Journal of Industrial Ecology (2026) 30:1285–1298
https://doi.org/10.1007/s44498-026-00010-6

RESEARCH

Forecasting future states of the environment with machine learning: a case study on water scarcity

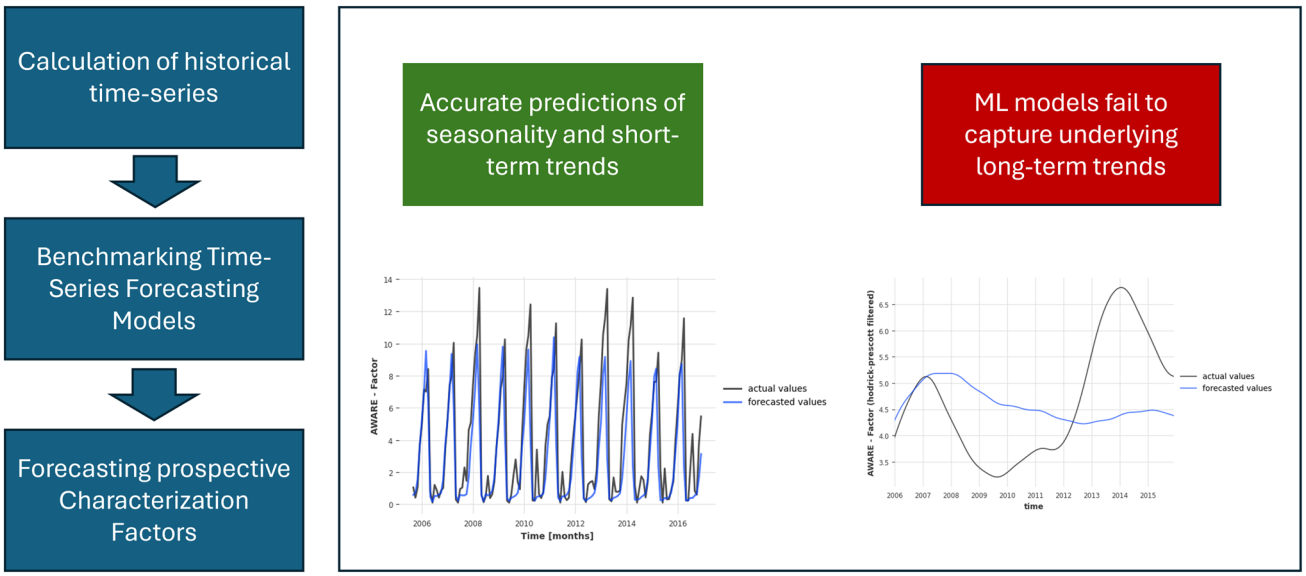
Niklas Engberg[1] · Carlos Felipe Blanco[2,3] · Valerio Barbarossa[2,4] · Conny Bakker[1] · Ruud Balkenende[1] · Justin Z. Lian[2] · Benjamin Sprecher[1]

Received: 3 December 2024 / Accepted: 20 October 2025 / Published online: 27 July 2026

Abstract

Recent research has emphasized the need to adapt life cycle inventories and life cycle impact assessments to account for changes in future scenarios. This is mostly achieved by using Integrated Assessment Models (IAMs), which combine environmental and economic data to develop prospective life cycle inventories (LCIs). However, running complex IAMs to simulate scenarios is computationally expensive, and the resulting data is not always aligned with the geographical scope of background inventories. This study explores the potential of machine learning (ML) to create prospective water-scarcity characterisation factors by forecasting the AWARE factor for ~ 9700 watersheds globally. Historical time series of water-scarcity characterisation factors are generated using the global freshwater model WaterGAP v2.2d and the AWARE method. Several ML models are trained on these historical datasets and benchmarked using symmetric mean absolute percentage error (sMAPE). The best-performing model, N-Beats, is then used to forecast AWARE values through 2032. The results demonstrate that ML can produce spatially and temporally explicit forecasts with reasonable accuracy (median sMAPE ~ 28%). However, the models primarily capture seasonal patterns rather than long-term structural trends and the results are sensitive to the quality and representativeness of the training data. This study highlights both the potential and the limitations of time series forecasting for developing prospective characterisation factors in life cycle assessment.

Graphical abstract



Extended author information available on the last page of the article

Keywords Industrial ecology · Life cycle impact assessment · Characterisation factors · Time series forecasting · Prospective LCA · AWARE

1 Introduction

The environmental benefits of adopting an emerging technology over an incumbent technology often become more salient when the future changes in the economy and environment are considered (Arvidsson et al., 2018, 2024; Thonemann et al., 2020; van der Giesen et al., 2020). Assessing the environmental impact of future technologies using prospective Life Cycle Assessment (LCA) is challenging, because numerous assumptions about uncertain socio-techno-economic pathways need to be made (Blanco et al., 2020; Cucurachi & Blanco et al., 2023; Cucurachi et al., 2018; Guinée et al., 2018; Sacchi et al., 2022).

Most prospective LCAs are solely transforming the inventory database to match those future scenarios. However, environmental impacts are determined when inventories are multiplied with characterisation factors (CFs) (Guinee, 2002). Therefore, determining prospective environmental impacts should ideally also include the use of prospective CFs, which has until now rarely been done.

When anticipating future environmental impacts and economic developments, practitioners generally rely on scenarios that are quantified in Integrated Assessment Models (IAMs), which offer a consistent platform to perform future-oriented analyses across multiple studies and system boundaries. However, in doing so, IAMs make many assumptions on technological development pathways. Furthermore, similar technologies and regions are grouped together, leading to aggregated and deterministic pathways. This does not always match well with LCAs, which aim for high detail and differentiation across technologies and locations (Sacchi et al., 2022).

This study explores the potential of machine learning to create future characterisation factors based on time series forecasting (Yuan et al., 2022). A recent study (Baustert et al., 2022) proposed a set of prospective characterisation factors (CFs) for future water scarcity using the output data from the IAM 'IMAGE' (Stehfest et al., 2014). We develop a set of prospective water scarcity characterisation factors based on ML, which allows us to explore the use of ML and compare it to IAM-based characterisation factors. It is our hypothesis that ML, as data-driven approach, can be less dependent on assumptions and will be able to generate spatially and temporally explicit characterisation factors based on historical time series data. This allows CFs to be developed for regions that are better aligned with the original scope of inventory databases like ecoinvent (Wernet et al., 2016).

Multiple studies have shown that water scarcity is strongly correlated to other climate variables, for example average surface temperature, vegetation growth and others (Jiao et al., 2021). These correlated variables—or covariates—can be obtained from well-established climate models. Training an ML model on historical data and historical covariates could allow prospective CFs to be calculated directly from climate models. With this study we aim to make a first step towards this goal by developing historical time series data of a water scarcity characterisation factor and training a deep learning ML model on it. By definition, mechanistic deviations from past trends cannot be predicted by forecasting models, because the necessary data is not present in historical datasets.

Machine learning (ML) is used in many scientific fields related to environmental impact assessment. It has for instance been used to predict climate- and water-related events and trends (Ahmed et al., 2022; Barbarossa et al., 2018; Khoi et al., 2022). Climate responses have been projected until the year 2100 (Kitsios et al., 2023). In prospective LCA, ML has been applied to address common challenges related to data availability, to derive environmental impacts from product properties and characteristics, and to optimize processes from an environmental perspective (Adedeji et al., 2020; Algren et al., 2020; Karka et al., 2022; Markovics & Mayer, 2022; Romeiko et al., 2020; von Borries et al., 2023; Zhu et al., 2020). An extensive review of ML in the context of prospective LCA was recently published by Blanco et al. (2024).

Time series forecasting aims at predicting future target values based on patterns and dependencies identified within sequences of historical data (Alsharef et al., 2022; Atabay et al., 2022; Kim et al., 2021; Montgomery et al., 2016). Traditionally, statistical methods such as autoregressive integrated moving average (ARIMA) models (Box, 1970) and state space models (Durbin & Koopman, 2012) are widely used for their interpretability and simplicity. Advanced machine learning techniques, such as Long Short-Term Memory networks (Hochreiter & Schmidhuber, 1997), convolutional neural networks (Borovykh et al., 2018), ensemble methods like Gradient Boosting (Chen & Guestrin, 2016), can also capture temporal patterns and dependencies with complex input data.

Deep learning methods, such as neural basis expansion analysis for interpretable time series forecasting (NBeats)

(Oreshkin et al., 2019) and Temporal Fusion Transformers (Lim et al., 2021), have outperformed classical methods on a variety of tasks (Salinas et al., 2017). In addition to improved accuracy, modern deep learning methods also perform well when trained on large numbers of time series sets, as opposed to classical methods like ARIMA, which typically require training one local model per individual time series. In this work, we refer to "local" models as individually trained on a single time series, and to "global" models as those that are fitted on multiple time series, as in multiple basins around the globe simultaneously.

Time series data typically have characteristics such as seasonality, trend and autocorrelation. Classical statistical models have traditionally focused on univariate forecasting, where the prediction of future values of a single variable is based on observations of past values of that variable alone. For climatic data, multivariate forecasting can increase the accuracy of ML-models by incorporating multiple input variables that are interdependent. Multivariate forecasting through deep learning techniques can often better capture the characteristics of the underlying data and are typically categorized into two types: many-to-one forecasting, where a single future value is predicted using several input variables, and many-to-many forecasting, where multiple future values are predicted based on multiple inputs. In time series forecasting, both the input and output consist of sequences of multiple values, which motivates the use of a many-to-many approach.

First, a brief discussion about computing historical time series of the AWARE factor is provided, based on the methodologies described by Boulay et al. (2018) and Baustert et al. (2022). These historical time series serve as training data for the machine learning models used in the present study. Second, preparing the data for input into ML-models is necessary to ensure robust model deployment. Model accuracy is assessed using metrics described in Sect. 2.2, which are used to select the best-performing algorithm on a sample set. Subsequently, the algorithm is deployed with the full dataset and the forecasts are analysed. The analysis entails a comparison to the IAM-based predictions to highlight the differences that result from calculating prospective CFs with different methodologies. Finally, we discuss the method, its limitations, and the general potential of time series forecasting to anticipate future changes in characterisation factors in comparison to IAMs (Fig. 1). To our knowledge, this is the first attempt to forecast CFs with ML-models and compare this approach to prospective characterisation factors based on data from IAMs.

2 Methods

2.1 Calculation of Historical Time Series: AWARE and Covariates

The water scarcity characterisation factor AWARE was selected as case study because hydrological data since the beginning of the twentieth century is available from the WaterGAP v2.2d model. The study by Baustert et al. (2022) offers prospective data for comparison with own results of forecasts. They chose the "Middle-of-the-road" scenario to calculate future CFs, following a narrative where "trends do not shift markedly from historical patterns" (van Vuuren et al., 2017). By comparing the forecasting results of this study with Baustert et al. (2022), differences between the methods to obtain prospective CFs can be identified.

The original AWARE characterisation factors are published and available online (www.wulca-waterlca.org). In the official dataset the AWARE-factor is available for the year 2010 on a monthly and annual timescale. However, training

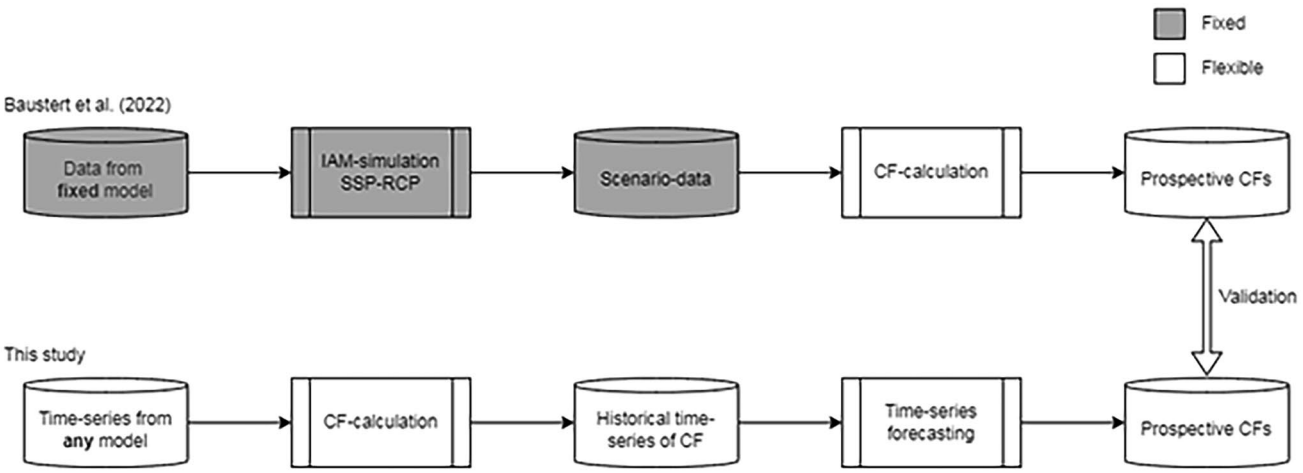


Fig. 1 Comparison of approaches to calculate prospective CFs from this study and IAM-based calculations as published by Baustert et al. (2022)

ML-models requires a significantly larger dataset. Therefore, time series data of the CF needed to be generated by applying the original AWARE methodology (Boulay et al., 2018) on historical data of the global freshwater model WaterGAP v2.2d (Müller Schmied et al., 2021). The method is well-described in several publications (Baustert et al., 2022; Boulay et al., 2018, 2021) and based on the "availability minus demand" (AMD) factor. AMD is calculated for each basin *i* at a monthly scale:

$$AMD_i = \frac{Availability_i - HWC_i - EWR_i}{Area_i} \quad (1)$$

with (see also Baustert et al., 2022):

Availability: actual water availability as runoff based on precipitation and evapotranspiration (in m^3/month);

HWC: human water consumption, which includes irrigation, domestic, industrial, and other sectorial demand (in m^3/month);

EWR: environmental water requirements (in m^3/month), which corresponds to the monthly natural water flow (virtual flow without any human influence) weighted depending on its value compared to the annual average of the natural water flows [weighting factor of 0.6 if the monthly value is below 40% of the average, 0.3 if it is above 80% of the average, and 0.45 for intermediary values, as defined by Pastor et al. (2014)];

Area: area of the geographical unit i (in m^2)

The final characterisation factor AWARE is the ratio between the consumption-weighted world average of AMD (AMD_{wa}) across all basins and the AMD of each individual basin (AMD_i). The AWARE-factor can be interpreted as the surface-time equivalent required to generate one m^3 of unused water in this region:

$$AWARE = \frac{AMD_{wa}}{AMD_i} \quad (2)$$

The WaterGAP datasets for the hydrological variables are available at a resolution of 30 arc minutes (~56 km at the equator). In combination with the basin-definitions, flow-direction and basin-filters as published in supplementary information of Boulay et al. (2018), data was aggregated to basin-level in order to compute AMD_i, AMD_{wa} and subsequently the AWARE for each year from 1960 to 2016.

In the original method, boundary conditions are set for the AWARE factor to define its minimum and maximum values. The maximum value of 100 is assigned when AMD_i is less than 1% of the yearly world average. Conversely, the minimum value of 0.01 is set when AMD_i exceeds the world average 100 times (Boulay et al., 2018). In this work the same boundary conditions were applied.

We used the output-data of WaterGAP v2.2d to calculate the monthly AWARE factors following the approach of Boulay et al. (2018). Discharge of a basin corresponded to the discharge of the most downstream cell within that basin, as all human demand has already been met before the waterflow leaves the basin in the outflow-cell. Data is cut off before 1960 because climate forcing data is modelled differently before that year (Müller Schmied et al., 2021). After filtering the basins according to Boulay et al. (2021) and removing basins in polar regions 9707 basins remained. The used covariates are: total water storage, total groundwater recharge, total runoff, and leaf area index. These variables stem from the ISIMIP3a simulation round (Gosling et al., 2024) and are aggregated to basin level, following the same approach as previously described.

2.2 Data Preparation

The AWARE-method was used to calculate monthly AMD, AMD world average, and AWARE values at the basin level for the period 1960–2016. As is common with ML methods, the data was prepared by applying a transformation pipeline consisting of log-transformation and scaling. Log-transformation was applied to reduce skewness of the data and the impact of outliers on the model. Scaling the values to the interval [0, 1] further aided the model to learn patterns from different basins and made them comparable. Normalization was achieved by applying

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

within each basin, where x_i is the original value, x_{min} and x_{max} are the minimum and maximum values of the dataset within the respective basin. x_i' is the resulting normalized value scaled to the range [0, 1]. The data was split into training-set and test-set with a ratio of 80% to 20%, the training dataset therefore ended in 2005. The transformation-pipeline was applied to both datasets, while only being fitted onto the training dataset (Hewamalage et al., 2023). That means if a value exceeds x_{max} and was in the test set it ended up as $x_{max\text{-}test} > 1$, because it was not considered during normalization.

Basins were clustered with k-means dynamic time warping (Petitjean et al., 2011) to inspect common historical subsets of time series behaviour. It was implemented with the *tslearn* toolkit (Tavenard et al., 2020) (Supplementary Materials SM2). K-means dynamic time warping is a clustering algorithm that groups time series data by aligning sequences, measuring similarity regardless of time shifts or distortions. As AWARE is a ratio and has cutoff values/boundaries that might hide important trends, AMD_i was selected for clustering the data. The results of the clustering can be found in Supplementary Materials SM1. No cluster showed a significant long-term

trend, meaning that the historical data did not significant changes in water scarcity over the course of 56 years present in the training dataset.

We compared the correlations between the published original AWARE-factors and our own calculations as well as the factors calculated by Baustert et al. (2022) (Fig. 2). IMAGE-based AWARE-values had a correlation-factor of only 0.53 to the original values in 2010. Many basins with originally low AWARE-factors had significantly higher values with IMAGE-data. Calculations of this study, which were based on WaterGAP v2.2d, had a correlation-factor of 0.81 to original AWARE-values. Perfect correlation was not reached because the original method (Boulay et al., 2018) used an older version of WaterGAP, which was not accessible. Most basins had higher AWARE-values with the new WaterGAP dataset, than originally calculated.

2.3 Model Prediction

The Darts library in python (Herzen et al., 2022) offered a framework to conduct machine learning with *sklearn* and *pytorch lightning* adjusted for time series forecasting. To save computational resources, 9707 basins were reduced to 1463 basins by selecting only those basins that consisted of at minimum three 0.5 × 0.5-degree grid cells. Small sized grid cells are typical for watersheds in coastal areas. Since water models are subject to higher uncertainties in those regions, we chose to focus on large basins in our study to assess the potential of machine learning and forecasting.

Multiple machine learning algorithms (Table 1) were benchmarked to identify the best-performing one for forecasting. XGBoost is effective and scalable but requires careful tuning and may struggle with sparse, high-dimensional data. LightGBM offers fast training and efficient feature handling but is prone to overfitting and has limited interpretability. NBeats provides interpretable time series predictions

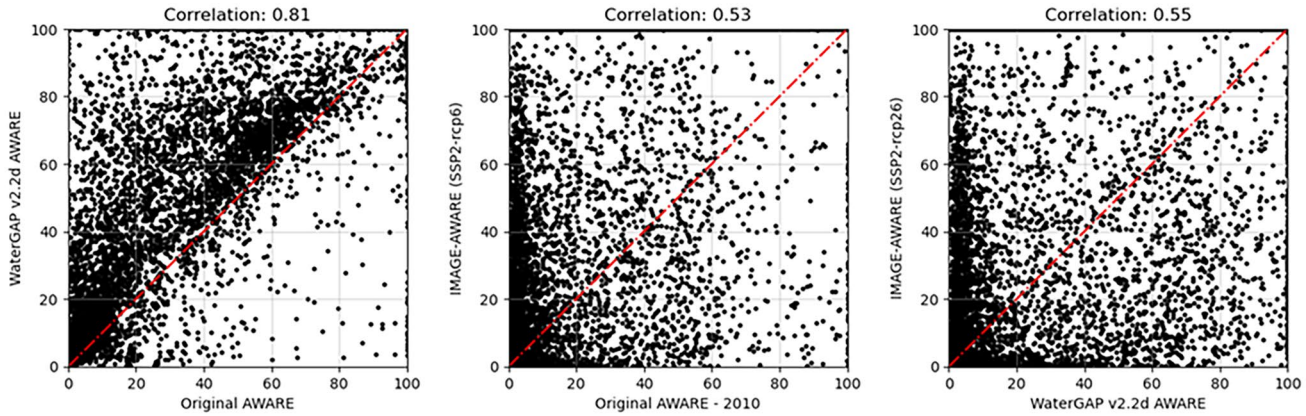


Fig. 2 Correlation of AWARE-factors in year 2010 per basin, left: original AWARE-values as published by WULCA against calculations of this study based on WaterGAP v2.2d, center: original AWARE-values against calculations of Baustert et al. (2022) based on IMAGE, right: correlation between this study's AWARE-values and IMAGE-AWARE. Underlying data for this figure is available in SM2

Table 1 Characteristics of the algorithms that are benchmarked in this study

Algorithm	Advantages	Disadvantages	Literature
XGBoost	• Highly effective and versatile • Efficient, scalable, suitable for large datasets • Built-in regularization, reduced risk of overfitting	• Sensitive to hyperparameter tuning, requires careful optimization • Might not perform well in high-dimensional, sparse feature spaces	Chen and Guestrin (2016)
LightGBM	• Comparatively fast training times and quick development of models • High accuracy through implementation of feature selection techniques • Native categorical feature selection without one hot encoding	• Prone to overfitting with small dataset or complex models • Limited interpretability • Not suited for small datasets	Ke et al. (2017)
NBeats	• Interpretable predictions by decomposing the time series into a series of basic functions • Modular and customizable	• Prone to overfitting • Potentially time-consuming hyper parameter tuning	Oreshkin et al. (2019)

and modularity but is susceptible to overfitting and may require extensive hyperparameter tuning. The dataset in this study included over a thousand time series of 684 monthly time steps each, which makes NBeats particularly suitable because it can learn shared patterns across multiple basins. XGBoost and LightGBM need more extensive feature engineering and can be less efficient when capturing temporal dependencies.

Benchmarking with accuracy, measured as median sMAPE across all basins, and the time taken to complete the forecasting task (Fig. 3) was based on the data from the sub selection of largest basins. Although the naïve models were the fastest to execute, their accuracy was significantly lower compared to more advanced models such as gradient boosting (XGB, LightGBM) and deep learning models (NBeats).

The NBeats deep learning model was selected, as it had the highest prediction accuracy and allows training a global model, considering data from all basins simultaneously. The deployed N-Beats model architecture consisted of 24 stacks, each composed of 2 blocks with 2 fully connected layers per block. Each layer contained 136 units, and the output coefficient dimension was set to 12. The model was trained with a learning rate of 0.001 and a batch size of 1024. These parameters were selected after sampling different configurations (Supplementary Materials SM2), changing input chunk length, output chunk length, number of stacks, blocks and layers, batch size and the number of maximum samples per time series used for model training.

To address overfitting and optimize model performance during training, we employed two common techniques:

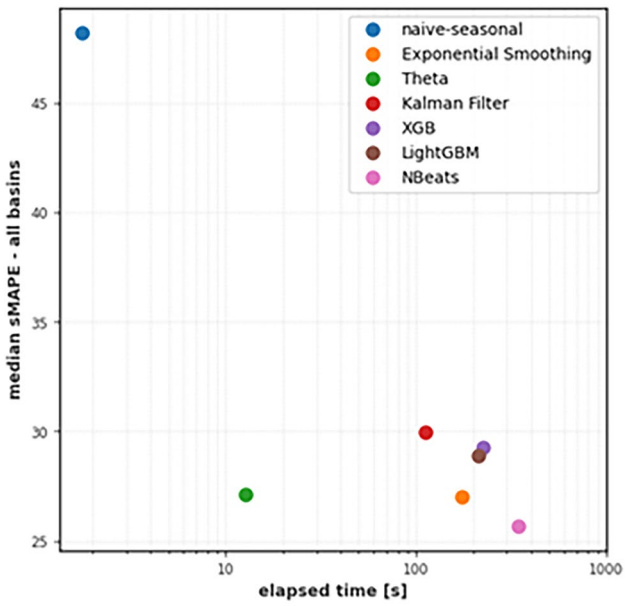


Fig. 3 Benchmarking seven algorithms along median sMAPE and elapsed training time. Underlying data for this figure is available in SM2

early stopping and learning rate scheduling. Early stopping was implemented using *PyTorch Lightning*'s EarlyStopping callback, which monitors validation loss and halts training when no improvement is observed for 5 consecutive epochs, thereby preventing the model from overfitting to the training data. The learning rate was initially set to 1e-3 and dynamically adjusted using *PyTorch Lightning*'s ReduceLROnPlateau scheduler, which reduces the learning rate by 50% when validation loss plateaus for 2 epochs. Both techniques were integrated into the NBeats model through the Darts *PyTorch Lightning* interface, where keyword arguments such as the EarlyStopping callback are passed directly to the Trainer class constructor.

The lookback window, or input chunk length, is the number of past time steps considered to make a prediction for the next time step. Considering the seasonality over 12 months of hydrological data a lookback window that is a multitude of 12 was chosen. In this case, 72 months or six years was picked as it captures long-term trends better than shorter windows. Even longer lookback windows increase the risk of overfitting and reduce generalizability. The model then becomes more complex and might include outdated information that is not relevant to most recent trends.

Because the historical time series data entails non-normality and non-stationarity (Supplementary Materials SM1) the symmetric mean absolute percentage error (sMAPE) is selected as performance metric (Hewamalage et al., 2023). sMAPE considers both the magnitude and direction of the error, over-forecasting and under-forecasting are treated equally. The value of sMAPE ranges between 0 and 200%, a lower value indicates a more accurate performance of the model with

$$\mathrm{SMAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|A_t - F_t|}{\frac{|A_t| + |F_t|}{2}} \tag{4}$$

where A_t is the actual value at timestep t and F_t represents the forecasted value at timestep t. In the nominator the difference between actual value and forecasted value is calculated. The denominator represents the average of the actual and forecasted value.

After cross-validating (Supplementary Materials SM1), the ML models were used to forecast the AWARE-factor until the year 2032, which was 16 years after the end of the dataset. The forecasting horizon of 16 years is a multiple of the output chunk length of 12 months, which corresponded to the timespan the ML model predicted in each iteration. Baustert et al. (2022) have calculated AWARE values based on IMAGE for 2030, which made direct comparison of the predictions in that year possible. Cross-validation was performed using two approaches. First, we

applied an expanding window strategy, commonly referred to as a time series split, which is a variation of k-fold cross-validation adapted for temporal data. In this method, the training window gradually expands over time: the first dataset covered 1960–1980, the second 1960–1990, and so on, until the full dataset was used. Each split took earlier data for training and later data for testing. Second, we used a rolling window approach, where each dataset covers a fixed 20-year period. The first window spans 1960–1980, the second 1970–1990, and so forth, moving forward in time while maintaining a constant window size. In total, ten cross-validation sets were constructed using these two methods (Supplementary Materials SM2). The rolling window validation is a complementary test, helping to understand whether the model degrades or potentially improves when it has access to less historical data over different periods.

3 Results

Applying the NBeats algorithm to the historical AWARE time series of the 1,463 largest basins resulted in a median sMAPE of 25% (Fig. 4). Forecast accuracy varied widely, with some basins achieving high precision (sMAPE ~ 3%), while others showed significantly poorer performance (sMAPE ≥ 50%). The NBeats algorithm outperformed other models tested in this study (Fig. 3). More than 60% of the basins showed an sMAPE below 30%, with a maximum value of approximately 78%. In comparison, other models, such as XGBoost and LightGBM, exhibited poorer performance, with maximum sMAPE values exceeding 100%.

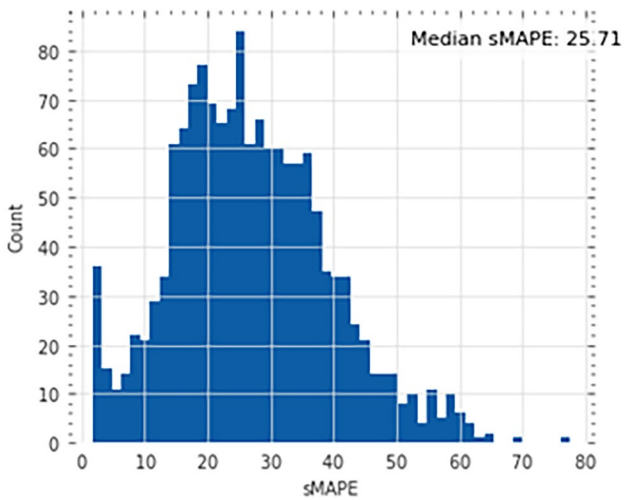


Fig. 4 Distribution of basins per sMAPE-value for the sub selection of basins with NBeats model. Underlying data for this figure is available in SM2

Furthermore, these models had a higher proportion of basins with sMAPE values above 50% compared to NBeats.

A visual inspection of 60 basin forecasts against the validation set indicated that the models generally predicted seasonality accurately. However, while seasonality was well captured, the 30-month moving averages often differed significantly from the validation set. Furthermore, forecasts for some basins with sMAPE values below 20% did not align closely with the validation set. Figure 5 shows both the monthly model predictions against the validation set and long-term 30-month moving averages of historical data as well as forecasts of three representative basins. For the monthly forecasts historical values were cut off before the year 2006.

Arguably, LCA practitioners are most interested in long-term trends and forecasts of regions in which water scarcity might change significantly over the course of the next decades. Forecasts of yearly values were therefore attempted, but did not result in satisfactory accuracy. We applied the moving-average over the monthly values to depict those long-term developments and trends. As result of analysing more long-term trends we observed less positive outcomes of the forecasts.

While the forecasts matched the validation set in 2006 and accurately captured initial trends during the validation period, the predictions for the following years did not match the validation values. In 2016, at the end of the validation period, actual values and forecasts were comparable again, in the basins located in Bolivia and Turkey. However, in the Wolga-basin, a significant difference of approximately 20 $m^3_{world\text{-}eq.}/m^3$ was prevalent. This type of errors was observed across many basins that were visually inspected, and were further discussed in the following sections.

In the final deployment of the NBeats algorithm, the machine learning (ML) model was trained on the whole period of the dataset, from 1960 to 2016 with a monthly temporal resolution. According to the forecasts of the NBeats model, in 2030 more than 70% of the basins are expected to have lower AWARE values than in the year 2010. Approximately 400 basins with factors over 60 $m^3_{world\text{-}eq.}/m^3$ in the baseline year were forecasted to further increase up to almost 100 $m^3_{world\text{-}eq.}/m^3$ (Fig. 6).

Figure 7 provides a global overview of the local differences within the basins between the NBeats forecasts and the IMAGE-based calculations. Positive values indicate that the NBeats forecasts were higher than the IMAGE-based forecasts, while negative values indicate the opposite. In South America, Europe and Central Africa, many basins had similar AWARE-factors for the year 2030, whereas basins near the equator, Eastern America and West Australia showed more saturated colours, indicating bigger discrepancies between the two approaches. More saturated colours seen in many basins, along with

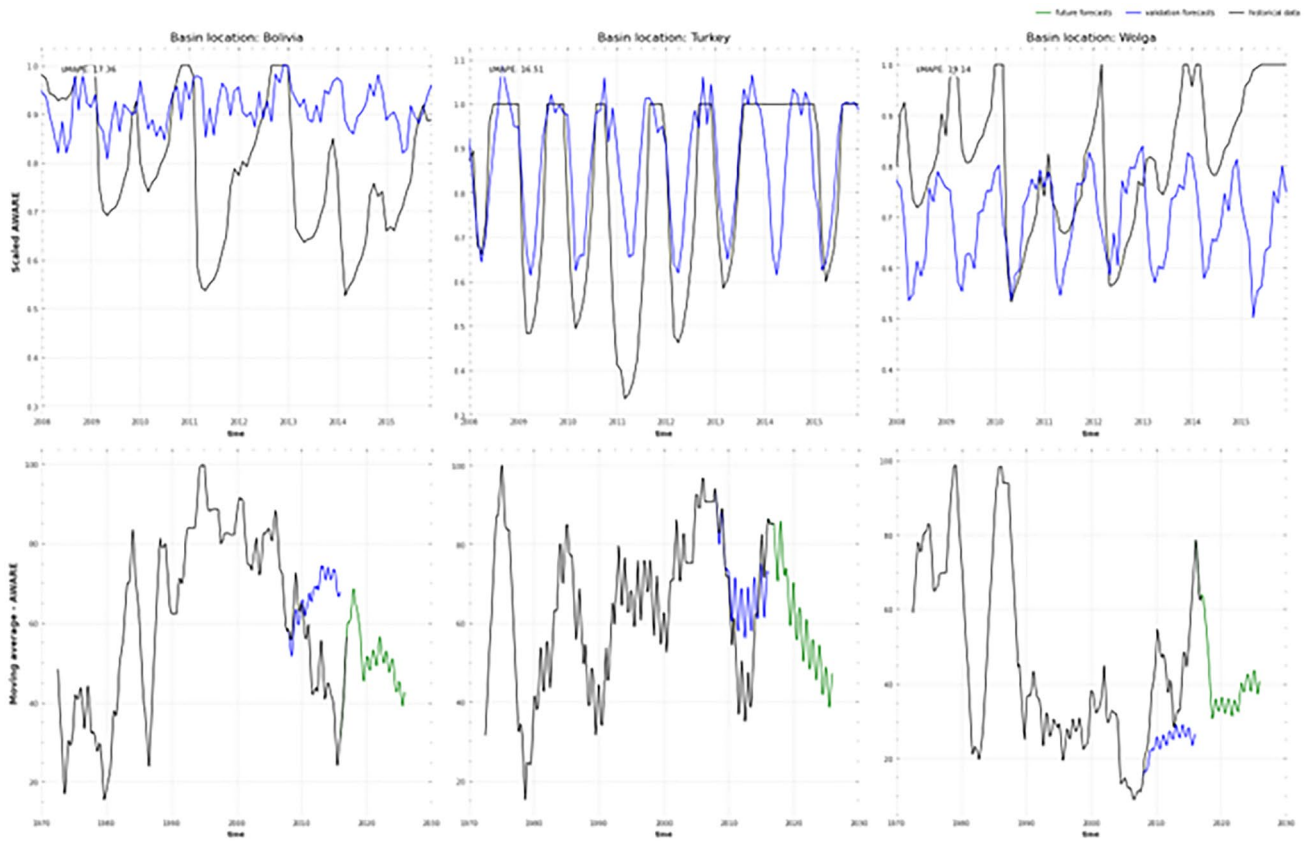


Fig. 5 Forecast evaluation of three basins located in Bolivia, Turkey and the Wolga; first row: Scaled AWARE values of validation forecasts against historical actual data in the timespan 2006–2016, second row: 30-month moving averages of historical actual AWARE-values 1960–2016, validation forecasts in timespan 2006–2016 and future forecasts in the timespan 2016–2026. Underlying data for this figure is available in SM2

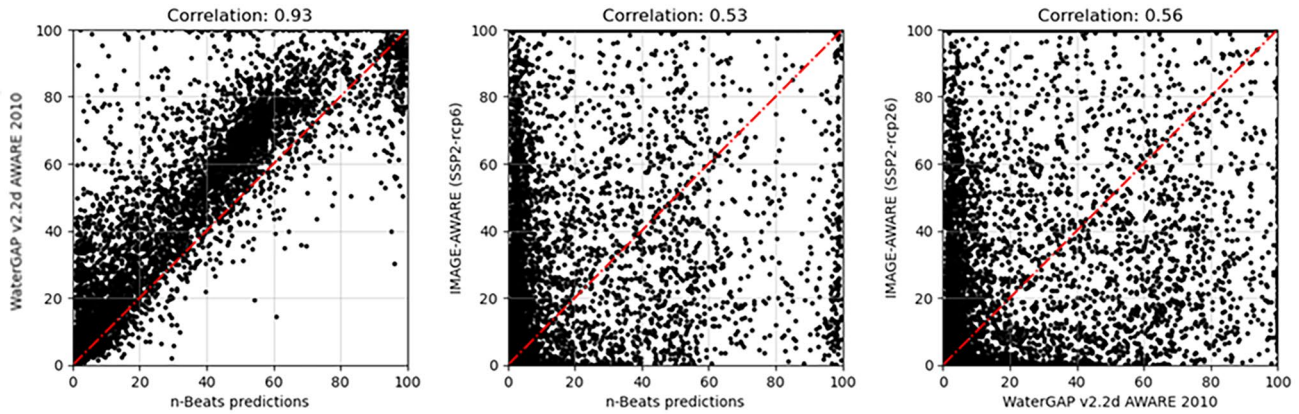


Fig. 6 Correlation of AWARE-factors in year 2030 per basin, left: NBeats-forecasts against calculations of this study based on WaterGAP v2.2d, center: NBeats-forecasts against calculations of Baustert et al. (2022) based on IMAGE, right: correlation between this study's AWARE-values in 2010 and IMAGE-AWARE. Underlying data for this figure is available in SM2

the correlation plots suggested that there were significant differences between the results from traditional IAM-based assessments and those based on machine learning (Supplementary Materials SM2).

Figure 7b shows the difference between the forecasts and the original AWARE-factor with the same colour code. As already pointed out, minor differences between forecasts and original values were noted in many basins across the globe.

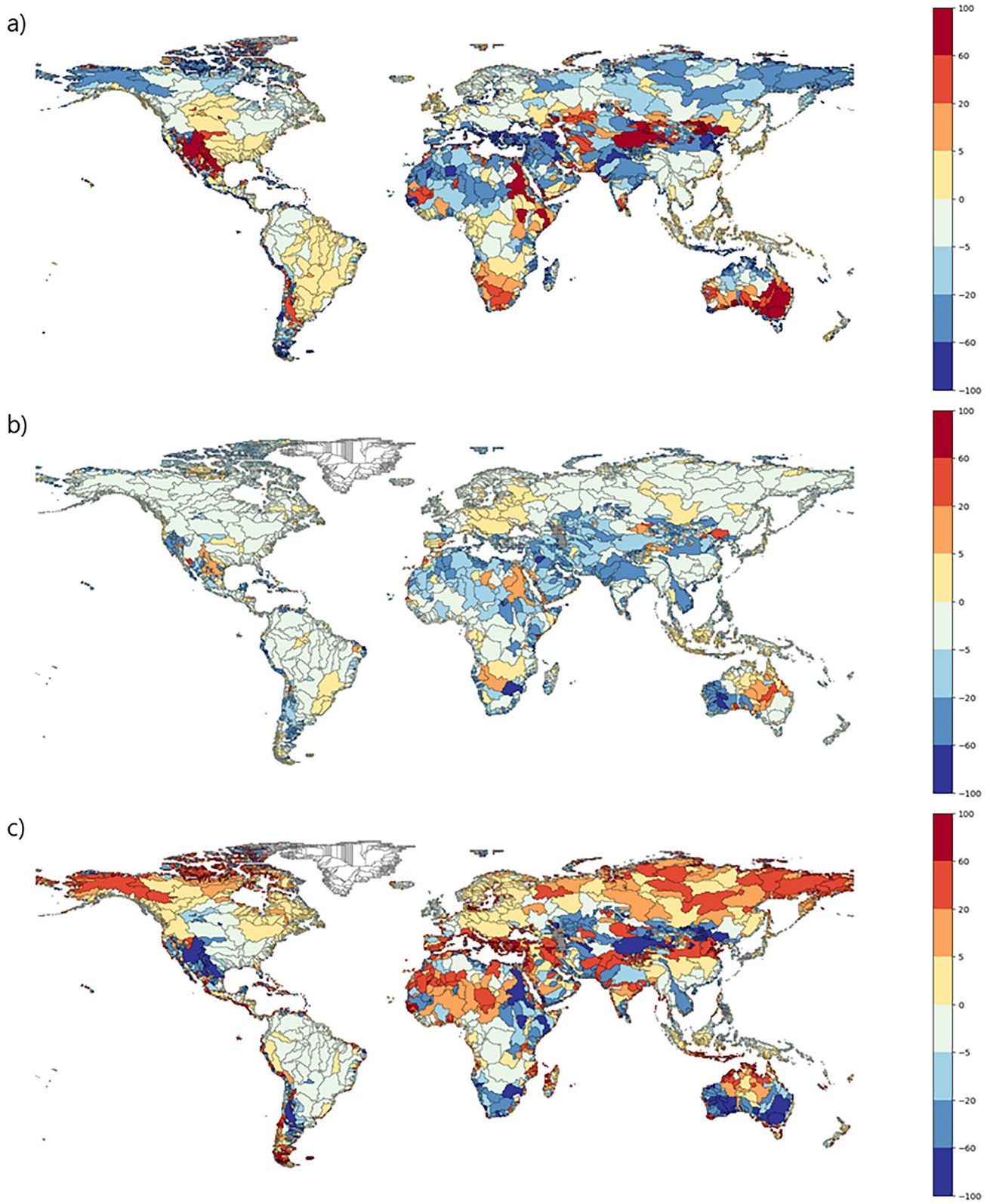


Fig. 7 Difference of AWARE-factors [m3 world eq./m3] with; **a** 2030 NBeats forecasts and 2030 SSP2 RCP2.6 IMAGE predictions; **b** 2030 NBeats forecasts and 2010 AWARE values of this study; **c** 2030 SSP2 RCP 2.6 IMAGE predictions and 2010 AWARE values of this study. IMAGE predictions are based on Baustert et al. (2022) and a similar colour scheme was applied. Underlying data for this figure is available in SM2

Approximately 100 basins showed a significant increase or decrease of over 20 m^3 world eq./m^3. For comparison, Fig. 7c shows the difference between IMAGE predictions and 2010 AWARE-factors, calculated with WaterGAP v2.2d. The differences in this case were more extreme, which can be attributed to discrepancies that were already present in the baseline year, also reported by Baustert et al. (2022).

4 Discussion and Conclusion

The aim of this study was to explore the extent to which machine learning can be used to forecast spatially and temporally explicit characterisation factors. We developed historical time series data of a water scarcity characterisation factor, based on the hydrological data from the WaterGAP-model and the characterisation factor AWARE. That data was used to train a deep learning ML-model that provided a set of prospective water scarcity CFs. These novel ML-based CFs, which represent individual basins improve on previous prospective CFs with a higher level of aggregation of underlying models and data. The ML-model also allows practitioners to use covariates that can be obtained from climate models (e.g., average surface temperature), opening the door to more accurate prospective CFs in the future, correlated with climate modelling.

Comparing our ML-based CFs with the original AWARE CFs (Boulay et al., 2018) for the baseline-year 2010 (the only year for which the original AWARE CF is available), we find that our model has a correlation factor of 0.8. Compared to the validation AWARE-time series from 2005 to 2016, we find that seasonality and short-term trends were accurately picked up by the ML-based models, but that long-term forecasts were less precise. In many basins the AWARE-factor seasonally fluctuates between maximum and minimum value over the course of a year. Big seasonal differences between maximum and minimum scarcity complicated forecasting. Additionally, outliers and uncertainties introduce noise, which can be mitigated in the future by aggregating on temporal (monthly to yearly) and geographical level (basin to country). The recently published AWARE2.0 (Seitfudem et al., 2025), with updated HWC and EWR calculations, might be more feasible for time-series forecasting and should be used as training data for follow-up research of the present work. Time-series specific data augmentation techniques (e.g., adding small amounts of noise, bootstrapping blocks of time series) can further mitigate the effect of outliers. Additional research should evaluate the sensitivity of forecasting accuracy to alternative pre-processing approaches, including Box-Cox transformations and different scaling methods beyond the min–max normalization used here.

Because of the different underlying water models, our results are not directly comparable to the IAM-based water scarcity CFs by Baustert et al. (2022). In contrast to the scenario-based projections of IAMs, our approach explores empirical forecasts that reflect the continuation of past dynamics, without embedding socio-economic pathways or policy assumptions. Nevertheless, similarities of ML-based and IAM-based projections were found on basin-level, with slightly decreasing or increasing water scarcity in many basins over an extended period.

Using synthetic data from projects such as the Inter-Sectoral Impact Model Intercomparison Project (Hempel et al., 2013) or Coupled Model Intercomparison Project (Taylor et al., 2012) is a promising avenue to improve model performance through inclusion of different covariate datasets that can enhance forecasting underlying variables to the AWARE factor. By exploring more data input configurations by following a similar approach as we did, more accurate predictions can likely be achieved. The methods presented in this paper could also be applied to other CF-datasets, for example land-use change, where local historical timeseries data are available or can be computed and potentially where IAM-based data is unavailable or too aggregated.

ML-based models can suffer from overfitting, that is, they may fit too closely to historical data from specific basins. To reduce this risk, a learning-rate monitor and early stopping were implemented. Additionally, the NBeats algorithm was trained simultaneously on multiple basins from different regions to improve generalizability and prevent the model from overfitting to patterns specific to a single basin. Further research is needed to more accurately determine when overfitting begins and to identify the optimal number of training epochs to achieve a balance between performance and generalization. We chose a relatively short lookback-window, to avoid the likelihood of overfitting and save computational resources. In future research, this should be addressed and accuracy improved.

Complex machine learning models lack inherent transparency and interpretability, acting as “black boxes”. To address this, so-called explainable ML techniques have been developed to improve trust in predictions and facilitate further model improvements (Yang et al., 2024). The NBeats architecture employed in this study, for example, is designed to produce interpretable outputs, contributing to the overall explainability of the forecasting process. We recommend future research to use these types of techniques to analyse trend and seasonality components of the model in more detail and track the response of model performance to input variables and parameters, during iterative model improvement. Through computationally expensive grid-search operations, optimal hyperparameters can be found. We did not have the computational resources to conduct such operations in this study, but finding the best hyperparameters

will contribute to more accurate and robust results and is achievable with the used python libraries.

Long-term structural changes in water scarcity are often driven by factors not directly observable in historical AWARE time series. Policy interventions (e.g., water management regulations), technological adoption (e.g., improved irrigation efficiency), demographic transitions, or unprecedented climate events are examples of such factors. Without explicit covariates that capture these drivers, models like N-Beats cannot anticipate their effects on future water scarcity patterns. For instance, a sudden policy change promoting water conservation or the implementation of large-scale desalination projects would create structural breaks in water scarcity trends that are invisible to models trained only on historical AWARE values.

One of the main benefits of training deep neural network models is that they allow for transfer learning, meaning a pretrained model can be used to make predictions from previously unseen sets of historical data from the same domain as training data (Ehrig et al., 2024; Ye & Dai, 2021). Experts could prepare models for specific characterisation factors (CFs) or other relevant time series. These models can then be used by LCA-practitioners to generate prospective CFs for specific areas that they have local data for. Requirements of data preparation, however, need to be well documented to discuss errors and biases. Thorough documentation of these methodological assumptions that are not directly visible to practitioners, including choice of input variables and data transformation operations, will improve critical understanding and consistent model deployment and development.

Our research also has implications for the development of characterisation factors in general, outside of ML based modelling. When creating the time series for training purposes, we found that consumption-weighted world average of AMD (AMD_{wa}) differed by about 30% between the baseline year 2010 and 2011. Since AMD_{wa} is the numerator in the Eq. (1), a 30% change in the numerator will also result in a 30% change in the AWARE-factor for individual basins, assuming the basin's AMD_i (denominator) remains constant, which was the case for most basins. These findings show that it is relevant to consider historical trends and longer time horizons.

In conclusion, we view ML-based time series forecasting as a potentially important tool for prospective LCA, because it allows researchers to utilize historical data to generate extrapolations into the future. In the context of characterisation factors, the approach can facilitate prospective LCA studies as localised as the underlying time series data allows and help to sidestep IAMs. Since this study was a first exploration of how to conduct an ML-based approach, we do not have conclusions on whether ML-based CFs are indeed a good alternative to IAM-based approaches. At this moment, they are not a standalone alternative to IAMs. While models like N-Beats excel at capturing seasonal patterns in water scarcity, they cannot reliably predict the long-term structural trends that matter most for prospective LCA. Although many ML-based approaches hold the promise of being easy to use, currently this is not the case. Specifically, the underlying data and methods to calculate characterisation factors need to support time-explicit assessments, and more work is needed to understand exactly which combinations of techniques and parameters yield the best results.

Supplementary Materials SM1: This supporting information provides additional figures visualizing characteristics of the historical data of AWARE and the different ML-models trained on that historical data.

Supplementary Materials SM2: This supporting information provides the code that was developed for this paper and files with the historical and forecasted AWARE-factor, as well as underlying code for the figures presented in the paper and supporting information S1. It is available at https://doi.org/10.5281/zenodo.14262678.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44498-026-00010-6.

Funding This research was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101101978.

Data availability The data that supports the findings of this study are available in the supporting information of this article. Sources for the data are described in the text.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Adedeji, P. A., Akinlabi, S. A., Madushele, N., & Olatunji, O. O. (2020). Potential roles of artificial intelligence in the LCI of renewable energy systems. In *Advances in manufacturing*

engineering (pp. 275–285). Springer. https://doi.org/10.1007/978-981-15-5753-8

Ahmed, A. N., Yafouz, A., Birima, A. H., Kisi, O., Huang, Y. F., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2022). Water level prediction using various machine learning algorithms: A case study of Durian Tunggal river, Malaysia. *Engineering Applications of Computational Fluid Mechanics, 16*(1), 422–440. https://doi.org/10.1080/19942060.2021.2019128

Algren, M., Fisher, W., & Landis, A. (2020). Machine learning in life cycle assessment. In J. Dunn & P. Balaprakash (Eds.), *Data science applied to sustainability analysis* (pp. 167–190). Elsevier. https://doi.org/10.1016/B978-0-12-817976-5.00009-7

Alsharef, A., Aggarwal, K., Sonia, K., & M., & Mishra, A. (2022). Review of ML and AutoML solutions to forecast time-series data. *Archives of Computational Methods in Engineering, 29*(7), 5297–5311. https://doi.org/10.1007/s11831-022-09765-0

Arvidsson, R., Svanström, M., Sandén, B. A., Thonemann, N., Steubing, B., & Cucurachi, S. (2024). Terminology for future-oriented life cycle assessment: Review and recommendations. *The International Journal of Life Cycle Assessment, 29*(4), 607–613. https://doi.org/10.1007/s11367-023-02265-8

Arvidsson, R., Tillman, A., Sandén, B. A., Janssen, M., Nordelöf, A., Kushnir, D., & Molander, S. (2018). Environmental assessment of emerging technologies: Recommendations for prospective LCA. *Journal of Industrial Ecology, 22*(6), 1286–1294. https://doi.org/10.1111/jiec.12690

Atabay, F. V., Pagkalinawan, R. M., Pajarillo, S. D., Villanueva, A. R., & Taylar, J. V. (2022). Multivariate time series forecasting using ARIMAX, SARIMAX, and RNN-based deep learning models on electricity consumption. In *2022 3rd international informatics and software engineering conference* (*IISEC*) (pp. 1–6). https://doi.org/10.1109/IISEC56263.2022.9998301

Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H., & Schipper, A. M. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Scientific Data, 5*(1), Article 180052. https://doi.org/10.1038/sdata.2018.52

Baustert, P., Igos, E., Schaubroeck, T., Chion, L., Mendoza Beltran, A., Stehfest, E., van Vuuren, D., Biemans, H., & Benetto, E. (2022). Integration of future water scarcity and electricity supply into prospective LCA: Application to the assessment of water desalination for the steel industry. *Journal of Industrial Ecology, 26*(4), 1182–1194. https://doi.org/10.1111/jiec.13272

Blanco, C. F., Cucurachi, S., Guinée, J. B., Vijver, M. G., Peijnenburg, W. J. G. M., Trattnig, R., & Heijungs, R. (2020). Assessing the sustainability of emerging technologies: A probabilistic LCA method applied to advanced photovoltaics. *Journal of Cleaner Production, 259*, Article 120968. https://doi.org/10.1016/j.jclepro.2020.120968

Blanco, C. F., Pauliks, N., Donati, F., Engberg, N., & Weber, J. (2024). Machine learning to support prospective life cycle assessment of emerging chemical technologies. *Current Opinion in Green and Sustainable Chemistry, 50*, Article 100979. https://doi.org/10.1016/j.cogsc.2024.100979

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2018). Conditional time series forecasting with convolutional neural networks. https://arXiv.org/1703.04691

Boulay, A.-M., Bare, J., Benini, L., Berger, M., Lathuillière, M. J., Manzardo, A., Margni, M., Motoshita, M., Núñez, M., Pastor, A. V., Ridoutt, B., Oki, T., Worbe, S., & Pfister, S. (2018). The WULCA consensus characterization model for water scarcity footprints: Assessing impacts of water consumption based on available water remaining (AWARE). *The International Journal of Life Cycle Assessment, 23*(2), 368–378. https://doi.org/10.1007/s11367-017-1333-8

Boulay, A.-M., Lesage, P., Amor, B., & Pfister, S. (2021). Quantifying uncertainty for AWARE characterization factors. *Journal of Industrial Ecology, 25*(6), 1588–1601.

Box, G. E. (1970). *GM Jenkins time series analysis: Forecasting and control*. Wiley

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, & R. Rastogi (Eds.), *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cucurachi, S., & Blanco, C. F. (2023). Practical solutions for ex-ante LCA illustrated by emerging PV technologies. In *Ensuring the environmental sustainability of emerging technologies* (pp. 149–167). Lausanne, EPFL. https://doi.org/10.5075/epfl-irgc-298445

Cucurachi, S., van der Giesen, C., & Guinée, J. (2018). Ex-ante LCA of emerging technologies. *Procedia CIRP, 69*, 463–468. https://doi.org/10.1016/j.procir.2017.11.005

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. OUP Oxford.

Ehrig, C., Sonnleitner, B., Neumann, U., Cleophas, C., & Forestier, G. (2024). The impact of data set similarity and diversity on transfer learning success in time series forecasting. https://doi.org/10.48550/arXiv.2404.06198

Gosling, S. N., Müller Schmied, H., Bradley, A., Burek, P., Chang, J., Ciais, P., Grillakis, M., Guillaumot, L., Hanasaki, N., Hartley, A., Huang, H., Kou-Giesbrecht, S., Koutroulis, A., Otta, K., Satoh, Y., Stacke, T., Zhu, Q., & Schewe, J. (2024). *ISIMIP3a Simulation Data from the Global Water Sector (v1.3). ISIMIP Repository.* https://doi.org/10.48364/ISIMIP.398165.3

Guinee, J. B. (2002). Handbook on life cycle assessment operational guide to the ISO standards. *The International Journal of Life Cycle Assessment, 7*(5), 311–313. https://doi.org/10.1007/BF02978897

Guinée, J. B., Cucurachi, S., Henriksson, P. J. G., & Heijungs, R. (2018). Digesting the alphabet soup of LCA. *The International Journal of Life Cycle Assessment, 23*(7), 1507–1511. https://doi.org/10.1007/s11367-018-1478-0

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction—The ISI-MIP approach. *Earth System Dynamics, 4*(2), 219–236. https://doi.org/10.5194/esd-4-219-2013

Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., & Grosch, G. (2022). Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*. https://doi.org/10.5555/3586589.3586713

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Mining and Knowledge Discovery, 37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jiao, W., Wang, L., Smith, W. K., Chang, Q., Wang, H., & D'Odorico, P. (2021). Observed increasing water constraint on vegetation growth over the last three decades. *Nature Communications, 12*(1), Article 3777. https://doi.org/10.1038/s41467-021-24016-9

Karka, P., Papadokonstantakis, S., & Kokossis, A. (2022). Digitizing sustainable process development: From ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design. *Chemical Engineering Science, 250*, Article 117339. https://doi.org/10.1016/j.ces.2021.117339

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach,

R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water, 14*(10), Article 1552. https://doi.org/10.3390/w14101552

Kim, Y. W., Kim, T., Shin, J., Go, B., Lee, M., Lee, J., Koo, J., Cho, K. H., & Cha, Y. (2021). Forecasting abrupt depletion of dissolved oxygen in urban streams using discontinuously measured hourly time–series data. *Water Resources Research, 57*(4), Article e2020WR029188. https://doi.org/10.1029/2020WR029188

Kitsios, V., O'Kane, T. J., & Newth, D. (2023). A machine learning approach to rapidly project climate responses under a multitude of net-zero emission pathways. *Communications Earth & Environment, 4*(1), 1–15. https://doi.org/10.1038/s43247-023-01011-0

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Markovics, D., & Mayer, M. J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renewable and Sustainable Energy Reviews, 161*, Article 112364. https://doi.org/10.1016/j.rser.2022.112364

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2016). *Introduction to time series analysis and forecasting* (2nd ed.). Wiley.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., & Döll, P. (2021). The global water resources and use model WaterGAP v2.2d: Model description and evaluation. *Geoscientific Model Development, 14*(2), 1037–1079. https://doi.org/10.5194/gmd-14-1037-2021

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019, May 24). *N-BEATS: Neural basis expansion analysis for interpretable time series forecasting*. https://arxiv.org/pdf/1905.10437

Pastor, A. V., Ludwig, F., Biemans, H., Hoff, H., & Kabat, P. (2014). Accounting for environmental flow requirements in global water assessments. *Hydrology and Earth System Sciences, 18*(12), 5041–5059. https://doi.org/10.5194/hess-18-5041-2014

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition, 44*(3), 678–693. https://doi.org/10.1016/j.patcog.2010.09.013

Romeiko, X. X., Guo, Z., Pang, Y., Lee, E. K., & Zhang, X. (2020). Comparing machine learning approaches for predicting spatially explicit life cycle global warming and eutrophication impacts from corn production. *Sustainability, 12*(4), Article 1481. https://doi.org/10.3390/su12041481

Sacchi, R., Terlouw, T., Siala, K., Dirnaichner, A., Bauer, C., Cox, B., Mutel, C., Daioglou, V., & Luderer, G. (2022). PRospective environMental impact asSEment (premise): A streamlined approach to producing databases for prospective life cycle assessment using integrated assessment models. *Renewable and Sustainable Energy Reviews, 160*, Article 112311. https://doi.org/10.1016/j.rser.2022.112311

Salinas, D., Flunkert, V., & Gasthaus, J. (2017, April 13). *DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks*. https://arxiv.org/pdf/1704.04110

Seitfudem, G., Berger, M., Schmied, H. M., & Boulay, A.-M. (2025). The updated and improved method for water scarcity impact assessment in LCA, AWARE2.0. *Journal of Industrial Ecology, 29*(3), 891–907. https://doi.org/10.1111/jiec.70023

Stehfest, E., van Vuuren, D., Kram, T., & Bouwman, L. (Eds.). (2014). *Integrated assessment of global environmental change with IMAGE 3.0: Model description and policy applications*. PBL Netherlands Environmental Assessment Agency.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., & Woods, E. (2020). Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research, 21*(118), 1–6.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-11-00094.1

Thonemann, N., Schulte, A., & Maga, D. (2020). How to conduct prospective life cycle assessment for emerging technologies? A systematic review and methodological guidance. *Sustainability, 12*(3), Article 1192. https://doi.org/10.3390/su12031192

van der Giesen, C., Cucurachi, S., Guinée, J., Kramer, G. J., & Tukker, A. (2020). A critical view on the current application of LCA for new technologies and recommendations for improved practice. *Journal of Cleaner Production, 259*, Article 120904. https://doi.org/10.1016/j.jclepro.2020.120904

van Vuuren, D. P., Riahi, K., Calvin, K., Dellink, R., Emmerling, J., Fujimori, S., Kc, S., Kriegler, E., & O'Neill, B. (2017). The shared socio-economic pathways: Trajectories for human development and global environmental change. *Global Environmental Change, 42*, 148–152. https://doi.org/10.1016/j.gloenvcha.2016.10.009

von Borries, K., Holmquist, H., Kosnik, M., Beckwith, K. V., Jolliet, O., Goodman, J. M., & Fantke, P. (2023). Potential for machine learning to address data gaps in human toxicity and ecotoxicity characterization. *Environmental Science & Technology, 57*(46), 18259–18270. https://doi.org/10.1021/acs.est.3c05300

Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E., & Weidema, B. (2016). The ecoinvent database version 3 (part I): Overview and methodology. *The International Journal of Life Cycle Assessment, 21*(9), 1218–1230. https://doi.org/10.1007/s11367-016-1087-8

Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., Li, X., Dasgupta, A., & Xiong, H. (2024). Interpretable machine learning for weather and climate prediction: A review. *Atmospheric Environment, 338*, Article 120797. https://doi.org/10.1016/j.atmosenv.2024.120797

Ye, R., & Dai, Q. (2021). Implementing transfer learning across different datasets for time series forecasting. *Pattern Recognition, 109*, Article 107617. https://doi.org/10.1016/j.patcog.2020.107617

Yuan, R., Rodrigues, J. F. D., Tukker, A., & Behrens, P. (2022). The statistical projection of global GHG emissions from a consumption perspective. *Sustainable Production and Consumption, 34*, 318–329. https://doi.org/10.1016/j.spc.2022.09.021

Zhu, X., Ho, C.-H., & Wang, X. (2020). Application of life cycle assessment and machine learning for high-throughput screening of green chemical substitutes. *ACS Sustainable Chemistry & Engineering, 8*(30), 11141–11151. https://doi.org/10.1021/acssuschemeng.0c02211

Authors and Affiliations

Niklas Engberg[1] · Carlos Felipe Blanco[2,3] · Valerio Barbarossa[2,4] · Conny Bakker[1] · Ruud Balkenende[1] · Justin Z. Lian[2] · Benjamin Sprecher[1]

✉ Niklas Engberg
n.r.engberg@tudelft.nl

Carlos Felipe Blanco
c.f.blanco@cml.leidenuniv.nl

Valerio Barbarossa
v.barbarossa@cml.leidenuniv.nl

Conny Bakker
c.a.bakker@tudelft.nl

Ruud Balkenende
a.r.balkenende@tudelft.nl

Justin Z. Lian
z.lian@cml.leidenuniv.nl

Benjamin Sprecher
b.sprecher@tudelft.nl

[1] Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

[2] Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands

[3] Circularity and Sustainability Impact Group, Netherlands Organisation for Applied Scientific Research (TNO), Utrecht, The Netherlands

[4] Global Sustainability, PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands