Neural Computing and Applications (2024) 36:21581–21605
https://doi.org/10.1007/s00521-024-10305-z

ORIGINAL ARTICLE

A multi-modal approach for mixed-frequency time series forecasting

Leopoldo Lusquino Filho[1,2] · Rafael de Oliveira Werneck[1] · Manuel Castro[1] · Pedro Ribeiro Mendes Júnior[1] · Augusto Lustosa[1] · Marcelo Zampieri[3] · Oscar Linares[1] · Renato Moura[1] · Elayne Morais[1] · Murilo Amaral[1] · Soroor Salavati[1] · Ashish Loomba[3] · Ahmed Esmin[1,4] · Maiara Gonçalves[3] · Denis José Schiozer[3,5] · Alexandre Ferreira[1] · Alessandra Davólio[3] · Anderson Rocha[1]

Received: 29 April 2024 / Accepted: 29 July 2024 / Published online: 4 September 2024

Abstract
This study proposes a novel multimodal approach for mixed-frequency time series forecasting in the oil industry, enabling the use of high-frequency (HF) data in their original frequency. We specifically address the challenge of integrating HF data streams, such as pressure and temperature measurements, with daily time series without introducing noise. Our approach was compared with existing econometric regression model mixed-data sampling (MIDAS) and with the data-driven models N-HiTS and a GRU-based network, across short-, medium-, and long-term prediction horizons. Additionally, we validated the proposed method on datasets from other domains beyond the oil industry. The experimental results indicate that our multimodal approach significantly improves long-term prediction accuracy.

Keywords Multimodal learning · Forecasting · Mixed-frequency time series · Pre-salt oil field

1 Introduction

Accurate forecasting in the oil industry is a paramount requirement for effective decision-making, resource allocation, and risk management. The complexities inherent in hydrocarbon reservoirs require the use of advanced forecasting techniques. Traditional methods, such as reservoir simulation models, are renowned for their precision in long-term predictions but come at the cost of intensive computational resources, particularly in intricate reservoir systems [1]. In contrast, machine learning-based approaches have emerged as promising alternatives, offering the potential for more agile short-term forecasts [2, 3].

One of the primary challenges faced in oil reservoir monitoring and production forecasting is the heterogeneous nature of data collection, a common factor in several classes of real-world problems that use varied sensor sets or multiple sources of information acquisition [4]. Sensors within reservoirs often operate at diverse sampling rates, ranging from daily measurements to high-frequency minute-by-minute readings. This disparate nature of data acquisition creates a significant hurdle in integrating and extracting valuable insights from these mixed-frequency time series.

In response to this challenge, our research focuses on the development and validation of novel multimodal architectures for time series forecasting in the oil industry, and such architectures refer to models whose topologies are designed to simultaneously deal with different types of data, such as image, audio, and video [5]. The central premise of our approach is to treat high-frequency (HF) data streams independently, recognizing that each HF series represents a unique source of information with its temporal characteristics.

✉ Leopoldo Lusquino Filho
leopoldo.lusquino@unesp.br

1 Artificial Intelligence Lab. (Recod.ai), Institute of Computing, Universidade Estadual de Campinas - UNICAMP, Campinas, SP 13083-852, Brazil

2 Institute of Science and Technology, São Paulo State University - UNESP, Sorocaba, SP 18087-180, Brazil

3 Center for Petroleum Studies (CEPETRO), Universidade Estadual de Campinas - UNICAMP, Campinas, SP 13083-970, Brazil

4 Department of Computer Science, Federal University of Lavras - UFLA, Centro, Lavras, MG 37200-000, Brazil

5 School of Mechanical Engineering, Universidade Estadual de Campinas - UNICAMP, Campinas, SP 13083-970, Brazil

These proposed multimodal architectures are designed to exploit the strengths of deep learning and Recurrent Neural Networks (RNNs) to model and forecast complex oil production patterns. By considering each HF series as a separate modality, our approach offers several advantages, including adapting to different periodicities, selectively emphasizing relevant sensors, and mitigating noise from less trustable sensors. Additionally, our proposed approach is flexible enough to be applied in different mixed-frequency time series forecasting domains.

We compared our approach with the traditional mixed-frequency time series regression technique MIDAS, the state-of-the-art time series forecasting method N-HiTS, and GRU2 [6], a recurrent neural network that has shown promising results for short-term oil production forecasting, as well as short-term prediction in other scenarios. The experimental results in different scenarios, with three different datasets, suggest that although there is no universal solution for mixed-time series forecasting, the solution proposed in this work is relevant and showed the best performance in several scenarios, especially in the long-term, with data whose attributes present behavior with a high degree of nonlinearity.

The structure of the paper is as follows: Sect. 2 provides an overview of the related research in mixed-frequency time series forecasting, multimodal machine learning, and the setup for multioutput forecast validation; Sect. 3 introduces the multimodal architectures proposed; Sect. 4 offers a detailed account of the experiments conducted, along with their analyses; and Sect. 5 presents the conclusions and outlines potential future research directions.

2 Related work

In the domain of time series forecasting, mixed-frequency time series analysis and multimodal learning techniques have emerged as pivotal areas of research, given the greater availability of heterogeneous data and the greater variety of data sources [7], in addition to the need to model rare and anomalous events, due to the advancement of climate change [8]. Mixed-frequency time series, characterized by diverse temporal resolutions within a single dataset, present unique challenges that demand specialized modeling approaches. Simultaneously, the fusion of information from multiple modalities allows for capturing complex dependencies and enhancing predictive capabilities. This paper explores the intersection of mixed-frequency time series and multimodal learning, specifically focusing on multi-output forecasting scenarios. All these themes are detailed in this section.

2.1 Mixed-frequency time series

Mixed-frequency time series data, characterized by observations collected at irregular intervals, have gained prominence across multiple domains, including economics, finance, and environmental monitoring. This section extensively reviews methodologies and models developed to tackle the complexities of mixed-frequency time series. The concept of mixed-frequency time series can be traced back to early econometric endeavors in modeling economic indicators. Initially, efforts were concentrated on synchronizing these irregularly sampled indicators using interpolation or aggregation techniques. However, this approach often led to the loss of critical temporal information, rendering it inadequate for capturing intricate temporal dependencies inherent in mixed-frequency data.

One of the significant complexities lies in handling data with features that are time series at different granularities, necessitating the development of novel modeling techniques to accommodate this irregularity. The autoregressive moving average (ARMA)-based models [9, 10], commonly used for time series analysis, face challenges when dealing with mixed-frequency data. In traditional ARMA models, all-time series are assumed to be observed at the same frequency, which contradicts the reality of mixed-frequency data. Consequently, adapting ARMA to mixed-frequency settings necessitates the introduction of time-varying parameters, missing data handling strategies, and synchronization methods to account for the different observation frequencies.

This intricate mathematical landscape underscores the need for sophisticated modeling approaches and computational methods to address the complexities of mixed-frequency time series forecasting. To overcome this limitation, the literature proposed models that specify conditional expectations as a distributed lag of regressors at some higher sampling frequencies, using a weighting function [11].

Among the consolidated solutions in the forecasting literature that can perform mixed-frequency time series analysis without the need to aggregate the data, interpolate it, or change its original frequency in some way, we have bridge equations, mixed-frequency vector autoregressive (VAR) models, mixed-frequency factor models, and the mixed-data sampling (MIDAS) models. Bridge equations [12] are particularly suitable for short-term forecasts that focus on capturing the immediate impact of changes in high-frequency variables on low-frequency ones. Mixed-frequency VARs [13] assume linear relationships between variables of different frequencies. They are often used for short- to medium-term forecasts, where the assumption of linearity is reasonable. However, their performance may

need to improve in capturing more complex and nonlinear relationships over longer forecasting horizons. Mixed factor models [14, 15] and MIDAS models construct lags between time series of different frequencies in a handcrafted way, which may degrade their performance in the long-term.

Mixed-frequency time series forecasting methods have been researched and applied in different domains, such as electrical power generation [16, 17], carbon emission monitoring [18] and econometrics [19]. While, most of these approaches are heavily based on autoregressive models, some of them are hybrids and have shown promise. As the use of sensors to capture vast amounts of data has increased and predictions about emerging behaviors of complex systems have become more critical, the development of robust methodologies for this type of time series analysis has become increasingly important.

2.1.1 MIDAS

Notable among the early developments is the mixed-data sampling (MIDAS) regression framework [20]. MIDAS addresses the interplay between high-frequency and low-frequency variables. By employing lagged high-frequency data, MIDAS allows for estimating long-term relationships between economic variables with varying sampling frequencies. It utilizes weighted least squares estimation and leverages the lag structure of the high-frequency data to capture the desired effects. MIDAS has since, become a cornerstone in econometric research, with various extensions and applications.

For a low-frequency dependent variable y_t and high-frequency explanatory variable x_t, the basic MIDAS (Mixed Data Sampling) specification can be written as follows:

$$y_t = \beta_0 + \beta_1 B\left(L^{\frac{1}{m}}; \Theta\right) x_t^{(m)} + \varepsilon_t^{(m)} \quad (1)$$

, where y_t is the low-frequency dependent variable, x_t is the high-frequency explanatory variable, m denotes the frequency ratio between y_t and x_t (e.g., if y_t is quarterly and x_t is monthly, $m = 3$), $\varepsilon_t^{(m)}$ is the error term corresponding to the low-frequency period, $B(L^{\frac{1}{m}}; \Theta)$ is a lag polynomial with parameters Θ, where $L^{\frac{1}{m}}$ represents the lag operator raised to the power $\frac{1}{m}$, indicating the aggregation of high-frequency lags into the low-frequency model.

Many MIDAS experiments indicate that it effectively incorporates high-frequency information into low-frequency models [21–23], enhancing the understanding of dynamic relationships. Despite this, MIDAS has known limitations, as discussed in [24, 25]. One of the primary limitations of MIDAS is its assumption of linearity in modeling relationships between high-frequency and low-frequency variables. This assumption may not hold in real-world scenarios where complex nonlinear interactions exist. MIDAS might not accurately capture these dynamics when the underlying relationship is inherently nonlinear, potentially leading to model inadequacy.

MIDAS models often require selecting the appropriate lag structure for high-frequency data. If the lag structure needs clarification, it can result in suboptimal model performance. Identifying the optimal number of lags is a non-trivial task, and errors in this selection can lead to biased parameter estimates and unreliable forecasts. Selecting the appropriate specification of a MIDAS model (e.g., lag structure, polynomial weighting schemes) can be challenging and often relies on heuristic approaches or model selection criteria, introducing subjectivity into the modeling process. Besides that, when fitting MIDAS models, there is a risk of overfitting, especially when many lags are used. Overfitting occurs when the model captures noise and idiosyncratic fluctuations in the data, leading to poor out-of-sample predictive performance. Balancing model complexity with generalization to unseen data is a critical challenge [26, 27].

One more limitation of MIDAS is the stationarity assumption [28]. Like many traditional time series models, MIDAS relies on the assumption of stationary, implying that the data's statistical properties do not change over time. Economic and financial time series data often exhibit non-stationary behavior, necessitating additional techniques such as differentiation or transformation to achieve stationary status. These transformations can introduce complexity and potential loss of information [29].

Another area for improvement with MIDAS is related to computational complexity. Estimating MIDAS models can be computationally intensive, particularly when dealing with many lags or high-frequency data with irregular patterns. This computational complexity can pose challenges for researchers and analysts, requiring substantial computational resources and time. Finally, MIDAS suffers from sensitivity to high-frequency data. The effectiveness of MIDAS models depends on the quality and reliability of high-frequency data. If the high-frequency data are noisy, contain outliers, or suffer from data quality issues, it can significantly impact the model's accuracy and stability [30].

Variations of MIDAS encompass (i) MS-MIDAS [31], which incorporates regime changes in MIDAS model parameters and permits the utilization of mixed-frequency data within Markov-switching models; (ii) tree-based MIDAS regressor [32], offering greater flexibility in sampling high-frequency lagged regressors compared to conventional MIDAS models characterized by tightly parameterized lag functions; (iii) TF-MIDAS [33], which,

instead of relying on the conventional distributed lag polynomial model, employs a transfer function; (iv) GARCH-MIDAS [34], a variant of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model designed for handling volatile time series and heteroskedasticity, that is, a violation of the assumptions for linear regression modeling, featuring a MIDAS augmentation to incorporate high-frequency data; and (v) U-MIDAS [35], an extension of the traditional model with unrestricted lag polynomials. Although all MIDAS variations expand their predictive capacity, none specifically make the model especially capable of dealing with long-term forecasting.

2.1.2 Neural network-based approaches

Although we have already mentioned different statistical solutions for dealing with mixed-frequency time series, notably MIDAS, they are generally based on approaches autoregressive. Aiming at approaches with greater potential for capturing nonlinear correlations between multiple time series, a few solutions based on neural networks have been proposed recently. ANN-(U-)MIDAS [36] uses a neural network with a single hidden layer, and between the input layer and the hidden layer, there is the application of a lag structure similar to that used by MIDAS to align time series of different frequencies. This is a hybrid solution between mixed data sampling and a neural network. In this model, each hidden layer node is obtained by applying a sigmoid transfer function to the inner product between the frequency alignment vector, the hidden layer weights, and the hidden layer bias. The same ANN-(U-)MIDAS framework was applied to a quantile regression neural network (QRNN) in order to increase its ability to explore the heterogeneous nonlinear relationships between variables, resulting in the QRNN-MIDAS model [37].

The TMID-DualAtt model [38] employs the frequency alignment approach to convert high-frequency variables into low-frequency observation vectors, with the attention mechanism selecting the most critical input features for the current prediction index. An encoder–decoder explores these temporal properties using LSTM-based attention.

These models are hybrid approaches and use some frequency alignment strategy, which is problematic, since the data distributions are artificially assumed. The temporal-attribute attention neural network (TAA-NN) model takes a distinct approach, unlike the methods mentioned earlier but similar to the architectures proposed in this paper, effectively addressing this issue. It accomplishes this by employing a sliding window strategy for mixed-frequency data to ascertain the quantity of data fed into the model from each data source. Subsequently, a set of convolutional neural networks with an identical number of filters is employed to extract and augment temporal characteristics from the hidden state of each data source.

Since, we could not find reproducible public implementations of these models, they were not used in the benchmarking in Sect. 4, which is future work.

2.1.3 N-HiTS

Neural hierarchical interpolation for time series forecasting (N-HiTS) [39] is an evolution of the N-BEATS [40] framework, introducing novel elements to enhance long-horizon forecasting capabilities. It employs a multilayered architecture comprising stacks and blocks, each featuring a multilayer perceptron (MLP). The methodology leverages multirate signal sampling, a distinctive feature, where a MaxPool layer with kernel size k' is applied to the input signal $y_{t-L:t}$ before feeding it to each block. This enables each block to focus on specific input scales, facilitating the analysis of high-frequency and low-frequency components.

Following multirate sampling, each block conducts nonlinear regression to obtain forward (Θ'_f) and backward (Θ'_b) interpolation coefficients. The coefficients are then employed to synthesize backcast ($\tilde{y'}_{t-L:t}$) and forecast ($\hat{y'}_{t+1:t+H}$) outputs. Hierarchical interpolation is a key feature, distributing expressiveness ratios across blocks and allowing for a structured hierarchy of interpolation granularity.

Incorporating the neural basis approximation theorem, N-HiTS ensures effective approximation of infinite/dense horizons. The theorem states that given certain conditions on the multi-resolution functions V_w and smooth variations in the forecast relationships, N-HiTS can accurately approximate the forecast mapping $Y(\cdot|y_{t-L:t})$.

Furthermore, N-HiTS introduces the concept of temporal interpolation, controlling the number of parameters per unit of output time through an expressiveness ratio r'. The hierarchical interpolation principle aligns with multirate sampling, with blocks closer to the input having smaller r' and larger k', facilitating more aggressive interpolation and focusing on larger-scale components. In contrast to MIDAS, a well-explored method in the existing literature, N-HiTS stands out as a novel approach. Its limitations have yet to become as apparent, given its recent application to a limited set of scenarios and datasets.

2.2 Multimodal learning

Unlike mixed-frequency time series analysis, which deals with multivariate time series with data collected at different time frequencies, multimodal learning [41] is a specific area of Machine Learning that aims to solve the intricate task of making the same model aiming to fulfill a certain

task learn from data from distinct natures belonging to the same domain. This approach has garnered substantial attention in recent years, driven by its versatility in real-world applications. From image captioning to video sentiment analysis, multimodal learning is valuable in seamlessly integrating information from different sources.

At its core, multimodal learning seeks to harness the joint information within data modalities to improve predictive and generative modeling. In mathematical terms, let us consider a multimodal dataset represented as $\{(X_1, X_2, \ldots, X_M), Y\}$, where $X_1, X_2, \ldots, X_M$ correspond to different modalities (e.g., images, text, audio) and Y represents the target variable. The fundamental challenge lies in effectively encoding the dependencies and relationships among these modalities.

One classical approach to multimodal learning involves the creation of a fused representation [42], which often entails transforming each modality into a shared feature space. This transformation can be achieved through techniques such as canonical correlation analysis or deep neural networks. Mathematically, this transformation can be expressed as $f_1(X_1), f_2(X_2), \ldots, f_M(X_M)$, where $f_i(\cdot)$ maps the data from modality i to a common space. One can employ conventional machine learning models or neural networks to make predictions in this shared latent space.

Furthermore, multimodal learning models can be divided into early and late fusion approaches [43]. Early fusion combines information from all modalities at the input level, where the fused representation can be expressed as $f(X_1, X_2, \ldots, X_M)$. A common formulation for early fusion involves weighted concatenation, where f is a function that incorporates these weights. Mathematically, it can be expressed as:

$$f(X_1, X_2, \ldots, X_M) = \sum_{i=0}^{M} w_i \cdot X_i, \tag{2}$$

where w_i represents the weights assigned to each modality, which can be learned during training.

Conversely, late fusion combines predictions made separately for each modality. In this case, the predictions from different modalities can be treated as additional features and combined using an ensemble approach [44, 45] or another model [46]. Mathematically, late fusion can be expressed as:

$$Y = g[h_1(X_1), h_2(X_2), \ldots, h_M(X_M)], \tag{3}$$

where $h_i(\cdot)$ represents the predictive model for modality i, and $g(\cdot)$ combines these predictions to obtain the final output.

An inherent challenge in multimodal learning is the heterogeneity of data types and potential mismatches in modalities' scales and dimensions. To mitigate these issues, usual strategies include employing specialized architectures, such as Siamese networks [47], cross-modal attention mechanisms [48], or graph-based fusion [49], which adapt to the particularities of multimodal data.

2.3 Multi-output forecasting

Werneck et al. [6] propose two validation paths that aim to be more realistic for multi-output forecasting than those traditionally found in the literature, while avoiding mixing different prediction confidences for long forecasts. The first approach is called first prediction, and its forecast considers the entire output of the first predicted window followed by only the last point of the following predictions. The second approach is called N-th Day and uses the same multi-output sliding window as the previous approach but only considers the last point of each output in its forecast for performance evaluation.

In Fig. 1, the ultimate forecast generated by the identical model is illustrated, employing the first prediction and N-th Day configurations, utilizing a sliding window encompassing four input points to forecast the subsequent three points. Within this visual representation, the yellow squares denote the input, the green ones signify the initial prediction, the blue squares represent the second prediction, and the red ones correspond to the third. To elaborate, the first prediction encompasses the initial complete forecast and the concluding square of subsequent predictions, whereas the N-th Day solely relies on the ultimate day of all predictions, thus constituting the most demanding prediction.

3 Proposed approach

Leopoldo et al. [50] presented an architecture for mixed-frequency short-term forecasting based on RNNs. In this paper, we detail, expand, and validate their proposal in a

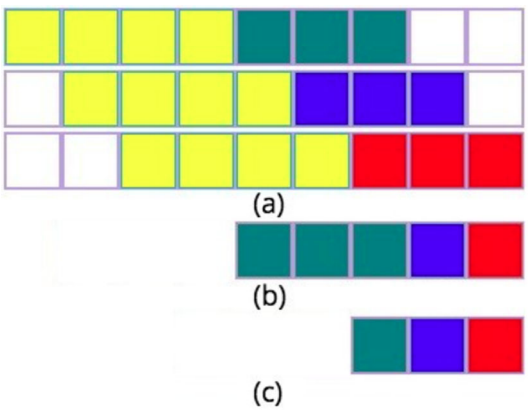


Fig. 1 Forecast setup: **a** multi-output forecast with a sliding window using four points to predict three; **b** first prediction; **c** N-th day

different scenario, for medium- and long-term forecasting. This paper also explores variations in the topologies of these architectures, using attention mechanisms and convolutional layers in multimodal neural networks, in addition to experimenting with both LSTM and GRU in their recurrent layers, as shown in Sect. 4.

3.1 Multimodal recurrent neural network

The proposed method is grounded on the fundamental notion that mixed-frequency time series inherently exhibit diverse underlying probability distributions akin to distinct modalities. To effectively harness this concept, we introduce a deep neural network architecture that leverages early fusion through a dedicated combination layer, facilitating the intricacies of multimodal learning. Within this framework, each HF time series, characterized by its potentially unique periodicity, is treated as an independent modality, while the entirety of features constituting the low-frequency (LF) series is treated collectively as a single modality.

Within this architectural construct, HF and LF time series enter the deep neural network via distinct initial input layers, traversing a hierarchy of stacked recurrent neural networks (RNNs). The HF and LF time series can be expressed, respectively, as:

$$Y_t^{HF_i} = f(X_t^{HF_i}, h_{t-1}), \tag{4}$$

$$Y_t^{LF} = f(X_t^{LF}, h_{t-1}), \tag{5}$$

, where Y_t is the output of the sequence in the time step t, X_t is the entry into the time step t, h_{t-1} is the hidden state of the last RNN in the stack at time $t-1$.

All these time series are aggregated, and the output of the aggregation layer is shown in Eq. 6:

$$X_t^{concat} = \mathsf{Aggregation}\left(Y_t^{HF_1}, Y_t^{HF_2}, \ldots, Y_t^{HF_N}, Y_t^{LF}\right) \tag{6}$$

These concatenated features are processed through dense layers, yielding the final forecasting outcome. In supervised training, targets are associated with the input data of the LF series during the training phase. During the aggregation step, the data from each HF series are uniformly distributed, preserving their sequential order concerning LF data. Notably, applying varying granularities in each HF pipeline is consistent with the correlation between training samples and targets. This approach also ensures that the same aggregation ratio between HF and LF data employed during training is upheld during the testing phase.

3.2 Architectural variations

Three distinct architectural variations are proposed to accommodate different data scenarios:

- *Mixed-Frequency Data* In cases, where HF data exhibit many frequencies, each HF series constitutes a unique modality. As shown in Fig. 2, this model accommodates diverse granularities among HF data and preserves the integrity of the training samples' correlations and targets. Each high-frequency time series $X_t^{HF_i}$ is processed by its respective function f_{HF_i} to produce the output $Y_t^{HF_i}$ (Eq. 7) and the low-frequency time series X_t^{LF} is processed by its function f_{LF} to produce the output Y_t^{LF} (Eq. 8). So, the outputs from all high-frequency and low-frequency series are concatenated into a single feature vector X_t^{concat} (Eq. 9) and the concatenated vector X_t^{concat} is passed through dense layers to produce the final prediction $\hat{Y}_t$ (Eq. 10).

$$Y_t^{HF_i} = f_{HF_i}(X_t^{HF_i}, h_{t-1}^{HF_i}), \quad i = 1, 2, \ldots, N \tag{7}$$

$$Y_t^{LF} = f_{LF}(X_t^{LF}, h_{t-1}^{LF}) \tag{8}$$

$$X_t^{concat} = \mathsf{Aggregation}\left(Y_t^{HF_1}, Y_t^{HF_2}, \ldots, Y_t^{HF_N}, Y_t^{LF}\right) \tag{9}$$

$$\hat{Y}_t = f_{dense}(X_t^{concat}) \tag{10}$$

- *Uniform-Frequency Data* When all HF data share a common frequency, the architecture simplifies to encompass only two modalities: LF data and HF data. This streamlined approach mitigates computational complexities associated with excessive high-frequency data, making the training process more computationally efficient. Notably, HF data can be downsampled to a granularity exceeding that of LF data, while preserving essential temporal characteristics. Figure 3 shows this architecture. Firstly the high-frequency data X_t^{HF} is processed by the function f_{HF} to produce the output Y_t^{HF} (Eq. 11), then the high-frequency and low-frequency outputs are concatenated into a single feature vector X_t^{concat} (Eq. 12). Finally, the concatenated vector X_t^{concat} is passed through dense layers to produce the final prediction $\hat{Y}_t$ (Eq. 13).

$$Y_t^{HF} = f_{HF}(X_t^{HF}, h_{t-1}^{HF}) \tag{11}$$

$$X_t^{concat} = \mathsf{Aggregation}(Y_t^{HF}, Y_t^{LF}) \tag{12}$$

$$\hat{Y}_t = f_{dense}(X_t^{concat}) \tag{13}$$

- *HF-Only Architecture* In domains, where LF series features exhibit limited influence or are particularly noisy relative to the target variable, a specialized architecture is proposed (Fig. 4). In this setup, HF data are employed as separate modalities, each potentially characterized by distinct granularities. The training process remains centered on the LF data, with their quantity guiding the amalgamation process within the

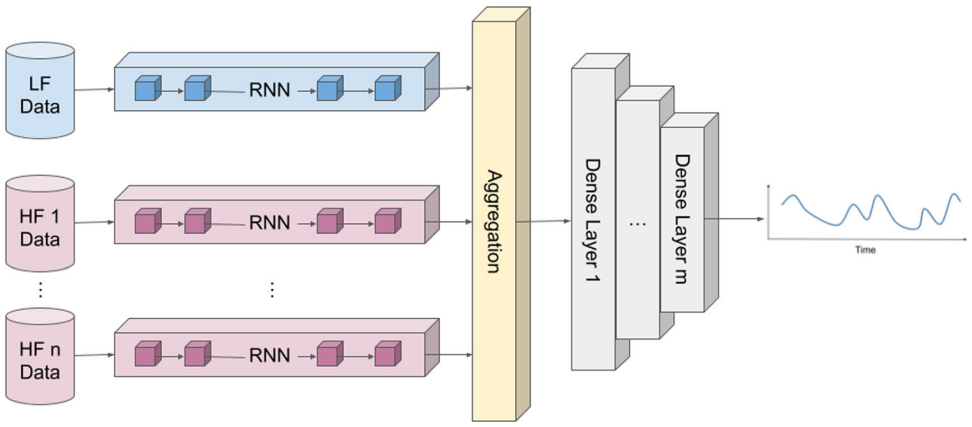


Fig. 2 Mixed-frequency data architecture. The blue color refers to the features with the lowest frequency in the dataset (which in all experiments shown in Sect. 4 have the same frequency as the target) and the pre-fusion layers, where they are processed. The pink color refers to the different HF features, which can have different frequencies and layers that process them. After the fusion of the different modalities (yellow layer), the network flow follows through a series of dense layers (there may be other mechanisms in this part of the topology), which will then make the final prediction (Color figure online)

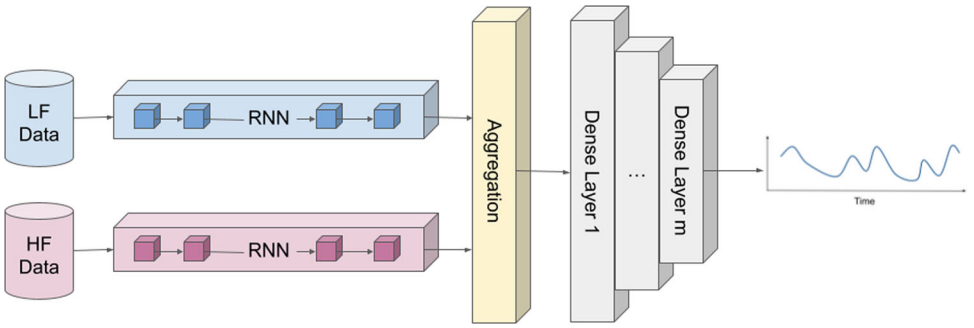


Fig. 3 Uniform-Frequency data. Here, the color schemes and topology details are the same as in Fig. 2 (Color figure online)

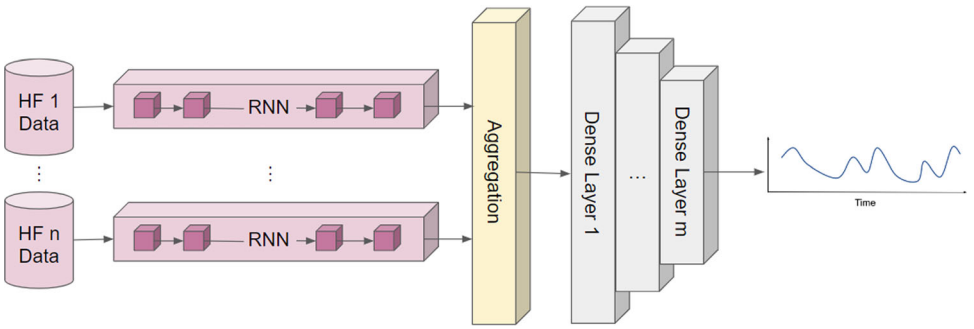


Fig. 4 HF-Only architecture. Here, the color schemes and topology details are the same as in Fig. 2 (Color figure online)

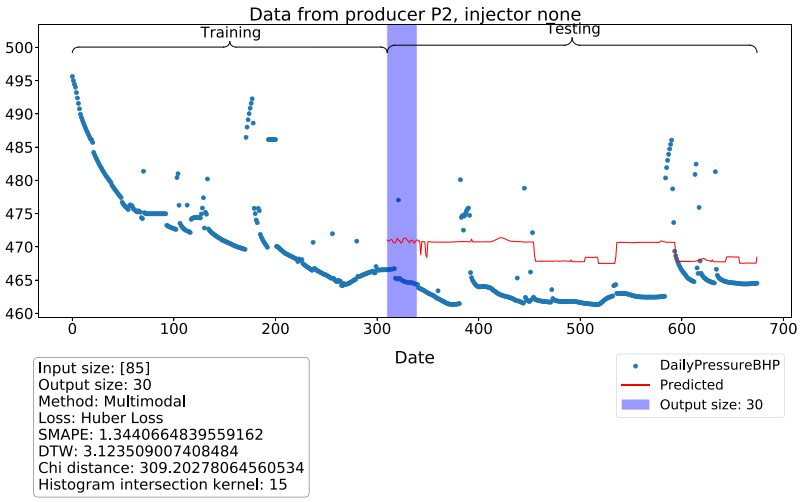
concatenation layer. Each high-frequency time series $X_t^{HF_i}$ is processed by its respective function f_{HF_i} to produce the output $Y_t^{HF_i}$ (Eq. 14), then the outputs from all high-frequency series are concatenated into a single feature vector X_t^{concat} (Eq. 15). Hence the concatenated vector X_t^{concat} is passed through dense layers to produce the final prediction $\hat{Y}_t$ (Eq. 16).

$$Y_t^{HF_i} = f_{HF_i}(X_t^{HF_i}, h_{t-1}^{HF_i}), \quad i = 1, 2, \ldots, N \tag{14}$$

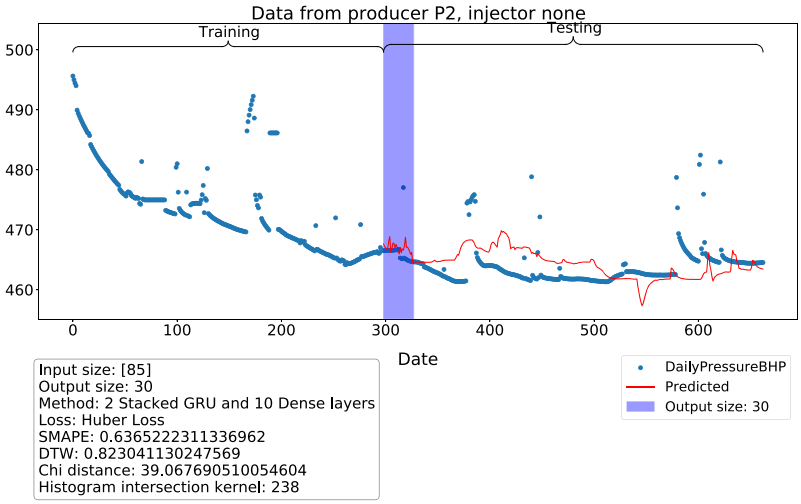
$$X_t^{concat} = \text{Aggregation}\left(Y_t^{HF_1}, Y_t^{HF_2}, \ldots, Y_t^{HF_N}\right) \tag{15}$$

$$\hat{Y}_t = f_{dense}(X_t^{concat}) \tag{16}$$

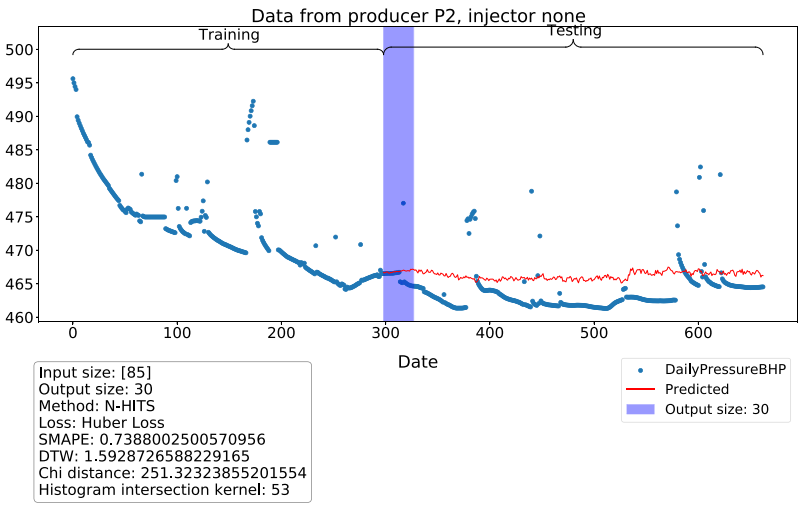
Mathematically, the aggregation operation can be described as the joining of vectors along a specified



(a) Our solution (multi-modal).



(b) Baseline (GRU2).



(c) Literature (N-HiTS).

Fig. 5 BHP short-term forecasting in P2 well

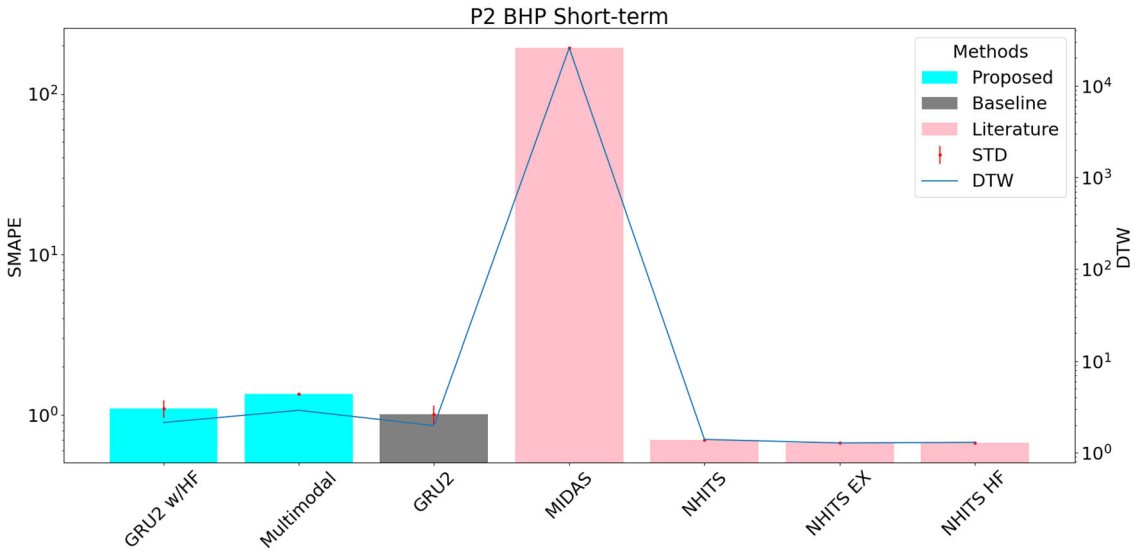


Fig. 6 Benchmarking in BHP short-term forecasting in P2 well

dimension (usually the last one). If each $\mathbf{Y}_t^{HF_i}$ is a vector of dimension d_{HF_i} and $\mathbf{Y}_t^{LF}$ is a vector of dimension d_{LF}, the output $\mathbf{X}_t^{concat}$ will have a total dimension of: $d_{concat} = d_{HF_1} \oplus d_{HF_2} \oplus \ldots \oplus d_{HF_N} \oplus d_{LF}$, where $\oplus$ denotes the concatenation operation.

These architectural variations offer flexibility in adapting to diverse data scenarios, with the ability to prioritize specific data modalities based on their relevance and influence within the context of the forecasting task.

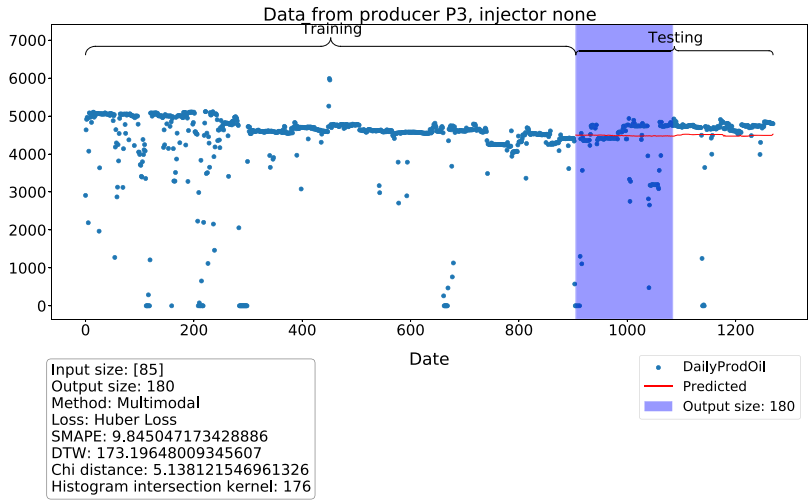
The advantages that this type of solution based on deep nets offers for the problem of mixed-frequency time series forecasting about traditional modeling based on autoregressive models, such as those shown in Sect. 2.1, include the ability of these neural networks to capture nonlinear relationships in data. This happens especially in multivariate time series [51] and in its way of automatically learning the relevant characteristics of the data rather than relying on assumptions about the structure of the data, as in autoregressive models, which can be especially valuable in open-world scenarios, where the composition, distribution, and relevance of features in a time series may change over time.

Solutions based on neural networks are also easier to deal with missing data and irregularities than traditional modeling [52], in addition to potentially having generalization power for different types of time series. In contrast, traditional autoregressive models are generally designed for specific types of time series. Additionally, a specific advantage of the multimodal architecture proposed here is that specific feature extraction mechanisms, such as attention mechanisms, can be applied to each modality before the aggregation layer so that the idiosyncrasies of components with different frequencies of a time series multivariate data can be better used to carry out the prediction.

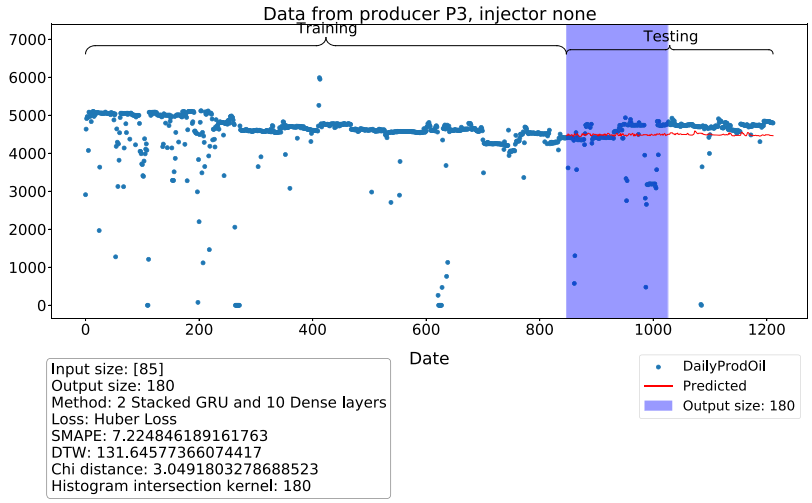
4 Experiments

This section describes the dataset used, the performed experiments, the validation used, and an analysis of the results obtained. The methodology was applied to assess the oil production rate (m^3/day) and BHP (Bottom Hole Pressure) ($kgf/cm^2 \times day$) in a Pre-salt carbonate field known for its challenging predictions due to its complex nature. The reservoir is highly heterogeneous, with fractures, faults, vugs, and karsts. Moreover, with a light oil composition and the presence of CO_2, when the reservoir is subject to further development by altering the initial conditions, such as pressure and temperature, fluids are modified over time, making forecasting even more difficult. We used two other datasets with different characteristics, such as different frequencies, quantities of data, and domains, to increase our insights into results obtained for forecasting in the oil dataset and their meaning about the predictive quality of models for mixed-frequency time series.

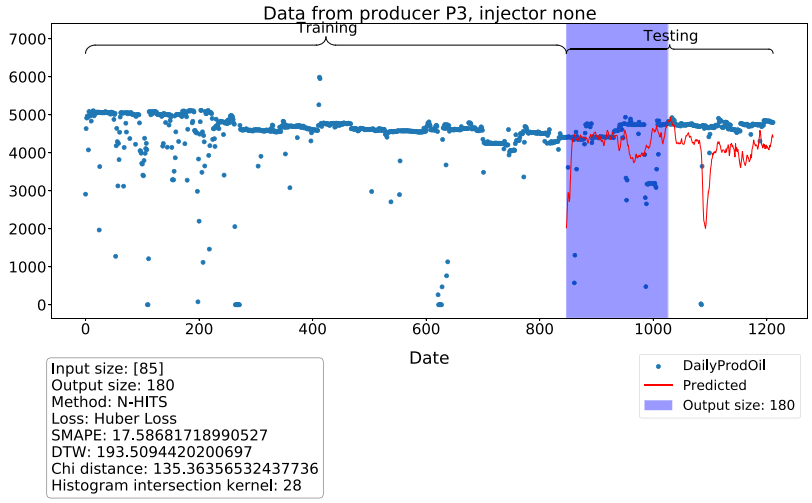
Regarding the graphical display of the results, Figs. 5, 7 and 9 are plots, where the blue points represent the ground truth, the purple vertical bar indicates the size of the prediction window and separates the training data from the test data, and the red line represents the prediction made. At the bottom of these figures, there is a box with metrics obtained from the validation of this model for this dataset. Figures 6, 8, 10, 11 and 12 compare the models' performance in terms of SMAPE for a given dataset and target. Our proposed approaches are represented by blue bars,



(a) Our solution (multi-modal).



(b) Baseline (GRU2).



(c) Literature (N-HiTS).

Fig. 7 Oil production medium-term forecasting in P3 well

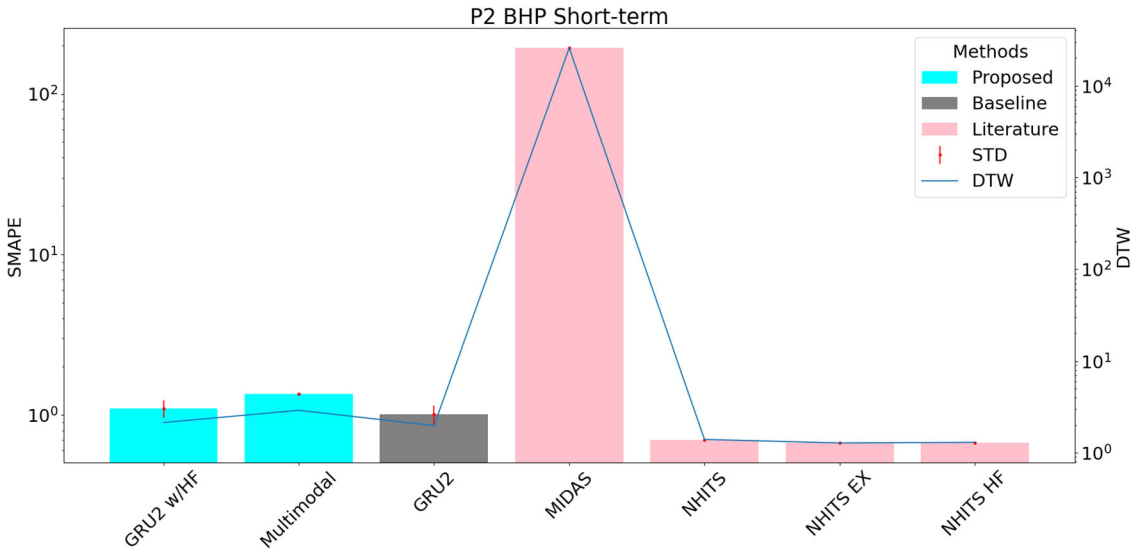


Fig. 8 Benchmarking in oil production medium-term forecasting in P3 well

established approaches in the literature are shown in pink, and the baseline is in gray. Each bar has a vertical red line indicating the size of the SMAPE standard deviation for each model, and a blue line crossing the bars corresponds to their performance in relation to DTW, allowing for visual correlation of the models' performance concerning both metrics. Finally, Tables 3, 4 and 5 and all tables in the Appendix 2 separately compare the performance of all benchmarking approaches in terms of SMAPE and DTW for short-, medium-, and long-term predictions, with the best result for each metric and prediction window highlighted in bold.

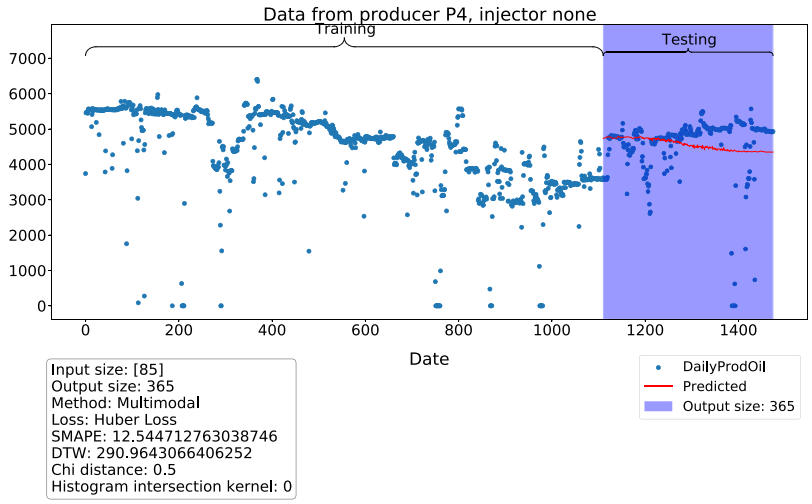
4.1 Datasets

For our experiments, we leveraged a private mixed-frequency time series dataset from a pre-salt oil reservoir and two established nonoil industry benchmarks (PPG-DaLiA and Jena Weather) to assess forecasting capabilities, with all these datasets having mixed-frequency time series. Including non-oil datasets demonstrates our methods' versatility, potentially extending beyond the oil and gas sector and presenting experiments that can be completely reproducible for the sake of the open science, since our oil dataset is private. We could not find a public oil dataset with a mixed-frequency time series.

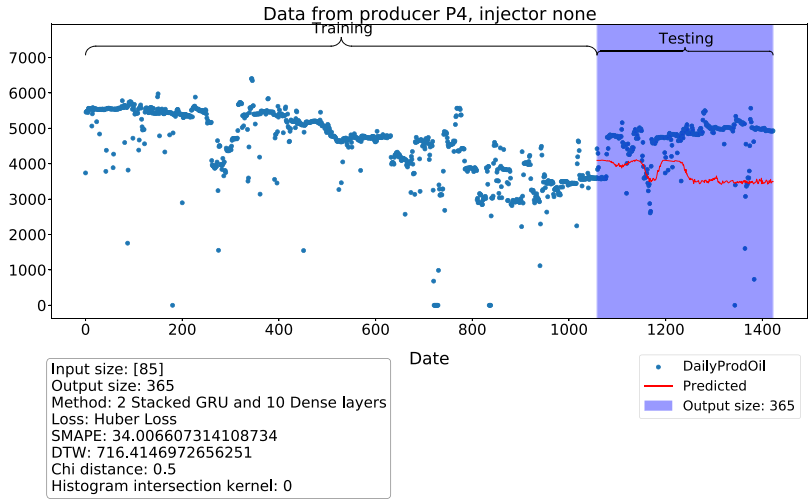
This study used private production data from a Brazilian pre-salt oil field with 1269 days of historical data. The reservoir contains 16 producers and 16 injector wells, divided into nine water and seven WAG injectors. The last year of the historical series was reserved for testing, while all other data were used for training. Four wells (P1–P4) were selected for this validation. All wells have the same low-frequency features: oil, water, and gas fluid productions, BHP, water cut, gas–oil ratio, and gas–liquid ratio values. Both daily oil production and daily BHP were the prediction targets for all wells. HF data vary from well to well in type and quantity but are related to temperature and pressure (P1, P2, and P4: 18 HF features; P3: 10 features).

The PPG-DaLiA dataset [53], which is publicly available, is a valuable resource for PPG-based heart rate estimation. This multimodal dataset encompasses physiological and motion data, meticulously collected from both wrist- and chest-worn devices, involving 15 participants engaged in a diverse range of activities closely resembling real-life scenarios. The dataset includes electrocardiogram data as the ground truth for heart rate estimation (1 Hz). The input features for this dataset consist of blood volume pulse (64 Hz), electrodermal activity (4 Hz), body temperature (4 Hz), and three-axis acceleration (32 Hz).

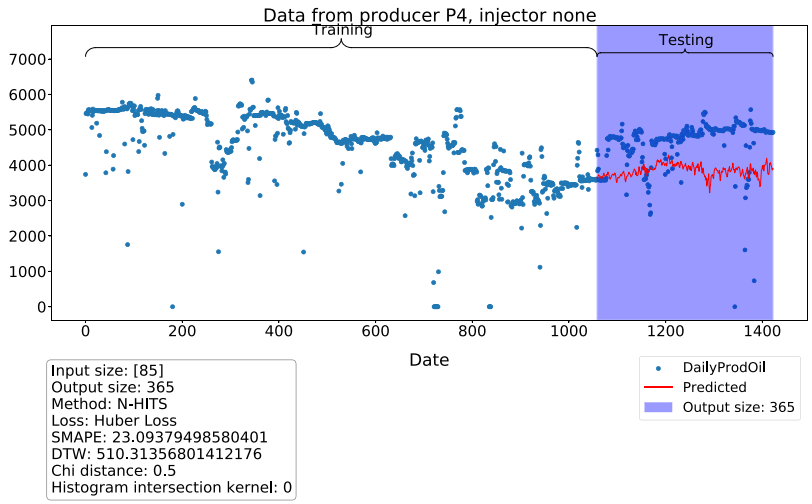
The Jena climate dataset [54] consists of weather time series meticulously recorded at the Weather Station of the Max Planck Institute for Biogeochemistry in Jena, Germany. This comprehensive dataset comprises 14 attributes, including air temperature, atmospheric pressure, humidity, wind direction, and more, which were meticulously logged at 10-minute intervals over several years. This dataset encompasses observations from January 1st, 2009, up to December 31st, 2016, and focuses on temperature measurements at an hourly frequency, provided in both Celsius degrees and Kelvin.



(a) Our solution (multi-modal).



(b) Baseline (GRU2).



(c) Literature (N-HiTS).

Fig. 9 Oil production long-term forecasting in P4 well

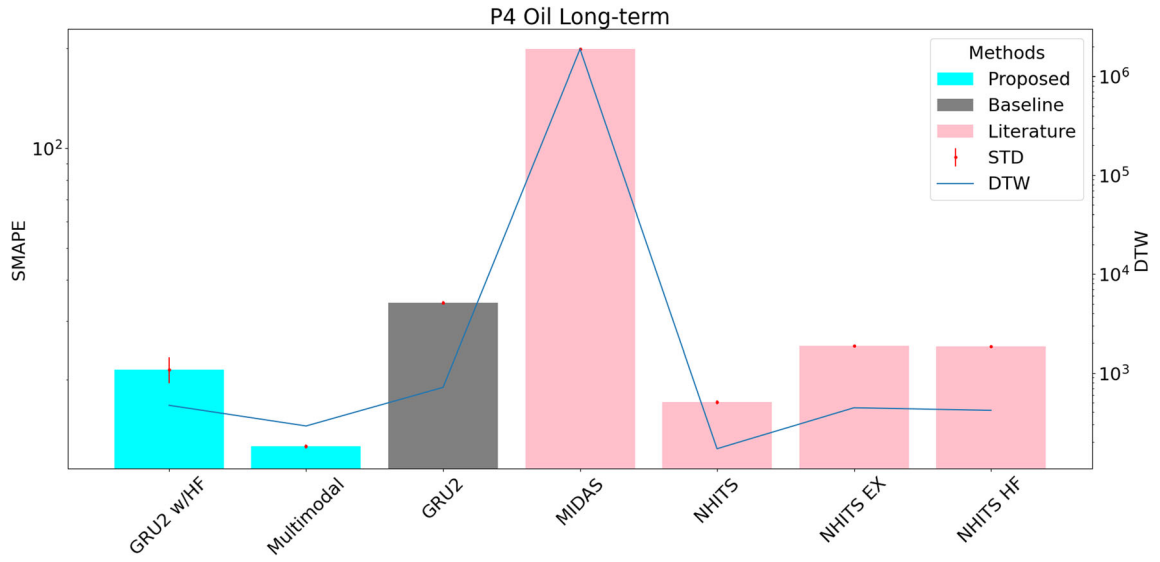


Fig. 10 Benchmarking in oil production long-term forecasting in P4 well

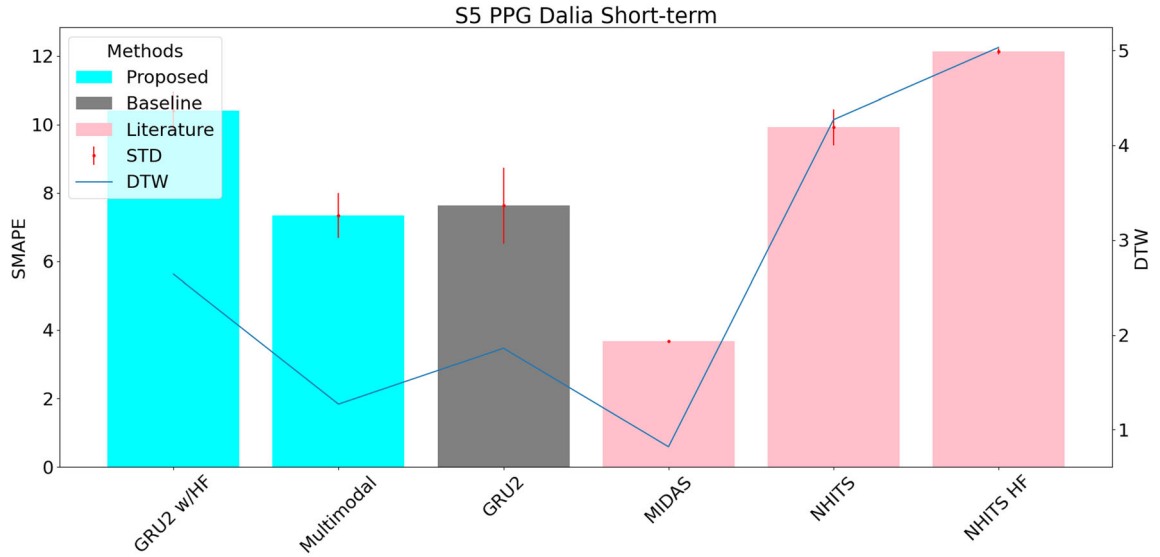


Fig. 11 Benchmarking in short-term forecasting in S5 patient

4.2 Experimental setup

Since, the baseline for this work is Werneck et al. [6] and it uses an output window of 30 days to perform its benchmarking for short-term time series forecasting, we decided to use the same amount of data and add outputs of 180 and 365 days to validate the medium- and long-term prediction capacity to investigate whether the use of HF data tends to become more useful over longer horizons. For the other datasets, we kept the same number of points (30, 180, and 365 for the short, medium, and long-term, respectively), although the granularity of the data in them is different.

Aiming to provide a more precise assessment of the model in short-term forecasts, a multi-output window with sliding steps from the past 85 days was used to predict the following days ahead, whose quantity depends on the output window size, considering the Nth-Day setup. The input window size is also a decision derived from [6]. However, it was not deemed applicable in all cases, with the exceptions listed below in the description of the experiment.

The predictions were validated using the symmetric mean absolute percentage error (SMAPE) metric, which measures the percentage error of the predicted values according to Eq. 17.

$$\text{SMAPE}(X,h) = \frac{100}{m}\sum_{i=1}^{m}\frac{|h(x_i) - y^i|}{\frac{(|y^i|+|h(x_i)|)}{2}}, \tag{17}$$

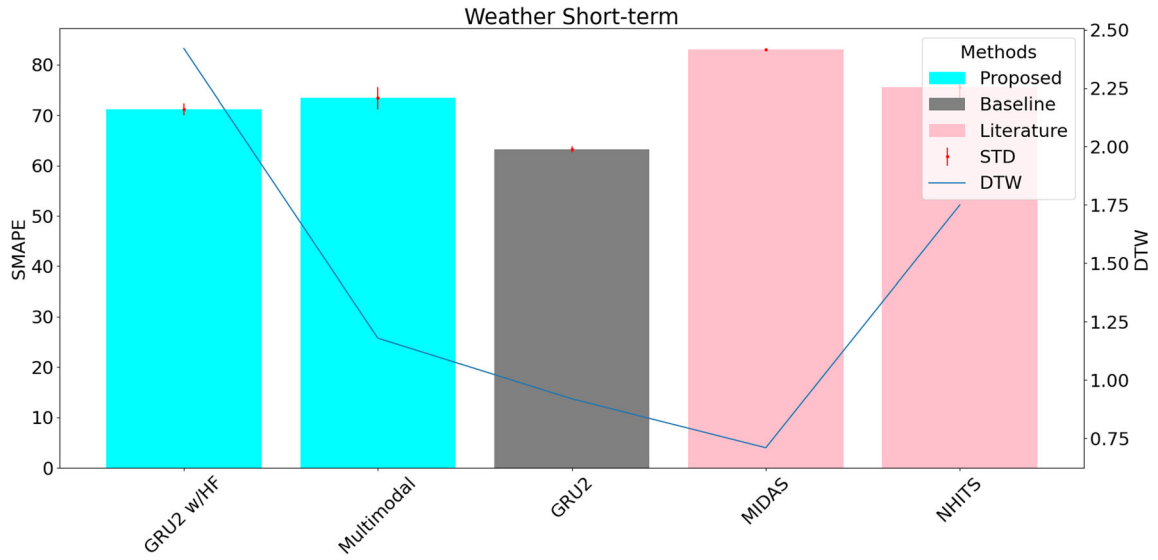


Fig. 12 Benchmarking short-term forecasting in Jena Weather dataset (temperature in Celsius degree)

Table 1 Comparison of the multimodal models' topologies in Mixed-Frequency Data architecture with GRU2 model (improvement about SMAPE in %)—P1 and P2 wells of the pre-salt dataset (input: 85 days; output: 30 days)

	QO		BHP	
	P1	P2	P1	P2
Multi LSTM	− 15.70	− 4.80	**14.37**	**21.74**
Multi GRU	− 9.47	12.36	− 52.38	− 58.70
Multi LSTM CNN	− 69.66	− 4.16	**14.37**	17.39
Multi GRU CNN	− 15.13	9.66	− 18.79	− 35.87
Multi LSTM Att	**2.83**	**14.30**	− 14.04	4.35
Multi GRU Att	− 19.38	1.39	**14.37**	**21.74**

Bold means the best result obtained for benchmarking a specific scenario

Table 2 Comparison of the multimodal architectural variations employing Multi LSTM topology with GRU2 model (improvement about SMAPE in %)—P1 and P2 wells of the pre-salt dataset (input: 85 days; output: 30 days)

	QO		BHP	
	P1	P2	P1	P2
MFD	**2.82**	**18.76**	8.29	**29.03**
UFD	− 30.05	17.44	**17.68**	0.00
HF-Only	− 14.38	− 9.88	− 119.61	− 50.54

Bold means the best result obtained for benchmarking a specific scenario

where $h(x)$ is the forecast value, y is the actual value, and m is the total number of observations made. Since, SMAPE is percentage-based, it is also scale-independent and, therefore, can be used to compare the accuracy of a predictive model across different datasets. It can also be used here to compare forecasting approaches across different wells. The lower the SMAPE value of a forecast, the better its accuracy.

Since, SMAPE and other error-based metrics do not necessarily capture the ability of the predicted time series to follow the ground truth trend, an additional measure of similarity between the time series was used for validation: dynamic time warping (DTW). This method measures the similarity between two sequences, especially time series data, by optimally aligning them. It is typically applied when sequences have different lengths or when there is a phase shift between corresponding elements.

Supposing we have two sequences: A of length n and B of length m, where $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_m)$. DTW constructs a $n \times m$ matrix, often denoted as D, where each element $D(i, j)$ represents the "cumulative distance" or "accumulated cost" to align a_i with b_j. This matrix is computed as follows:

1. Initialize the matrix D as follows: $D(0,0) = 0$ and $D(i,0) = \infty$ for $i > 0$ and $D(0,j) = \infty$ for $j > 0$.
2. Compute the cumulative distances for the remaining cells in the matrix. This is done iteratively using dynamic programming: $D(i,j) = \mathsf{dist}(a_i, b_j) + \mathsf{min}[D(i-1,j), D(i,j-1), D(i-1,j-1)]$, where $\mathsf{dist}(a_i, b_j)$ is the distance (e.g., Euclidean distance) between a_i and b_j.
3. The final DTW distance between sequences A and B is $D(n, m)$, which represents the minimum cumulative distance after alignment.

Table 3 Percentage gain in relation to baseline in oil dataset

		Model	Short	Medium	Long
SMAPE	QO	GRU2 w/HF (ours)	**0.93**	**13.76**	25.91
		Multimodal (ours)	− 22.56	− 10.49	**40.00**
		MIDAS	− 982.48	− 600.87	− 765.65
		N-HiTS	− 63.50	− 9.62	− 8.80
		N-HiTS w/Ex	− 43.66	− 5.58	− 9.41
		N-HiTS w/Ex + HF	− 45.12	− 6.16	− 8.14
	BHP	GRU2 w/HF (ours)	17.91	28.64	38.60
		Multimodal (ours)	− 92.04	35.56	43.08
		MIDAS	− 5291.04	− 1525.78	− 1928.85
		N-HiTS	35.82	67.54	78.66
		N-HiTS w/Ex	32.84	**71.12**	78.13
		N-HiTS w/Ex + HF	**37.31**	66.11	**81.29**
DTW	QO	GRU2 w/HF (ours)	**4.78**	23.05	26.68
		Multimodal (ours)	− 28.33	0.58	40.75
		MIDAS	− 394110.41	− 309623.54	− 2359567.65
		N-HiTS	− 19.72	**50.09**	**51.37**
		N-HiTS w/Ex	− 28.10	32.44	30.81
		N-HiTS w/Ex + HF	− 26.05	32.50	33.80
	BHP	GRU2 w/HF (ours)	31.38	44.68	41.66
		Multimodal (ours)	− 103.85	49.67	45.64
		MIDAS	− 207570.65	− 20580.45	− 829230.01
		N-HiTS	**63.36**	85.66	**88.66**
		N-HiTS w/Ex	56.48	**85.73**	85.89
		N-HiTS w/Ex + HF	56.68	83.56	88.13

Bold means the best result obtained for benchmarking a specific scenario

Table 4 Percentage gain in relation to baseline in PPG-DaLiA

	Model	Short	Medium	Long
SMAPE	GRU2 w/HF (ours)	− 53.93	− 6.25	− 6.89
	Multimodal (ours)	27.53	− 7.81	− 8.83
	MIDAS	**84.18**	**94.11**	**95.17**
	N-HiTS	− 19.76	21.86	26.11
	N-HiTS w/HF	− 32.30	14.11	19.06
DTW	GRU2 w/HF (ours)	− 20.81	23.59	14.63
	Multimodal (ours)	63.66	46.25	40.08
	MIDAS	**91.30**	**97.03**	**97.19**
	N-HiTS	− 9.63	41.72	42.76
	N-HiTS w/HF	− 5.59	45.00	46.69

Bold means the best result obtained for benchmarking a specific scenario

DTW effectively searches for the alignment path through this matrix, which connects the starting point (0, 0) to the endpoint (n, m), while minimizing the total accumulated cost. This alignment path characterizes the optimal warping between the two sequences, capturing their similarities and accounting for local variations and phase shifts.

Table 5 Percentage gain in relation to baseline in Jena Weather

	Model	Short	Medium	Long
SMAPE	GRU2 w/HF (ours)	− 8.69	− 5.23	0.56
	Multimodal (ours)	3.96	6.83	3.71
	MIDAS	− 36.70	16.39	13.92
	N-HiTS	**8.08**	**27.23**	**25.46**
DTW	GRU2 w/HF (ours)	− 2.90	− 76.19	− 12.75
	Multimodal (ours)	70.60	**32.74**	54.67
	MIDAS	**76.19**	− 89.88	**66.01**
	N-HiTS	46.58	− 34.52	35.69

Bold means the best result obtained for benchmarking a specific scenario

4.3 Models' architectures

This subsection describes the models used in this benchmarking and discusses how to set their hyper-parameters.

4.3.1 GRU2

The model employed was the gated recurrent unit (GRU)-based architecture GRU2, introduced by [6], and validated

in the same oil dataset of this study in another historical period and additional prediction datasets. All experiments were run using 100 epochs, with an early stopping method for those networks that did not improve the validation loss after ten epochs. The architecture of GRU2 comprises two layers of GRUs, each consisting of 128 neurons and return sequences as True/False, respectively. Additionally, there are ten Dense layers with neuron counts of 128/128/64/64/64/32/32/32/30/1, respectively. The optimization uses the Adam optimizer with a learning rate of 0.001.

4.3.2 GRU2 with HF

Additionally, high-frequency data models were used in the benchmark: data whose granularity is greater than the daily periodicity. These features are related to temperature and pressure, and the number of such variables and the magnitude and frequency of their measurements depend on the sensors used in each well. To use them alongside the daily data, their granularity was reduced to the daily one through aggregation performed by the median or autoencoders (AE). When using AE, the encoder and the decoder have only one hidden layer (Mean Absolute Error as loss, Adam as optimizer, and ten epochs in training). Only the original series without any augmentation was used in training models that used HF data, and only 1269 days of the historical series have HF features. Models that used aggregation by AE were, on average, 7.78% and 13.33% better than models that used median forecasting of oil production and BHP, respectively, considering outputs of 30, 180, and 365 days. In the following experiments, we will only report the results of models with HF data interpolated by AE.

4.3.3 Proposed multimodal architectures

Various topologies were tested for the multimodal strategy, as seen in Table 6. The first line of each model indicates the low-frequency branch, the second line the high-frequency branches, and the last line the model topology after the concatenation layer. The mixed-frequency data architecture has a low-frequency branch and as many high-frequency branches as there are non-minimum frequencies in the mixed-frequency time series. The Uniform Frequency Data architecture will only have one high-frequency branch, and the HF-Only architecture will not have a low-frequency branch.

Regarding the application of these models in the pre-salt dataset, since the periodicity of these HF data is high, they were aggregated through AE for the hourly frequency, even though the multimodal architecture proposed in Sect. 3.1 can deal with series of any frequencies. This option was adopted to prevent the models from becoming time-consuming.

The choice of topology to be used in the other experiments was made based on a validation of all topologies shown in Table 6 (Appendix 1) with the mixed-frequency architecture of wells P1 and P2 of the pre-salt dataset for the QO (oil rate) and BHP targets. The SMAPE of these topologies was compared with that obtained by GRU2 in these same wells for short-term forecasting, using the same setup shown in Sect. 4.3.1. Table 1 shows the comparison of improvement about GRU's SMAPE and, as can be seen, no model was specifically better for all scenarios, but the Multi LSTM Att topology had the best average improvement for QO (8.57%). Multi LSTM and Multi GRU Att outperformed the other topologies for BHP forecasting, with an average improvement of 18.06%, becoming a real candidate to be used in the other experiments.

The Multi LSTM Att and Multi GRU Att architectures were then validated in all architecture variations for the same setup to predict QO and BHP, respectively. The results of this experiment are shown in Table 2. As can be seen, only the mixed-frequency data architecture improved compared to GRU2 in all scenarios, also having the best performance in three of four scenarios.

4.3.4 MIDAS

The MIDAS fit parameters were: bounds = $(-\infty, \infty)$, optimization method = Trust Region Reflective, stopping tolerance for optimization (*ftol*) = stopping tolerance for optimization in parameters (*xtol*) = stopping tolerance for optimization on derivatives (*gtol*) = 10^{-9}, loss function = linear, scale for the objective function (*f_scale*) = 1, maximum number of objective function evaluations (*max_nfev*) = 500.

For MIDAS to operate in a multi-output manner, the size of the weight matrix and the loss function were adjusted to accommodate output size, with the loss function receiving additional terms according to the size of the output. The different predictions were concatenated and evaluated using the N-th Day approach. Owing to the inherent limitation of MIDAS in configuring lags effectively when confronted with substantial disparities in the dimensionality between the input and output, the input size was equated to the output size in experiments involving MIDAS for medium and long-term forecasting, using the same setup as GRU2 for short-term forecasting.

In our investigation, we observed that all Python implementations of MIDAS utilized forthcoming data from exogenous variables for time series forecasting. However, based on our interpretation, this approach may diminish the forecasting task to a mere regression. Therefore, in our MIDAS implementation, we deliberately avoided

employing any information from the dataset that followed the training set.

4.3.5 N-HiTS

Even when using interpolations of the same endogenous time series internally and operating them with different frequencies in their components, N-HiTS cannot solve the problem of predicting mixed-frequency time series, since the model input will necessarily be a time series of the same frequency. However, N-HiTS has proven to be an efficient model for long-term forecasting and a state-of-the-art solution for general time series forecasting, with good training and inference performance. Therefore, we used it in these experiments, and its comparison with other solutions that effectively predict time series with different granularities is interesting to validate in which scenarios the use of high-frequency data is necessary.

Since, N-HiTS cannot handle mixed-frequency time series as input, three different types of inputs were offered to it in the following experiments: (i) only the endogenous variable, (ii) the endogenous variable, and the exogenous variables with the same target frequency, and (iii) the endogenous variable and all exogenous variables, aggregated to the same target frequency. The parameters used are the default ones from the neuralforecast library.[1]

4.4 Experimental results

This subsection compares the approaches proposed in this work and those in the literature for short-, medium-, and long-term forecasting for the three datasets already mentioned.

4.4.1 Oil dataset

This subsection presents the results of the experiments for the oil dataset. Figure 5 shows the plots of the multimodal approach, baseline (GRU2), and N-HiTS for BHP short-term forecasting in well P2. Figure 6 follows with a benchmarking comparison of all models for the same scenario. Figure 7 displays the plots for the selected models in medium-term oil forecasting in well P3, while Figure 8 provides the benchmarking comparison for this scenario. well P4, and Figure 10 shows the benchmarking results for this case. Table 3 summarizes the models' performance based on SMAPE and DTW improvements. The complete benchmark results are shown in the Appendix 2.

Regarding oil production forecasting, in the short-term, the GRU2 with HF data approach was prominent and achieved better performance in three of the four scenarios (wells P1, P2, and P3), with the multimodal approach being the best for well P4. GRU2 with HF also stood out in the medium-term, having the best SMAPE for two scenarios (wells P1 and P2), while N-HiTS had the best DTW for two scenarios (wells P1 and P4). Considering the long-term scenario, the multimodal solution stood out with the best SMAPE for two scenarios (wells P1 and P4). Except for the best DTW for well P4 (N-HiTS), all other models with better long-term performance used HF data.

In BHP's short-term forecasting, the N-HiTS family performed better in three of the four scenarios (wells P1, P2, and P4). In the medium-term, N-HiTS using exogenous variables had better SMAPE in two wells (P2 and P4) and better DTW in three wells (P1, P2, and P3), with traditional N-HiTS obtaining the best DTW in well P4, so the N-HiTS family was also prominent for this output window. In the long-term, there was no outstandingly superior model, with the baseline, our strategies, and the N-HiTS family obtaining better performance in at least one scenario each. MIDAS did not have any relevant results for any of these predictions.

4.4.2 PPG-DaLiA dataset

In this dataset, MIDAS was prominent and stood out in 13 of the 15 scenarios, both about SMAPE and DTW. The only models that outperformed in any scenario other than MIDAS were GRU2 (baseline) and the multimodal approach (for patients S11 and S14). Figure 11 compares all models for patient S5 in short-term forecasting, which presents similar behavior to most others. Table 4 compares the models about SMAPE and DTW improvements.

4.4.3 Jena Weather dataset

For target T (deg C), GRU2 outperformed all models for all output windows. For the target Tpot (K), MIDAS had better performance in the medium and long-term forecasts, with N-HiTS having the best average percentage gain about the baseline. Figure 12 shows a comparison of all models in the short-term for target T (deg C). Table 5 compares the models to SMAPE and DTW improvements.

4.5 Discussion

In light of our experimental findings, a discernible conclusion emerges, underscoring the advantageous performance of MIDAS in scenarios characterized by short intervals between points in the time series. This is notably evident in the PPG-DaLiA dataset, where the target

[1] https://nixtla.github.io/neuralforecast/models.nhits.html, accessed October 26th, 2023.

frequency is set at 1 Hz, and the range of the HF variables varies between 4 and 64 Hz. It becomes apparent, however, that the efficacy of MIDAS experiences a noteworthy decline when confronted with the Jena Weather dataset. In this particular dataset, the frequency of HF variables extends to 10 min, and the target is measured at hourly intervals. In the oil dataset featuring a daily target, MIDAS lags behind its counterparts across all scenarios. This discrepancy may be attributed to the handcrafted lag structure aligning HF data with low-frequency (LF) data, where noises in alignment accumulate with greater distance between LF variables, particularly in long-term forecasting.

In contrast to MIDAS, our proposed models, including GRU2 with HF data and the multimodal architecture, alongside the baseline model (GRU2) and N-HiTS, consistently demonstrated robust performances across the benchmarking scenarios. Both our multimodal architecture and the utilization of GRU2 with interpolated HF data exhibited competitiveness, not only with the baseline model but also with solutions established in the literature. Notably, no singular model acted as a comprehensive solution, as each exhibited superior performance in specific scenarios. However, a noteworthy observation emerged, suggesting that incorporating HF data into models yields increased advantages as the output window extends. This assertion is substantiated by the results obtained in our models, particularly in the prediction of daily oil production across all wells, daily BHP prediction for well P2, as well as patients S2, S6, S9, and S11 in the PPG-DaLiA dataset, and for the target Tpot (K) in the Jena Weather dataset. Similarly, for N-HiTS, this pattern is evident in the BHP prediction for wells P2 and P4 and patients S2, S4, and S8 of PPG-DaLiA.

In the oil dataset, which initially inspired the introduction of the multimodal architecture and the utilization of GRU2 with HF data, our proposed models showcased commendable performances in the benchmarking. Specifically, for oil production forecasting, the proposed solutions consistently outperformed alternative models in all short-term scenarios. Furthermore, in the medium term, our proposed models achieved superior SMAPE for half of the wells, and in the long-term, they exhibited better SMAPE for all wells. Additionally, our models demonstrated enhanced DTW for one well in both medium and long-term forecasting, further establishing their efficacy in capturing the underlying patterns of the oil dataset.

As illustrated by the comprehensive findings shown in the Appendix 2, while certain models exhibited superiority across particular prediction horizons, it's evident that the performance of each model is significantly influenced by the unique attributes of low- and high-frequency time series data and their interrelations. Hence, the development of a unified system that integrates the diverse models introduced in this study, coupled with a model selection algorithm, emerges as a compelling avenue for future exploration. Such an integrated approach holds promise for mitigating the inherent sensitivity to data characteristics and optimizing forecasting accuracy across various temporal horizons

5 Conclusion and future work

The primary objective of this study was to enhance the efficacy of time series forecasting across short-, medium-, and long-term horizons, leveraging high-frequency data-, particularly in the context of fluid rates and bottom-hole pressure for hydrocarbon reservoirs—through data-driven methodologies. Various multimodal architectures were introduced, conceptualizing each distinct frequency of the data as a separate modality within the neural networks. Rigorous experimentation was conducted using proprietary data from Brazilian pre-salt wells and publicly available datasets in the cardiological and climate fields. The main advantage of these architectures in relation to traditional mixed-frequency sampling techniques is in their treatment of HF data, which are not aligned according to lag structures based on artificial assumptions of time series distributions since, in their topologies, the time series of the same frequency are treated as distinct modalities, with the extraction of patterns and characteristics implicit to them occurring regardless of each other.

The benchmarking framework encompassed our proposed solutions (the GRU2 recurrent neural network) known for their promising results in short-term forecasting, the MIDAS technique for mixed-frequency time series, and the state-of-the-art data-driven solution, N-HiTS. The experiments' results underscored the proposed solutions' competitiveness, particularly in scenarios characterized by lower data frequency and extended prediction windows, positioning them favorably within the contemporary landscape of forecasting methodologies, as demonstrated by their performance in long-term oil forecasting with data from Brazilian pre-salt fields.

Our multimodal approach not only provides practical advancements in time series forecasting but also offers theoretical insights that could deepen our understanding of

data dynamics. By treating distinct data frequencies as separate modalities within the neural architecture, our approach suggests a novel framework for handling heterogeneous time series data. This methodology prompts a reevaluation of the relationships between different components of a time series and their implications for forecasting future trends. Furthermore, our emphasis on high-frequency data representation highlights the significance of temporal granularity in predictive modeling. These theoretical underpinnings hold the potential to stimulate further theoretical inquiries and methodological refinements in time series analysis, contributing to the broader discourse within the machine learning and forecasting communities.

However, it is important to note a limitation of our approaches: the dependency on data quality and availability, especially concerning proprietary datasets from Brazilian pre-salt wells. Access to such datasets may be restricted, hindering the reproducibility and generalizability of our findings. Moreover, the performance of our models may vary depending on the characteristics of the datasets, such as noise levels, missing values, and outliers. Addressing these challenges requires robust data preprocessing techniques and the development of more resilient forecasting models. Another limitation is the reliance on neural network architectures, which can be computationally intensive and require substantial computational resources for training and inference. This may limit the scalability and practicality of deploying our models in real-time forecasting applications, particularly in resource-constrained environments.

In addition to analyzing the effectiveness of our proposed multimodal approach and the utilization of GRU2 with interpolated HF data, this study shed light on key characteristics of other models tested, insights that are uniquely gleaned from this comprehensive benchmarking. Specifically, our findings underscored the suitability of MIDAS for forecasting when the frequency of LF data is low, while highlighting N-HiTS as a viable option for univariate time series. Moreover, our investigation revealed the distinct advantage of employing N-HiTS with exogenous data and interpolated HF data, particularly in medium and long-term forecasting scenarios characterized by significant differences in the granularity of LF and HF data. By elucidating these nuances, our study not only contributes to the understanding of our proposed methodologies but also enriches the discourse surrounding established forecasting approaches.

Furthermore, while our methodologies have shown promising results in forecasting fluid rates and bottom-hole pressure, their applicability to other domains and datasets remains to be fully explored. Future research should focus on validating the effectiveness of our approaches across diverse datasets and forecasting scenarios, ensuring their broader applicability and robustness.

5.1 Future work

Looking ahead, several avenues for future research emerge. One crucial direction involves the development of meta-learning strategies for model selection tailored to specific forecasting tasks. Given, the absence of a universally superior solution in our experiments, these strategies could dynamically adapt to the nuances of each scenario. Additionally, implementing a classifier within the meta-learning framework could discern the utility of high-frequency data for a given scenario.

Another promising avenue for future work entails delving into multimodal learning techniques based on late fusion to address the intricacies of mixed-frequency time series. Ensembles employing predictors with data of the same frequency could be explored to harness the potential synergies within diverse data sources. These future endeavors contribute to the refinement of time series forecasting methodologies and pave the way for adaptive, context-aware models that enhance performance across varied scenarios.

Appendix 1: Proposed multimodal architectures

Table 6 presents the architecture specifications of various multimodal models designed for our time series forecasting approach. Each model consists of three branches: a low-frequency branch, a high-frequency branch, and a post-concatenation topology. The low-frequency branch processes data with lower temporal granularity, while the high-frequency branch handles data with higher temporal resolution. The post-concatenation topology integrates the outputs from both branches to generate the final forecast. Architectural components such as long short-term memory (LSTM), gated recurrent unit (GRU), 1-dimensional convolutional layer (Conv1D), sequential self-attention (SeqSelfAttention), and dense layers are utilized to capture temporal dependencies and extract features from the input data.

This table provides detailed insights into the configurations of multimodal architectures, including variations

Table 6 Multimodal models (the first line refers to the low-frequency branch, the second line refers to the high-frequency branches, and the third line refers to the post-concatenation topology)

Multi LSTM	Multi GRU	Multi LSTM CNN	Multi GRU CNN	Multi LSTM Att	Multi GRU Att
LSTM_daily (units = 128, return_seq = True), LSTM (units = 128) 2x Dense (units = 8/4, activation = 'relu')	GRU_daily (units = 200), 2x Dense (units = 8/4, activation = 'relu')	2x Conv1D (filters = 64, kernel_size = 5, activation = 'relu'), LSTM_daily (units = 200), 2x Dense (units = 8/4, activation = 'relu')	2x Conv1D (filters = 64, kernel_size = 5, activation = 'relu'), GRU_daily (units = 200), 2x Dense (units = 8/4, activation = 'relu')	LSTM_daily (units = 128, return_seq = True), SeqSelfAttention (att_activation = 'sigmoid'), LSTM (units = 128), 2x Dense (units = 8/4, activation = 'relu')	GRU_daily (units = 128, return_seq = True), SeqSelfAttention (att_activation = 'sigmoid'), GRU(units = 128), 2x Dense (units = 8/4, activation = 'relu')
LSTMs_HF (units = 128, return_seq = True), LSTM (units = 128), 2x Dense (units = 8/4, activation = 'relu')	GRUs_HF (units = 200), 2x Dense (units = 8/4, activation = 'relu')	2x Conv1D (filters = 64, kernel_size = 5, activation = 'relu'), LSTMs_HF (units = 200), 2x Dense (units = 8/4, activation = 'relu')	2x Conv1D (filters = 64, kernel_size = 5, activation = 'relu'), GRUs_HF (units = 200), 2x Dense (units = 8/4, activation = 'relu')	LSTMs_HF (units = 128, return_seq = True), SeqSelfAttention (att_activation = 'sigmoid'), LSTM (units = 128), 2x Dense (units = 8/4, activation = 'relu')	GRUs_HF (units = 128, return_seq = True), SeqSelfAttention (att_activation='sigmoid'), GRU(units=128), 2x Dense (units = 8/4, activation = 'relu')
9x Dense (units = 128/64/32, activation = 'linear'), Dense (units = size_output, activation = 'linear')	Dense (units = size_output, activation = 'linear')	Dense (units = size_output, activation = 'linear')	Dense (units = size_output, activation = 'linear')	9x Dense (units = 128/64/32, activation = 'linear'), Dense (units = size_output, activation = 'linear')	9x Dense (units = 128/64/32, activation = 'linear'), Dense(units = size_output, activation = 'linear')

such as Multi LSTM and Multi GRU (simple RNNs), Multi LSTM CNN and Multi GRU CNN (RNNs with convolutional layers), Multi LSTM Att and Multi GRU Att (RNNs with attention mechanism). Each model's architecture is shown in terms of the layers and their configurations, including the number of units, activation functions, and specific operations performed at each stage.

Appendix 2: Complete results for oil dataset per well

The Tables 7, 8, 9 and 10 show the SMAPE and DTW results for wells P1, P2, P3 and P4, respectively, with GRU2 (without and with interpolated HF data), the multimodal approach, MIDAS and N-HiTS (only with the target time series, using the multivariate approach and using all-time series, including interpolated HF data). These results show that although we can derive some general behaviors from these models, as shown in Sects. 4.4.1 and 4.5, we can see how the performance of

Table 7 SMAPE and DTW results for P1 well

		Model	Short	Medium	Long
SMAPE	QO	GRU2	24.65	35.66	43.09
		GRU2 w/HF (ours)	**20.63**	**23.31**	32.68
		Multimodal (ours)	36.18	45.21	**27.02**
		MIDAS	68.55	170.87	192.13
		N-HiTS	50.03	48.86	48.17
		N-HiTS w/Ex	27.96	28.65	28.52
		N-HiTS w/Ex + HF	28.15	27.97	28.68
	BHP	GRU2	5.27	12.20	19.66
		GRU2 w/HF (ours)	3.55	7.46	10.07
		Multimodal (ours)	8.18	5.21	10.29
		MIDAS	6.13	3.39	64.39
		N-HiTS	2.72	2.66	2.70
		N-HiTS w/Ex	**2.20**	2.23	2.24
		N-HiTS w/Ex + HF	2.24	**2.22**	**2.21**
DTW	QO	GRU2	273.51	833.18	846.60
		GRU2 w/HF (ours)	**212.38**	388.14	600.34
		Multimodal (ours)	549.17	999.06	478.20
		MIDAS	371.50	5849.21	103009.60
		N-HiTS	418.32	373.98	407.33
		N-HiTS w/Ex	285.15	308.74	**302.43**
		N-HiTS w/Ex + HF	299.06	**296.63**	308.18
	BHP	GRU2	14.02	43.41	49.93
		GRU2 w/HF (ours)	8.12	20.03	24.03
		Multimodal (ours)	21.22	12.47	24.59
		MIDAS	8.40	5.74	348. 26
		N-HiTS	**2.91**	**3.64**	**3.64**
		N-HiTS w/Ex	3.64	3.72	4.01
		N-HiTS w/Ex + HF	4.02	3.74	3.67

Bold means the best result obtained for benchmarking a specific scenario

Table 8 SMAPE and DTW results for P2 well

		Model	Short	Medium
SMAPE	QO	GRU2	8.75	8.22
		GRU2 w/HF (ours)	**8.17**	**7.49**
		Multimodal (ours)	11.78	15.66
		MIDAS	199.40	21.83
		N-HiTS	9.97	9.99
		N-HiTS w/Ex	10.19	10.21
		N-HiTS w/Ex + HF	10.13	10.15
	BHP	GRU2	1.01	2.01
		GRU2 w/HF (ours)	1.10	2.41
		Multimodal (ours)	1.36	0.87
		MIDAS	192.47	1.99
		N-HiTS	0.70	0.70
		N-HiTS w/Ex	**0.67**	**0.67**
		N-HiTS w/Ex + HF	**0.67**	**0.67**
DTW	QO	GRU2	149.55	154.91
		GRU2 w/HF (ours)	**137.55**	**134.87**
		Multimodal (ours)	171.82	233.36
		MIDAS	3642575.81	342.23
		N-HiTS	158.15	158.29
		N-HiTS w/Ex	164.01	157.76
		N-HiTS w/Ex + HF	159.38	153.94
	BHP	GRU2	1.98	5.43
		GRU2 w/HF (ours)	2.14	6.12
		Multimodal (ours)	2.91	1.93
		MIDAS	25742.20	2.00
		N-HiTS	1.40	1.41
		N-HiTS w/Ex	**1.28**	1.30
		N-HiTS w/Ex + HF	1.30	**1.29**

Bold means the best result obtained for benchmarking a specific scenario

Table 9 SMAPE and DTW results for P3 well

		Model	Short	Medium	Long
SMAPE	QO	GRU2	9.18	**7.40**	7.11
		GRU2 w/HF (ours)	**8.77**	8.11	**6.81**
		Multimodal (ours)	10.75	9.88	7.63
		MIDAS	187.64	189.85	200
		N-HiTS	16.31	16.13	17.09
		N-HiTS w/Ex	23.38	23.91	23.12
		N-HiTS w/Ex + HF	23.92	24.87	22.50
	BHP	GRU2	**0.68**	1.33	2.51
		GRU2 w/HF (ours)	1.17	1.32	2.82
		Multimodal (ours)	2.47	1.73	1.13
		MIDAS	37.92	76.83	197.69
		N-HiTS	1.00	1.04	1.08
		N-HiTS w/Ex	1.55	**0.99**	1.77
		N-HiTS w/Ex + HF	1.16	1.80	**1.08**

Table 9 (continued)

		Model	Short	Medium	Long
DTW	QO	GRU2	165.93	**135.52**	164.65
		GRU2 w/HF (ours)	**131.36**	142.41	**158.06**
		Multimodal (ours)	201.52	163.54	176.03
		MIDAS	51397.65	69115.69	21334014.45
		N-HiTS	160.55	181.41	171.00
		N-HiTS w/Ex	306.80	301.93	280.92
		N-HiTS w/Ex + HF	284.85	301.75	262.80
	BHP	GRU2	**1.24**	1.33	5.97
		GRU2 w/HF (ours)	1.67	1.32	6.70
		Multimodal (ours)	6.61	1.73	2.74
		MIDAS	109.66	76.83	65733.68
		N-HiTS	1.42	1.04	1.43
		N-HiTS w/Ex	1.93	**1.15**	2.21
		N-HiTS w/Ex + HF	1.35	2.25	**1.27**

Bold means the best result obtained for benchmarking a specific scenario

Table 10 SMAPE and DTW results for P4 well

		Model	Short	Medium	Long
SMAPE	QO	GRU2	9.18	**7.40**	7.11
		GRU2 w/HF (ours)	**8.77**	8.11	**6.81**
		Multimodal (ours)	10.75	9.88	7.63
		MIDAS	187.64	189.85	200
		N-HiTS	16.31	16.13	17.09
		N-HiTS w/Ex	23.38	23.91	23.12
		N-HiTS w/Ex + HF	23.92	24.87	22.50
	BHP	GRU2	**0.68**	1.33	2.51
		GRU2 w/HF (ours)	1.17	1.32	2.82
		Multimodal (ours)	2.47	1.73	1.13
		MIDAS	37.92	76.83	197.69
		N-HiTS	1.00	1.04	1.08
		N-HiTS w/Ex	1.55	**0.99**	1.77
		N-HiTS w/Ex + HF	1.16	1.80	**1.08**
DTW	QO	GRU2	165.93	**135.52**	164.65
		GRU2 w/HF (ours)	**131.36**	142.41	**158.06**
		Multimodal (ours)	201.52	163.54	176.03
		MIDAS	51397.65	69115.69	21334014.45
		N-HiTS	160.55	181.41	171.00
		N-HiTS w/Ex	306.80	301.93	280.92
		N-HiTS w/Ex + HF	284.85	301.75	262.80
	BHP	GRU2	**1.24**	1.33	5.97
		GRU2 w/HF (ours)	1.67	1.32	6.70
		Multimodal (ours)	6.61	1.73	2.74
		MIDAS	109.66	76.83	65733.68
		N-HiTS	1.42	1.04	1.43
		N-HiTS w/Ex	1.93	**1.15**	2.21
		N-HiTS w/Ex + HF	1.35	2.25	**1.27**

Bold means the best result obtained for benchmarking a specific scenario

the models depends greatly on the specific scenario and its characteristics, such as the data mean, its dispersion, skewness and kurtosis, the stability and seasonality of the time series that describes the target, the presence of outliers, the size of the input window and the forecast horizon. Therefore, possibly all of these models presented in this benchmark, including our approaches, would be interesting options in a prediction system composed of a pool of models with a meta-learning model selection algorithm.

Acknowledgements This research was carried out as part of the finished R&D project with Shell Brasil Petróleo Ltda, registered as ANP number 21373-6 as "Machine-Learning Development for Analysis of Complex Production Data in a Pre-Salt Carbonate Field''- (UNICAMP/Shell Brazil/ANP) and of the ongoing R&D project with Shell Brasil Petróleo Ltda, registered as ANP number 24282-6 as "Artificial intelligence and machine learning for analyzing complex and multispectral production data, establishing causal links, effects of interventions and hybrid modeling in Pre-Salt fields – Phase 2" -(UNICAMP/Shell Brazil/ANP), both of them funded by Shell Brazil Technology, under the ANP R &D levy as ''Compromisso de Investimentos com Pesquisa e Desenvolvimento''. We thank Shell Brazil for permission to publish this work.

Author Contributions LLF—Conception; data acquisition and analysis; code development; writing a draft of the paper. ROW—Conception; discussion of results; code development; paper review. MCA; PRMJ; AF—Conception; discussion of results; paper review. AL; MZ; OL; IN; BAD; SS; AL; AE; MG; DJS; AD—Discussion of results; article review. RM—Code development. AR—Conception; discussion of results; paper review; project coordination.

Funding This research was carried out within the scope of the project "Shell-ML: Machine-Learning Development for Analysis of Complex Production Data in a Pre-Salt Carbonate Field", funded by Shell Brasil Petróleo LTDA.

Data availability Some datasets used in this work are private and cannot be shared. We try to describe the characteristics of this data as best as possible. The public datasets used have their repository links described in the text. The pre-salt oil dataset used in this study is not publicly available due to its proprietary nature, as it belongs to a private company. However, the another datasets utilized for the experiments are publicly accessible. The PPG-DaLiA dataset can be found at https://archive.ics.uci.edu/dataset/495/ppg+dalia, and the Jena Weather Dataset is available at https://www.bgc-jena.mpg.de/wetter/. Researchers interested in accessing the proprietary dataset may contact the respective company for further information and possible data sharing arrangements.

Declarations

Conflict of interest The authors declare that they have no conflict of interest

References

1. Ertekin T, Sun Q (2019) Artificial intelligence applications in reservoir engineering: a status check. Energies 12(15):2897
2. Liu W, Liu WD, Gu J (2020) Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network. J Petrol Sci Eng 189:107013
3. Sun J, Ma X, Kazi M (2018) Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells. In: SPE western regional meeting, p 11
4. Xu C, Qu Y, Xiang Y, Gao L (2023) Asynchronous federated learning on heterogeneous devices: a survey. Comput Sci Rev 50:100595
5. Gao J, Li P, Chen Z, Zhang J (2020) A survey on deep learning for multimodal data fusion. Neural Comput 32(5):829–864
6. Oliveira Werneck R, Prates R, Moura R, Goncalves MM, Castro M, Soriano-Vargas A, Júnior PRM, Hossain MM, Zampieri MF, Ferreira A et al (2022) Data-driven deep-learning forecasting for oil production and pressure. J Petrol Sci Eng 210:109937
7. Chen M, Mao S, Liu Y (2014) Big data: a survey. Mobile Netw Appl 19:171–209
8. Ren X, Li X, Ren K, Song J, Xu Z, Deng K, Wang X (2021) Deep learning-based weather prediction: a survey. Big Data Res 23:100178
9. Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. Wiley, Hoboken
10. Gurland J (1954) Hypothesis testing in time series analysis. J Am Stat Assoc 49:197
11. Ghysels E, Santa-Clara P, Valkanov R (2006) Predicting volatility: getting the most out of return data sampled at different frequencies. J Econom 131(1–2):59–95
12. Baffigi A, Golinelli R, Parigi G (2004) Bridge models to forecast the euro area GDP. Int J Forecast 20(3):447–460
13. Mariano RS, Murasawa Y (2010) A coincident index, common factors, and monthly real GDP. Oxford Bull Econ Stat 72(1):27–46
14. Frale C, Marcellino M, Mazzi GL, Proietti T (2010) Survey data as coincident or leading indicators. J Forecast 29(1–2):109–131
15. Frale C, Marcellino M, Mazzi GL, Proietti T (2011) EUROMIND: a monthly indicator of the euro area economic conditions. J R Stat Soc Ser A Stat Soc 174(2):439–470
16. Yu H, Wang Z, Xie Y, Wang G (2024) A multi-granularity hierarchical network for long- and short-term forecasting on multivariate time series data. Appl Soft Comput 157:111537
17. Babii A (2022) High-dimensional mixed-frequency IV regression. J Bus Econ Stat 40(4):1470–1483
18. Chang T, Hsu C-M, Chen S-T, Wang M-C, Wu C-F (2023) Revisiting economic growth and CO_2 emissions nexus in Taiwan using a mixed-frequency var model. Econ Anal Policy 79:319–342
19. Kamolthip S (2021) Macroeconomic forecasting with LSTM and mixed frequency time series data. arXiv preprint arXiv:2109.13777
20. Ghysels E, Santa-Clara P, Valkanov R (2004) The MIDAS touch: mixed data sampling regression models. https://escholarship.org/uc/item/9mf223rs
21. Kuck K, Schweikert K (2021) Forecasting Baden-Württemberg's GDP growth: Midas regressions versus dynamic mixed-frequency factor models. J Forecast 40(5):861–882
22. Bai J, Ghysels E, Wright JH (2013) State space models and MIDAS regressions. Econom Rev 32(7):779–813
23. Kuzin V, Marcellino M, Schumacher C (2011) MIDAS vs. mixed-frequency var: nowcasting GDP in the euro area. Int J Forecast 27(2):529–542
24. Foroni C, Marcellino MG (2013) A survey of econometric methods for mixed-frequency data. Available at SSRN 2268912
25. Wohlrabe K (2009) Forecasting with mixed-frequency time series models. PhD thesis, LMU

26. Ghysels E, Qian H (2019) Estimating MIDAS regressions via OLS with polynomial parameter profiling. Econom Stat 9:1–16
27. Audrino F, Kostrov A, Ortega J-P (2019) Predicting US bank failures with MIDAS logit models. J Financ Quant Anal 54(6):2575–2603
28. Liu Y (2019) Statistical methods for mixed frequency data sampling models. PhD thesis, Michigan Technological University
29. Hecq A, Götz T, Urbain J (2012) Forecasting mixed frequency time series with ECM-MIDAS models. METEOR, Maastricht University School of Business and Economics, Maastricht
30. Hamid A (2015) Prediction power of high-frequency based volatility measures: a model based approach. RMS 9:549–576
31. Guérin P, Marcellino M (2013) Markov-switching MIDAS models. J Bus Econ Stat 31(1):45–56
32. Qiu Y (2020) Forecasting the consumer confidence index with tree-based MIDAS regressions. Econ Model 91:247–256
33. Bonino-Gayoso N, Garcia-Hiernaux A (2021) TF-MIDAS: a transfer function based mixed-frequency model. J Stat Comput Simul 91(10):1980–2017
34. Engle RF, Ghysels E, Sohn B (2013) Stock market volatility and macroeconomic fundamentals. Rev Econ Stat 95(3):776–797
35. Foroni C, Marcellino MG, Schumacher C (2011) U-MIDAS: MIDAS regressions with unrestricted lag polynomials. Bundesbank Series 1 Discussion. Paper No. 2011,35
36. Xu Q, Zhuo X, Jiang C, Liu Y (2019) An artificial neural network for mixed frequency data. Expert Syst Appl 118:127–139
37. Xu Q, Liu S, Jiang C, Zhuo X (2021) QRNN-MIDAS: a novel quantile regression neural network for mixed sampling frequency data. Neurocomputing 457:84–105
38. Li X, Yu H, Xie Y, Li J (2021) Attention-based novel neural network for mixed frequency data. CAAI Trans Intell Technol 6(3):301–311
39. Challu C, Olivares KG, Oreshkin BN, Ramirez FG, Canseco MM, Dubrawski A (2023) N-HiTS: Neural hierarchical interpolation for time series forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol 37, pp 6989–6997
40. Oreshkin BN, Carpov D, Chapados N, Bengio Y (2019) N-beats: neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437
41. Baltrušaitis T, Ahuja C, Morency L-P (2018) Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell 41(2):423–443
42. Zhang C, Yang Z, He X, Deng L (2020) Multimodal intelligence: representation learning, information fusion, and applications. IEEE J Sel Top Signal Process 14(3):478–493
43. Ramachandram D, Taylor GW (2017) Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process Mag 34(6):96–108
44. Morvant E, Habrard A, Ayache S (2014) Majority vote of diverse classifiers for late fusion. In: Structural, syntactic, and statistical pattern recognition: joint IAPR international workshop, S+ SSPR 2014, Joensuu, Finland, August 20–22, 2014. Proceedings, pp. 153–162. Springer
45. Shutova E, Kiela D, Maillard J (2016) Black holes and white rabbits: Metaphor identification with visual features. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 160–170
46. Glodek M, Tschechne S, Layher G, Schels M, Brosch T, Scherer S, Kächele M, Schmidt M, Neumann H, Palm G et al (2011) Multiple classifier systems for the classification of audio-visual emotional states. In: Affective computing and intelligent interaction: fourth international conference, ACII 2011, Memphis, October 9–12, 2011, Proceedings, Part II. Springer, pp 359–368
47. Chen L, Li Z, Xu T, Wu H, Wang Z, Yuan NJ, Chen E (2022) Multi-modal siamese network for entity alignment. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 118–126
48. Song C, Ning N, Zhang Y, Wu B (2021) A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Inf Process Manag 58(1):102437
49. Angelou M, Solachidis V, Vretos N, Daras P (2019) Graph-based multimodal fusion with metric learning for multimodal classification. Pattern Recogn 95:296–307
50. Lusquino Filho LAD, Werneck RDO, Mendes Júnior PR, Castro M, Santos Pereira E, Moura R, Sousa Ferreira VH, Ferreira AM, Gomes AD, Rocha A (2022) Oil production and pressure multimodal forecasting integrating high-frequency production data. In: Rio oil & gas expo and conference. IBP, pp 308–309
51. Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Philos Trans R Soc A 379(2194):20200209
52. Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A (2021) Deep learning for time series forecasting: a survey. Big Data 9(1):3–21
53. Reiss A, Indlekofer I, Schmidt P (2019) PPG-DaLiA. UCI Machine Learning Repository. https://doi.org/10.24432/C53890
54. Institute MP (2016) Jena Climate Dataset. https://www.bgc-jena.mpg.de/wetter/. Weather time series dataset recorded at the Weather Station of the Max Planck Institute for Biogeochemistry in Jena, Germany