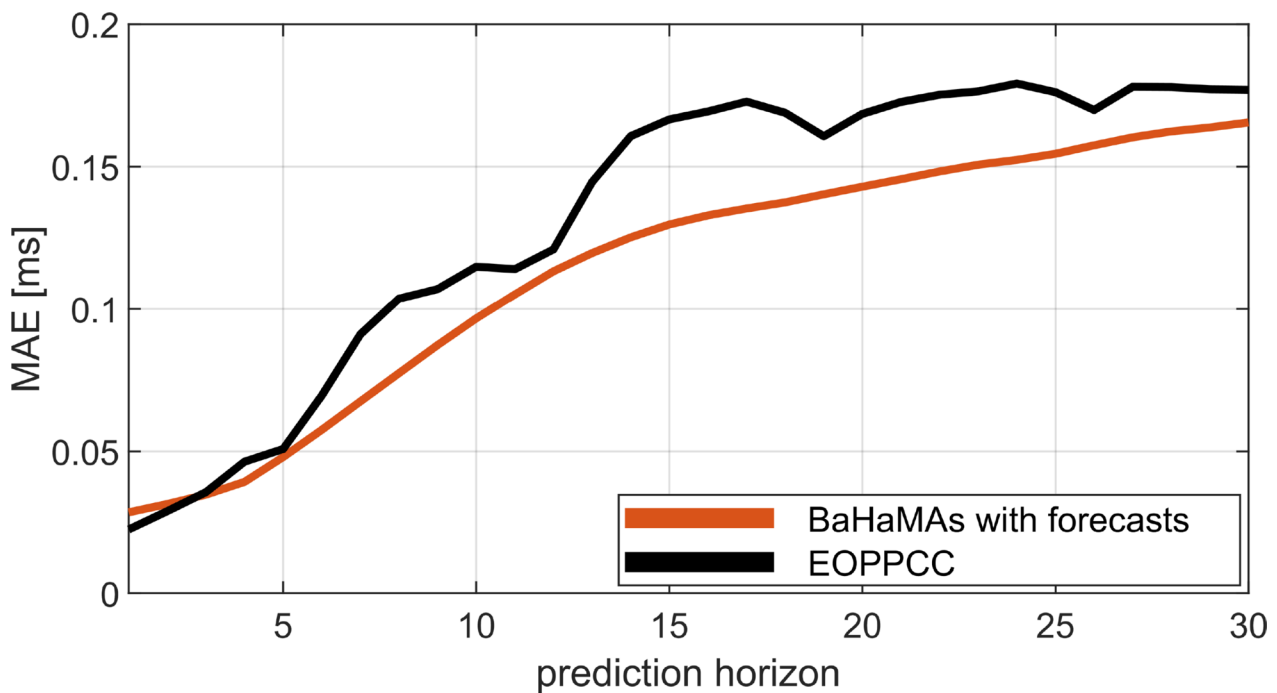


Fig. 15 Comparison of the prediction accuracy (in MAE) of LOD for the best results obtained during the second EOPPCC and those from our approach. The unit is ms

of LOD (Gou et al. 2023), we proceed to use 14-day-ahead forecasts provided by ETH Zurich, quantify their uncertainty, and use them in the prediction of LOD. As observed in Fig. 14(a), the inclusion of AAM and OAM forecasts with their uncertainties derived from BaHaMAs significantly improves the LOD prediction accuracy, on average by 30% and 31% for studies 1 and 2, respectively. In addition, to compare the prediction accuracy of our algorithm with an independent data provider, we use the LOD predictions of SYRTE (Bizouard et al. 2019). Since we have archived these operational predictions only since May 20, 2021, we only present the results for a part of the Test 2 time frame. The results are presented in Fig. 14(b). Comparisons with the predictions of SYRTE from May 2021 until the end of 2022 reveal that BaHaMAs predictions are on average 66% more accurate, demonstrating the high accuracy of our algorithm.

We also compare our results with the best result obtained in the second Earth Orientation Parameters Comparison Campaign (EOPPCC; Śliwińska et al. 2023, Śliwińska-Bronowicz et al. 2024), under the campaign conditions. The results are shown in Fig. 15. As it can be seen, we can improve the prediction performance of LOD by 12% on average (we note that these results are before various filtering applied by EOPPCC). This further confirms that using the uncertainties derived by BaHaMAs in the prediction of LOD is advantageous.

5 Conclusions

We present BaHaMAs, a method for quantifying the aleatoric uncertainty in time series using autoencoders based on Bayesian machine learning through Hamiltonian Monte Carlo sampling. In case a time series is provided without the information regarding uncertainty, BaHaMAs can be used to approximate the aleatoric uncertainty in the time series from itself by analyzing how predictable or stochastic the time series is. We apply the method to the AAM and OAM time series of GFZ and use the derived uncertainties alongside the AAM and OAM themselves for the short-term prediction of LOD. The prediction performance is improved by around 17% on average. Furthermore, BaHaMAs is more accurate than VAEs or DEs. Among the parameters that can impact the accuracy of BaHaMAs the most important is the sampling size (around 50000 is recommended). In terms of data, the most significant contribution to the improvement of prediction accuracy of LOD is from 14-day-ahead forecasts of AAM and OAM.

Prediction of LOD is important for a host of applications such as spacecraft navigation and orientation of deep space telescopes (Dobslaw and Dill 2018). As a result, the second EOPPCC (Śliwińska-Bronowicz et al. 2024) was organized to compare the prediction accuracy of different methods under operational settings. In the 30-day horizon our predictions were the most accurate

in terms of MAE, providing us with incentive to enhance our methodology to further improve the LOD predictions. With BaHaMAs we have made a step toward this goal and hope to motivate other researchers to consider incorporating the uncertainty quantification in their methods as well. We note, however, that one of the main shortcomings of BaHaMAs is that there is no universal, independent approach to compare the results of BaHaMAs against. Therefore, for various applications different ways of comparison should be designed based on prior knowledge. One way, as we have shown in this paper, is to compare the data of different institutions provided that the modelling approach is the same for all data centers. Finally, the uncertainties derived by BaHaMAs are approximate and do not necessarily correspond exactly to the 'true' yet unknown uncertainties. However, we derived conditions under which the uncertainties derived by BaHaMAs correspond to the actual uncertainties. Hence, these uncertainties are useful for the prediction purposes, as we have demonstrated in this paper.

Appendix A

Proof of proposition 1

According to Jennrich (1969) and Wu (1981), an estimator $\hat{\theta}$ is said to be consistent (weakly or strongly) if the quantity $\frac{1}{N}D_N(\theta, \theta')$ uniformly converges to a continuous function $D(\theta, \theta')$, and $D(\theta, \theta_0) = 0 \iff \theta = \theta_0$. D has the form as in Eq. (11):

$$D_N(\theta, \theta') = \sum_{i=1}^N \left(f(X_i, \theta) - f(X_i, \theta') \right)^2 \tag{11}$$

Now, we note that f is an autoencoder and thus, we can write

$$\begin{aligned} f(X_i, \theta) &= X_i + \xi_i \\ f(X_i, \theta') &= X_i + \eta_i \end{aligned} \tag{12}$$

in which ξ_i and η_i are functions of θ and θ' , as well as X . Putting Eq. (12) to (11) we get Eq. (13):

$$D_N(\theta, \theta') = \sum_{i=1}^N \left(\xi_i - \eta_i \right)^2 \tag{13}$$

As Wu (1981) states, the condition $\lim_{N \rightarrow \infty} D_N(\theta, \theta') \rightarrow \infty$ should be satisfied. According to theorems on divergence of infinite series (see e.g., Bourchtein and Bourchtein (2022)) if we assume $|\xi_i - \eta_i| \sim N^{-p}$, then $D_N(\theta, \theta') \sim N^{-2p+1}$ and therefore $p \leq 1$. On the other hand, $\frac{1}{N}D_N(\theta, \theta') \sim N^{-2p}$ should be convergent and therefore, $p > \frac{1}{2}$.

Therefore, the $D_N(\theta, \theta')$ should in general scale as N^{-p} , $\frac{1}{2} < p \leq 1$. Turning to second condition, if $\theta = \theta_0$ then $\xi_i = \eta_i$, implying $D_N(\theta, \theta') = 0$ and therefore $D(\theta, \theta') = \lim_{N \rightarrow \infty} \frac{1}{N}D_N(\theta, \theta') \rightarrow 0$. On the other hand, if $D(\theta, \theta') = 0$, then since $D_N(\theta, \theta') > 0$ we conclude that $\xi_i = \eta_i, \forall i$. Considering that ξ_i and η_i are continuous functions in terms of θ and θ' , the only possible way to satisfy $\xi_i = \eta_i, \forall i$ is $\theta = \theta_0$.

Appendix B

Numerical values of MAE and improvement in Fig. 9

The numerical values of the MAE and improvement in Fig. 9 are given in the following Table 1.

Table 1 Numerical values of the MAE (in ms) and improvement (in percentage) related to Fig. 9

Forecast horizon	MAE study 1	Improvement study 1	MAE study 2	Improvement study 2
1	0.015	5.4	0.014	6.0
2	0.039	8.6	0.038	7.4
3	0.059	11.2	0.059	9.5
4	0.074	12.2	0.074	10.8
5	0.085	12.1	0.084	11.2
6	0.095	11.9	0.094	11.2
7	0.102	12.0	0.103	10.8
8	0.109	12.2	0.111	9.9
9	0.115	12.3	0.118	8.5
10	0.121	12.7	0.124	7.5
11	0.125	13.6	0.129	7.4
12	0.128	14.7	0.134	7.6
13	0.131	15.7	0.138	7.7
14	0.133	17.1	0.140	8.2
15	0.134	18.5	0.142	9.2
16	0.136	19.4	0.143	10.1
17	0.138	19.5	0.146	10.5
18	0.140	19.3	0.147	10.9
19	0.142	19.4	0.149	11.2
20	0.143	20.0	0.150	11.9
21	0.143	20.7	0.151	12.4
22	0.144	21.0	0.152	12.3
23	0.145	21.3	0.154	12.2
24	0.146	21.7	0.155	12.2
25	0.147	21.9	0.156	12.0
26	0.148	22.1	0.158	12.0
27	0.149	22.2	0.159	11.9
28	0.150	22.2	0.160	11.7
29	0.151	22.2	0.162	11.5
30	0.154	22.3	0.165	11.5

Abbreviations

BaHaMAs Bayesian Hamiltonian Monte Carlo Autoencoders

DEs	Deep Ensembles
EOPs	Earth Orientation Parameters
IERS	International Earth Rotation and Reference Systems Service
MAE	Mean Absolute Error
mas	Milliarcseconds
SYRTE	Systèmes de Référence Temps-Espace
VAEs	Variational Auto Encoders

Acknowledgements

Data centers the data of which we have used in this study are greatly appreciated. These include IERS for LOD data, SYRTE for LOD predictions, and GFZ for AAM and OAM series.

Author contributions

Mostafa Kiani Shahvandi conceived the research, implemented the method, analyzed the results, and wrote the original draft. All the authors read the manuscript and commented upon it.

Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich.

Data availability

The IERS 20 C04 series can be accessed via <https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html>. GFZ AAM and OAM data are available at <http://rz-vm115.gfz-potsdam.de:8080/repository>. ETH Zurich AAM and OAM forecasts can be accessed at <https://gpc.ethz.ch/EAM/>. SYRTE LOD predictions can be accessed at <https://hpiers.obspm.fr/eop-pc/index.php>. EOPCC predictions can be accessed at <https://doi.org/10.5880/GFZ.1.3.2023.001>.

Declarations

Competing interests

The authors declare that they do not have Conflict of interest.

Author details

¹Institute of Geodesy and Photogrammetry, ETH Zürich, Zürich, Switzerland.

²Seminar for Applied Mathematics, Department of Mathematics, and ETH AI Center, ETH Zürich, Zürich, Switzerland.

Received: 12 April 2024 Accepted: 4 September 2024

Published online: 04 October 2024

References

- Altamimi Z, Rebischung P, Collilieux X, Métivier L, Chanard K (2023) ITRF2020: an augmented reference frame refining the modeling of nonlinear station motions. *J Geodesy* 97:47 <https://doi.org/10.1007/s00190-023-01738-w>
- Ba JL, Kiros JR, Hinton GE (2016) Layer normalization. *Advances in NIPS 2016 deep learning symposium*
- Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. *J Mach Learn Res* 27:37–50
- Bengio Y (2000) Gradient-based optimization of hyperparameters. *Neural Comput* 12(8):1889–1900
- Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *J Mach Learn Res* 13(2):281–305
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer Information Science and Statistics, Cham
- Bizouard C, Lambert S, Gattano C, Becker O, Richard JY (2019) The IERS EOP 14C04 solution for Earth orientation parameters consistent with ITRF 2014. *J Geodesy* 93:621–633
- Bourchtein L, Bourchtein A (2022) *Theory of infinite sequences and series*. Birkhäuser Cham
- Brzeziński A, Bizouard C, Petrov SD (2002) Influence of the atmosphere on Earth rotation: What new can be learned from the recent atmospheric angular momentum estimates? *Surv Geophys* 23:33–69
- Chen M, Li T (2018) Why 1986 El Niño and 2005 La Niña evolved different from a typical El Niño and La Niña. *Clim Dyn* 51:4309–4327
- Dill R, Dobsław H, Thomas M (2019) Improved 90-day Earth orientation predictions from angular momentum forecasts of atmosphere, ocean, and terrestrial hydrosphere. *J Geodesy* 93(3):287–295
- Dill R, Dobsław H, Thomas M (2023) ESMGFZ EAM products for EOP prediction: towards the quantification of time-variable EAM forecast errors. *Artif Satellites* 58:4
- Dobsław H, Dill R (2018) Predicting earth orientation changes from global forecasts of atmosphere-hydrosphere dynamics. *Adv Space Res* 61(4):1047–1054
- Dobsław H, Dill R (2019) Effective angular momentum functions from Earth system modelling at GeoForschungsZentrum in Potsdam. GFZ Product Description Document
- Dobsław H, Dill R, Grötzsch A, Brzeziński A, Thomas M (2010) Seasonal polar motion excitation from numerical models of atmosphere, ocean, and continental hydrosphere. *Journal of Geophysical Research: Solid Earth* 115(B10)
- Gou J, Kiani Shahvandi M, Hohensinn R, Soja B (2023) Ultra-short-term prediction of LOD using LSTM neural networks. *J Geodesy* 97:52
- Gross RS (2015) Earth rotation variations-long period. *Treatise Geophys* 3:239–294
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Hüllermeier E, Waegeman W (2021) Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn* 110:457–506
- Jaynes ET (2012) *Probability theory: the logic of science*. Cambridge University Press, Cambridge
- Jennrich RI (1969) Asymptotic properties of non-linear least squares estimators. *Ann Math Stat* 40(2):633–643
- Kalarus M, Schuh H, Kosek W, Akyilmaz O, Bizouard C, Gambis D, Gross R, Jovanovic B, Kumakshev S, Kutterer H, Mendes Cerveira PJ, Pasynok S, Zotov L (2010) Achievements of the earth orientation parameters prediction comparison campaign. *J Geodesy* 84:587–596
- Kiani Shahvandi M, Soja B (2022a) Inclusion of data uncertainty in machine learning and its application in geodetic data science, with case studies for the prediction of Earth orientation parameters and GNSS station coordinate time series. *Adv Space Res* 70(3):563–575
- Kiani Shahvandi M, Schartner M, Soja B (2022b) Neural ODE differential learning and its application in polar motion prediction. *Journal of Geophysical Research: Solid Earth* 127(11):e2022JB024775
- Kiani Shahvandi M, Gou J, Schartner M, Soja B (2022c) Data driven approaches for the prediction of Earth's effective angular momentum functions. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*: 6550–6553
- Kiani Shahvandi M, Soja B (2022d) Small geodetic datasets and deep networks: attention-based residual LSTM autoencoder stacking for geodetic time series. *Machine Learning, Optimization, and Data Science. LOD 2021. Lect Notes Comput Sci* 13163:296–307
- Kiani Shahvandi M, Dill R, Dobsław H, Kehm A, Bloßfeld M, Schartner M, Mishra S, Soja B (2023a) Geophysically informed machine learning for improving rapid estimation and shortterm prediction of Earth orientation parameters. *Journal of Geophysical Research: Solid Earth* 128(10):e2023JB026720
- Kiani Shahvandi M, Schartner M, Gou J, Soja B (2023b) Operational 14-day-ahead prediction of Earth's effective angular momentum functions with machine learning. *XXVIII General Assembly of the International Union of Geodesy and Geophysics (IUGG)*
- Kiani Shahvandi M, Belda S, Karbon M, Mishra S, Soja B (2024a) Deep ensemble geophysics-informed neural networks for the prediction of celestial pole offsets. *Geophys J Int* 236(1):480–493
- Kiani Shahvandi M, Belda S, Mishra S, Soja B (2024b) Short-term prediction of celestial pole offsets with interpretable machine learning. *Earth Planets Space* 76:1
- Kiani Shahvandi M, Schartner M, Gou J, Soja B (2024c) Operational forecasting of effective angular momentum functions fourteen days ahead. In: *International Association of Geodesy Symposia*. Springer, Berlin, Heidelberg

- Kiani Shahvandi M, Adhikari S, Dumberry M, Mishra S, Soja B (2024d) The increasingly dominant role of climate change on length of day variations. *Proc Natl Acad Sci* 121:30
- Kingma DP, Welling M (2014) Auto-encoding variational Bayes. *International Conference on Representation Learning*. <https://doi.org/10.48550/arXiv.1312.6114>
- Koot L, de Viron O, Dehant V (2006) Atmospheric angular momentum time-series: characterization of their internal noise and creation of a combined series. *J Geodesy* 79:663–674
- Kur T, Dobsław H, Śliwińska J, Nastula J, Wińska M, Partyka A (2022) Evaluation of selected short-term predictions of UT1-UTC and LOD collected in the second Earth orientation parameters prediction comparison campaign. *Earth Planets Space* 74:191
- Lakshminarayanan B, Pritzel A, Blundell C (2016) Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems (NeurIPS 2016)*
- Lehmann E, Névir P (2012) Uncertainties in relative atmospheric angular momentum computed from zonal winds in reanalysis data. *J Geophys Res Atmos* 117(D9)
- Li X, Wu Y, Yao D, Liu J, Nan K, Ma L, Cheng X, Yang X, Zhang S (2023) Research on UT1-UTC and LOD prediction algorithm based on denoised EAM dataset. *Remote Sensing* 15:19
- Loffe S, Szegedy C (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on International Conference on Machine Learning* 37: 448–456
- Lowenstein JH (2012) *Remote Sensing*. Cambridge University Press, Cambridge
- Lukman AF, Ayinde K, Aladeitan B, Bamidele R (2020) An unbiased estimator with prior information. *Arab J Basic Appl Sci* 27(1):45–55
- Neal RM (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC. <https://doi.org/10.1201/b10905>
- Petit G, Luzum B (2010) IERS Conventions (IERS Technical Note 36) Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie
- Prechelt L (2012) Early Stopping - But When?. *Lecture Notes in Computer Science*: 7700
- Quinn KJ, Ponte RM, Heimbach P, Fukumori I, Campin JM (2019) Ocean angular momentum from a recent global state estimate, with assessment of uncertainties. *Geophys J Int* 216(1):584–597
- Robert CP (2015) The Metropolis-Hastings algorithm. In: Robert CP (ed) *Statistics Reference Online*. Wiley StatsRef, Hoboken
- Śliwińska-Bronowicz J, Kur T, Wińska M, Dobsław H, Nastula J, Partyka A, Belda S, Bizouard C, Boggs D, Bruni S, Chen L, Chin M, Dhar S, Dill R, Ferrandiz JM, Gou J, Gross R, Guessoum S, Han S, Heinkelmann R, Irrgang C, Kiani Shahvandi M, Li J, Ligas M, Liu L, Lu W, Mayer V, Michalczak M, Modiri S, Otten M, Ratcliff T, Raut S, Saynisch-Wagner J, Schartner M, Schoenemann E, Schuh H, Soja B, Su X, Thaller D, Thomas M, Wang G, Wu Y, Xu X, Yang X, Zhao X, Zhou Z (2024) Assessment of length-of-day and universal time predictions based on the results of the second earth orientation parameters prediction comparison campaign. *J Geodesy* 98:22
- Śliwińska J, Dobsław H, Kur T, Nastula J, Wińska M, Partyka A, Belda S, Bizouard C, Boggs D, Bruni S, Chen L, Chin M, Dhar S, Dill R, Duan P, Ferrandiz JM, Gou J, Gross R, Guessoum S, Han S, Heinkelmann R, Huang C, Irrgang C, Kiani Shahvandi M, Kudrys J, Li J, Ligas M, Liu L, Lu W, Mayer V, Miao W, Michalczak M, Modiri S, Otten M, Ratcliff T, Raut S, Saynisch-Wagner J, Schartner M, Schoenemann E, Schuh H, Soja B, Su X, Thaller D, Thomas M, Wang G, Wu Y, Xu C, Xu X, Yang X, Zhao X, Zhou Z (2023) EOP predictions collected during the operational phase of the Second Earth Orientation Parameters Prediction Comparison Campaign. *GFZ Data Services*
- Soja B, Kiani Shahvandi M, Schartner M, Gou J, Kłopotek G, Crocetti L, Awadaljeed M (2022) The new geodetic prediction center at ETH Zurich. *EGU General Assembly 2022*
- Sullivan TJ (2015) *Introduction to uncertainty quantification*. Springer, Texts in Applied Mathematics 63
- Thomopoulos NT (2013) *Essentials of Monte Carlo simulation; statistical methods for building simulation models*. Springer Stochastic Modelling and Applied Probability
- Ulyanov D, Vedaldi A, Lempitsky V (2016) Instance normalization: the missing ingredient for fast stylization. *arXiv*
- Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J, (2005) The ERA-40 re-analysis. *Quarter J Royal Meteorol Soc* 131(612):2961–3012
- Wolter K, Timlin MS (1993) Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proceedings of the 17th Climate Diagnostics Workshop*: 7700
- Wu CF (1981) Asymptotic theory of nonlinear least squares estimation. *Ann Stat* 9(3):501–513
- Xu P, Yin L, Yue Z, Zhou T (2019) On predictability of time series. *Phys A* 523(1):345–351
- Yang L, Meng X, Karniadakis GE (2021) B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *J Comput Phys* 425:109913
- Yong BX, Brintrup A (2022) Do autoencoders need a bottleneck for anomaly detection?. *IEEE Access*: 10
- Yu N, Liu H, Chen G, Chen W, Ray J, Wen H, Chao N (2021) Analysis of relationships between ENSO events and atmospheric angular momentum variations. *Earth and Space Science* 8(12)
- Zheng D, Ding X, Zhou Y, Chen Y (2003) Earth rotation and ENSO events: combined excitation of interannual LOD variations by multiscale atmospheric oscillations. *Global Planet Change* 36:89–97
- Zhou Y (2022) Rethinking reconstruction autoencoder-based out-of-distribution detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.