Article

Short- and Mid-Term Forecasting of Pan-Arctic Sea Ice Volume Using Variational Mode Decomposition and Bidirectional Long Short-Term Memory

Aymane Ahajjam [1] , Jaakko Putkonen [2] , Timothy J. Pasch [3,*] and Xun Zhu [3]

1   School of Electrical Engineering and Computer Science, University of North Dakota,
    Grand Forks, ND 58202, USA; mohamed.ahajjam@und.edu
2   Harold Hamm School of Geology and Geological Engineering, University of North Dakota,
    Grand Forks, ND 58202, USA; jaakko.putkonen@und.edu
3   Department of Communication, University of North Dakota, Grand Forks, ND 58202, USA;
    xun.zhu@und.edu
*   Correspondence: timothy.pasch@und.edu

Abstract: The well-documented decrease in the annual minimum Arctic sea ice extent over the past few decades is an alarming indicator of current climate change. However, much less is known about the thickness of the Arctic sea ice. Developing accurate forecasting models is critical to better predict its changes and monitor the impacts of global warming on the total Arctic sea ice volume (SIV). Significant improvements in forecasting performance are possible with the advances in signal processing and deep learning. Accordingly, here, we set out to utilize the recent advances in machine learning to develop non-physics-based techniques for forecasting the sea ice volume with low computational costs. In particular, this paper aims to provide a step-wise decision process required to develop a more accurate forecasting model over short- and mid-term horizons. This work integrates variational mode decomposition (VMD) and bidirectional long short-term memory (BiLSTM) for multi-input multi-output pan-Arctic SIV forecasting. Different experiments are conducted to identify the impact of several aspects, including multivariate inputs, signal decomposition, and deep learning, on forecasting performance. The empirical results indicate that (i) the proposed hybrid model is consistently effective in time-series processing and forecasting, with average improvements of up to 60% compared with the case of no decomposition and over 40% compared with other deep learning models in both forecasting horizons and seasons; (ii) the optimization of the VMD level is essential for optimal performance; and (iii) the use of the proposed technique with a divide-and-conquer strategy demonstrates superior forecasting performance.

Keywords: pan-Arctic sea ice volume forecasting; multi-horizon forecasting; time-series forecasting; short-term forecasting; mid-term forecasting; deep learning

1. Introduction

Current global climate models predict a continuing increase in global air temperatures [1]. Ongoing and impending climate change is predicted to affect the whole climate system, including ocean temperatures, ocean currents, and sea ice extent and thickness [2]. The Arctic ice pack is especially critical to the climate system, as the shrinking sea ice includes a feedback mechanism to further enhance warming. Instead of the snow-covered sea ice reflecting the solar radiation back into the atmosphere and space, the open ocean absorbs a large fraction of the insolation, thus leading to further warming and accelerated ice melting.

The ice thickness, in addition to the aerial extent of the sea ice, is important for the longevity of the ice itself and general shipping and emerging economic activities in the Arctic. While the aerial extent of the sea ice is routinely observed by polar-orbiting

satellites (e.g., NASA's Aqua and Terra satellites and the European Space Agency's CryoSat-2 satellite), the thickness of the sea ice is much more challenging to determine on a pan-Arctic scale and with reasonable frequency (i.e., daily, weekly). The spatial heterogeneity of the pan-Arctic ice thickness is further complicated by the age of the ice. First-year ice is thin, whereas multiyear ice is thicker as it has had more time to build up. In the Arctic Ocean, multiyear ice is preferentially collected in the Beaufort Gyre, located off the north-facing Arctic coasts of Alaska and Canada. This has led to a generally increasing sea ice thickness profile across the Arctic Ocean from the East Siberian Sea (thin) to the Beaufort Sea (thick). In combination, this results in a complex system where the sea ice thickness is controlled by its age, the wind- and current-driven movement of the ice pack, and seasonal freezing and thawing (see Figure 1) [3]. The best thickness data have come from rare submarine transits under the Arctic ice pack. Physics-based sea ice modeling has been the most common means of obtaining sea ice thickness estimates. More recently, data-driven techniques have been successfully employed in sea ice modeling. A recent research study [4] showed that a deep learning probabilistic model outperformed the latest dynamical model in forecasting the summer sea ice extent. Similarly, a model based on convolutional neural networks [5] surpassed the predictive capability of the leading thermodynamic-dynamical model in predicting the sea ice horizontal movement. Moreover, functional data analysis approaches [6], as demonstrated in [7,8], have also been employed to understand sea ice melting patterns. However, the sea ice thickness is challenging to model due to the inherent lack of routine field or remote sensing observations. Nevertheless, the recent results in [9] showed that a statistical two-stage model generating probabilistic forecasts of Arctic sea ice thickness three months into the future met or exceeded the currently available forecasts. Accordingly, we explore the best approaches and the capability of AI-based models to forecast sea ice volume (SIV) time series. Given the relatively direct links between insolation, air temperature, ice melting, and ocean currents, there is a high likelihood that a multivariable correlative model based on historical atmospheric data can be used for ice volume forecasting with relatively minor computational costs.
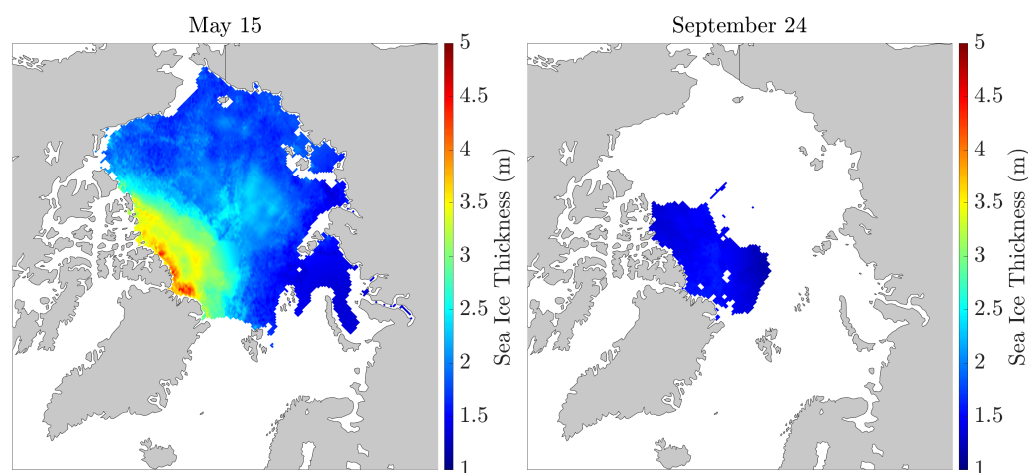
Figure 1. Maxima and minima estimates of the sea ice thickness in the Northern Hemisphere averaged over a 5-year period from 2019 to 2023, derived from the ESA CryoSat-2 Synthetic Aperture Interferometric Radar Altimeter [10].

To achieve these goals, in this paper, we explore the underlying relations between historical atmospheric variables and related sea ice thickness observations in order to build a correlative forecast model for sea ice thickness. In particular, we study various related aspects such as the types and number of inputs (e.g., single or multivariate forecasting), outputs (i.e., single-horizon, multi-horizon), and forecasting models (e.g., machine learning). As a proof of concept, we focus on short-term (i.e., next seven days) to mid-term (i.e., next 30 days) horizons. However, we envision that this step-wise decision process can be used

as a roadmap in the future to investigate signal-processing and deep learning techniques and develop accurate forecasting techniques over longer forecasting horizons (e.g., next six months, next year) using inputs and outputs reflecting other granularities (e.g., weekly) and lengths.

1.1. Background of Time-Series Forecasting

Multivariate forecasting involves using both historical and non-targeted variables for predictions, the latter offering supplementary information to enhance forecast accuracy [11].

Multi-horizon forecasting, in contrast to single-horizon forecasting, aims to predict a phenomenon's future state at multiple time steps simultaneously (see Figure 2). Thus, it provides a timely and efficient approach to decision making in real-world problems, such as Arctic sea ice volume predictions for ship navigation safety [12].
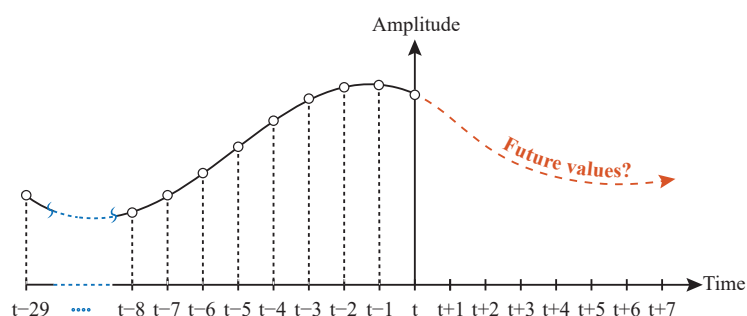
Figure 2. Conceptual representation of multi–horizon forecasting, where the aim is to predict a phenomenon's state in multiple future time instances (e.g., seven steps ahead).

There are three main strategies in multi-horizon forecasting [13]. An iterative method uses a single model repeatedly for multiple predictions, but it risks accumulating errors. A direct approach uses multiple models for each horizon, avoiding error accumulation but at a higher computational cost. Both methods, however, may produce biased forecasts due to the omission of complex stochastic dependencies between past and future values [14,15]. The third strategy, called multi-input multi-output forecasting, simultaneously forecasts all horizons with a single model, reducing computational costs. However, this approach adds complexity, as it needs to account for various time scales, dependencies, and potential interactions between inputs and outputs [16].

1.2. Related Works

In the literature, time-series forecasting has been traditionally addressed using two distinct methods: statistical models and conventional machine learning techniques. Statistical methods include exponential smoothing [17], autoregressive integrated moving average (ARIMA) [18], and their variations (e.g., seasonal ARIMA [19]). Although these models have been extensively used in the literature, they are usually employed for univariate forecasting of linear time series [20]. On the other hand, conventional machine learning techniques have proven effective in handling non-linear and non-stationary data in the literature. Such techniques include support vector regression (SVR) [21], gradient-boosting decision tree (GBDT) [22], and random forest [23]. They are generally computationally efficient and can offer high performance using limited training data. However, they usually require manual feature engineering and can struggle to learn intricate patterns from complex time series.

Recently, the development of deeper machine learning models has proved beneficial in learning from large and complex time-series datasets. Particularly, deep neural networks have become increasingly popular for multi-horizon forecasting thanks to their significant performance improvements over conventional forecasting techniques [24,25]. For instance, the authors of [26] proposed a hybrid model integrating deep belief networks with a state-dependent autoregressive model for single-horizon forecasting of non-linear time series.

In [27], the authors proposed a one-dimensional convolutional neural network (CNN) for single-horizon time-series forecasting. However, recurrent neural networks have become one of the most popular deep models for time-series forecasting thanks to their distinctive recursive architecture. Long short-term memory (LSTM) and gated recurrent unit (GRU) are variations of this type of network. For instance, the authors of [28] proposed a dual-stage two-phase recurrent neural network model for long-term multivariate time-series forecasting. In [29], the authors developed a CNN-LSTM model to forecast multi-step-ahead short-term time series. In [30], the authors proposed the integration of a bidirectional GRU (BiGRU) and the sparrow search algorithm for recursive multi-horizon forecasting. However, sophisticated forecasting models are not sufficient to attain optimal forecasts. This is especially the case with processes that are complex and volatile.

Recent research has underscored the potential of signal-decomposition techniques, such as variational mode decomposition (VMD), in improving the performances of forecasting processes for different research problems [31–34]. Particularly, VMD can generate relatively stationary K-decomposed sequences without suffering from issues such as modal aliasing found with other techniques such as empirical mode decomposition (EMD), thereby simplifying complex patterns in the initial time series and enabling the forecasting model to produce more accurate forecasts. For instance, the authors of [31] developed a hybrid short-term wind-speed forecasting technique consisting mainly of VMD, wavelet transform, and a radial basis function neural network. The authors employed a VMD decomposition level of $K = 7$ after implicit preliminary observations of the decomposition effects of different decomposition levels to achieve improved forecasting performance. In [32], the authors developed a hybrid short-term daily runoff forecasting technique using VMD ($K = 8$) and deep learning. The results showcased the superiority of the proposed hybrid technique over some of its variations. In [33], the authors proposed a hybrid VMD-BiGRU ($K = 11$) for recursive short-term forecasting of natural rubber's price. Both studies employed the central frequency method [34,35] to determine the optimal decomposition level, K, and avoid mode mixing.

1.3. Limitations and Contributions

The current literature holds major limitations. First of all, there is a general lack of studies tackling the problem of pan-Arctic SIV forecasting, as most standard approaches rely on physics-based numerical sea ice modeling, which can be limited in terms of complexity, parameterization requirements, uncertainty in initial conditions and forcing data, limited spatial resolution, and the scarcity of real-world observational data for validation purposes.

The second limitation relates to the incorporation of signal-decomposition techniques in the forecasting process. In particular, although VMD-based techniques provide improvements to the overall performance of forecasting models, the related decomposition level is usually selected arbitrarily or follows techniques that can limit modal mixing (e.g., central-frequency method). However, depending on the specific application and the nature of the data used, this might not be the optimal solution. In particular, while modal mixing can make the decomposition process more complex, it can also introduce additional spatial or temporal correlations between different components. Hence, disregarding the impact of different decomposition levels on the performance of the forecasting model can limit its generalizability and robustness, especially when dealing with complex and highly variable data.

The third limitation is that most studies investigate single-horizon forecasting, which cannot provide a comprehensive overview of future trends, particularly in applications that necessitate a broader perspective through forecasts over multiple future points. Furthermore, among the studies that do implement a multi-horizon forecasting scheme, many rely on recursive approaches that limit the accuracy and reliability of the forecasts.

This paper aims to address these issues, and its main contributions are as follows:

- A hybrid VMD-bidirectional LSTM (BiLSTM) technique is proposed for multi-horizon multivariate forecasting of the pan-Arctic SIV. The use of VMD enables the extraction

of spectral and temporal information reflecting various hidden patterns in the raw volume sequences. On the other hand, the bidirectional structure of the deep learning model can effectively process and learn from future and past instances within the input sequence. The multi-horizon aspect reflects the technique's ability to produce predictions for several points in the future (i.e., lead times) rather than just the next one. The multivariate aspect indicates its ability to process multiple input features, in addition to past values of SIV, to generate its forecasts. Thus, the proposed forecasting model can better learn and memorize the complex patterns and dependencies between different features and over multiple time steps to provide accurate forecasts.

- The proposed technique is investigated in two aspects: one is related to the decomposition technique and one to the implementation strategy. In particular, the impact of multiple VMD decomposition levels on its performance is analyzed to identify the optimal decomposition level for achieving the best results. The proposed technique is further analyzed by investigating the performance of two strategies affecting the way the decomposed inputs are handled within the learning process, namely "all-in-one" and "divide & conquer".
- Technical experiments are carried out using the pan-Arctic Sea Ice Ocean Modeling and Assimilation System (PIOMAS) dataset, which reflects pan-Arctic SIV time series spanning more than four decades. The performance of the proposed technique is studied for short-term and mid-term forecasting horizons, with distinctions between the fall/winter and spring/summer seasons. In addition, multiple benchmarks are conducted under different inputs, conventional and deep forecasting models, and decomposition techniques.

The remainder of this paper is organized as follows. Section 2 provides the mathematical backgrounds of VMD and BiLSTM. Section 3 provides a preliminary exploratory analysis of the used dataset. In Section 4, the main technical aspects employed in this study are described, namely the proposed technique and the conducted experiments. The performance of the proposed hybrid forecasting approach is reported in Section 5. The results of the benchmarks, impact of the VMD decomposition level, and robustness check are also reported and discussed in this section. Finally, Section 6 concludes this research.

2. Methods

In this section, the foundational theories and concepts that underpin this study are described.

2.1. Variational Mode Decomposition

Signal-decomposition techniques are popularly employed in time-series processing tasks such as removing noise and extracting time-frequency information. VMD is a mathematically framed non-recursive decomposition technique. It can simultaneously decompose the input time series into a discrete number of stationary and narrow-band sequences called intrinsic mode functions (IMFs) [36]. The summation of all IMFs and the residual produces the original signal.

Let $\{TS(t)\}_{t=1}^{H}$ represent a certain non-stationary sequence of discrete values sampled over time. The decomposition of this time series using VMD is as follows:

$$TS(t) = \sum_{k=1}^{K} F_k(t) + r(t) \tag{1}$$

where $F_k(t)$ is the extracted kth IMF sequence, K is the chosen decomposition level, and $r(t)$ is the residual.

Based on [36], an IMF is an amplitude- and frequency-modulated signal expressed as:

$$F_k(t) = A_k(t) \cos \phi_k(t), \quad A_k(t) \geq 0 \tag{2}$$

where $\phi_k(t)$ is the phase and $A_k(t)$ is the slowly changing envelope corresponding to the kth *IMF*. It also has a non-decreasing instantaneous frequency that varies slowly and is mostly compact around a center frequency, ω_k.

The VMD algorithm performs several computations to find the K IMFs and their corresponding central frequencies concurrently via an optimization technique called the Alternate Direction Method of Multipliers (ADMM) [37]. According to the ADMM, VMD can decompose the input $TS(t)$ into K F_k and ω_k using these equations [36]:

$$\mathcal{F}\{F_k^{n+1}\} = \frac{\mathcal{F}\{TS(\omega)\} - \sum_{i \neq k} \mathcal{F}\{F_i^{n+1}(\omega)\} + \frac{\mathcal{F}\{\lambda^n(\omega)\}}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left|\mathcal{F}\{F_k^{n+1}(\omega)\}\right|^2 d\omega}{\int_0^\infty \left|\mathcal{F}\{F_k^{n+1}(\omega)\}\right|^2 d\omega} \tag{4}$$

where n is the number of iterations; λ is the Lagrangian multiplier; and $\mathcal{F}\{F_k^{n+1}\}$, $\mathcal{F}\{TS(\omega)\}$, $\mathcal{F}\{F(\omega)\}$, and $\mathcal{F}\{\lambda^n(\omega)\}$ correspond to the Fourier transform of F_k^{n+1}, $TS(t)$, $F(t)$, and λ^n, respectively. The initial value of n and those of other parameters, λ^1, $\mathcal{F}\{F_k^1\}$, and ω_k^1, are set to 0.

This signal-decomposition technique can be distinguished from others by its decomposition level K. If K is chosen at random, there may be issues associated with noise and overlap, producing sub-optimal decompositions. It is, therefore, necessary to identify a suitable K for the context of the task at hand.

2.2. Bidirectional LSTM

LSTM is a popular deep learning neural network that was first proposed back in 1997 [38]. It is a type of recurrent neural network that works as a composition of long- and short-term memory to make future predictions. More specifically, LSTM networks provide an efficient solution to the vanishing gradient problem, as they can retain long-term input dependencies [39]. In addition, an LSTM cell can use or ignore the retained information based on the assigned weights from its learning process. This decision is conducted using its forget gate. These factors make LSTMs optimal in handling sequential data, outperforming other deep learning models but usually at the expense of higher computational costs.

Let $x = (x_1, x_2, ..., x_T)$ represent an input sequence of T elements and h_t represent the LSTM memory at time t. Usually, this gate employs a sigmoid function σ to make the decision f_t at time t, as follows:

$$f_t = \sigma(w_{f_h}[h_{t-1}], w_{f_x}[x_t], b_f) \tag{5}$$

Two other types of gates comprise this model. An input gate determines if the information will be added to the memory. This gate consists of two layers: a sigmoid layer that decides whether values are to be updated i_t, and a *tanh* layer that creates a new vector of candidate values $\tilde{c}_t$ to be added to the memory, as follows:

$$i_t = \sigma(w_{i_h}[h_{t-1}], w_{i_x}[x_t], b_i) \tag{6}$$

$$\tilde{c}_t = tanh(w_{c_h}[h_{t-1}], w_{c_x}[x_t], b_c) \tag{7}$$

Memory can be updated with the vector c_t as follows:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{8}$$

where $*$ is the element-wise multiplication and c_{t-1} is the vector of the old values.

The output gate determines whether the value in the cell passes to the output using a sigmoid layer o_t, as follows:

$$o_t = \sigma(w_{o_h}[h_{t-1}], w_{o_x}[x_t], b_o) \tag{9}$$

In addition, it maps the values within the range $[-1, 1]$ using the tanh function h_t, as follows:

$$h_t = o_t * tanh(c_t) \tag{10}$$

In all these equations, b_f, b_i, b_c, and b_o are the biases and w_{f_h}, w_{f_x}, w_{i_h}, w_{i_x}, w_{c_h}, w_{c_x}, w_{o_h}, and w_{o_x} are the weight matrices.

In this work, we use BiLSTMs (see Figure 3). Such a network was first proposed in [40] using a standard recurrent cell. The LSTM cell's neurons are divided into forward and backward states in such a network. Following this, information from past and future instances can be obtained as forward and backward hidden layers connected to the same output. In comparison with other recurrent neural networks, the BiLSTM's structure allows it to take into account not only the preceding but also the succeeding context. This aspect also makes it more robust to noise and hence able to make more accurate predictions. Such capabilities make BiLSTMs especially well suited for processing and learning patterns found in the input time series (e.g., SIV) to produce accurate forecasts.
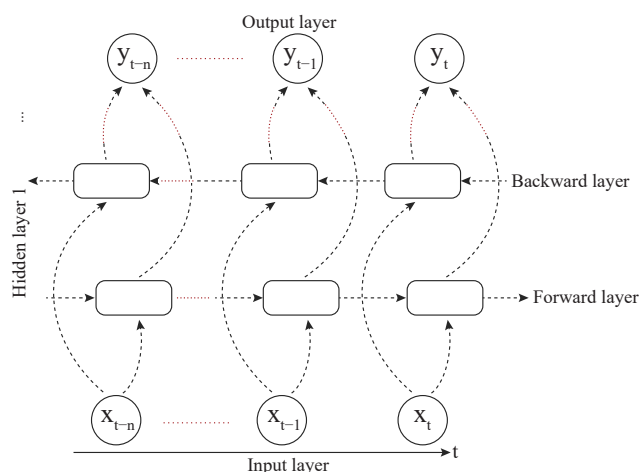
Figure 3. Architecture of the bidirectional LSTM network, incorporating two LSTMs: one processing the data in a forward direction along the temporal dimension, and the other processing the data in a backward direction. This allows the model to have access to both past (backward) and future (forward) contexts simultaneously, enhancing its ability to capture long-term patterns in the sequence data.

3. Dataset Description

The PIOMAS dataset (The PIOMAS dataset can be accessed at the following link: https://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/data (accessed on 25 October 2023)) consists of a 12-category thickness and enthalpy distribution sea ice model coupled with a parallel ocean program [41]. It is the product of the Polar Science Center within the Applied Physics Laboratory at the University of Washington. It relies on the daily sea ice concentration and sea surface temperature acquired from satellites to produce several daily sea ice and ocean features. The ocean model is based on an enhanced Bryan–Cox–Semtner ocean model. The sea ice model is a dynamic thermodynamic system that relies on atmospheric forcing of surface winds and air temperature, humidity, downwelling long- and short-wave radiative fluxes, precipitation, and evaporation [42]. This model incorporates Thorndike thickness distribution theory and enthalpy distribution theory [43]. The PIOMAS dataset contains daily pan-Arctic sea ice thickness and volume data from 1979 onward (see Table 1 for more details). For the remainder of this study, we use the PIOMAS' averaged pan-Arctic SIV time-series data. We assume that the retrieved time series is of

sufficient quality, although previous studies in the literature have indicated some errors in the PIOMAS dataset [43]. The corresponding training set chronologically covers 80% of the data from 1 January 1979 to 17 October 2013, whereas the testing set covers 20% of the data from 18 October 2013 to 30 June 2022.

Table 1. Basic details of the PIOMAS dataset [42]. Both pan-Arctic sea ice thickness and volume time series do not have any missing values.

| Dataset	Resolution	Time Range	# Samples	Features	Units
PIOMAS	Daily	1 January 1979–30 June 2022	15,876	pan-Arctic sea ice thickness pan-Arctic SIV	m km^3

Table 2 summarizes some statistical properties of the pan-Arctic SIV time series for daily, monthly, and yearly periods. Figure 4 showcases seasonal variations over similar time frames. As can be seen, seasonal variations are present in the SIV data. When comparing the SIV between the fall/winter and the spring/summer seasons, it is apparent that the average standard deviation and dynamic range are significantly higher during the latter season. This implies volume values that are more spread out and volatile during the spring/summer season compared to the fall/winter season (as seen in Figure 4b). Two seasonal periods with similar volume patterns can be distinguished: a spring/summer period that corresponds to the thawing of ice (22 March–21 September), characterized by a decreasing volume, and a winter period that corresponds to the icing of ice (22 September–21 March) with an increasing volume. This observation suggests dividing the pan-Arctic SIV time series into two seasonal subsets and developing a separate model for each. Such a division can help the forecasting models better learn each season's characteristics. In addition, daily time series over a whole month all have a similar linear pattern (see Figure 4c). We note that the latter days of the month (i.e., days 29, 30, and 31) have fewer instances in the time series compared to the other days, hence the slight differences in the figure.

Although similar seasonal patterns are present throughout the time series, another distinction can be made between the years (see Figure 4a). The effects of global warming can be observed, with significant decreases in the volume values from one decade to another. In particular, the lowest annually averaged pan-Arctic SIV was 12.863 km^3 in 2017, whereas the highest was 25.390 km^3 in 1979. Hence, capturing these intra-monthly, monthly, and yearly variations could improve forecasting performance.

Table 2. Descriptive statistics of the pan-Arctic SIV times series data from January 1979 to June 2022, with distinctions between the fall/winter and spring/summer seasons (all values are in km^3). Distinct seasonal SIV patterns, shown by notable variations in the ranges and standard averages between seasons, may impact the forecasting models' efficacy (highest values and largest ranges are in bold).

Season	Length	Daily			Monthly			Yearly		
		Avg.	Std.	Min–Max	Avg.	Std.	Min–Max	Avg.	Std.	Min–Max
Fall/winter	8819	18.2	6.7	3.6–32.7	17.9	6.8	3.7–32.0	18.3	3.7	12.3–24.3
Spring/summer	8400	**19.4**	**7.9**	**3.6–33.0**	**20.1**	**7.8**	**3.7–32.9**	**19.4**	**4.0**	**12.8–25.9**
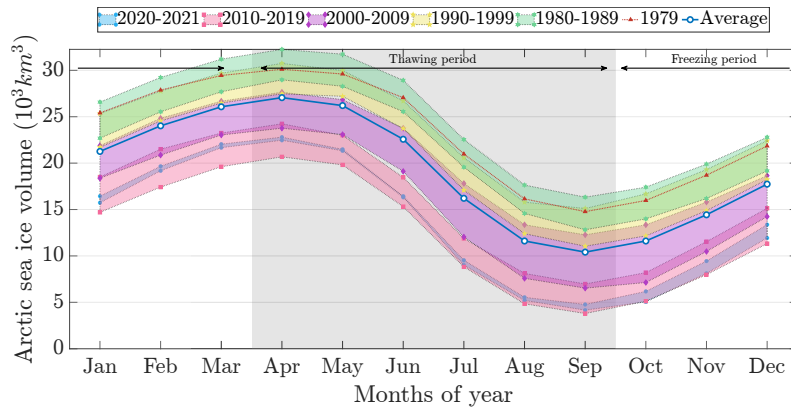
(**a**) Plot of monthly averages in a year with highlighted decades.
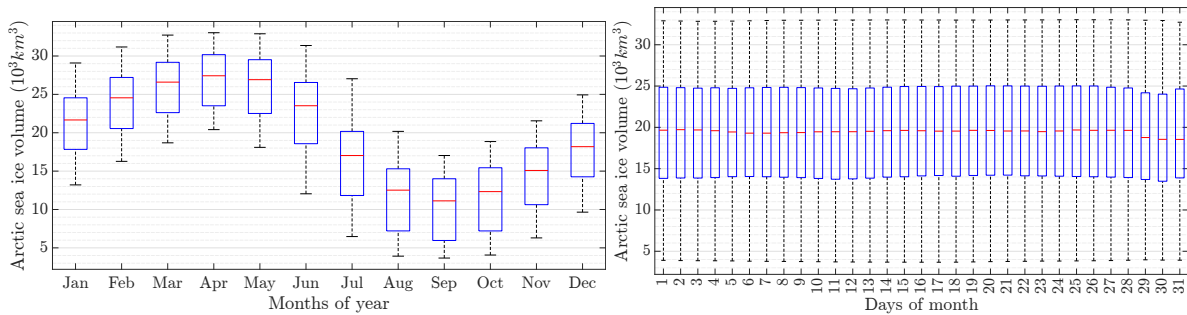
(**b**) Box plot of monthly averages.

(**c**) Box plot of daily averages.

Figure 4. Seasonal cycles of pan-Arctic SIV from 1979 to 2021 over different time frames. The volume patterns distinguish two seasonal periods: the spring/summer thawing period (from around April to September) with decreasing volume, and the winter icing period (from around October to March) with increasing volume.

4. Technical Implementation

This section focuses on the technical aspects of the study forecasting the pan-Arctic SIV. The proposed hybrid forecasting technique is described first. Figure 5 showcases its architecture. Then, three experiments are outlined that offer benchmarks and investigations of the impact of different inputs, decomposition techniques, and deep learning models on forecasting performance. The last subsection summarizes the benchmark models and how their performance is assessed throughout this study.

4.1. Proposed Hybrid VMD-BiLSTM Model

Let $\{I(t)\}_{t=1}^{L}$ denote an Arctic sea SIV sequence (of L samples) that can represent the ice volume $\{V(t)\}_{t=1}^{L}$. Multi-horizon forecasting of an SIV sequence, $I_{t-S+1|t} = \{I(t-S+1), ..., I(t)\}$, of $S-1$ elements and a time index t, can be written as:

$$\hat{I}_{t+S|t+1} = f\left(I_{t-S+1|t}, \epsilon\right) \tag{11}$$

where $\hat{I}_{t+S|t+1}$ is the forecasted SIV sequence over S-steps (i.e., horizons), f is the forecasting model, S is the horizon, and ϵ is the error.

Two forecasting tasks are considered in this paper: forecasting $\hat{I}_{t+S|t+1}$ over short-term horizons (i.e., $S = 7$) and over mid-term horizons (i.e., $S = 30$). Raw volume daily time series are processed to multivariate continuous sequences of $30 + S$ in length, where the first 30 elements are historical values (i.e., inputs) and the S-elements are the targeted future values (i.e., targets). Generally, two main types of sequences are extracted:

1. Pan-Arctic SIV sequences, $V_{t-S+1|t}$, to provide direct historical 30-day volume data (i.e., a sequence of 30 elements).

2. Time sequences to provide recurrent temporal and contextual information, comprising the corresponding values for the day of the month, $D_{t-S+1|t} = \{D(t-S+1), ..., D(t)\}$, ranging from 1 to 31; the month, $M_{t-S+1|t} = \{M(t-S+1), ..., M(t)\}$, ranging from 1 to 12; and the year, $Y_{t-S+1|t} = \{Y(t-S+1), ..., Y(t)\}$, ranging from 1979 to 2022, all with the same granularity (i.e., three sequences of 30 elements each).
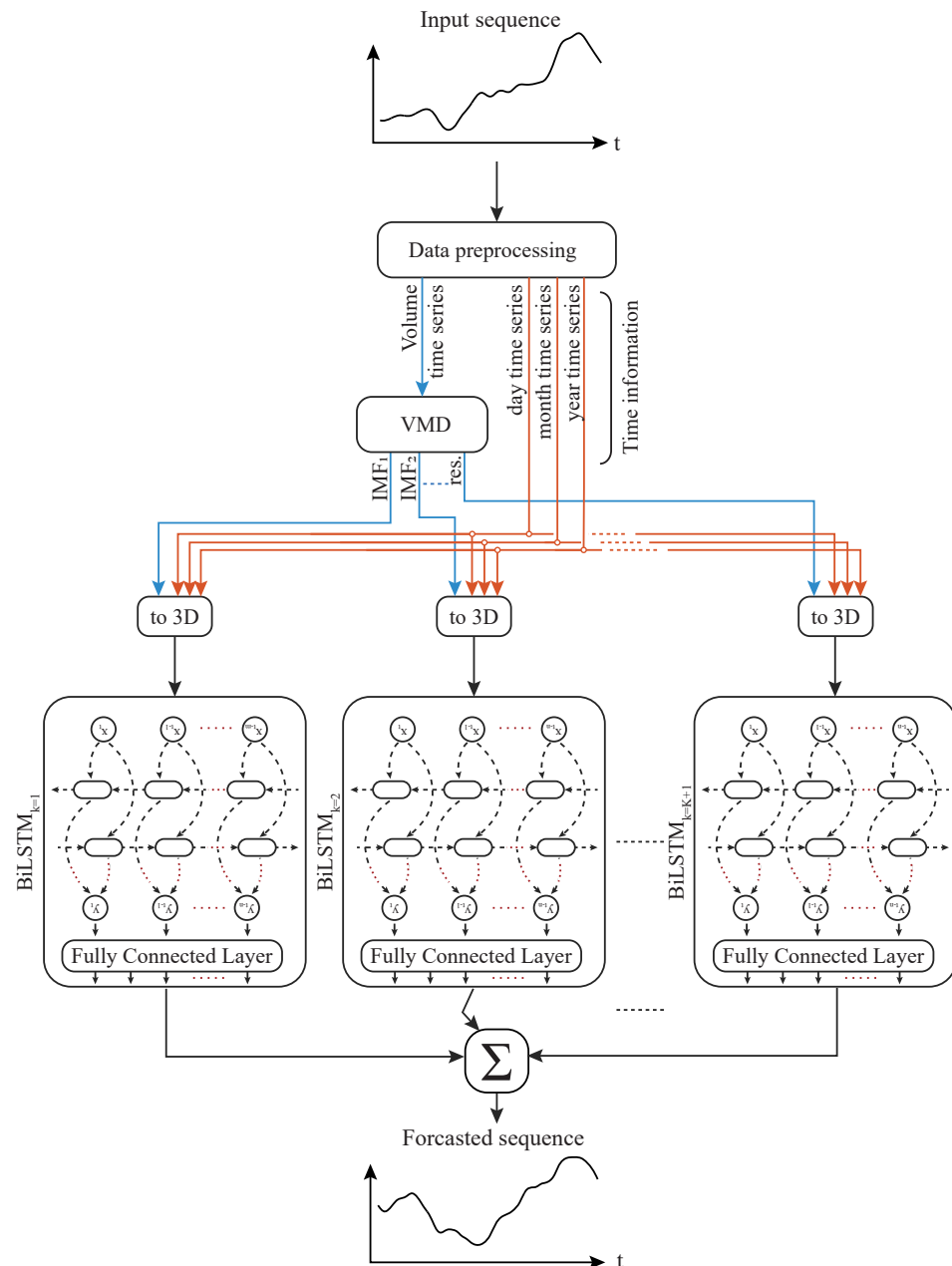
Figure 5. A detailed overview of the proposed hybrid VMD-BiLSTM technique for multi-horizon forecasting of the pan-Arctic SIV, detailing how data are preprocessed, transformed, fed to the BiLSTM model, and then aggregated to generate the final forecasts.

The raw daily SIV time series are processed using a technique called a sliding window. This involves moving a 'window' of a fixed size (i.e., $30 + S$) along the time series to create continuous sequences of data. This is carried out with a unit stride, meaning the window moves one data point at a time. We repeat this process until the entire time series is covered, resulting in a set of overlapping sequences.

Table 3. Properties of the sets of input sequences of the SIV forecasting models. A decomposed SIV input sequence using EMD (i.e., V^{EMD}) has the same size as one using VMD (i.e., $V^{VMD,k}$).

Horizon	Season	Input	Size
Short-term	Fall/winter	Time	$5936 \times 3 \times 37$
		V + Time	$5936 \times 4 \times 37$
		$V^{VMD,k}$ + Time	$5936 \times 4 \times 37$
	Spring/summer	Time	$6816 \times 3 \times 37$
		V + Time	$6816 \times 4 \times 37$
		$V^{VMD,k}$ + Time	$6816 \times 4 \times 37$
Mid-term	Fall/winter	Time	$4671 \times 3 \times 60$
		V + Time	$4671 \times 4 \times 60$
		$V^{VMD,k}$ + Time	$4671 \times 4 \times 60$
	Spring/summer	Time	$5804 \times 3 \times 60$
		V + Time	$5804 \times 4 \times 60$
		$V^{VMD,k}$ + Time	$5804 \times 4 \times 60$

Next, time information sequences (i.e., day, month, year) are read from each input SIV sequence. The resulting SIV and time datasets are chronologically split into training and testing sets based on 80/20 ratios.

Within the proposed hybrid framework, a BiLSTM model with five layers was developed, containing an input layer, forward hidden layers, backward hidden layers, an output layer, and a fully connected layer. For short-term horizons in spring/summer, the proposed model has three hidden layers; in fall/winter, it has five hidden layers; and for mid-term horizons, it has eight hidden layers in both seasons. The dimensions of the input and target reflect the number of nodes in the input and output layers. The fully connected layer outputs the corresponding multi-horizon forecasts (i.e., 7 outputs for short-term horizons and 30 outputs for mid-term horizons). The Adam optimizer was used to train the model with a batch size of 512 and a learning rate of 0.0032 for short-term forecasting, 0.0014 for mid-term forecasting in spring/summer, and around 0.002 for both forecasting horizons in fall/winter. In addition, dropout rates were employed on the outputs of each LSTM layer, except for the last one and the fully connected layer. We note that the learning rate determines how much a model changes in response to the estimated error each time its weights are updated during training, whereas the dropout rate is a regularization technique that prevents overfitting by randomly setting a fraction of the input to 0 at each update.

The performance of the proposed hybrid model is studied under different inputs and compared with other deep learning models. Three main input cases are considered: time only, SIV and time, and decomposed SIV and time. The VMD technique is benchmarked against EMD. A decomposition level of $K = 31$ (and a residual) is set for this last input case.

The first input case comprises time data only. All three time sequences are concatenated to form a 3D array of size $[X \times 3 \times 30 + S]$. The first dimension represents the training set size (i.e., X instances), the second dimension represents the number of input sequences, and the third dimension represents the length of the input sequences (see Table 3 for more details).

Each SIV sequence in the second input case is concatenated with time sequences to form a 3D array of size $[X \times 4 \times 30 + S]$. Consequently, the second dimension reflects a single SIV sequence and three time sequences.

SIV sequences are decomposed for the last two cases using EMD (i.e., third case) and VMD (i.e., fourth case). Each SIV sequence is processed using EMD into a specific number of IMFs (and a residual). While using VMD, they are decomposed into $K = 31$ IMFs (and a residual). For both cases, each decomposed sequence is concatenated with three time sequences to form a 3D array similar to the second case. The second dimension, in this case, reflects a single decomposed sequence in addition to three time sequences.

Thus, in the context of the third case, Equation (11) becomes:

$$\hat{V}_{t+S|t+1} = \Sigma f\left(V^{EMD}_{t-S+1|t}, D_{t-S+1|t}, M_{t-S+1|t}, Y_{t-S+1|t}, \epsilon\right) \qquad (12)$$

and in the context of the fourth case, it becomes:

$$\hat{V}_{t+S|t+1} = \Sigma f\left(V^{VMD,k}_{t-S+1|t}, D_{t-S+1|t}, M_{t-S+1|t}, Y_{t-S+1|t}, \epsilon\right) \qquad (13)$$

where k is a decomposition level ($k \in [1, K+1]$), $V^{VMD,k}_{t-S+1|t}$ and $V^{EMD}_{t-S+1|t}$ are single decomposed SIV sequences using VMD and EMD, and $S = 7$ is used to forecast the next seven days (i.e., short term) or $S = 30$ to forecast next month (i.e., mid term). In addition, other machine learning models are considered. Mainly, convolutional neural networks with InceptionTime and Transformers with time-series transformers (TST). For more details about these two models, refer to Section 4.2.

All data preprocessing is conducted using MATLAB 2023a. The parameters of EMD and VMD are set as follows. For EMD, the sifting relative tolerance is 0.2, the maximum sifting iteration is 100, the maximum number of extracted IMFs is 31, and the maximum number of extrema in the residual signal is 1. For VMD, the penalty parameter is 1000, the number of IMFs is $K = 31$ (in addition to the residual), the initial center frequency is 0, and the convergence criterion is 5×10^{-6}. In addition, optimal sets of hyperparameters of all deep learning models are used in this study. More details about the considered models and their hyperparameters are reported in Section 4.2.

However, the optimal performance of the proposed hybrid model relies not only on the inputs and the employed forecasting model but also on the parameters defining the decomposition and the forecasting processes. Hence, additional investigations were conducted, as described below.

4.1.1. Experiment I: Impact of VMD Decomposition Level on Forecasting Performance

In this experiment, the proposed hybrid model is empirically studied to identify the impact of its VMD decomposition level on its forecasting performance. A divide-and-conquer strategy is employed for all cases of K, where each decomposed sequence and the three time sequences are injected into a single forecasting model. Consequently, $K + 1$ models forecast the K IMFs and a residual. Their forecasts are then summed to produce the intended SIV forecast. Such a strategy enables a timely investigation of the best K value for each forecasting horizon.

4.1.2. Experiment II: Evaluating the Robustness of VMD-Based Forecasting Strategies

Robustness checks are conducted here to identify the best VMD-based forecasting strategy. Using the proposed hybrid model with results from previous experiments, we consider two different strategies (see Figure 6). The first is referred to as the "all-in-one" strategy and represents the most straightforward approach, where all sequences from the best input case (i.e., $K + 1$ decompositions and time sequences) are injected directly into the forecasting model to forecast the SIV. In the second strategy, referred to as "divide and conquer", $K + 1$ instances of the best model are employed. Each model takes as input a single VMD-decomposed sequence, in addition to its corresponding time sequences, and is assigned to forecast the future state of that VMD decomposition. All forecasts are then summed to obtain the actual model's SIV forecast.

(**a**) "All-in-one" strategy   (**b**) "Divide & conquer" strategy
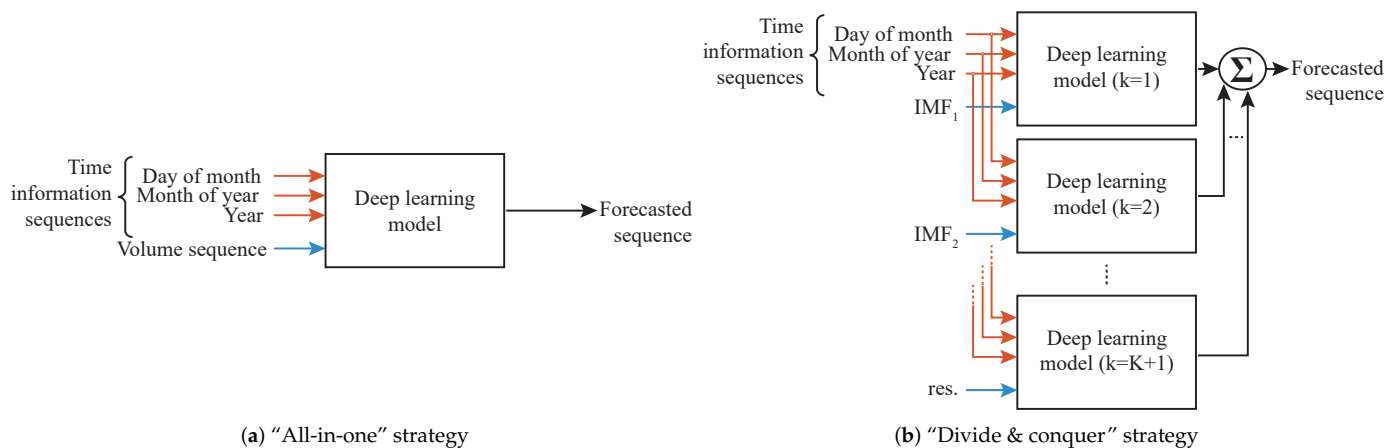
Figure 6. The strategies investigated for multi-horizon pan-Arctic SIV forecasting: 'All-in-one' benchmarking strategy versus the proposed 'Divide and conquer' approach. The former injects all optimal input sequences directly into the model, whereas the latter operates an individual model on one VMD-decomposed sequence, aggregating their forecasts at the end.

4.2. Performance Evaluation

4.2.1. Baseline Models

We consider two deep learning models for benchmarking purposes: TST and InceptionTime. In addition, we employ five other statistical and conventional machine learning as baselines. A brief description of each of these models follows.

- Historical mean: A simple baseline model relying on the values from the previous week (i.e., last 7 elements) or month (i.e., last 30 elements) to provide short-term or mid-term forecasts. Other variations of this model were considered (e.g., same week averaged over the past three months, same month averaged over the past three years), but the proposed historical model proved the best one.
- GBDT [44]: This is an iterative ensemble model of multiple decision trees. The output of the GBDT is the accumulation of the outputs of all its comprised decision trees.
- SVR [45]: This is a popular conventional machine learning model for regression. We employ SVR under two kernels: sigmoid and radial basis functions.
- ARIMA [18]: This is a well-known statistical model for forecasting. ARIMA is generally applicable to non-stationary time series. Its difference transformation can effectively transform non-stationary data into stationary data.
- TST [46]: This is a recent deep neural network that handles long-term dependencies while tracking relationships in sequential input to learn context and meaning. This model was initially proposed in 2017 for translation tasks in Natural Language Processing in [47], but it has now become a state-of-the-art model for various tasks in that field. Multihead self-attention is the core component of TST that makes it suitable for processing time-series data. This mechanism identifies multiple types of dynamic contextual information (i.e., past values, future values) of every element in a sequence, with every attention head. In the recent literature, attention-based deep learning has been effectively employed for uni- and multivariate time-series forecasting problems [48–50].
- InceptionTime [51]: This is a state-of-the-art time-series classification model that was first introduced in 2019. The inception module serves as its main structural component. An inception module consists of (i) a bottleneck layer (one-dimensional convolutional layer) to reduce the dimensionality of the inputs; (ii) three one-dimensional convolutional layers with kernel sizes of 10, 20, and 40, which are fed the output of the bottleneck layer; (iii) a max pooling layer, which is fed the input of the inception module; and (iv) a depth concatenation layer [12], which is the final layer, where the four convolutional layers' outputs are concatenated along the depth dimension. Each

inception module (i.e., convolutional layer) comes with 32 filters by default, which are simultaneously applied to the input time series. To the best of the authors' knowledge, this model has yet to be investigated for the task at hand.

4.2.2. Model Hyperparameter Tuning

Deep learning models have multiple hyperparameters that affect their learning processes and overall performance. The use of non-optimal values can result in model underperformance. For this reason, several hyperparameters of each of these models were subject to optimization.

As hyperparameter tuning requires increased time and computational costs, all deep learning models were tuned using only the non-decomposed pan-Arctic SIV and time information. The tuned hyperparameters from this case were used for the other cases (i.e., time only, EMD-decomposed volume and time, VMD-decomposed volume and time). The optimization trials were conducted using the Optuna framework [52] with the tree-structured Parzen estimator as the sampling algorithm for five epochs (with no early stopping) and over 100 trials (with no pruning). All the considered deep learning models were implemented using the tsai library [53] in Python 3.6 and trained using batch sizes and epochs equal to 512 and 25, respectively.

4.2.3. Evaluation Metrics

The forecasting performance of all models is assessed in this study using five popular metrics: RMSE, Mean Absolute Percentage Error (MAPE), Coefficient of Variance (CV), Anomaly Correlation Coefficient (ACC), and climatology forecast skill score (SS). RMSE and MAPE are accuracy measures. CV indicates the degree of variability in the forecasts. ACC assesses the ability of the model to capture deviations from the mean, whereas the SS metric compares the model's performance to the baseline reference (climatology method). These metrics are defined over the entirety of both forecasting horizons using Equations (14)–(18):

$$RMSE = \sqrt{\frac{1}{N \times S} \sum_{i=1}^{N} \sum_{h=1}^{S} \left(y_i(h) - f_i(h) \right)^2} \tag{14}$$

$$MAPE = \frac{1}{N \times S} \sum_{i=1}^{N} \sum_{h=1}^{S} \left| \frac{y_i(h) - f_i(h)}{y_i(h)} \right| \times 100 \tag{15}$$

$$CV = \frac{\sqrt{\frac{1}{N \times (S-1)} \sum_{i=1}^{N} \sum_{h=1}^{S} \left(y_i(h) - f_i(h) \right)^2}}{\frac{1}{N} \sum_{i=1}^{N} y_i(h)} \times 100 \tag{16}$$

$$ACC = \frac{\sum_{i=1}^{N} \sum_{h=1}^{S} \left((y_i(h) - \bar{y}_i) \cdot (f_i(h) - \bar{f}_i) \right)}{\sqrt{\left(\sum_{i=1}^{N} \sum_{h=1}^{S} \left(y_i(h) - \bar{y}_i \right)^2 \right) \cdot \left(\sum_{i=1}^{N} \sum_{h=1}^{S} \left(f_i(h) - \bar{f}_i \right)^2 \right)}} \tag{17}$$

$$SS = 1 - \frac{\frac{1}{N \times S} \sum_{i=1}^{N} \sum_{h=1}^{S} (y_i(h) - f_i(h))^2}{\frac{1}{N \times S} \sum_{i=1}^{N} \sum_{h=1}^{S} \left(y_i(h) - \frac{1}{N} \sum_{i=1}^{N} y_i(h) \right)^2} \tag{18}$$

where y_i is the hth actual value of a sequence, i; f_i is the hth forecasted value of the same sequence; N is the number of sequences; S is the number of horizons to be forecasted; and $\bar{y}_i$ and $\bar{f}_i$ are the means of the actual sequence and the forecasted sequence, i, respectively.

All volume input sequences in all training and testing sets are standardized to have a zero mean and a unit standard deviation. This simplifies the calculation and amplifies the forecasting model's convergence speed. Accordingly, the forecasted sequences will have to be reversely standardized to provide the actual predictions.

5. Results and Discussion

This section highlights the results of the different experiments investigating different aspects of the hybrid VMD-BiLSTM model in short- and mid-term forecasting horizons.

5.1. Correlation Analysis

This first section examines the relationship between the input and output sequences using the Spearman correlation coefficient, incorporating both the correlation coefficient and the statistical significance. The Spearman coefficient (ρ) is defined by:

$$\rho = 1 - \frac{6\Sigma_{i=1}^{n}d_i^2}{n(n^2-1)} \tag{19}$$

where d_i is the rank difference of two sequences, and n is the length of each sequence.

Figure 7 showcases the average Spearman's coefficient between the historical SIV input and output sequences. Different lengths from the previous 10 to 90 days were constructed as input sequences. The output sequences are the next 7 days and the next 30 days. As can be seen, the immediate historical sequences had the highest correlation with the output sequences. The further back the inputs go, the lower the correlation. In addition, this analysis revealed statistically significant correlations, particularly for the short-term output sequences (i.e., the next seven days), with p-values effectively at zero. For the mid-term output sequences (i.e., the next 30 days), although the correlation remained statistically significant with an average p-value of 0.028, the weaker correlation suggests that the influence of the historical inputs may be attenuated over time.
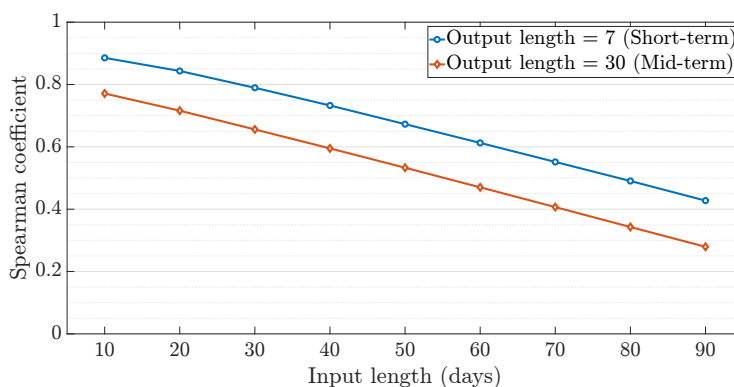
Figure 7. Spearman's correlation coefficients between varying lengths of SIV input sequences (ranging from 10 to 90 elements) and output sequences, reflecting short- and mid-term horizons. The input sequences were linearly interpolated to match the lengths of the output sequences prior to correlation calculation. The correlation results were found to be statistically significant, with p-values of less than 0.001 for short-term outputs and 0.02 for mid-term outputs.

Hence, the use of long durations of inputs (e.g., 40 to 90 days) is not optimal, especially since they would necessitate higher computational costs during all steps of model development. In addition, such durations may not always be available for use. However, very short durations (e.g., 10 days) may not provide enough information for the subsequent processes. This is especially important for signal-decomposition techniques that rely on time-frequency processing to decompose the input or deep learning models that can extract complex patterns from the input. Hence, the choice of the past 30 days as the input length is reasonable for this work, as such an input length can provide a balance between the amount of historical data and computational costs. Nevertheless, further experiments are necessary to quantify the influence of the input lengths on forecasting performance.

5.2. Stationarity Analysis

The stationarity of a time series can be assessed using the Augmented Dickey–Fuller (ADF) [54] and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) [55] tests. The ADF test assumes a unit root in the series (null hypothesis), rejecting this if the test statistic is below a critical value, thereby indicating stationarity. Conversely, the KPSS test assumes trend stationarity (null hypothesis), rejecting it for a non-stationary unit root process if the test statistic exceeds the critical value.

Table 4 presents the average results of the ADF and KPSS tests run on the pre-processed SIV sequences (i.e., the main input data to the forecasting model) for the fall/winter and spring/summer seasons. The tests were run using sequences without decomposition, as well as with EMD and VMD decomposition. We investigated the stationarity of the decomposed sequences using different multiples of eight VMD decomposition levels (K) corresponding to $7, 15, 23, 31, 39, 47, 55, 63, 71,$ and 79 IMFs in addition to the residual (only specific K values are reported in the table). All reported p-values and test static values were averaged over all the sequences of the same case (e.g., test static was averaged over 8819 sequences corresponding to the no-decomposition case for fall/winter). The results show that raw volume sequences were usually not stationary (refer to test results on sequences from the fall/winter season using both tests and on spring/summer sequences using the KPSS). Additionally, decomposing using VMD can provide stationary sequences. For instance, the test statistics using ADF were lower than their corresponding critical values for all considered K values using the fall/winter sequences and for $K > 7$ using the spring/summer sequences. Using KPSS, the test statistics were lower than the critical values for $K \geq 55$ for both seasons. Further, the higher the K, the wider the difference between the two values until a certain K. This demonstrates that greater degrees of stationarity can result from higher decomposition levels.

Table 4. ADF and KPSS tests ($\alpha = 1\%$) on the processed pan-Arctic SIV sequences with or without decomposition (smallest values in each case are indicated in bold). All results were averaged over all related input sequences.

Season	Sequences	ADF			KPSS		
		p-Value	Test Statistic	Critical Value	*p*-Value	Test Statistic	Critical Value
Fall/winter	no decomposition	0.999	29.422444	**−2.64298**	**0.013149**	0.466233	**0.216**
	EMD	0.46391	−0.932447	**−2.64298**	0.028428	0.463015	**0.216**
	VMD, K = 7	0.210504	**−2.743359**	−2.64298	0.038997	0.35998	**0.216**
	VMD, K = 31	**0.133635**	−6.358259	−2.64298	0.053761	0.274097	**0.216**
	VMD, K = 55	0.110459	**−12.187335**	−2.64298	0.069083	**0.192984**	0.216
Spring/summer	no decomposition	0.100008	−17.627617	−2.64298	**0.010001**	0.655554	**0.216**
	EMD	0.467808	−0.990105	**−2.64298**	0.029544	0.438946	**0.216**
	VMD, K = 7	0.197314	−2.438189	**−2.64298**	0.039224	0.353945	**0.216**
	VMD, K = 31	**0.112083**	−6.126092	−2.64298	0.06529	0.228267	**0.216**
	VMD, K = 55	0.114006	**−6.546266**	−2.64298	0.069402	**0.197404**	0.216

In the following subsections, the results of the reported forecasting performance from all conducted experiments are discussed.

5.3. Benchmarking the Proposed Hybrid Model

As can be seen in Table 5, the historical mean model generally achieved superior forecasting performance compared to all considered baseline conventional models. In particular, the GBDT and SVR models yielded the worst performance among all models for both forecasting horizons. In addition, the historical mean outperformed all deep learning forecasting models in the absence of information about the SIV, as evidenced in Table 6 (i.e., time only). We note that this specific case, where the deep learning models were trained using time-only inputs, can be considered another baseline that more advanced models should always outperform. Among these deep learning models, TST and InceptionTime consistently exhibited the poorest performance for short-term and mid-term forecasting,

respectively. The superiority of the historical mean model highlights the usefulness of immediate past values in predicting future ones for both forecasting horizons.

Table 5. Overall performance values of the baseline models for multi-horizon pan-Arctic SIV forecasting (best results are indicated in bold). All metric values are averages, computed over the entirety of the testing set and across the two specified forecasting horizons. Metric units are represented in m^3 for the RMSE and as a % for all other metrics.

Model	Next 7 Days					Next 1 Month				
	RMSE	MAPE	CV	ACC	SS	RMSE	MAPE	CV	ACC	SS
Historical Mean	**922.39**	**6.65**	**6.47**	**36.43**	**97.92**	**3416.56**	**26.76**	**24.02**	**72.25**	**71.64**
SVR-Sigmoid	9089.28	1398.22	63.83	<0	<0	9145.84	97.15	64.5	<0	<0
SVR-RBF	10,109.84	1498.07	71.00	7.26	<0	10,154.48	107.59	71.61	17.60	<0
GBDT	10,212.71	444.54	71.72	0.00	<0	10,178.06	49.41	71.78	0.00	<0
SARIMAX	7070.50	776.15	49.65	8.99	<0	7060.88	51.84	49.79	16.79	<0

Table 6. Overall performance values of the proposed technique, benchmarked against other inputs and models for multi-horizon pan-Arctic SIV forecasting. All metric values are averages, computed over the entirety of the testing set and across the two specified forecasting horizons for each season (best results are indicated in bold). Metric units are represented in m^3 for the RMSE and as a % for all other metrics.

Season	Model	Input	Next 7 Days					Next 1 Month				
			RMSE	MAPE	CV	ACC	SS	RMSE	MAPE	CV	ACC	SS
Fall/winter	TST	Time only	3275.06	19.51	20.11	74.19	48.56	3108.40	19.04	18.88	96.89	39.88
		Volume and Time	293	1.68	1.80	76.47	99.59	825.67	4.41	5.01	97.82	95.71
		V^{EMD} and Time	288.12	1.58	1.77	75.70	99.60	453.31	2.33	2.75	98.23	98.70
		$V^{VMD,31}$ and Time	235.94	1.18	1.45	79.24	99.73	421.13	2.33	2.56	98.39	98.87
	InceptionTime	Time only	2057.75	10.90	12.63	43.99	79.67	3570.47	19.51	21.68	85.82	20.40
		Volume and Time	535.85	3.02	3.29	73.00	98.61	745.62	3.87	4.53	82.34	96.49
		V^{EMD} and Time	430.03	2.3	2.64	61.96	99.10	775.66	4.01	4.71	85.07	96.24
		$V^{VMD,31}$ and Time	355.57	1.86	2.18	62.42	99.40	664.46	3.59	4.04	86.54	97.24
	BiLSTM	Time only	2222.17	13.28	13.64	96.40	76.31	1908.08	10.5	11.59	98.94	77.3
		Volume and Time	257.23	1.50	1.58	97.54	99.68	375.96	2.12	2.28	99.11	99.10
		V^{EMD} and Time	304.79	1.77	1.87	97.75	99.55	354.12	1.92	2.15	98.99	99.19
		$V^{VMD,31}$ and Time	**173.03**	**0.97**	**1.06**	**99.44**	**99.86**	**250.07**	**1.43**	**1.52**	**99.84**	**99.61**
Spring/summer	TST	Time only	5022.15	57.62	39.88	81.32	44.94	4094.32	42.5	33.35	96.47	56.70
		Volume and Time	350.05	2.94	2.78	80.73	99.73	762.13	7.12	6.21	97.1	98.38
		V^{EMD} and Time	614.78	6.37	4.88	84.38	99.17	735.09	7.15	5.99	96.59	98.52
		$V^{VMD,31}$ and Time	265.91	2.56	2.11	78.80	99.84	222.62	**1.99**	1.81	98.62	99.87
	InceptionTime	Time only	3825.53	43.10	30.38	52.08	68.05	5286.05	59.4	43.05	94.49	27.81
		Volume and Time	755.37	7.18	6.00	61.78	98.75	1778.18	19.69	14.48	94.81	91.6
		V^{EMD} and Time	710.16	6.03	5.64	59.07	98.90	1684.15	18.93	13.72	94.5	92.29
		$V^{VMD,31}$ and Time	522.47	4.86	4.15	65.19	99.40	1092.40	12.62	8.90	94.39	96.94
	BiLSTM	Time only	3961.78	44.15	31.46	91.91	65.75	4877.53	51.05	39.73	95.83	38.84
		Volume and Time	455.37	4.16	3.62	96.51	99.55	806.78	8.00	6.57	99.18	98.25
		V^{EMD} and Time	485.95	5.32	3.86	94.09	99.48	769.15	6.90	6.26	97.46	98.37
		$V^{VMD,31}$ and Time	**164.42**	**1.43**	**1.31**	**99.29**	**99.94**	**208.39**	2.16	**1.7**	**99.86**	**99.89**

In the absence of a decomposition technique, better performance can be achieved using the SIV and time information. In this case (i.e., volume and time), all deep learning models significantly outperformed all baselines. Regardless of the forecasting horizon and season, these models could learn from the patterns in the historical SIV and time data, thus achieving overall average improvements of 81.61%, 82.7%, 81.61%, 7.89%, and 114.58% in terms of the RMSE, MAPE, CV, ACC, and SS compared with the previous time-only case. More specifically, BiLSTM performed the best with these input data during the fall/winter season, with overall average improvements of 88.42%, 88.7%, and 88.42% in terms of the RMSE, MAPE, and CV for short-term forecasting and 80.30%, 79.81%, and 80.33% in terms of the same metrics for mid-term forecasting. During the spring/summer season, TST performed the best, with overall average improvements of 93.03%, 94.90%, and 93.03% in

terms of the RMSE, MAPE, and CV for short-term forecasting and 81.39%, 83.25%, and 81.38% in terms of the same metrics for mid-term forecasting.

When processing the SIV with a decomposition technique, even better performance was generally achieved. The overall average improvements achieved using either decomposition technique were 19.52%, 18.15%, and 19.53% in terms of the RMSE, MAPE, and CV over all the forecasting horizons, seasons, and models. Nevertheless, the use of EMD added few to no improvements to the forecasting performance, generally yielding comparable performance with the previous case (i.e., volume and time). This was better seen with short-term forecasting, as the fewest overall average improvements were achieved, with 0.97%, 3.93%, and 1.02% in terms of the RMSE, MAPE, and CV during fall/winter over all three models. During the spring/summer season, deteriorations from the same previous case were observed, with overall averages of 25.45%, 42.84%, and 25.39% in terms of the RMSE, MAPE, and CV over all three models. However, longer forecasting horizons seemed to better exploit the EMD-decomposed input, as generally better improvements were achieved, with 15.63%, 17.66%, and 15.61% during the fall/winter season and 4.50%, 5.73%, and 4.50% during the spring/summer season in terms of the same metrics over the three models. Nevertheless, errors often accumulated with longer forecasting horizons. These underwhelming performance values can be attributed to EMD's drawbacks, such as noise sensitivity and the inability to extract closely spaced high-frequency components in the input data. Nevertheless, the extracted IMFs still provided valuable information to the forecasting model for further forecasting horizons.

The best performance was achieved using the VMD-processed inputs (i.e., $V^{VMD,31}$ and time). Similar to the second case (i.e., volume and time), all models could better learn the variations in the SIV. Particularly, compared with the same second case, overall average improvements of 40.13%, 40.19%, 40.12%, 0.44%, and 1.32% were achieved in terms of the RMSE, MAPE, CV, AC, and SS over all the forecasting horizons, seasons, and models. These results showcase the capabilities of VMD in capturing intrinsic time-frequency patterns, which is essential for better forecasts. Similar to the V^{EMD} and time case, better overall performance was observed with longer forecasting horizons. In particular, the overall average improvements for short-term forecasting were 28.62%, 34.50%, and 28.70% in terms of the RMSE, MAPE, and CV during fall/winter and 39.59%, 36.95%, and 39.58% during spring/summer over all three models. Mid-term forecasting saw overall average improvements of 31.12%, 28.98%, and 31.02% during fall/winter and 61.18%, 60.32%, and 61.17% during spring/summer in terms of the RMSE, MAPE, and CV over all three models. Nevertheless, higher errors were associated with longer horizons.

Specifically, the proposed BiLSTM model consistently achieved the best performance in this case ($V^{VMD,31}$ and time) for both seasons and forecasting horizons. In particular, BiLSTM achieved the highest overall average improvements compared with the first case (i.e., time only), with 94.26% for short-term horizons and 91.23% for mid-term horizons in terms of the RMSE, MAPE, and CV metrics. TST and InceptionTime followed BiLSTM with smaller improvements, where TST achieved average improvements of 94.08% for short-term forecasting and 90.85% for mid-term forecasting for the same metrics and over all seasons. InceptionTime achieved average improvements of 84.97% for short-term forecasting and 80.29% for mid-term forecasting for the same metrics and seasons. In addition, both of these models appeared to struggle to efficiently assimilate patterns from the decomposed sequences and accurately forecast SIV anomalies, yielding ACC averages that were consistently less than 85% for TST and 75% for InceptionTime for all considered input cases. A variety of factors contributed to these lower performance values. For instance, although it incorporates a multihead self-attention mechanism that can process past and future information in the input, TST usually requires a larger amount of training data than that required by LSTM to learn the underlying patterns and make accurate predictions. Nonetheless, these results emphasize the importance of the BiLSTM network's memory cells in capturing and retaining long-term information from the input.

These observations can made about the data in Figure 8, where different examples of forecasting the SIV over short- and mid-term horizons using the proposed hybrid VMD-BiLSTM are showcased. Particularly, the forecasts in both seasons seem to closely follow all target variations over multiple horizons. We note that in Figure 8a, the forecasts using the EMD variation of the proposed technique seem to follow the general trend of the target SIV sequence for September 2016. However, the proposed technique still qualitatively performs better, as it was able to produce forecasts following the change in the slope of the target.

In this study, we only considered the historical data of the pan-Arctic SIV and time information. However, as daily weather features such as air temperature and humidity are readily available in open datasets, such data can be employed as additional inputs to the forecasting model to improve its performance over shorter or longer horizons. Nevertheless, the proposed hybrid model has proven to be more accurate than other forecasting techniques. However, the VMD decomposition level used within it ($K = 31$) is yet to be optimized. The next subsection discusses the experiment's results, identifying the most optimal K value.
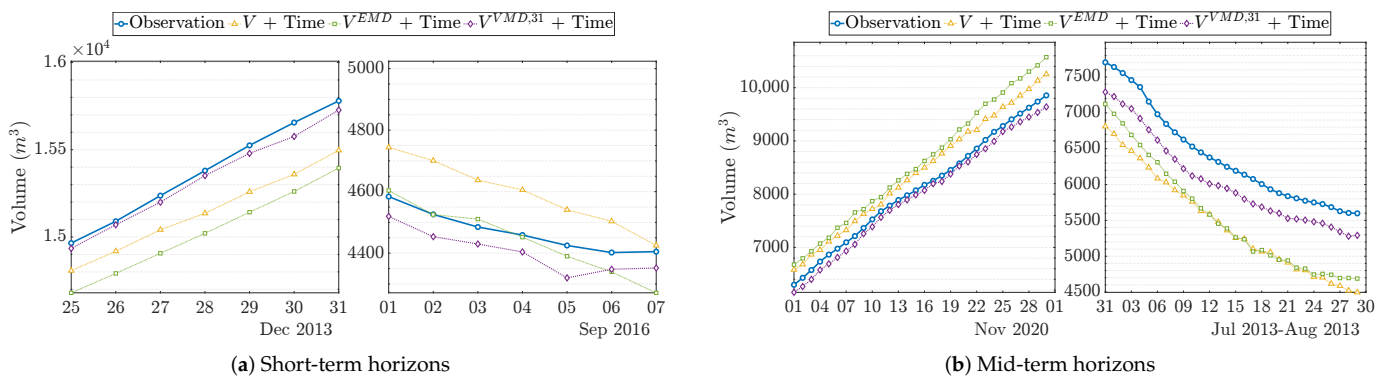

Figure 8. Plot of the BiLSTM model's forecasts based on two instances from the test set of each season, employing three distinct inputs: $V^{VMD,31} + Time$, $Volume + Time$, and $V^{EMD} + Time$. The target SIV sequence is included in the plot for comparison. Generally, the proposed hybrid technique is able to provide forecasts that can closely follow the actual variations in the SIV, even with unusual values (such as the 2016 minima seen on 6 September).

5.4. Experiment I: Impact of VMD Decomposition Level on Forecasting Performance

As can be seen in Figure 9, the forecasting performance of the proposed hybrid model was impacted by the VMD decomposition level (i.e., K). The forecasting performance generally improved with higher values of K up to a certain value, after which it worsened. This pattern can be seen for every forecasting horizon and season. The best forecasting performance during the fall/winter season necessitated higher values than $K = 31$, which was employed in the corresponding case in Experiment I (i.e., $V^{VMD,31}$ and time), with $K = 39$ for short-term forecasting and $K = 63$ for mid-term forecasting. These decomposition levels enabled overall average improvements of 25.93%, 40.81%, and 25.83% for short-term forecasting and 28.52%, 42.20%, and 28.51% for mid-term forecasting in terms of the RMSE, MAPE, and CV compared with the original case of $K = 31$ (i.e., $V^{VMD,31}$ and time). The opposite trend can be seen in the other season, where fewer decomposition levels were sufficient to achieve the best performance. In particular, in spring/summer, a decomposition level of $K = 15$ for short-term forecasting achieved overall average improvements of 42.45%, 19.86%, and 39.37%, compared with the same corresponding case (i.e., $V^{VMD,31}$ and time) in terms of the RMSE, MAPE, and CV. For mid-term forecasting during the same season, the initial $K = 31$ was sufficient to achieve the best performance.

Forecasting longer horizons generally necessitated higher VMD decomposition levels. The forecasting performance was improved with a smaller value of K for short-term forecasting, with $K = 39$ and $K = 15$ for the fall/winter and spring/summer seasons,

respectively. Almost double those values were needed for mid-term forecasting, with $K = 63$ and $K = 31$ for the fall/winter and spring/summer seasons, respectively. This observation can be attributed to the increased uncertainty related to increased forecasting horizons. Thus, further horizon forecasting requires more detailed time-frequency patterns that can be extracted using VMD to achieve optimal performance.

(**a**) Short-term horizon

(**b**) Mid-term horizon
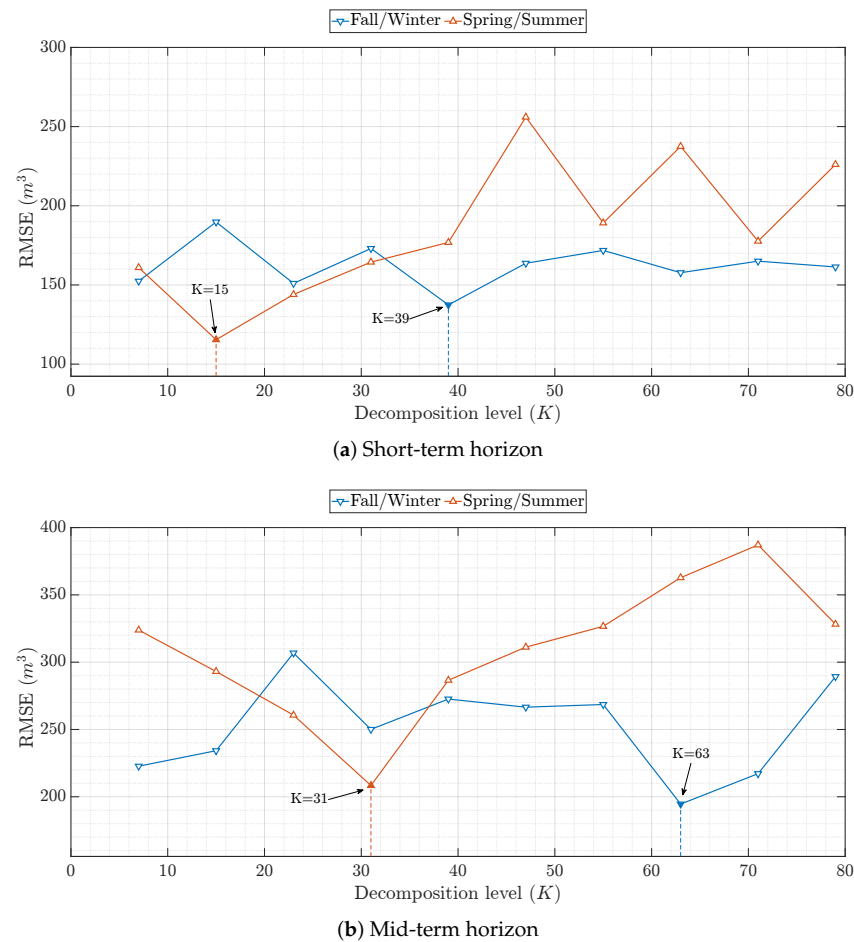
Figure 9. Overall RMSE values of the proposed hybrid technique developed using different VMD decomposition levels, averaged over the whole test set of each season (Experiment I). The lowest RMSE values are highlighted.

As higher decomposition levels were investigated, this experiment required more time and computational resources. In particular, the time and computation costs needed to train the model with no decomposition involved were multiplied by $K + 1$, the number of VMD decompositions (and residual) chosen, to train the same model with VMD-decomposed inputs using the "divide & conquer" strategy (refer to the results of Experiment II for more details). Nevertheless, this experiment must be conducted whenever such a decomposition technique is integrated into a learning process to ensure optimal results. Moreover, the choice of K values is critical. In this experiment, eight multiples were selected based on previous works in the literature and several preliminary trials. Specifically, $K < 7$ values were initially investigated but were dropped after observing few to no improvements in forecasting performance. Hence, multiples of eight provided comprehensible results with big enough improvements following a steady increase in the decomposition levels.

The proposed hybrid forecasting model can provide better forecasting performance with an optimized VMD decomposition level. However, it is important to conduct a robustness check to quantify the usefulness of employing the "divide and conquer" strategy.

5.5. Experiment II: Evaluating the Robustness of VMD-Based Forecasting Strategies

From Table 7 and Figure 10, it is apparent that using the "divide and conquer" strategy produced better overall forecasting performance compared to the "all-in-one" strategy. Particularly, average degradations of 237.09%, 232.99%, and 236.98% in terms of the RMSE, MAPE, and CV were observed, regardless of the forecasting horizon and season. These results provide empirical proof of the usefulness of dividing the forecasting task into sub-forecasting tasks using separate deep learning models. Moreover, the errors over each horizon showcase how the model using the "all-in-one" strategy struggled to deal with the increased uncertainty in further horizons (e.g., mid-term forecasting during the spring/summer season). Although a similar observation can be made for the other case, it was significantly attenuated, as minor errors accumulated over the horizons. Such outcomes prove that dividing the forecasting task into specific and separate tasks can reduce the related uncertainty and generate more accurate forecasts.

However, the proposed hybrid technique implements the "divide and conquer" strategy, which requires more time for model learning. More specifically, training the hybrid model using this strategy can be seen as equivalent to training the same model multiple times using the "All-in-one" strategy. As seen in Figure 6, this increase in time and complexity is tied with $K + 1$, the number of decompositions (and residual), reflecting the training of the $K + 1$ models with partial input data (e.g., an ith IMF with a corresponding time sequence). Nevertheless, the granularity of data (e.g., daily, in the case of this study) and the availability of higher-performance computation devices can enable the successful application of such a forecasting strategy in real-world scenarios.

Further results are reported in Table 8, which displays the performance of the proposed hybrid model, computed and averaged for each month of the year over the whole testing set. Some notable patterns and intriguing trends can be observed. In the case of short-term forecasts, the model exhibited superior performance for January, as evidenced by the lowest RMSE, MAPE, and CV values. This result indicates high accuracy and consistency in the model's forecasts for this month. However, it is noteworthy to highlight March and September when the model's performance dipped. During these transitional months, when sea ice begins to melt in March and starts to form in September, forecasting can be particularly challenging due to the high variability and dynamics of the sea ice processes. The model, while experiencing relatively higher errors during these periods, still managed to provide reasonably accurate forecasts, illustrating its resilience in handling complex Arctic conditions.

For mid-term forecasts, similar trends can be observed. The model continued to demonstrate the strongest performance for January with the lowest RMSE and CV values, whereas for February, it yielded the lowest MAPE, signifying minimal relative errors. Despite the elevated error values for August, September, November, and December, the model still maintained high ACC and SS values, especially for June and July.

The high ACC values for the majority of months indicate that the model correctly captured the directionality of the anomalies. Moreover, the SS values suggest that the proposed technique performed well compared to climatology-based reference forecasts, especially during the summer months when the sea ice melt is in its advanced stage.

Table 7. Experiment II: overall performance values of the optimized proposed hybrid technique (VMD-BiLSTM) using the divide-and-conquer and all-in-one forecasting strategies (best results are indicated in bold). Metric units are represented in m^3 for the RMSE and as a % for all other metrics.

Season	Strategy	Short-Term Horizon					Mid-Term Horizon				
		RMSE	**MAPE**	**CV**	**ACC**	**SS**	**RMSE**	**MAPE**	**CV**	**ACC**	**SS**
Fall/winter	Divide-and-conquer	**137.39**	**0.69**	**0.84**	**99.53**	**99.91**	**194.57**	**1.00**	**1.18**	**99.68**	**99.76**
	All-in-one	395.536	2.033	2.428	97.00	99.20	909.80	3.45	5.52	99.30	94.80
Spring/summer	Divide-and-conquer	**115.42**	**0.86**	**0.91**	**99.21**	**99.97**	**208.38**	**2.15**	**1.69**	**99.86**	**99.88**
	All-in-one	391.72	3.91	3.11	97.70	99.70	528.31	5.17	4.30	99.50	99.30
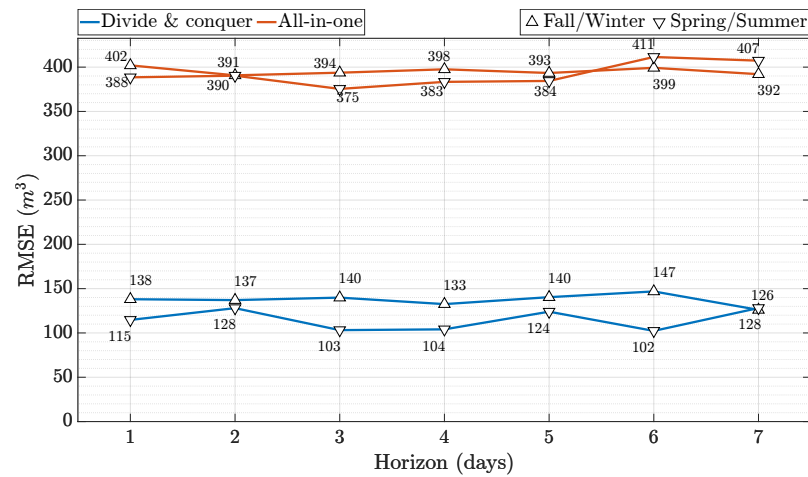
(**a**) Short-term horizon
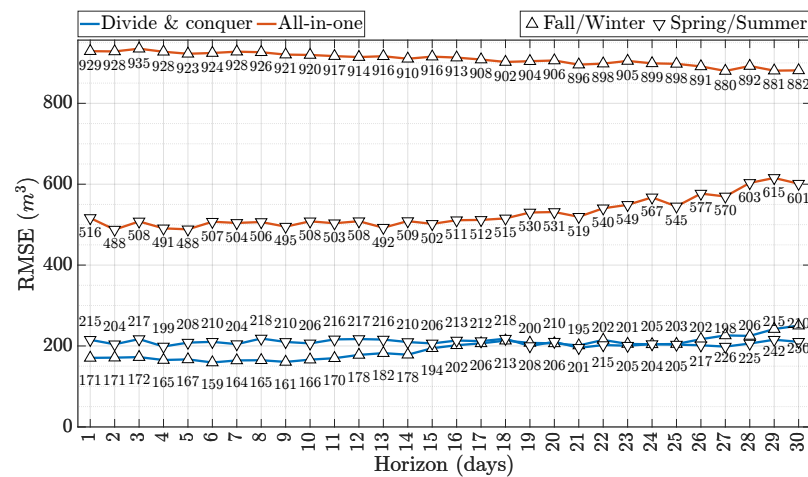
(**b**) Mid-term horizon

Figure 10. RMSE values by horizon of the hybrid VMD-BiLSTM model implemented using two different VMD-based forecasting strategies, averaged over the whole test set of each season (Experiment II).

Table 8. Average performance values of the proposed hybrid technique (VMD-BiLSTM) on the whole test set, computed for each month of the year (best results are indicated in bold). Each data sequence was considered to belong to a specific month based on the start date of that sequence. Metric units are represented in m^3 for the RMSE and as a % for all other metrics.

Month	Short Term					Mid Term				
	RMSE	**MAPE**	**CV**	**ACC**	**SS**	**RMSE**	**MAPE**	**CV**	**ACC**	**SS**
January	**47.74**	**0.22**	**0.33**	99.61	99.87	**102.72**	0.51	**0.73**	99.82	99.72
February	79.19	0.30	0.55	99.42	99.59	110.54	**0.46**	0.78	99.71	99.43
March	211.29	0.87	1.48	98.94	96.16	123.68	0.47	0.87	99.61	98.85
April	131.03	0.55	0.92	78.58	99.94	193.88	0.72	1.37	99.24	94.89
May	165.83	0.63	1.16	97.32	98.31	177.54	0.72	1.26	99.83	98.88
June	123.25	0.52	0.86	**99.68**	99.74	148.55	0.74	1.05	**99.91**	**99.82**
July	50.50	0.41	0.35	**99.68**	**99.95**	155.72	1.52	1.10	**99.91**	99.76
August	78.44	1.09	0.55	98.85	99.62	255.92	3.81	1.81	99.75	97.32
September	105.16	1.73	0.73	85.52	98.54	304.16	5.15	2.15	98.69	89.79
October	169.90	1.72	1.19	99.60	97.71	264.20	2.31	1.87	99.65	94.40
November	168.76	1.45	1.18	99.60	98.79	328.33	2.41	2.32	99.56	95.63
December	102.15	0.65	0.71	99.63	99.67	254.40	1.57	1.80	99.67	98.96

6. Conclusions

Based on satellite imagery over the past few decades, ongoing climate change has reduced the area of the Arctic sea ice pack. However, much less is known about the

changes in the thickness and volume of the Arctic sea ice. To alleviate the current lack of a straightforward and low computational cost model, we set out to develop a hybrid VMD-BiLSTM technique for forecasting the pan-Arctic SIV over multiple horizons. With the integration of VMD with the bidirectional LSTM, innate patterns and fluctuations can be captured and learned for more accurate short-term and long-term forecasts. Empirical experiments highlighted the benefits of multivariate inputs (i.e., previous 30 days of time and volume information) in forecasting accuracy. In addition, the experiments showcased how a tuned decomposition level in VMD-empowered forecasting techniques can further enhance the model's forecasting accuracy. Although cumbersome, such a fine-tuning task is necessary to achieve optimal forecasting results. Further, this paper provides evidence that a forecasting approach based on a signal decomposition must rely on the "divide and conquer" strategy to optimize the usage of the extracted time-frequency patterns from the input data.

We find that the main benefits of these results are twofold: (1) They are proof of concept that an advanced deep learning-based forecasting technique can produce better results with fewer computational costs compared to conventional forecast modeling for Arctic sea ice thickness. (2) Given the increasing importance of thawing the Arctic Ocean for economic, military, and recreational shipping, future machine learning-based sea ice thickness modeling focused on subsections of the Arctic Ocean may help operational navigation in the Arctic.

References

1. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.; et al. Climate change 2021: The physical science basis. *Contrib. Work. Group Sixth Assess. Rep. Intergov. Panel Clim. Chang.* **2021**, *157*.
2. World Economic Forum. The Global Risks Report 2022, 17th Edition. 2022. Available online: https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2022.pdf (accessed on 26 August 2023).
3. Shalina, E.V.; Khvorostovsky, K.; Sandven, S. Arctic Sea Ice Thickness and Volume Transformation. In *Sea Ice in the Arctic*; Springer Polar Sciences; Springer: Cham, Switzerland, 2020; pp. 167–246.
4. Andersson, T.R.; Hosking, J.S.; Pérez-Ortiz, M.; Paige, B.; Elliott, A.; Russell, C.; Law, S.; Jones, D.C.; Wilkinson, J.; Phillips, T.; et al. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.* **2021**, *12*, 5124. [CrossRef] [PubMed]
5. Zhai, J.; Bitz, C.M. A machine learning model of Arctic sea ice motions. *arXiv* **2021**, arXiv:2108.10925.
6. Shah, I.; Muhammad, I.; Ali, S.; Ahmed, S.; Almazah, M.M.; Al-Rezami, A. Forecasting day-ahead traffic flow using functional time series approach. *Mathematics* **2022**, *10*, 4279. [CrossRef]
7. Mcdonald, S.; Koulis, T.; Ehn, J.; Campbell, K.; Gosselin, M.; Mundy, C. A functional regression model for predicting optical depth and estimating attenuation coefficients in sea-ice covers near Resolute Passage, Canada. *Ann. Glaciol.* **2015**, *56*, 147–154. [CrossRef]
8. Das, P.; Lahiri, A.; Das, S. Understanding sea ice melting via functional data analysis. *Curr. Sci.* **2018**, *115*, 920–929. [CrossRef]
9. Gao, P.A.; Director, H.M.; Bitz, C.M.; Raftery, A.E. Probabilistic forecasts of Arctic sea ice thickness. *J. Agric. Biol. Environ. Stat.* **2022**, *27*, 280–302. [CrossRef]
10. Kurtz, N.; Harbeck, J. CryoSat-2 Level-4 Sea Ice Elevation, Freeboard, and Thickness, Version 1 [Data Set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. Available online: https://nsidc.org/data/rdeft4/versions/1 (accessed on 26 August 2023). [CrossRef]

11. Woodward, W.A.; Sadler, B.P.; Robertson, S. *Time Series for Data Science: Analysis and Forecasting*; Multivariate Time Series; CRC Press: Boca Raton, FL, USA, 2022; Chapter 5, p. 417.

12. Taieb, S.B.; Atiya, A.F. A bias and variance analysis for multistep-ahead time series forecasting. *IEEE Trans. Neural Networks Learn. Syst.* **2015**, *27*, 62–76. [CrossRef]

13. Ye, R.; Dai, Q. MultiTL-KELM: A multi-task learning algorithm for multi-step-ahead time series prediction. *Appl. Soft Comput.* **2019**, *79*, 227–253. [CrossRef]

14. Bontempi, G. Long term time series prediction with multi-input multi-output local learning. In Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08, Porvoo, Finland, 17–19 September 2008; pp. 145–154.

15. Ahajjam, M.A.; Licea, D.B.; Ghogho, M.; Kobbane, A. Experimental investigation of variational mode decomposition and deep learning for short-term multi-horizon residential electric load forecasting. *Appl. Energy* **2022**, *326*, 119963. [CrossRef]

16. Taieb, S.B.; Sorjamaa, A.; Bontempi, G. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* **2010**, *73*, 1950–1957. [CrossRef]

17. Billah, B.; King, M.L.; Snyder, R.D.; Koehler, A.B. Exponential smoothing model selection for forecasting. *Int. J. Forecast.* **2006**, *22*, 239–247. [CrossRef]

18. Newbold, P. ARIMA model building and the time series analysis approach to forecasting. *J. Forecast.* **1983**, *2*, 23–35. [CrossRef]

19. Chen, P.; Niu, A.; Liu, D.; Jiang, W.; Ma, B. Time series forecasting of temperatures using SARIMA: An example from Nanjing. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2018; Volume 394, p. 052024.

20. Yin, J.; Rao, W.; Yuan, M.; Zeng, J.; Zhao, K.; Zhang, C.; Li, J.; Zhao, Q. Experimental study of multivariate time series forecasting models. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2833–2839.

21. Ceperic, E.; Ceperic, V.; Baric, A. A strategy for short-term load forecasting by support vector regression machines. *IEEE Trans. Power Syst.* **2013**, *28*, 4356–4364. [CrossRef]

22. Zhou, F.; Zhang, Q.; Sornette, D.; Jiang, L. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Appl. Soft Comput.* **2019**, *84*, 105747. [CrossRef]

23. Mei, J.; He, D.; Harley, R.; Habetler, T.; Qu, G. A random forest method for real-time price forecasting in New York electricity market. In Proceedings of the 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, USA, 27–31 July 2014; pp. 1–5.

24. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [CrossRef]

25. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]

26. Xu, W.; Peng, H.; Zeng, X.; Zhou, F.; Tian, X.; Peng, X. Deep belief network-based AR model for nonlinear time series forecasting. *Appl. Soft Comput.* **2019**, *77*, 605–621. [CrossRef]

27. Sayeed, A.; Choi, Y.; Eslami, E.; Lops, Y.; Roy, A.; Jung, J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Netw.* **2020**, *121*, 396–408. [CrossRef]

28. Liu, Y.; Gong, C.; Yang, L.; Chen, Y. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Syst. Appl.* **2020**, *143*, 113082. [CrossRef]

29. Ferreira, L.B.; da Cunha, F.F. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput. Electron. Agric.* **2020**, *178*, 105728. [CrossRef]

30. Li, X.; Ma, X.; Xiao, F.; Xiao, C.; Wang, F.; Zhang, S. Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *J. Pet. Sci. Eng.* **2022**, *208*, 109309. [CrossRef]

31. Zhang, Y.; Chen, B.; Pan, G.; Zhao, Y. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. *Energy Convers. Manag.* **2019**, *195*, 180–197. [CrossRef]

32. Xie, T.; Zhang, G.; Hou, J.; Xie, J.; Lv, M.; Liu, F. Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China. *J. Hydrol.* **2019**, *577*, 123915. [CrossRef]

33. Zhu, Q.; Zhang, F.; Liu, S.; Wu, Y.; Wang, L. A hybrid VMD–BiGRU model for rubber futures time series forecasting. *Appl. Soft Comput.* **2019**, *84*, 105739. [CrossRef]

34. Wang, J.; Cui, Q.; He, M. Hybrid intelligent framework for carbon price prediction using improved variational mode decomposition and optimal extreme learning machine. *Chaos Solitons Fractals* **2022**, *156*, 111783. [CrossRef]

35. Huang, N.; Chen, H.; Cai, G.; Fang, L.; Wang, Y. Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier. *Sensors* **2016**, *16*, 1887. [CrossRef]

36. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

37. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]

38. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

39. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.

40. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]

41. Schweiger, A.; Lindsay, R.; Zhang, J.; Steele, M.; Stern, H.; Kwok, R. Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res. Ocean.* **2011**, *116*. [CrossRef]
42. Zhang, J.; Rothrock, D.A. Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Weather Rev.* **2003**, *131*, 845–861. [CrossRef]
43. Wang, X.; Key, J.; Kwok, R.; Zhang, J. Comparison of Arctic Sea Ice Thickness from Satellites, Aircraft, and PIOMAS Data. *Remote Sens.* **2016**, *8*, 713. [CrossRef]
44. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
45. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1996**, *9*.
46. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A Transformer-Based Framework for Multivariate Time Series Representation Learning. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21, New York, NY, USA, 14–18 August 2021; pp. 2114–2124. [CrossRef]
47. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
48. Lara-Benítez, P.; Gallego-Ledesma, L.; Carranza-García, M.; Luna-Romera, J.M. Evaluation of the Transformer Architecture for Univariate Time Series Forecasting. In Proceedings of the Conference of the Spanish Association for Artificial Intelligence, Malaga, Spain, 22 September 2021; Springer: Cham, Switzerland, 2021; pp. 106–115.
49. Hu, Y.; Xiao, F. Network self attention for forecasting time series. *Appl. Soft Comput.* **2022**, *124*, 109092. [CrossRef]
50. He, X.; Shi, S.; Geng, X.; Xu, L. Dynamic Co-Attention Networks for multi-horizon forecasting in multivariate time series. *Future Gener. Comput. Syst.* **2022**, *135*, 72–84. [CrossRef]
51. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962. [CrossRef]
52. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
53. Oguiza, I. Tsai—A State-of-the-Art Deep Learning Library for Time Series and Sequential Data. 2022. Available online: https://github.com/timeseriesAI/tsai (accessed on 24 August 2023).
54. Said, S.E.; Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **1984**, *71*, 599–607. [CrossRef]
55. Kwiatkowski, D.; Phillips, P.C.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econom.* **1992**, *54*, 159–178. [CrossRef]