# Highlights

## Sediment Prediction in the Great Barrier Reef using Vision Transformer with Finite Element Analysis

Mohammad Jahanbakht,Wei Xiang,Mostafa Rahimi Azghadi

- Finite element analysis is incorporated in a vision Transformer network to form a novel FE-Transformer model for next-frame prediction.
- Sediment distribution in the Great Barrier Reef oceanic region, which is considered as a frame, is forecasted using the proposed FE-Transformer.
- Physics-informed neural network is employed to merge sediment partial differential equation solutions with in-situ measured sediment data.
- The proposed model produces highly accurate unblurred output frames, which cannot be achieved using previous next-frame predictors.

# Sediment Prediction in the Great Barrier Reef using Vision Transformer with Finite Element Analysis

Mohammad Jahanbakht[a], Wei Xiang[b,*] and Mostafa Rahimi Azghadi[a,*]

[a]College of Science and Engineering, James Cook University, 1 James Cook Dr, Douglas 4811, QLD, Australia
[b]School of Engineering and Mathematical Sciences, La Trobe University, Plenty Rd, Bundoora 3086, VIC, Australia

ARTICLE INFO

*Keywords*:
Deep Neural Networks
Vision Transformer
Finite Element Analysis
Partial Differential Equation
Total Sediment Forecasting
Next-frame Prediction
Great Barrier Reef
Physics-informed Neural Network
eReefs Modeling Suite

ABSTRACT

Suspended sediment is a significant threat to the Great Barrier Reef (GBR) ecosystem. This catchment pollutant stems primarily from terrestrial soil erosion. Bulk masses of sediments have potential to propagate from river plumes into the mid-shelf and outer-shelf regions. Existing sediment forecasting methods suffer from the problem of low-resolution predictions, making them unsuitable for wide area coverage. In this paper, a novel sediment distribution prediction model is proposed to augment existing water quality management programs for the GBR. This model is based on the state-of-the-art Transformer network in conjunction with the well-known finite element analysis. For model training, the emerging physics-informed neural network is employed to incorporate both simulated and measured sediment data. Our proposed Finite Element Transformer (FE-Transformer) model offers accurate predictions of sediment across the entire GBR. It provides unblurred outputs, which cannot be achieved with previous next-frame prediction models. This paves a way for accurate forecasting of sediment, which in turn may lead to improved water quality management for the GBR.

## 1. Introduction

The Great Barrier Reef (GBR) in Australia is a world heritage site and the world's largest coral ecosystem. Coral reefs provide jobs and foods for over half a billion people worldwide, protect coastlines from storms and erosion, offer opportunities for recreation, and source new medicines (NOAA, 2021). Unfortunately, these ecosystems are under great pressure both by natural events and human activities. These threats include storms, floods, man-made pollution, and warmer ocean temperatures that can stress corals, leading to coral bleaching or physical damages (Baird et al., 2021).

The GBR water quality has been declining over the last 150 years (Margvelashvili et al., 2018). Sediments, nutrients, and pesticides are the main coral reef pollutants (Waterhouse et al., 2020), which result in poor water quality that is believed to be one of the main contributors to the current adverse situation in the GBR coastal and marine areas (Waterhouse et al., 2020).

More specifically, sediments have significant impacts on the GBR ecosystem. Suspended sediments eventually settle down on seafloor in a process called sedimentation. Then, nutrients and other pollutants attached to sediment release in the underwater environment, which in turn, limits coral recruitment, health, and productivity (Howarth and Marino, 2006). Suspended sediments can also reduce sunlight available for seagrasses, algae, and other plants' photosynthesis and reduce their growth.

GBR sedimentation is primarily attributed to the land-based sediments from 35 catchments into the GBR lagoon (Margvelashvili et al., 2018). Among all these catchments, those with higher levels of land clearing show more contribution to the total GBR sedimentation. To elaborate, 70% of river-based sediments stem from 20% of GBR catchment. This suggests that much of the sedimentation problem can be mitigated by managing a relatively small area (McKergow et al., 2005).

Management of sediment and nitrogen loads to the GBR and improving its water quality has been the focus of major investments for many years (Kroon et al., 2016; Coggan et al., 2021). According to literature, the coastal waters of the GBR have noticeably lost their quality due to changes in their catchment land use since European settlement (McCloskey et al., 2021). Besides, GBR sedimentation has been particularly elevated during flood events (Bhattacharya and Solomatine, 2006), which has been linked primarily to the land-use practices and reduced vegetation cover in QLD catchments (Margvelashvili et al., 2018).

These sediments include minerals, muds, dusts, and granite from soil erosion, as well as white calcium carbonate ($CaCO_3$) from coral erosion (Baird et al., 2020). The sum of all these components is termed the Total Sediment (TS). Any TS value over the 2 mg/L threshold can adversely affect marine ecosystems and corals in open coastal and mid-shelf waters (R. Reichelt *et al.*, 2010). To better control adverse sediment erosion, a long-term water quality improvement plan was established by the Australian Government, dubbed *The Reef 2050 Plan* (Reef-2050, 2021). This plan aims to better manage the relevant catchments and to reduce run-off pollutants into the GBR.

This work is funded by the Australian Government Research Training Program Scholarship.

*Corresponding authors.
mohammad.jahanbakht@my.jcu.edu.au ( Mohammad Jahanbakht); w.xiang@latrobe.edu.au ( Wei Xiang); mostafa.rahimiazghadi@jcu.edu.au ( Mostafa Rahimi Azghadi)
ORCID(s): 0000-0002-3609-9677 ( Mohammad Jahanbakht); 0000-0002-0608-065X ( Wei Xiang); 0000-0001-7975-3985 ( Mostafa Rahimi Azghadi)

However, implementing an accurate TS predictor is not an easy task, attributed to the high dynamics of sediment processes in the wide GBR region. These processes start from catchments in the vicinity of river discharge areas, where bulk masses of sediments settle on the sea bed of each river mouth, extending tens of kilometers along the coastline. These sediments are then dispersed by waves and relocated by ocean currents to remote GBR areas, and partly buried into deeper benthic sediments (Margvelashvili et al., 2018).

To understand the TS dynamics and to model its complex behaviour, a few nonlinear ocean models have been developed. For example, *eReefs* is a comprehensive suite of hydrodynamic, biogeochemical, and sediment transport models, which was especially designed for the GBR (Stevena et al., 2019). This suite of models provide access to historical environmental variables, and offer limited management options to mitigate risks associated with its predefined scenarios (Margvelashvili et al., 2016). The sediment transport model in *eReefs* can analyse the fate of suspended sediments in the GBR (Stevena et al., 2019).

However, *eReefs* is a computationally expensive model, which has a very complex calibration process and may take weeks to simulate any desired ocean sedimentation. It also lacks the much-desired capability of forecasting. This means that *eReefs* cannot be used to forecast future values of TS distribution (Margvelashvili et al., 2016). Therefore, *eReefs* and other similar modelling suits have limited capacities to underpin decision-support for the spatially vast GBR. This motivates demands for a data-driven deep learning model to predict future sediment distributions as fast and accurate as possible.

To meet the above demand, many studies have been carried out to design an accurate sediment prediction model (Bhagat et al., 2021; Mehri et al., 2021; Ali-Khan et al., 2019). For example, Bhagat et al. (2021) designed a shallow neural network for sediment lead (Pb) prediction for two Australian bays. In other examples, Mehri et al. (2021) and Ali-Khan et al. (2019) developed an Adaptive Neuro Fuzzy Inference System (ANFIS) and a shallow neural network model to predict the sediment distribution in river open-channel flows, respectively.

Thus far, all the aforementioned models have avoided the problem of next-frame sediment distribution prediction. These models simply average sediment in their study areas, making them the well-studied timeseries prediction models. However, many applications require an accurate distribution prediction of their parameters of interest across an area. This distribution forecasting is known in parlance as *next-frame* prediction, which has evolved from the former timeseries prediction models (Zhou et al., 2020).

The existing next-frame forecasting models mainly utilize 3D Convolutional Neural Networks (Conv3D) (Mathieu et al., 2016) and Convolutional Long Short-Term Memory (ConvLSTM) (Guen and Thome, 2020). These models result in blurred prediction frames with low coefficient of determination. To address this problem, we propose a novel Deep Neural Network (DNN) to predict the next-frame suspended sediment distribution in the GBR. This model is based on the state-of-the-art Transformer network (Vaswani et al., 2017).

Transformers are a new class of DNNs that were initially developed for tackling timeseries and sequential problems (Sergio and Lee, 2021). Traditional timeseries processing models such as the Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) suffer from the following drawbacks:

- *Long-term dependencies*: RNNs are unable to explore long-term correlations within data, due to the so-called vanishing gradient phenomenon. This problem is alleviated by the gating mechanism in LSTMs and GRUs to some extent. However, RNN-based networks are still incapable of addressing wide time-gaps in long-term dependencies (Cossu et al., 2021).

- *Parallelization*: All RNN-based models are limited to directional data, where the value in time $t$ is strictly dependent on the value in time $t - \Delta t$. This directional property limits the model's capability for parallel processing, distributed processing, etc. (Su et al., 2020).

To address the above two significant limitations, an *attention* mechanism was initially proposed by Bahdanau et al. (2014). This mechanism models long-term dependencies without considering the entity location in the sequence, which can better extract long-term dependencies than RNN-based models (Zeyer et al., 2020). Additionally, this non-recurrent mechanism consists of multi-head attention and Multi-Layer Perceptron (MLP) layers, which make it a good candidate for parallelization. This initial self-attention architecture was used in the first Transformer model for natural language processing, by Vaswani et al. (2017). Dosovitskiy et al. (2020) later extended the Transformer network to process images by proposing the so-called Vision Transformer (ViT). The ViT is the model of choice in this paper, which is used in conjunction with concepts of Finite Element Analysis (FEA) to accurately predict the next sediment frame in the entire GBR.

The proposed FE-Transformer takes as input the TS distribution frames (matrices) in the past $N$ days, and forecasts the TS distribution frame in $F$ days later. However, the observational TS data in the wide GBR are scarce and insufficient for training data-hungry DNNs. To address this problem, we employ a technique to merge sparse TS measurements with the readily available simulated TS data (e.g., *eReefs*). This approach proves effective in training our proposed FE-Transformer network, which demonstrates an excellent TS prediction performance across the GBR, when compared with state-of-the-art next-frame prediction approaches.

The remainder of this paper is organized as follows. Section 2 proposes the FE-Transformer as a novel solution to the problem of sediment distribution prediction in the GBR. The physics-informed neural network technique for measured and simulated data fusion is described in the remaining of this section. We evaluate the accuracy of our

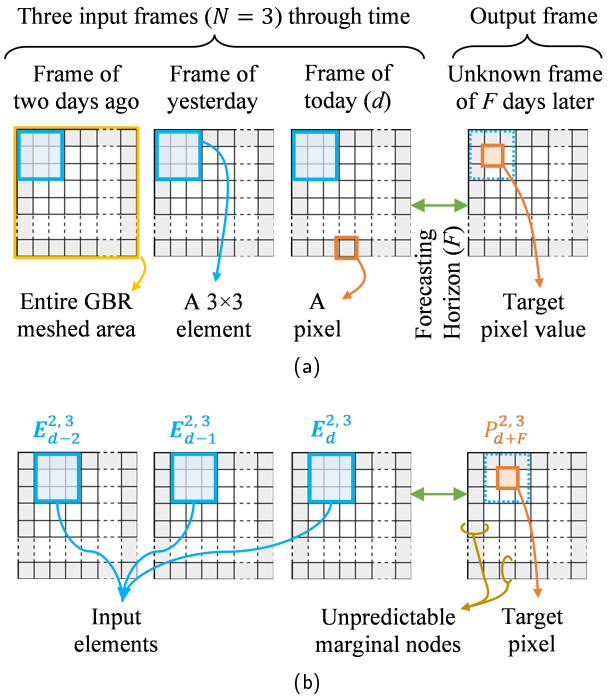



**Figure 1**: Three input frames ($N = 3$) are used to predict the next TS frame in $F$ days after today ($d$). Towards this end, the pixels of historical TS values in the GBR are grouped into 3×3 elements $\boldsymbol{E}^{x,y}_{d-2}$, $\boldsymbol{E}^{x,y}_{d-1}$, and $\boldsymbol{E}^{x,y}_{d}$. The central coordinate $(x, y)$ of these historical elements will be adjusted to predict TS at (a) pixel (2, 2) (i.e., $\boldsymbol{P}^{2,2}_{d+F}$); and (b) pixel (2, 3) (i.e., $\boldsymbol{P}^{2,3}_{d+F}$), while leaving a 1-pixel margin of unpredictable TS values. The $3 \times 3$ element size is used here for illustration purposes only. This size can be optimized to better address the requirements of TS prediction in the GBR.

FE-Transformer model for next sediment frame forecasting in Section 3, where a detailed investigation of the computational complexity and an ablation study are also carried out. The paper is concluded in Section 4.

## 2. Proposed Model

As discussed in the preceding section, increased sedimentation since European settlement reduces water quality and imposes a major threat to the health of GBR ecosystems (MacNeil et al., 2019). This makes TS improvement important through local management actions, which in turn improves the resilience of GBR ecosystems. To support such actions, it is important to be able to monitor and accurately predict sediment on the scale of the whole GBR.

The problem of accurate TS distribution prediction in GBR requires a sophisticated next-frame predicting model. The state-of-the-art next-frame predictors are surveyed by Zhou et al. (2020), where a collection of convolutional neural networks, long short-term memory, generative adversarial network, and gated recurrent unit DNNs along with their possible combinations are benchmarked.

A common weakness of all these next-frame predictors lies in their inability to learn input variations (Zhou et al., 2020), which results in blurred predicted frames. In other words, these models try to reduce output frame dynamics with the objective of minimizing prediction errors. As a result, the forecasted output frame suffers from low variance and smoothed details. To address this problem, we propose to combine the FEA mechanism with the state-of-the-art Vision Transformer network to create a highly-accurate TS distribution predictor.

### 2.1. Finite Element Analysis

The FEA is a numerical method for solving Partial Differential Equations (PDE) in engineering. This method is extensively used in simulating physical phenomena, by finding current transient responses to the current transient inputs. A FEA system is intrinsically unable to address forecasting scenarios in which one predicts future responses to the given historical inputs. In other words, FEA-based forecasting systems rely heavily on their coupled surrogate models to estimate future boundary conditions. However, our proposed FE-Transformer integrates the forecasting DNN techniques with the FEA to accurately predict unknown future TS values.

Similar to other hydrodynamic phenomena, sediments dynamics can be described and modeled by some governing PDEs when being suspended, transported, and deposited by ocean waves and flows (Margvelashvili et al., 2016). As illustrated in Fig. 1a, the TS distribution for each day forms a *frame* that covers the entire GBR. By knowing the past $N$ frames before today ($d$), the goal is to predict the future TS frame in $F$ days later (forecasting horizon).

To accurately predict a future frame, the FEA concept can be employed by dividing the large GBR area into *pixels* and *elements*. Every pixel $\boldsymbol{P}$ retains the average TS in a 16 km$^2$ area in the GBR. For each given coordinate $(x, y)$, the pixel in that coordinate and several surrounding pixels are grouped into a small square termed an *element*. The historical elements across $\boldsymbol{E}^{x,y}_{d-N+1}$, $\boldsymbol{E}^{x,y}_{d-N+2}$, … , $\boldsymbol{E}^{x,y}_{d}$ constitute the known inputs to the model, which yields the unknown pixel $\boldsymbol{P}^{x,y}_{d+F}$ as output. As shown in Figs. 1a and 1b, the input elements overlap and sweep to cover all pixels $\boldsymbol{P}$ inside the TS distribution frame in the GBR, leaving the half element size unpredictable in margins, e.g., one pixel for $3 \times 3$ elements and three pixels for $7 \times 7$ elements.

After discretizing the study area into elements, the FEA formulation results in the following abstract system of algebraic equations (Surana and Reddy, 2017)

$$\boldsymbol{P} = f\,(\boldsymbol{K},\, \boldsymbol{E})\,, \tag{1}$$

where $\boldsymbol{K}$ is the *stiffness matrix*. This matrix is a core parameter in FEA, which relies on scientific understanding of the underlying physics of the targeted system. In Section 2.2, we propose a new method to numerically calculate $\boldsymbol{K}$ using the training dataset, where no prior knowledge of the TS dynamics in the GBR is required.

Furthermore, unlike the FEA that resorts to matrix algebra to solve (1), we feed both $\boldsymbol{K}$ and $\boldsymbol{E}$ to our proposed

machine learning model. As a result, the governing equation in (1) changes to

$$\boldsymbol{P} = \boldsymbol{W}\, f\,(\boldsymbol{K},\, \boldsymbol{E})\,, \tag{2}$$

where $\boldsymbol{W}$ represents the unknown weights of the supervised machine learning model. Using this innovative approach, the model will learn TS variations in each element, using the historical TS values for that element and its corresponding stiffness matrix. It is worth mentioning that the recursive technique for solving (1) is very similar to the gradient descent technique used by machine learning to solve (2).

## 2.2. Stiffness Matrices

In the FEA, a stiffness matrix denotes an approximate solution to underlying PDEs by relating nodal responses to external nodal forces (Surana and Reddy, 2017). Similarly, the stiffness matrix of an element in the GBR represents its physical properties and relates TS variations to its historical values. Having said that, solving the stiffness matrix for ocean sedimentation requires a deep understanding of GBR multiphysical behaviours, including fluid flow, mass transport, etc. (Stevena et al., 2019). To address this challenge, an innovative approach is proposed to numerically compute the stiffness matrix $\boldsymbol{K}$.

Consider a simple linear spring problem illustrated in Fig. 2a, where a linear spring with $k = 10$ acts against external forces $F_1$ and $F_2$, and it compresses to the $X_1$ extent. We simulate this system for two seconds with 40 samples (20 samples per second) using sine waves and random Gaussian noise. By dividing this simulation time into four segments ($r$), we have 10 samples per segment. This way, the true stiffness matrix for each segment can be numerically estimated by

$$\begin{aligned} &\boldsymbol{K}_r \boldsymbol{X} = \boldsymbol{F} \Rightarrow \\ &\boldsymbol{K}_r = \frac{\boldsymbol{F}}{\boldsymbol{X}} \xrightarrow{\text{In matrix algebra}} \boldsymbol{K}_r = \boldsymbol{F}\,\boldsymbol{X}^T \left(\boldsymbol{X}\,\boldsymbol{X}^T\right)^{-1}, \end{aligned} \tag{3}$$

where $\boldsymbol{F}_{2\times 10}$ is the matrix of 10 $F_1$ and $F_2$ values, and $\boldsymbol{X}_{1\times 10}$ is the matrix of 10 $X_1$ values in one segment. The final stiffness matrix $\boldsymbol{K}$ of the linear spring problem can then be calculated by averaging $\boldsymbol{K}_r$ as follows

$$\boldsymbol{K} = \frac{1}{4}\sum_{r=1}^{4} \boldsymbol{K}_r\,. \tag{4}$$

The simulation results in Fig. 2a suggest 96% accuracy in numerical estimation of $\boldsymbol{K}$. However, adapting this technique to our TS prediction problem requires the following adjustments:

1. As stated earlier, the number of input elements to our model is $N$, which is composed of elements from today ($d$) to $N$ days ago. Consequently, we must have $N$ stiffness matrices for each of these input elements;
2. To better capture the TS distribution, we mesh the GBR into small spatial elements. Furthermore, to properly capture temporal TS changes, we need to take a further step by dividing the study period into smaller time steps, e.g., months. This temporal division leads to a dedicated set of stiffness matrices for each month, which results in a more accurate prediction of marine sedimentation. In this regard, smaller time steps will increase memory consumption, while larger time steps will reduce accuracy;
3. Different from (1) in conventional FEA that places the stiffness matrix $\boldsymbol{K}$ in an equation with $\boldsymbol{E}$, we propose to feed $\boldsymbol{K}$ as an additional input to our DNN. This way, the variations in each TS element will be learned in line with its corresponding stiffness matrix; and
4. The Conv2D layers in Fig. 3 implement $\boldsymbol{E}-\boldsymbol{K}$. This is in contrast to the original FEA implementation in (3), where $\boldsymbol{X} = \boldsymbol{F}/\boldsymbol{K}$. Therefore, in our proposed model, $\boldsymbol{K}$ is calculated by subtracting the known output $\boldsymbol{P}$ from element $\boldsymbol{E}$, i.e., $\boldsymbol{E} - \boldsymbol{P}$.

Considering the above adjustments, we can rewrite and merge (3) and (4) into

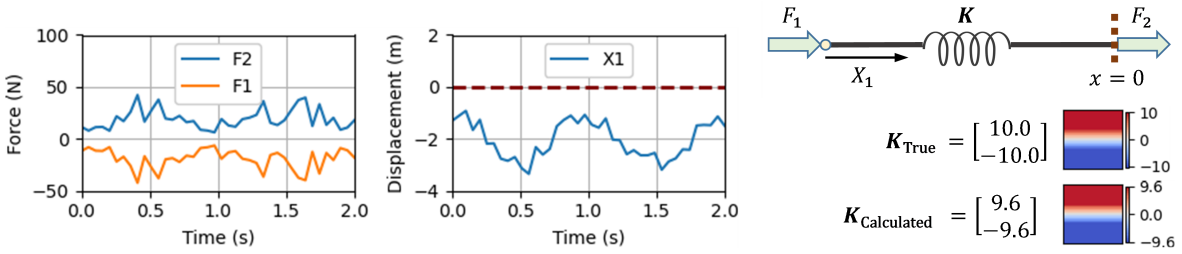
$$\boldsymbol{K}^{x,\,y}_{n,\,M} = \frac{1}{N_y}\frac{1}{N_d}\sum_{y}\sum_{d}\left(\boldsymbol{E}^{x,\,y}_{d-n} - \boldsymbol{P}^{x,\,y}_{d+F}\,\boldsymbol{J}\right) \qquad \forall n \in [0,\, N-1], \tag{5}$$

where $M$ represents the corresponding month of the year, $y$ sweeps all $N_y$ years in our training dataset, $d$ sweeps all $N_d$ days of month $M$, and $\boldsymbol{J}$ is an all-one matrix of the same size as our elements. After calculation of the stiffness matrices using the training dataset, the resulting $\boldsymbol{K}$ matrices will become independent of both year and day, and will be a function of coordinates $(x,\, y)$, month $M$, and their time difference $n \in [0,\, N-1]$ from today.

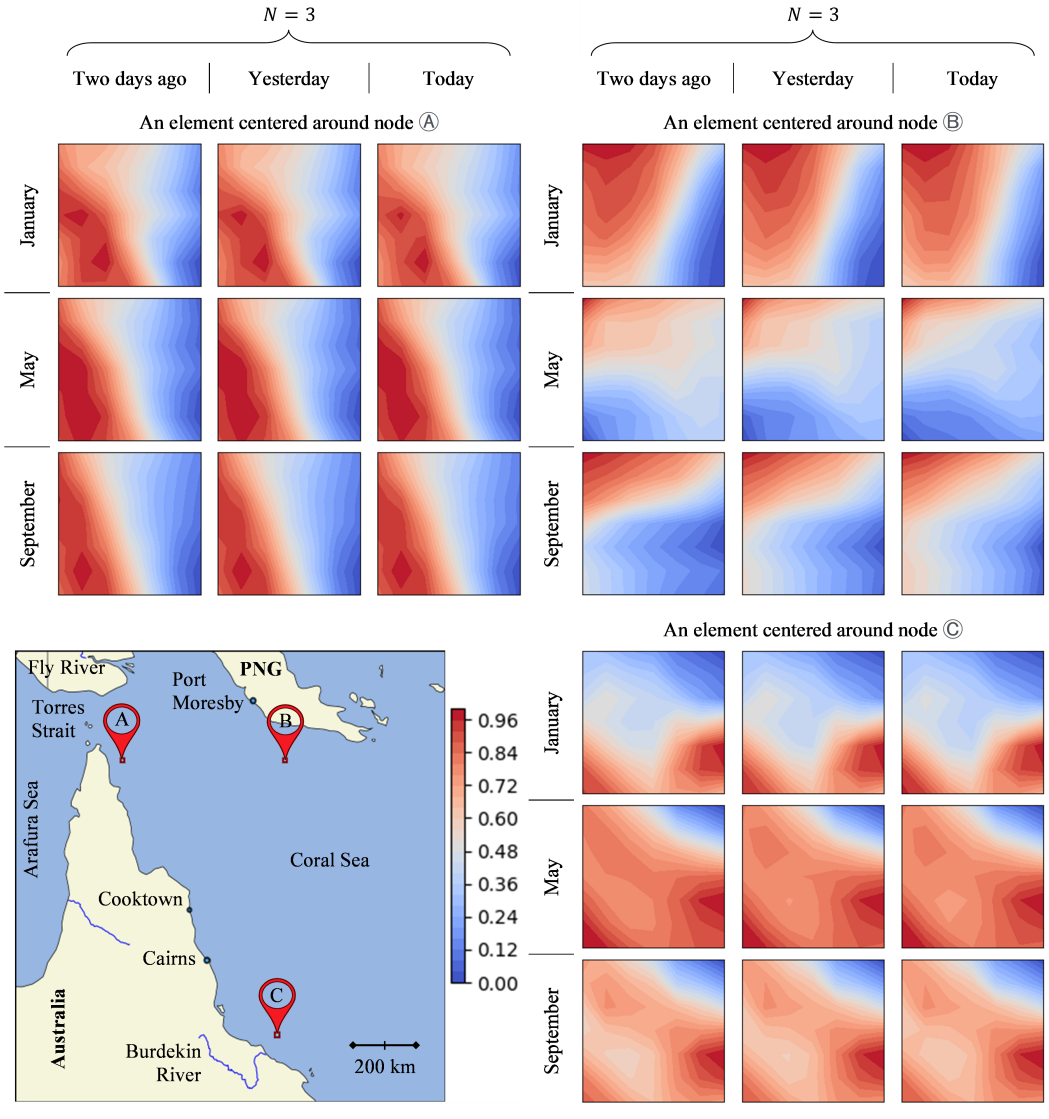
Using the new equation in (5), the stiffness matrices of January, May, and September are plotted in Fig. 2b for three random locations in the GBR. For these matrices, the number of inputs elements considered is 3, i.e., $N = 3$, while the forecasting horizon is $F = 1$. These settings result in three matrices $\boldsymbol{K}^{x,\,y}_{2,\,M}$, $\boldsymbol{K}^{x,\,y}_{1,\,M}$, $\boldsymbol{K}^{x,\,y}_{0,\,M}$ per location per month, denoted by *two days ago*, *yesterday*, and *today*, respectively. In the normalized contour plots of this figure, 0 means the maximum negative difference from the central pixel today, 1 means the maximum positive difference, and 0.5 means no difference.

As a result, the stiffness matrices such as those shown in Fig. 2b reveal the overall behavior of TS in the GBR. For instance:

- TS in point Ⓐ is influenced by its vicinity to the water flows in the Torres Strait;
- TS in point Ⓑ is influenced by both the Fly River and Papua New Guinea (PNG) catchments; and
- TS in point Ⓒ is under influence of the Burdekin River.




(a)




(b)

**Figure 2**: (a) Stiffness matrix of the classic linear spring problem in the presence of Gaussian noise. (b) Normalized logarithmic scale stiffness matrices of three typical geolocations: Ⓐ at (11.18°S, 143.25°E), Ⓑ at (11.18°S, 148.17°E), and Ⓒ at (19.10°S, 147.93°E), where the element size is $7 \times 7$.

## 2.3. FE-Transformer Architecture

In the previous section we explained that, in our proposed FE-Transformer the GBR is meshed into distinct pixel-level geolocations $(x, y)$, and in each location there are $N$ elements representing historical TS values around that location (i.e., $\boldsymbol{E}_{d-n}^{x,y}$ for $n \in [0, N-1]$). We then calculate $N$ relevant stiffness matrices $\boldsymbol{K}_{n,M}^{x,y}$ for each month $M$. Next, we use these two inputs in (2) to predict the TS distribution $\boldsymbol{P}$. This prediction is conducted in a pixel-by-pixel manner, by calculating the TS in each coordinate $(x, y)$ (i.e., $\boldsymbol{P}_{d+F}^{x,y}$).

The $f(.)$ operator in (2) is implemented by using a novel DNN structure shown in Fig. 3. To elaborate, we

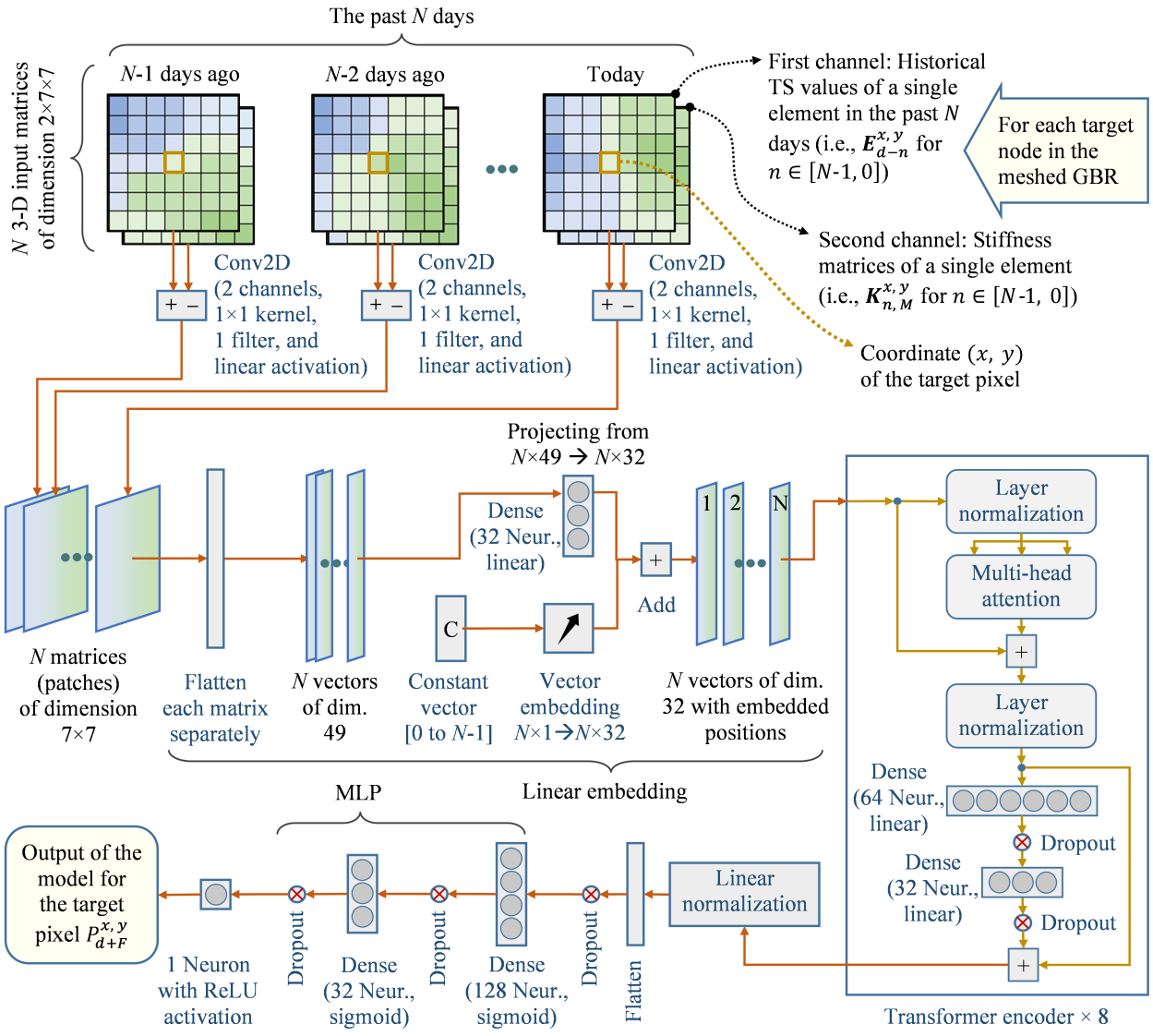



**Figure 3**: Architecture of the proposed FE-Transformer network to predict pixel $P^{x,y}_{d+F}$ in $F$ days after today ($d$), where the $N$ input elements are illustrated in Fig. 1, and the $N$ stiffness matrices for month $M$ are calculated by (5). The shown kernel sizes, filter numbers, activation functions, Transformer layers, etc. are optimized for TS prediction in the GBR.

employ the state-of-the-art ViT (Dosovitskiy et al., 2020) to propose a new FEA-inspired DNN dubbed FE-Transformer for next-frame prediction of sediment in the GBR. As shown in Fig. 3, this process starts by merging $N$ historical TS elements with their corresponding stiffness matrices using $N$ two-dimensional convolutional layers (Conv2D). Feeding $\boldsymbol{K}$ as a separate input to the DNN model is in contrast to the conventional FEA that multiplies an inversed stiffness matrix with the excitation TS element values. Besides, the attention mechanism in Transformer networks have limited capacity to exploit the sequential nature of the input. Using $N$ Conv2D layers to merge historical elementwise TS values with their corresponding stiffness matrices will help the subsequent Transformer layers by capturing part of the temporal relations inside the elements.

The resulting $N$ patches are then flattened before being fed to the next layer, where they are projected onto $N$ new vectors with a reduced size of 32. These vectors are then embedded (added) with $N$ position vectors. This allows the model to learn the day-by-day ordinal behavior, as well as the natural moving directions in the TS distribution.

Next, the resulting sequence of positional embedded vectors enter eight cascaded standard Transformer layers, as originally described in (Vaswani et al., 2017). The output of the last Transformer layer is flattened and enters an MLP network to predict the TS value of the target pixel. To improve the convergence of weight learning and to avoid overfitting, we have

- Multiple dropout points with a 20% dropping rate are placed in the layer intervals in Fig. 3;
- All the dense layers are equipped with the Ridge regularization of $L2 = 0.01$; and

- The learning rate is set to 0.001.

### 2.4. Physics-Informed Neural Network

To train our FE-Transformer model, we need daily TS data, i.e., $P_{\text{Measured}}$, measured for each pixel in the GBR. However, the existing $P_{\text{Measured}}$ datasets are intermittently observed and sparsely distributed. This scarce dataset is not suitable to sufficiently train DNNs and our Transformer network.

Raissi et al. (2019) recently introduced a novel Physics-Informed Neural Network (PINN), which can be used to merge measurements with PDE simulations for DNN training. PINN works by summing two separately calculated measured and simulated loss functions. To elaborate, suppose that we use the scarcely measured data to calculate the Mean Squared Error (MSE) loss function as follows

$$\text{MSE}_{\text{Measured}} = \frac{1}{N_{\text{measured data}}} \sum (P - P_{\text{Measured}})^2, \quad (6)$$

where $P$ is the output of our model in Fig. 3, and $N_{\text{measured data}}$ is the number of used data points. At the same time, PDEs can be solved to simulate the TS concentration distribution in the GBR (i.e., $P_{\text{PDE}}$) for the following loss function

$$\text{MSE}_{\text{PDE}} = \frac{1}{N_{\text{PDE data}}} \sum (P - P_{\text{PDE}})^2 . \quad (7)$$

The MSE loss functions in (6) and (7) are then combined to create the following overall loss metric

$$\text{MSE} = (1 - \lambda)\, \text{MSE}_{\text{Measured}} + (\lambda)\, \text{MSE}_{\text{PDE}} , \quad (8)$$

where $\lambda$ is an adjustable hyperparameter that is used to implement a weighted sum to balance between the measured and PDE simulated data.

The MSE loss function in (8) is used to train the proposed FE-Transformer. It is worth mentioning that whenever $\text{MSE}_{\text{Measured}}$ is not available, it is removed from (8).

## 3. Results and Discussions

This section starts by introducing the employed TS datasets in the GBR. We will then proceed with evaluating the accuracy of the proposed FE-Transformer for predicting the TS distribution. We will also study the computational complexity of our model and undertake an ablation study for a block-level evaluation of our model's structure.

### 3.1. Data Sources

According to (6) and (7), two sets of measured ($P_{\text{Measured}}$) and simulated ($P_{\text{PDE}}$) data are required to train the FE-Transformer in conjunction with PINN methodology. The measured TS values stem from the Australian Institute of Marine Science (AIMS) Marine Monitoring Program (MMP) (AIMS-MMP, 2021). This data is sparsely gathered, both temporally and spatially, and thus is insufficient for DNN training purposes.

Hence, simulated data (from solving PDEs) is used to compensate for the scarcity of the measured data. To obtain the PDE solutions for the TS distribution in the GBR, the *eReefs* modeling suite is employed, which is a suite of models designed to deliver a near real-time coupled biogeochemical, hydrodynamic, and sediment transport simulation of the GBR (Baird et al., 2020). This raw simulation data has been cleaned and interpolated onto a regular grid and is downloadable from the AIMS servers (AIMS-eReefs, 2021).

**Table 1**
Statistics of the TS Values from *eReefs* ($P_{\text{PDE}}$)

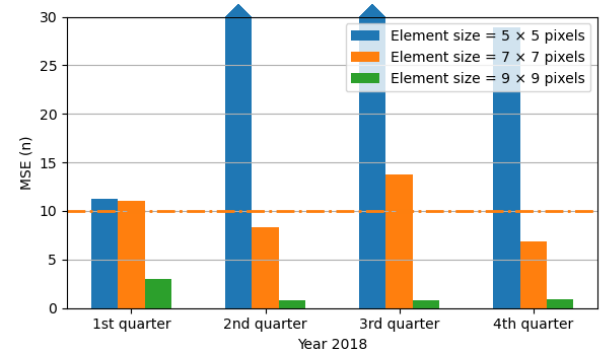
| Dataset | Year | Min | Max | Mean (μ) | STD (m) |
|---|---|---|---|---|---|
| Training and Validation | 2011 | 0 | 0.34 | 104.05 | 1.19 |
| | 2012 | 0 | 0.26 | 80.44 | 1.25 |
| | 2013 | 0 | 0.28 | 122.26 | 1.82 |
| | 2014 | 0 | 0.32 | 91.21 | 1.48 |
| | 2015 | 0 | 0.30 | 84.12 | 1.53 |
| | 2016 | 0 | 0.20 | 67.13 | 1.10 |
| | 2017 | 0 | 0.50 | 104.08 | 1.88 |
| Testing | 2018 | 0 | 0.30 | 91.22 | 1.39 |

The sediment simulation in *eReefs* is conducted for multiple layers, from which the accumulated sedimentation in the top 3 m surface is used in this paper. In addition, the AIMS version of *eReefs* is calculated on a daily basis during 2011 to 2018. We divide this period into three timespans. The first one from 2011 to 2016 is used as the training dataset, 2017 is used as the validation dataset, and 2018 is the testing dataset. The model weights are first calibrated using the training dataset and, without changing the calibrated weights, accuracy is evaluated by the validation datasets. Finally, the overall accuracy of the model is measured by the testing dataset, which is not *a priori* known to the model.

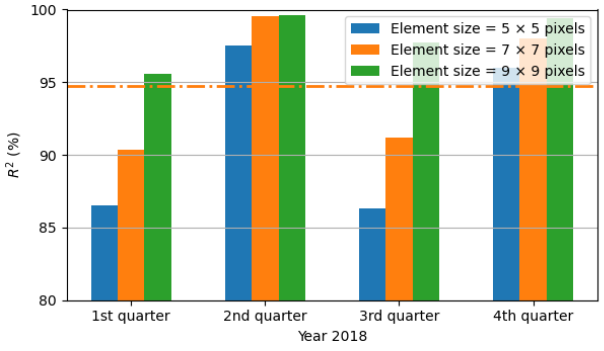
Overall, after meshing the whole GBR into $150 \times 122$ pixels, we would have 18300, 17760, and 16704 overlapping $1 \times 1$, $3 \times 3$, and $7 \times 7$ elements, respectively, in each day. Accumulating these elements in 365 days of 7 years in our training dataset results in 47, 45, and 43 million training patch images (i.e., elements). According to Dosovitskiy et al. (2020), ViT-based models can benefit from this large dataset to gain higher accuracy, compared to traditional CNN architectures.

To better understand the nature of TS simulation in *eReefs*, the statistics of TS values are presented in Table 1. As can be seen from this table, the maximum TS values are far greater than their average values. This is because the TS concentrations within river deltas are usually very high, while in many other places these concentrations can be very low. As a result, the ecologically important TS variations throughout the near-shore regions and out to the Mid-shelf waters may look insignificant. To capture these small variations, we use the logarithmic scale for model training as follows

$$TS_{\log} = \log_{10} \left( TS + 10^{-7} \right), \quad (9)$$




(a)




(b)

**Figure 4**: Effect of element size on the metrics of (a) MSE and (b) $R^2$ with 3-day forecasting horizons and $N = 3$. The results are averaged per quarter of the testing year of 2018.

where the constant $10^{-7}$ in (9) is added to avoid $\log(0)$.

### 3.2. Prediction Accuracy

The performance of the proposed FE-Transformer on TS distribution prediction in the GBR is studied in this section. We show that our model provides accurate predictions, as well as unblurred output frames, which are two desirable properties of any next-frame predictor.

To measure accuracy, we use the MSE metric in (8). Meanwhile, the coefficient of determination ($R^2$) is adopted as an indicator of our unblurred output, which is calculated by

$$R^2 = 1 - \frac{\sum(P - P_{\text{True}})^2}{\sum(P_{\text{True}} - \mu_{P_{\text{True}}})^2} \tag{10}$$

where $P_{\text{True}}$ stands for either $P_{\text{Measured}}$ or $P_{\text{PDE}}$, and $\mu_{P_{\text{True}}}$ is the average of $P_{\text{True}}$. Here, $R^2$ shows how much of the true TS variation has been captured in the predicted output frame. This metric is also known as *goodness of fit*, where the larger is better, meaning even low levels of TS variation within a frame are captured.

As stated earlier in Section 2, the element size is a hyperparameter that requires adjustment. Towards this end, the effect of the element size on the MSE and $R^2$ metrics is investigated in Figs. 4a and 4b, respectively. As a rule of thumb, increasing the element size will lead to an improved model performance, but it also increases the number of unpredictable marginal pixels in Fig. 1. For consistency, 7×7 elements are used in our simulations. The horizontal dash lines in Fig. 4 indicate the average metric values for the 7×7 elements.

**Table 2**
MSE and $R^2$ of TS Prediction for the Testing Dataset in 2018, using Elements of Size 7×7 Pixels

| Month | $F = 1$ | $F = 3$ | $F = 7$ | $F = 1$ | $F = 3$ | $F = 7$ |
|---|---|---|---|---|---|---|
| | MSE (n) | | | $R^2$ (%) | | |
| Jan | 2.95 | 7.58 | 14.30 | 98.05 | 88.44 | 97.12 |
| Feb | 0.24 | 0.71 | 7.11 | 98.67 | 94.31 | 76.78 |
| Mar | 3.82 | 27.68 | 23.59 | 99.15 | 87.85 | 97.55 |
| Apr | 0.05 | 2.33 | 5.16 | 99.96 | 98.99 | 97.79 |
| May | 9.34 | 21.92 | 15.60 | 99.95 | 99.81 | 99.85 |
| Jun | 0.35 | 0.70 | 11.19 | 99.96 | 99.91 | 99.74 |
| Jul | 0.42 | 1.22 | 4.73 | 99.39 | 96.34 | 79.13 |
| Aug | 0.35 | 2.39 | 3.27 | 99.06 | 87.03 | 96.04 |
| Sep | 17.34 | 37.66 | 22.23 | 99.26 | 90.11 | 98.42 |
| Oct | 0.03 | 0.15 | 1.75 | 99.68 | 98.86 | 76.20 |
| Nov | 0.14 | 0.77 | 1.37 | 99.52 | 97.04 | 87.24 |
| Dec | 59.37 | 19.63 | 27.55 | 98.87 | 98.22 | 94.65 |
| Mean | 7.87 | 10.23 | 11.49 | 99.29 | 94.74 | 91.71 |

In addition to the element size, other hyperparameters of the proposed FE-Transformer are also optimized. These hyperparameters include but are not limited to the Conv2D kernel sizes and filter numbers, neural network activation functions, number of Transformer layers and their internal variables, etc. The hyperparameter optimization helps the model accurately forecast the TS in the GBR. This also tunes the number of model weights to the optimum value, which in turn, helps the model avoid overfitting and underfitting.

Some typical logarithmic-scale TS prediction results of the FE-Transformer model in 2018 (i.e., the test dataset) are plotted in Fig. 5. In the cases of the 1-day, 3-day, or 7-day forecasting horizons, the ($7^{\text{th}}$, $8^{\text{th}}$, $9^{\text{th}}$), ($5^{\text{th}}$, $6^{\text{th}}$, $7^{\text{th}}$), or ($1^{\text{st}}$, $2^{\text{nd}}$, $3^{\text{rd}}$) days of each month are fed to the model, while the $10^{\text{th}}$ day of the months are forecasted. This figure includes the predicted frames, ground-truths, and their absolute differences (i.e., the prediction errors), which are mostly close to zero. The prediction errors of larger $F$ values have further spread geographically, resulting in a larger MSE. Besides, due to the adoption of the element size of $7 \times 7$, three unpredictable marginal pixels are imposed along the shoreline. These pixels are predicted with a linear regression model.

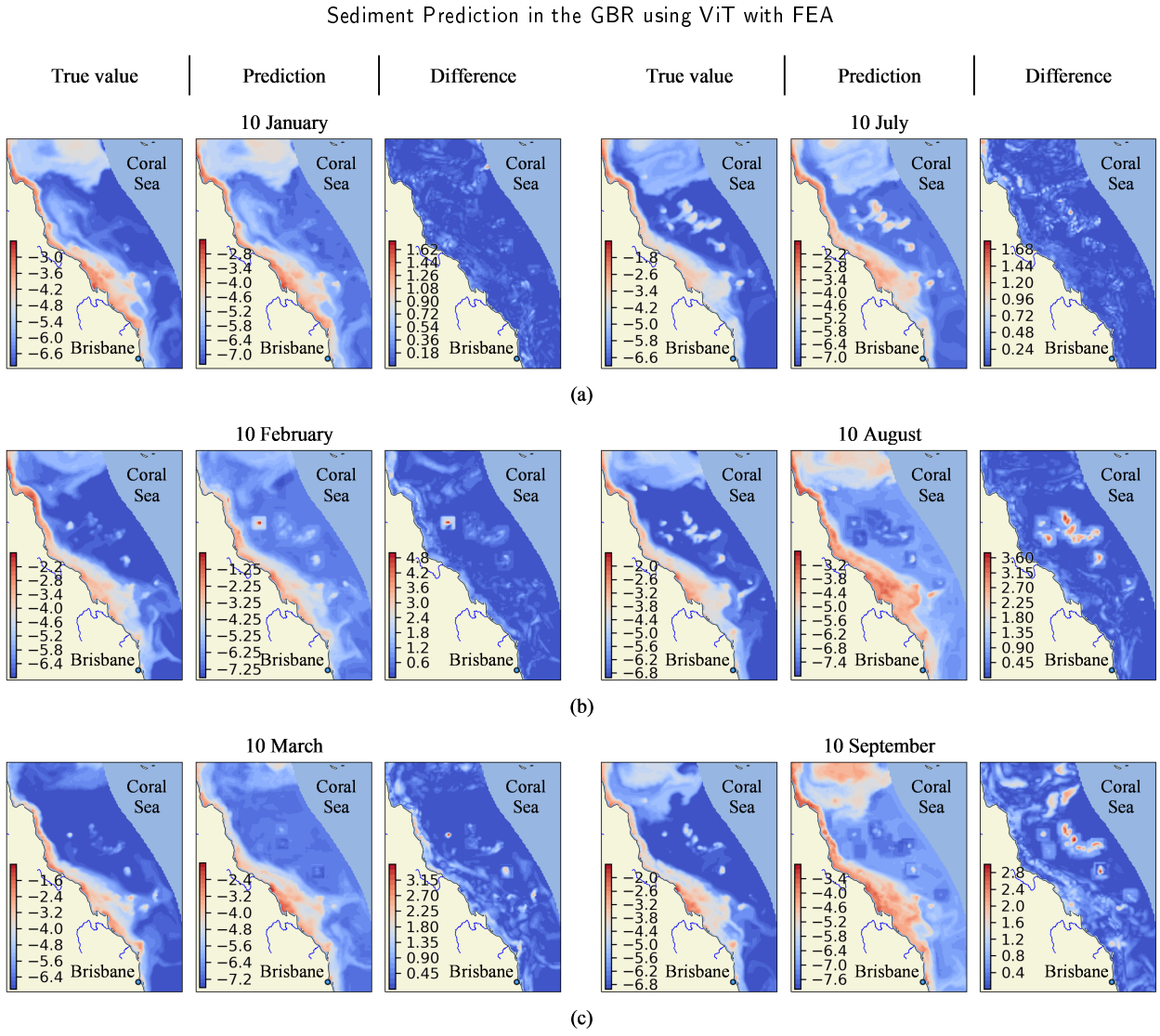



**Figure 5**: True values, predictions, and their absolute differences for $TS_{\log}$ in the GBR at 2018 test dataset, for (a) 1-day, (b) 3-day, and (c) 7-day forecasting horizons, using $7 \times 7$ elements with $N = 3$.

It is worth mentioning that classic image processing techniques for compensating for the unpredictable margins (e.g., padding, flipping, etc.) do not suit the TS distribution prediction problem considered in this paper. This mainly is due to the fact that current TS values in the GBR are dependent on their neighbors' historical TS values, which are asymmetric, chaotic, and dynamic (Skerratt et al., 2019). Therefore, if these historical neighboring values are not available, we should not simply fill them with constant padding, symmetrical flipping, or other normal paddings. Besides, the unpredictable margin is rather small and has a minimum effect on the overall modeling performance.

Both the MSE and $R^2$ metrics of the proposed FE-Transformer model are investigated in Table 2. The reported metrics in this table do not include the unpredictable margins. Here, the average MSE increases from $7.87 \times 10^{-9}$ for the 1-day forecasting horizon to $11.49 \times 10^{-9}$ for the 7-day forecasting horizon. The $R^2$ values remain above 90% for all $F$ values, which imply that our model is able to accurately capture the majority of the TS variations. Let us emphasize that this highly-accurate and unblurred prediction is achieved throughout the wide GBR with more than 35 river discharges (Stevena et al., 2019), which makes the prediction even more challenging.

Since our model is a next-frame predictor, in Table 3, we compare its performance with two of the recent next-frame prediction methods, i.e., a Conv3D-based model (Mathieu et al., 2016) and a Convolutional LSTM model named PhyD-Net (Guen and Thome, 2020). The comparison is conducted on the 2018 test dataset with $7 \times 7$ elements. At the time of writing, PhyDNet (Guen and Thome, 2020) is ranked as the best next-frame video predicting model in multiple categories (papers with code). Having said that, the MSE and Mean Absolute Error (MAE) of the previous models are always greater than those of the proposed FE-Transformer, which results in their $R^2 < 60\%$ for all $F$ values.

Another performance metric in Table 3 is the scatter index, which is calculated by dividing the root-MSE by the average TS values in each day. The low percentage of the

**Table 3**
Comparing the performance of our FE-Transformer model with two recently published works in literature.

| DNN Model | Forecasting horizon ($F$) | | |
|---|---|---|---|
| | 1 | 3 | 7 |
| | MSE | | |
| Conv3D by Mathieu et al. (2016) | 58.2 µ | 67.2 µ | 64.7 µ |
| PhyDNet by Guen and Thome (2020) | 28.6 µ | 27.8 µ | 45.8 µ |
| The Proposed FE-Transformer | 7.87 n | 10.23 n | 11.49 n |
| | MAE | | |
| Conv3D by Mathieu et al. (2016) | 120 µ | 163 µ | 177 µ |
| PhyDNet by Guen and Thome (2020) | 142 µ | 174 µ | 205 µ |
| The Proposed FE-Transformer | 5.25 µ | 8.54 µ | 15.8 µ |
| | Scatter Index | | |
| Conv3D by Mathieu et al. (2016) | 8.6% | 23.6% | 45.6% |
| PhyDNet by Guen and Thome (2020) | 9.3% | 22.7% | 45.8% |
| The Proposed FE-Transformer | 0.6% | 1.3% | 4.7% |
| | $R^2$ | | |
| Conv3D by Mathieu et al. (2016) | 57% | 29% | 19% |
| PhyDNet by Guen and Thome (2020) | 51% | 27% | 19% |
| The Proposed FE-Transformer | 99% | 95% | 92% |
| | F-test (p-value) | | |
| Conv3D by Mathieu et al. (2016) | 0.45 | 0.50 | 0.47 |
| PhyDNet by Guen and Thome (2020) | 0.45 | 0.48 | 0.47 |
| The Proposed FE-Transformer | 0.01 | 0.01 | 0.13 |

scatter index of the FE-Transformer indicates a small relative error with respect to the average TS.

We also conduct an analysis of variance (F-test) in Table 3 to find out whether the TS predictions and their relevant true values have the same variance. Larger p-values reject the null hypothesis (H0), indicating that the variances are not equal. Based on this test, our model perfectly captures the TS variances with probabilities greater than 99%, 99%, and 85% for $F = 1$, $F = 3$, and $F = 7$, respectively. These p-values are acceptable in the wide GBR with high-dynamic sediment distributions, especially when the other comparing models are barely close to the 55% probability at their best.

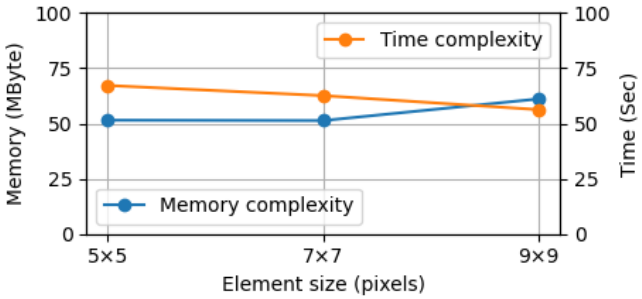



**Figure 6**: Time and memory complexities with respect to the element size for one-day forecasting horizon with $N = 3$.

As can be seen from Table 3, neither Conv3D nor PhyDNet can accurately predict the next TS frame in the wide GBR. This is mainly because both of these next-frame predictors treat each frame as a whole, compared to FE-Transformer that borrows the finite element concept and uses the modified stiffness matrices to produce accurate predictions in meshed GBR frames. Additionally, according to Dosovitskiy et al. (2020), ViT-based models perform better than their convolutional counterparts on larger training datasets. As discussed in Section 3.1, we have a huge number of training image patches (i.e., elements), which justify the employment of the ViT in the proposed FE-Transformer model.

### 3.3. Computational Complexity

The proposed FE-Transformer model was implemented by TensorFlow 2.4.1 in Python, and the DNN layers were employed from the relevant Keras APIs. The model was trained on a machine with Intel® Core i7-7700HQ CPU, NVIDIA® GeForce 1050 GPU, and 16 GB RAM.

The computational complexity of the proposed FE-Transformer model is shown in Fig. 6, where both the memory consumption and running time are plotted versus the element size. The memory consumption refers to the allocated RAM for the model, which excludes the dataset memory and GPU memory. The running time on the other hand, is the time required for inference. The slight reduction in running time versus the element size in this figure is due to the increased number of unpredictable marginal pixels. Overall, the plot reveals an $O(1)$ constant demand for both time and memory, which suggests our model is efficient.

### 3.4. Ablation Study

The ablation study in machine learning is the mechanism of removing a component of a model to study its importance in the overall system. This study is conducted in Table 4, where each row corresponds to the elimination of a given component from the proposed FE-Transformer in Fig. 3. The first row includes all the construction blocks, the second row

**Table 4**
Evaluating the Elimination of Selected Blocks on Accuracy of the Model

| Ablation study | MSE (n) | MSE increment |
|---|---|---|
| Including all the blocks | 7.87 | 0% |
| Excluding stiffness matrix | 10.47 | 33% |
| Excluding the transformer layer | 9.52 | 21% |
| Excluding the MLP layer | 9.29 | 18% |

excludes the stiffness matrix channel from input, the third row excludes linear embedding as well as the Transformer encoder, and the fourth row excludes two dense layers of the MLP. In each case, the MSE is averaged over the test dataset in year 2018. As expected, the MSE increases in subsequent rows, by removing functional components of the structural body.

Comparing all the MSE values in this table reveals a high contribution of the stiffness matrix to the TS next-frame prediction. Removing this FEA-originated parameter reduces the model accuracy by 33%. The stiffness matrix is a core component of the proposed FE-Transformer model. Not only does it approximate the underlying PDE solutions, but it also relates TS variations to its historical values. Therefore, removing the stiffness matrix has the most adverse effect on the model performance.

The transformer layer is another key component playing a significant role in our next-frame predicting model. In short, the transformer makes the model capable of addressing wide time-gaps in TS long-term dependencies. Eliminating this layer drops the model accuracy by 21%. Finally, the MLP module consists of two hidden layers and 160 neurons. This module is responsible for producing the TS prediction at the Transformer output for a given pixel. Similar to conventional DNN architectures, the MLP after the flattened values in Fig. 3 helps reduce the dimension of the transformer output to one final neuron for the target pixel. Without the MLP, the dimension reduction does not happen and an 18% drop in prediction accuracy would occur.

### 3.5. Limitations and Future Directions

Similar to other DNN-based models, the main limitation of the proposed FE-Transformer is the availability of training data. After dividing a study area into elements and pixels, training data must be collected for each pixel in regular time-steps. In other words, this model is only applicable to spatio-temporal problems with regularly measured or interpolated data in both the time and spatial domains. Additionally, computational resources are another limiting factor of the proposed model. To overcome this limitation, training data must be downsampled into widely separated pixels. Another workaround can be using a more powerful computing unit or even a distributed training scheme.

Accordingly, future research can both address the above limitations and enhance the capabilities of the proposed model. One such enhancement would involve combining the FE-Transformer with remote sensing data. While environmental data is hard to gather in remote areas, remotely sensed data in multiple frequency bands are readily available. These data can be merged with scarce measurements using the PINN method to train the model with more data, potentially improving its prediction accuracy.

## 4. Conclusion

Distribution of TS in the oceans is highly spatially and temporally dynamic, and is usually modeled using a complex set of PDEs. In this paper, we proposed a novel FEA-inspired DNN to predict the sediment distribution in the wide GBR. In our proposed method, the GBR is meshed into small elements. For each element, a stiffness matrix is calculated using the historical training dataset. This stiffness matrix along with its relevant TS element are fed to the proposed FE-Transformer to predict the next TS frame in a forecasting horizon between 1 to 7 days. For training the proposed FE-Transformer model, we employed the PINN technique to merge measured TS data from AIMS MMP with PDE simulated data from *eReefs*. Through integrating the FEA concept with the state-of-the-art Transformer network, the trained FE-Transformer model was able to efficiently learn long-term TS dependencies. This resulted in highly accurate TS frame predictions with a small MSE. Besides, the model's output frame had great $R^2$, which is an indicator of the model's unblurred prediction. Our proposed model will be beneficial in reliably predicting TS concentration and distribution across the GBR. This will aide in better managing and planning water quality targets set for the GBR.

## Software and Data Availability

The observational TS values are gathered from the GBR Marine Park Authority MMP, which is led by Australian institute of marine science (AIMS-MMP, 2021). The PDE solutions for the TS distribution in the GBR are obtained from the *eReefs* modeling suite. To elaborate, the *eReefs* regional GBR4 biogeochemical simulation data are downloaded from the AIMS website (AIMS-eReefs, 2021). This data is publicly available as NetCDF files (nc-files). The enthusiast reader might programmatically select a geographical region, a spatial period, and a desirable parameter (i.e., hydrodynamical, sedimental, biogeochemical, or optical variables) before downloading the relevant data. The proposed FE-Transformer model is implemented by Keras APIs of TensorFlow 2.5.0 in Python 3.8.

## Acknowledgment

The *eReefs* model simulations were produced as part of the *eReefs* project (eReefs.info), a collaboration between the Science Industry Endowment Fund (SIEF), the Commonwealth Scientific Industrial Research Organisation (CSIRO),

the Australian Institute of Marine Science (AIMS), the Bureau of Meteorology (BOM), and the Great Barrier Reef Foundation (GBRF), with support from BHP Billinton Mitsubishi Alliance, the Australian and Queensland governments, and with observations obtained through the Integrated Marine Observing System (IMOS). The marine monitoring program is funded by the GBR Marine Park Authority, and data provided by AIMS.

## References

AIMS-eReefs, 2021. A mirror eReefs model data, with custom aggregations performed by the eAtlas team. www.aims.gov.au.

AIMS-MMP, 2021. Australian institute of marine science: water quality particulate and dissolved nutrient data. www.aims.gov.au.

Ali-Khan, M.Y., Tian, F., Hasan, F., Chakrapani, G.J., 2019. Artificial neural network simulation for prediction of suspended sediment concentration in the River Ramganga, Ganges Basin, India. International Journal of Sediment Research 34, 95–107.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv .

Baird, M.E., Mongin, M., Skerratt, J., Margvelashvili, N., Tickell, S., Steven, A.D.L., Robillot, C., Ellis, R., Waters, D., Kaniewska, P., Brodie, J., 2021. Impact of catchment-derived nutrients and sediments on marine water quality on the Great Barrier Reef: an application of the eReefs marine modelling system. Marine Pollution Bulletin 167, 112297–112313.

Baird, M.E., Wild-Allen, K.A., Parslow, J., Mongin, M., Robson, B., Skerratt, J., Rizwi, F., Soja-Woźniak, M., Jones, E., Herzfeld, M., Margvelashvili, N., Andrewartha, J., Langlais, C., Adams, M.P., Cherukuru, N., Gustafsson, M., Hadley, S., Ralph, P.J., Rosebrock, U., Schroeder, T., Laiolo, L., Harrison, D., Steven, A.D.L., 2020. CSIRO Environmental Modelling Suite (EMS): scientific description of the optical and biogeochemical models (vB3p0). Geoscientific Model Development 13, 4503–4553.

Bhagat, S.K., Tiyasha, T., Awadh, S.M., Tung, T.M., Jawad, A.H., Yaseen, Z.M., 2021. Prediction of sediment heavy metal at the Australian Bays using newly developed hybrid artificial intelligence models. Environmental Pollution 268, 115663.

Bhattacharya, B., Solomatine, D.P., 2006. Machine learning in sedimentation modelling. Neural Networks 19, 208–214.

Coggan, A., Thorburn, P., Fielke, S., Hay, R., Smart, J.C.R., 2021. Motivators and barriers to adoption of improved land management practices: a focus on practice change for water quality improvement in Great Barrier Reef catchments. Marine Pollution Bulletin 170, 112628.

Cossu, A., Carta, A., Lomonaco, V., Bacciu, D., 2021. Continual learning for recurrent neural networks: An empirical evaluation. Neural Networks 143, 607–627.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: transformers for image recognition at scale. arXiv .

Guen, V.L., Thome, N., 2020. Disentangling physical dynamics from unknown factors for unsupervised video prediction, in: Proc. Computer Vision and Pattern Recognition (CVPR), Virtual. pp. 11474–11484.

Howarth, R.W., Marino, R., 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: evolving views over three decades. Limnology and oceanography 51, 364–376.

Kroon, F.J., Thorburn, P., Schaffelke, B., Whitten, S., 2016. Towards protecting the Great Barrier Reef from land-based pollution. Global change biology 22, 1985–2002.

MacNeil, M.A., Mellin, C., Matthews, S., Wolff, N.H., McClanahan, T.R., Devlin, M., Drovandi, C., Mengersen, K., Graham, N.A.J., 2019. Water quality mediates resilience on the Great Barrier Reef. Nature Ecology & Evolution 3, 620–627.

Margvelashvili, N., Andrewartha, J., Baird, M., Herzfeld, M., Jones, E., Mongin, M., Rizwi, F., Robson, B.J., Skerratt, J., Wild-Allen, K., Steven, A., 2018. Simulated fate of catchment-derived sediment on the Great Barrier Reef shelf. Marine Pollution Bulletin 135, 954–962.

Margvelashvili, N.Y., Herzfeld, M., Rizwi, F., Mongin, M., Baird, M.E., Jones, E., Schaffelke, B., King, E., Schroeder, T., 2016. Emulator-assisted data assimilation in complex models. Ocean Dynamics 66, 1109–1124.

Mathieu, M., Couprie, C., LeCun, Y., 2016. Deep multi-scale video prediction beyond mean square error, in: Proc. 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico. pp. 1–14.

McCloskey, G.L., Baheerathan, R., Dougall, C., Ellis, R., Bennett, F.R., Waters, D., Darr, S., Fentie, B., Hateley, L.R., Askildsen, M., 2021. Modelled estimates of fine sediment and particulate nutrients delivered from the Great Barrier Reef catchments. Marine pollution bulletin 165, 112163.

McKergow, L.A., Prosser, I.P., Hughes, A.O., Brodie, J., 2005. Sources of sediment to the Great Barrier Reef world heritage area. Marine Pollution Bulletin 51, 200–211.

Mehri, Y., Nasrabadi, M., Omid, M.H., 2021. Prediction of suspended sediment distributions using data mining algorithms. Ain Shams Engineering Journal 1, 1.

NOAA, 2021. United States National Oceanic and Atmospheric Administration: Coral reef ecosystems. www.noaa.gov.

R. Reichelt *et al.*, 2010. Water quality guidelines for the Great Barrier Reef marine park. Second ed., Great Barrier Reef Marine Park Authority.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707.

Reef-2050, 2021. Australian and Queensland government: The Reef 2050 plan. environment.gov.au.

Sergio, G.C., Lee, M., 2021. Stacked DeBERT: All attention in incomplete data for text classification. Neural Networks 136, 87–96.

Skerratt, J.H., Mongin, M., Baird, M.E., Wild-Allen, K.A., Robson, B.J., Schaffelke, B., Davies, C.H., Richardson, A.J., Margvelashvili, N., Soja-Wozniak, M., Steven, A.D.L., 2019. Simulated nutrient and plankton dynamics in the Great Barrier Reef (2011–2016). Journal of Marine Systems 192, 51–74.

Stevena, A.D.L., Baird, M.E., Brinkman, R., Car, N.J., Cox, S.J., Herzfeld, M., Hodge, J., Jonesa, E., King, E., Margvelashvilia, N., Robillotd, C., Robson, B., Schroedera, T., Skerratt, J., Tickella, S., Tutejae, N., Wild-Allena, K., Yu, J., 2019. eReefs: an operational information system for managing the Great Barrier Reef. Journal of Operational Oceanography 12, 12–28.

Su, X., Wang, G., Li, Q., 2020. Prediction method for transformer state based on GRU network, in: Proc. Industrial and Commercial Power System Asia, Weihai, China. pp. 1751–1755.

Surana, K.S., Reddy, J.N., 2017. The finite element method for boundary value problems: mathematics and computations. First ed., CRC press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. arXiv .

Waterhouse, J., Brodie, J., Tracey, D., Smith, R., VanderGragt, M., Collier, C., Petus, C., Baird, M., Kroon, F., Mann, R., et al., 2020. Land use impacts on the Great Barrier Reef water quality and ecosystem condition. The Scientific Consensus Statement by Queensland Government .

Zeyer, A., Bahar, P., Irie, K., Schlüter, R., Ney, H., 2020. A comparison of transformer and LSTM encoder decoder models for ASR, in: Proc. Automatic Speech Recognition and Understanding Workshop (ASRU), Singapore. pp. 8–15.

Zhou, Y., Dong, H., El Saddik, A., 2020. Deep learning in next-frame prediction: a benchmark review. IEEE Access 8, 69273–69283.