

REQUEST FOR AN EMI R&D PROJECT 2027–2029

Country/Organisation: The Netherlands
Principal Investigator¹: Paulina Kindermann (p.e.kindermann@tudelft.nl)

Affiliation: Delft University of Technology

Address:
Stevinweg 1
2628 CN Delft
The Netherlands

Other researchers:
Henk van den Brink (henk.van.den.brink@knmi.nl)
José A. A. Antolínez (j.a.a.antolinez@tudelft.nl)

Project Title: Storm surge and wave forecasting for the North Sea from meteorological characteristics using the SEAS5–WAQUA–HurryWave dataset

To make changes to an existing project please submit an amended version of the original form

If this is a continuation of an existing project, please state the computer project account assigned previously.	spnlkind	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2027	
Would you accept support for 1 year only, if necessary?	YES <input checked="" type="checkbox"/>	NO <input type="checkbox"/>

Computer resources required for project year:	2027	2028	2029
High Performance Computing Facility [SBU]	500,000	-	-
Graphics Processing Unit Cluster-A [GBU]	1,000,000	-	-
Graphics Processing Unit Cluster-B [GBU]	-	-	-
Accumulated data storage (total archive volume) ² [GB]	20 TB	20 TB	-

¹ The Principal Investigator will act as contact person for this EMI R&D Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.

² These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.

³ The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

EWC resources required for project year:		2027	2028	2029
Number of vCPUs	[#]	-	-	-
Total memory	[GB]	-	-	-
Storage	[GB]	-	-	-
Number of vGPUs ³	[#]	-	-	-

Principal Investigator:

Paulina Kindermann

Project Title:

Storm surge and wave forecasting for the North Sea from meteorological characteristics using the SEAS5–WAQUA–HurryWave dataset

Extended abstract

All EMI R&D Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The completed form should be submitted/uploaded at <https://www.ecmwf.int/en/research/special-projects/special-project-application/special-project-request-submission>.

Following submission by the relevant Member State the EMI R&D Project requests will be published on the ECMWF website and evaluated by ECMWF and its Scientific Advisory Committee. The requests are evaluated based on their scientific and technical quality, and the justification of the resources requested. Previous EMI R&D Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.

Requests exceeding 5,000,000 SBU should be more detailed (3-5 pages).

Introduction

Coastal flood defences in the Netherlands must be designed to withstand hydraulic boundary conditions with return periods of up to 10,000 years. These conditions result from the combined effects of astronomical tides, wind-driven surges, and waves, which are all primarily forced by wind over the North Sea. Reliably characterising extremes in these variables, and understanding their mutual dependencies, is therefore essential for robust flood defence design.

Traditional approaches based on statistical extrapolation of observational records face well-known limitations: records are short (~100 years), interactions between physical processes are often neglected, and results are restricted to observation sites. To overcome these limitations, KNMI has previously used ECMWF's SEAS5 seasonal ensemble forecast to drive the WAQUA-DCSMv5 surge model, producing a large, consistent dataset of synthetic wind and sea level conditions equivalent to approximately 9,000 years in the current climate. This dataset has already been applied for extreme value analysis of sea levels and wind speed along the Dutch coast (de Valk & van den Brink, 2023).

The first phase of this project (approved as an ECMWF Special Project in 2026), focused on extending this dataset with wave conditions by running the HurryWave wave model (van der Lugt et al., 2024) driven by the SEAS5-WAQUA wind and sea level fields. Preliminary validation has confirmed that HurryWave produces sufficiently accurate results at offshore locations with water depths greater than 15 metres.

In parallel, the ECMWF resources from the first project phase are being used to begin developing a machine learning model for storm surge forecasting, trained on the SEAS5–WAQUA dataset. This model is currently able to predict storm surges at Dutch coastal locations with an R^2 of 0.9 and an RMSE of approximately 15–20 cm.

Scientific goals

This extension requests resources to complete the two remaining objectives of the overall project:

1. Integrate wave conditions into the machine learning forecasting model. Once the HurryWave-SEAS5 wave dataset is complete, it will be incorporated into the training dataset alongside the existing wind and surge fields. The goal is to develop a model capable of jointly forecasting extreme storm surges and waves from meteorological input alone. This requires extending the current architecture from point-based surge predictions to

spatial fields, and exploring architectures that can capture the physical relationships between wind forcing, surge, and wave generation, such as Graph Neural Networks (GNNs) and Fourier Neural Operators.

2. Analyse statistical dependencies between meteorological drivers, storm surges, tides, and waves. The combined SEAS5-WAQUA-HurryWave dataset, spanning approximately 9,000 synthetic years, is very suitable to characterise the joint statistical behaviour of these variables under extreme conditions. Key questions include: which meteorological characteristics (storm track, size, duration, wind direction) drive the most severe combined surge-wave events? What are the exceedance probabilities of joint extremes? How do surge and wave extremes co-vary spatially along the Dutch coast and beyond?

Together, these two objectives will result in a physically consistent, data-driven framework for deriving hydraulic boundary conditions that accounts for the full range of interactions between meteorological forcing, surges, tides, and waves.

Methodology

The machine learning framework will be trained on the full SEAS5-WAQUA-HurryWave dataset. Storm-active periods are selected from the dataset to focus training on the conditions most relevant for extremes. Model architectures to be tested include CNNs, CNN-LSTMs, GNNs, and Fourier Neural Operators, with physics-informed components incorporated where possible (e.g. wind stress–surge relationships). Model performance will be evaluated against observational records at Dutch coastal locations.

Dependency analysis will be carried out using the complete synthetic dataset, combining extreme value statistics, copula-based methods, and data-driven approaches to characterise joint distributions of surge, wave height, wave period, and meteorological drivers.

Justification of resources

The primary computational demand of this extension is GPU-intensive machine learning model training on the full combined dataset. Experience from the first project phase (first half of 2026) has shown that the machine learning model performs significantly better on GPU than CPU. Given that the HurryWave-SEAS5 wave dataset is still being completed, it is unlikely that all planned work can be finished within 2026. We therefore request a partial reallocation of remaining 2026 resources to 2027, to ensure sufficient resources are available for full model training and dependency analysis once the complete dataset is in place.

So far, approximately 150,000 GBU-A have been used in 2026, and model training has only just begun. In 2027, we plan to incorporate wave conditions into the model, for which we expect GPU demand comparable to the 2026 allocation (1,000,000 GBU-A). Regarding SBU, it has become clear in the first half of 2026 that the available budget exceeds what will be needed, partly because GPU resources proved more efficient than anticipated, and partly because some activities were carried out using other infrastructure. We therefore expect that only a relatively small amount of SBU will be required in 2027 to finalise postprocessing (estimated 500,000 SBU). Finally, we request continued access to data storage through the end of 2028, as the combined dataset needs to remain accessible throughout the analysis and reporting phase.

Application

The resulting framework will directly support the derivation of hydraulic boundary conditions for Dutch flood defence design, enabling a more robust treatment of joint surge-wave extremes than currently possible. The spatial extent of the dataset also makes results relevant for coastal applications in Belgium, the United Kingdom, and Denmark.