REQUEST FOR A SPECIAL PROJECT 2024–2026

MEMBER STATE:	CROATIA
Principal Investigator[1]**:**	Cléa Denamiel
Affiliation:	Ruđer Bošković Institute (RBI)
Address:	Bijenička cesta 54 10 000 Zagreb, Croatia
Other researchers:	Frédéric Dias (University College Dublin - UCD, Ireland), Serge Guillas (University College London - UCL, UK), Xun Huan (University of Michigan, USA), Anne Mangeney (University Paris Cité, France) and Iva Tojčić (RBI, Croatia)
Project Title:	Exploring the potential of uncertainty quantification and machine learning techniques to forecast rare extreme events

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.	**SPCRDENA**	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2024	
Would you accept support for 1 year only, if necessary?	YES ☒	NO ☐

Computer resources required for project year:	**2024**	**2025**	**2026**
High Performance Computing Facility [SBU]	20,000,000	20,000,000	20,000,000
Accumulated data storage (total archive volume)[2] [GB]	25,000	50,000	75,000

EWC resources required for project year:	**2024**	**2025**	**2026**
Number of vCPUs [#]	/	/	/
Total memory [GB]	/	/	/
Storage [GB]	/	/	/
Number of vGPUs[3] [#]	/	/	/

Continue overleaf.

---

[1] The Principal Investigator will act as contact person for this Special 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.

Principal Investigator: Cléa Denamiel

Project Title: Exploring the potential of uncertainty quantification and machine learning techniques to forecast rare extreme events

Extended abstract

1) Motivation and Problem Identification

Rare extreme events are particularly difficult to forecast and can cause serious infrastructure damage and human casualties when occurring without warning. For example, the 18th of August 2022 severe storms occurred in a swath from Menorca (Balearic Islands, Spain) through Corsica (France), northern Italy, Slovenia, Austria, and southern Czechia (European Severe Storms Laboratory, 2022). In total, 12 people died and 106 people were injured by wind and hail (up to 11 cm of diameter). All fatalities, and most of the injuries, were caused by a long-lived convective system, also known as a Derecho, that produced extremely severe wind gusts (up to 62.2 m/s recorded in Corsica) and rapidly moving showers (Fig. 1). In Corsica, this Derecho, which was not forecasted by Meteo France, resulted in the death of 5 people and, at one point, in up to 350 people being reported missing as pleasure boats had capsized or been thrown adrift.

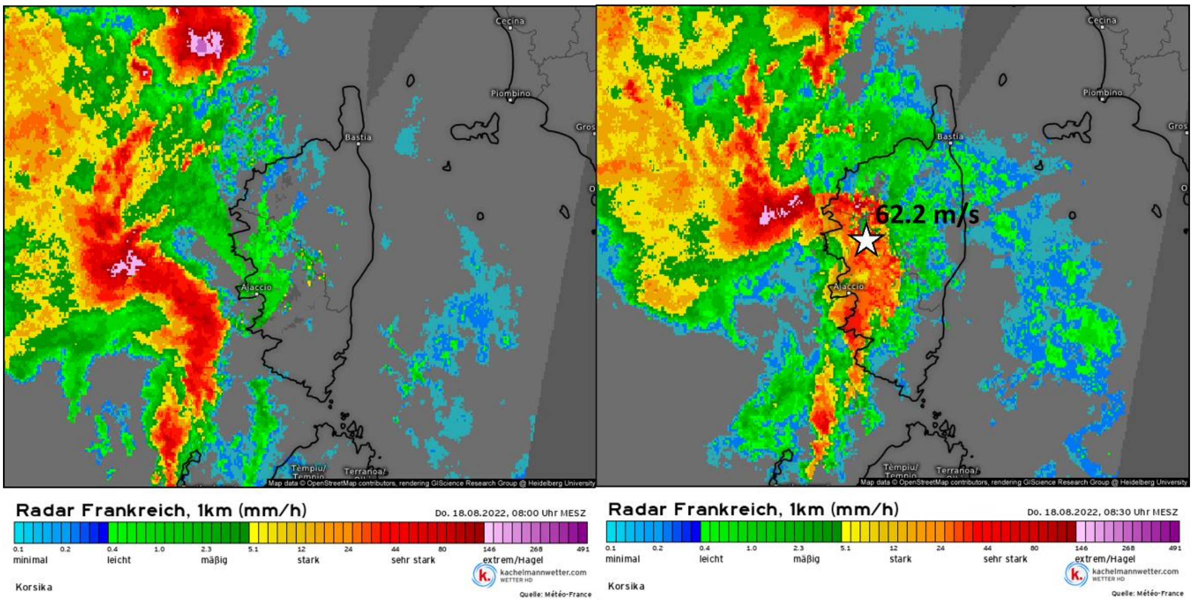

Figure 1. Time evolution of the Derecho over Corsica as shown by radar between 06:00 and 06:30 UTC. The location of the most severe wind gust is shown by the star. Sources: kachelmannwetter.com, meteologix.com, essl.org/cms/the-derecho-and-hailstorms-of-18-august-2022.

Another type of rare events difficult to forecast are tsunamis driven by atmospheric acoustic-gravity waves – or sonic boom related meteotsunami waves – such as the ones generated after the explosive eruption of the Hunga Tonga–Hunga Ha'apai (HTHH) volcano in January 2022 (e.g., Adam, 2022; Amores et al., 2022; Harrison, 2022; Matoza et al., 2022; Omira et al., 2022; Wright et al., 2022; Winn et al., 2023). Only rare catastrophic events such as volcano explosions (Choi et al., 2003) or asteroid impacts (Chapman and Morrison, 1994; Morgan et al., 2022), even occurring inland (e.g., Tunguska explosion; Chyba et al., 1993), can produce sonic booms capable of generating worldwide acoustic-gravity waves (Yeh and Liu, 1974) driving meteotsunamis. The most prominent of these acoustic-gravity waves are the Lamb waves (Lamb, 1911; Bretherton, 1969) that propagate horizontally in the atmosphere with a speed close to the mean sound speed (about 318 m/s; Dragoni and Santoro, 2020). They can circle the globe multiple times (Press and Harkrider, 1966) and are associated with surface pressure oscillations of several hectopascals (hPa) per minute driving

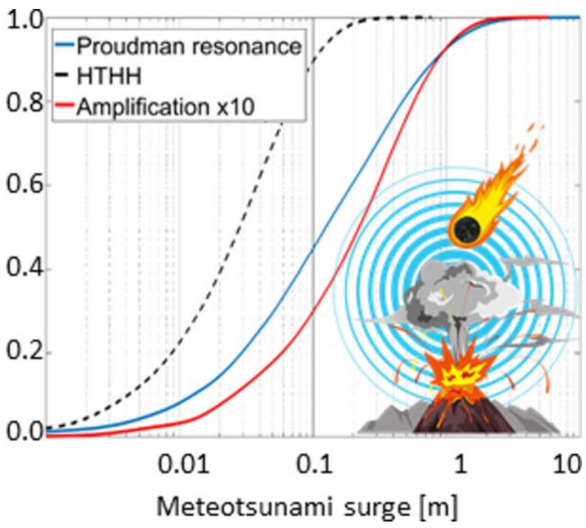

Figure 2. Cumulative density functions of the meteotsunami surges obtained under deep-Proudman resonance and for tenfold amplification of the HTHH Lamb wave amplitudes. From Denamiel et al. (2023).

planetary meteotsunami waves that could generate surges reaching 1 to 10 m along more than 7 % of the worldwide coastlines for intense explosions generating higher Lamb waves than the HTHH event (Fig. 2).

Finally, tsunamis triggered by submarine landslides are also extremely difficult to forecast in real-time as the precise location and volume of such landslides are not easy to predict or observe. As explained and investigated in Poulain et al. (2022), since May 2018, Mayotte Island has been experiencing seismo-volcanic activities that could potentially trigger submarine landslides driving tsunamis. Further, in this area, the tsunami travel time to the coast is very short (a few minutes) and the tsunami is not necessarily preceded by a sea withdrawal. Fast tsunami simulations are thus needed for evacuation plans and early-warning systems in Mayotte.

However, the forecast of such rare events (e.g., Derechos, meteotsunamis, tsunamis) with traditional ensemble methods (if even possible) would be extremely costly in terms of numerical resources and would reach the limit of what state-of-the-art numerical models can simulate in forecast mode. Consequently, the question of whether or not the cost of (not) forecasting these events is acceptable (i.e., human casualties vs. modelling capabilities and efforts) can be raised. In this project, we propose to explore the potential of using uncertainty quantification (UQ) and machine learning (ML) techniques (e.g., surrogate models/emulators) for the forecast of rare extreme events and to leverage the costs and benefits of such an approach.

2) Previous Numerical Modelling Efforts

For rare extreme events, a good balance between model accuracy (with high resolution and detailed physics) and real-time stochastic forecasts is hard to achieve (Veeramony et al., 2012). This is why the interest in applying surrogate models and emulators based on UQ/ML techniques has recently been growing within the geoscience community (Formaggia et al., 2013; Wang et al., 2016; Sraj et al., 2014; Giraldi et al., 2017; Bulthuis et al., 2019; Salmanidou et al., 2021). Based on these studies, Denamiel et al. (2021) derived a surrogate-based early warning system framework (Fig. 3) which generalizes, optimizes and promotes the use of UQ/ML for the forecast of rare extreme events.

In terms of the different ways to build the surrogate models/emulators, several avenues have already been explored. On one hand, the use of Gaussian Processes (GPs; Rasmussen and William, 2006) is extremely common in geosciences. In particular, Ming and Guillas (2021) implemented an iterative procedure to construct linked Gaussian processes as surrogate models for any feed-forward systems of computer models. They also introduce an adaptive design algorithm that could increase the approximation accuracy of linked Gaussian process surrogates with reduced computational costs on running expensive computer systems, by allocating runs and refining emulators of individual sub-models based on their heterogeneous functional complexity. On the other hand, Denamiel et al. (2019, 2020) developed, within a prototype early warning system, a meteotsunami surge surrogate model based on generalized Polynomial Chaos Expansions (gPCE; Xiu and Karniadakis, 2002) which has proven to be extremely useful and reliable during recent events spanning between 2014 and 2020. In particular, during the 2020 multi-meteotsunami event, if the prototype early warning system had been fully operational, the Croatian coastal communities would have been warned, at least a day in advance, of the meteotsunami surges forecasted with the surrogate model (Tojčić et al., 2021).

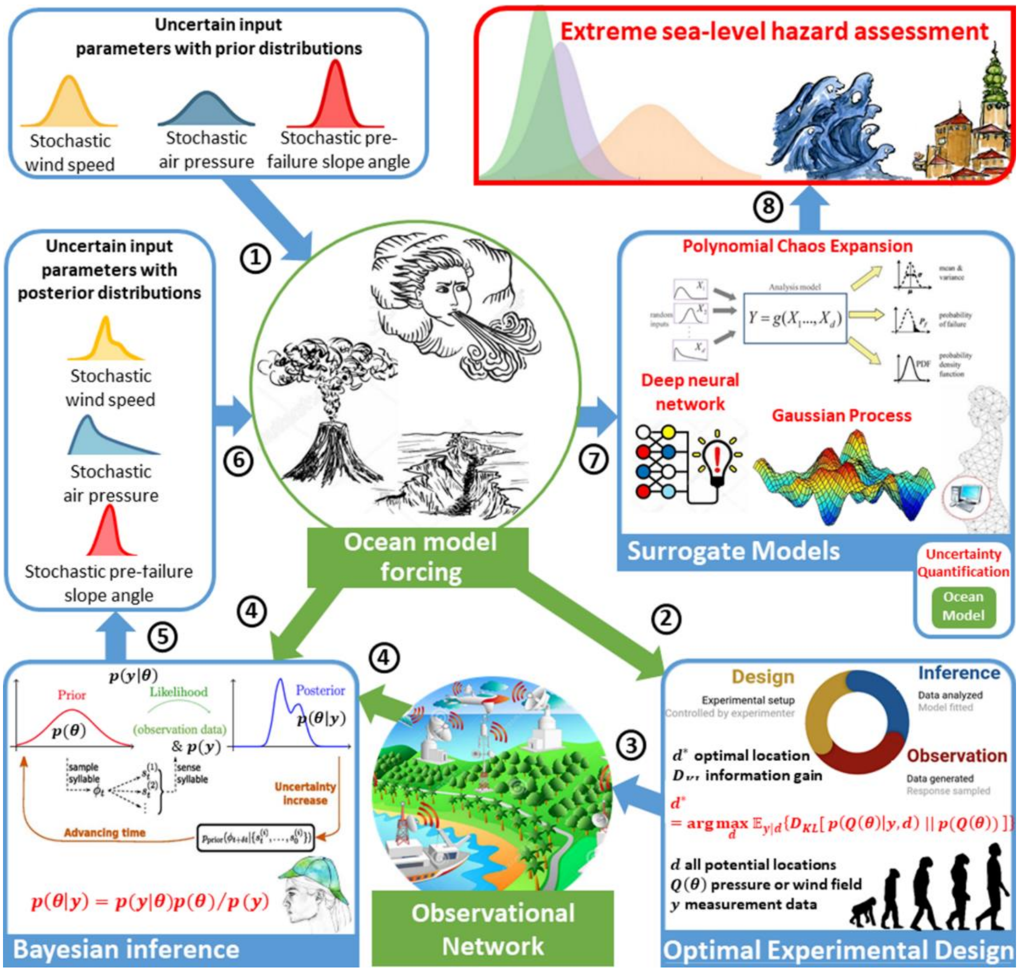

Figure 3. Rare extreme event forecast based on uncertainty quantification and optimization engineering methods. Drawing of the flooded city adapted from Frits Ahlefeld: https://fritsahlefeldt.com/2019/01/24/not-ready-city-facing-flooding. From Denamiel et al. (2021).

In terms of the capability of state-of-the-art models to forecast or even to hindcast rare extreme events, the 2022-08-18 Derecho over Corsica perfectly illustrates the many challenges faced by numerical modelers.

Table 1. Summary of the Corsica island Sea and Coast (CoriSC) modelling suite set-up.

Module	Coupling	Domain	Model	Horizontal resolution	Region	Initial and boundary conditions (frequency)
Basic	COAWST online	Atmosphere	WRF	15 km	Western Mediterranean Sea	ERA5 or IFS (6-hourly)
			WRF	3 km	Corsica-Adriatic	Two-way nesting (30 sec)
		Ocean	ROMS SWAN	3 km	Corsica-Adriatic	MEDSEA (daily)
			ROMS SWAN	1 km	Wider Corsica	One-way nesting (50 s)
Nearshore	Offline	Atmosphere	WRF	1 km	Wider Corsica	WRF 3-km (hourly)
		Ocean	ADCIRC SWAN	up to 50 m	Corsica	ROMS-SWAN 1-km (hourly)

First, the modular approach successfully used in the Adriatic Sea and Coast (AdriSC; Denamiel et al., 2019) modelling suite has been replicated in the Corsica island Sea and Coast (CoriSC) modelling suite to downscale the atmosphere-ocean dynamics from global models to (sub)-kilometre-scale (Table 1). The CoriSC modelling suite is thus composed of a basic module which provides hourly kilometre-scale atmosphere-ocean-wave dynamics and of a nearshore module only ran during atmospherically driven extreme events. In the basic module (Table 1), the kilometre-scale coastal circulation is derived using the Coupled Ocean-Atmosphere-Wave-Sediment Transport (COAWST) modelling system (Warner et al., 2010). Hourly results are produced at resolutions up to 3-km for the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005) in the atmosphere and 1-km for the Region Ocean Modeling System (ROMS; Shchepetkin & McWilliams, 2009) and the Simulating WAve Nearshore (SWAN) model in the ocean (Table 1). The nearshore module (Table 1) further downscales the results of the CoriSC basic module and is based on the offline coupling between the WRF model at 1 km resolution and the Advanced CIRCulation model (ADCIRC-SWAN; Luettich et al., 1991) at up to 50 m resolution along the Corsican coasts.

Second, several simulations have been performed in both forecast mode (forced by ECMWF IFS) and hindcast mode (forced by ERA5; Hersbach et al., 2020) and starting the simulations at different dates: 2022-08-14 00:00:00 UTC, 2022-08-15 00:00:00 UTC, 2022-08-16 00:00:00 UTC and 2022-08-17 00:00:00 UTC. The WRF 1 km atmospheric results over Corsica extracted from the "best simulations" in hindcast (starting 2022-08-15) and forecast (starting 2022-08-14) mode are compared to ground-based meteo-station wind and rain observations in Figure 4 which perfectly illustrates the difficulty to reproduce the exact location, timing and strength of the Derecho, even in hindcast mode.

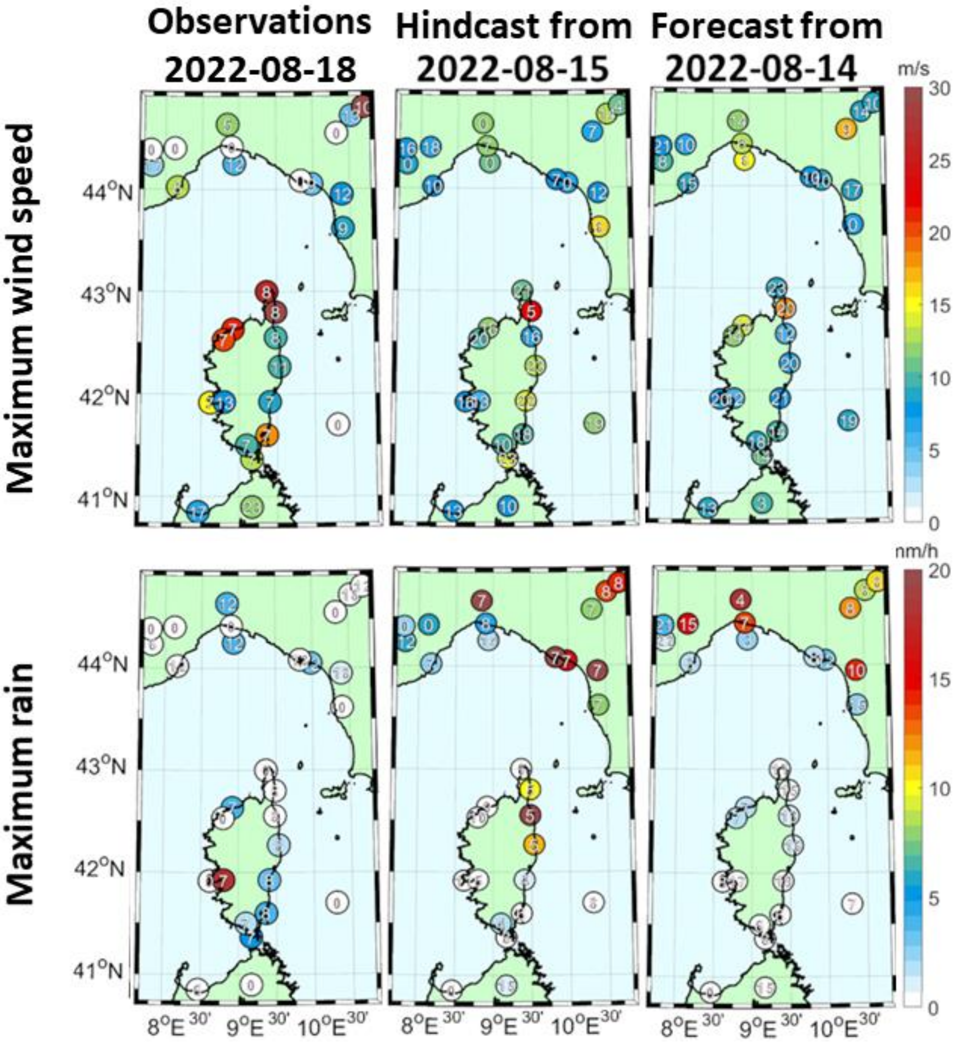

Figure 4. Comparison of the values and timing (number inside the circles in hour) of the maximum wind speed and rain during the 2022-08-18 Derecho event over Corsica: ground-based meteo-station observations (left panels) and best simulation results of WRF 1 km CoriSC model in both hindcast (middle panels) and forecast (right panels) modes.

3) Proposed Work

The main advantage of surrogate models/emulators is that they can be used to account for the uncertainty of the forcing (e.g., location, timing, strength of the Derecho over Corsica). Their main drawback is that their accuracy highly dependents on both the available training data (i.e. good for interpolation but not extrapolation) and their unique mathematical approximation forms (Laloy and Jacques, 2019).

Up to three surrogate families will thus be tested in this project: generalized polynomial chaos expansions and Gaussian processes (UQ techniques) as well as deep neural networks (DNNs; Goodfellow et al., 2016; ML techniques). For instance, it is anticipated that gPCEs and GPs work well when our training dataset is small, and that DNNs excel as the dataset grows over time. Selection will also be dependent on the nature of the quantity of interest, for example gPCEs would not be suitable for discontinuous quantities but DNNs may achieve good accuracy, again caveat on having sufficient training data. When possible (pending available numerical resources), multiple surrogates/emulators will be combined and compared for each of the selected rare extreme event: Derechos in Corsica, sonic boom related planetary meteotsunami waves and tsunamis caused by landslides in Mayotte. The reliability (in terms of the accuracy of the results) and the efficiency (in terms of numerical cost) of the UQ/ML techniques will be leveraged and, when possible, the surrogate model with the “best performance” will be selected depending of the rare extreme event.

Given the challenging nature of the proposal objectives, the numerical simulations needed to build the surrogates/emulators will be spread on the resources of the different partners of the project. For example, the simulations for the Mayotte landslides will be run by Anne Mangeney (no cost to ECMWF) and some of the simulations for the Corsica Derechos can also be run on resources provided by Serge Guillas. However, all simulations dealing with the sonic boom related meteotsunami events will be run on the ECMWF HPC. The targeted use of the ECMWF resources is thus 20 000 SBUs and 25 000 GB of storage per year and the modelling strategy of the project consists in two points:

Surrogate models/emulators for Derechos over Corsica

For this task, the surrogate models/emulators will be built with the CoriSC WRF 1 km model. Based on previous experience with the new ATOS HPC at ECMWF, the following system should be run with a maximum of 2 Cores in order to achieve reasonable time of execution.

Computing resources needed: About **2 Cores * 1/2 day * 86400 s * P * 672 simulations = ~ 30 000 000 SBUs** and up to **37 500 GB** are planned to be used in the framework of the ECMWF special project.

Surrogate models/emulators for sonic boom related planetary meteotsunami events

For this task, the surrogate models/emulators will be built with the barotropic version of the Transient Inertia-Gravity And Rossby wave dynamics (TIGAR) model (Vasylkevych and Žagar, 2021) and the Atmospheric Tsunamis Associated with Lamb waves (ATAL) model in the ocean (Denamiel et al., 2023). A T170 computational grid is used in TIGAR while ATAL is based on an unstructured mesh with resolutions ranging from 20 km in the deep ocean to 1.5 km along the coastlines. Based on the previous experience with the old CRAY HPC at ECMWF, the following system should be run with 2 Cores in order to achieve reasonable time of execution.

Computing resources needed: Following our first estimate, **2 Cores * 1/12 day * 86400 s * P * 4032 simulations = ~ 30 000 000 SBUs** and up to **37 500 GB** will be used in the framework of the ECMWF special project to fully cover these runs.

References

Adam, D. (2022). Tonga volcano eruption created puzzling ripples in Earth’s atmosphere. Nature, 601, 497. https://doi.org/10.1038/d41586-022-00127-1

Amores, A., Monserrat, S., Marcos, M., Argüeso, D., Villalonga, J., Jordà, G., Gomis, D. (2022). Numerical simulation of atmospheric Lamb waves generated by the 2022 Hunga-Tonga volcanic eruption. Geophys. Res. Lett., 49, e2022GL098240. https://doi.org/10.1029/2022GL098240

Bretherton, F. (1969). Lamb waves in a nearly isothermal atmosphere. Quart. J. Roy. Meteor. Soc., 95, 754–757. https://doi.org/10.1002/qj.49709540608

Bulthuis, K., Arnst, M., Sun, S., Pattyn, F. (2019). Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change. The Cryosphere, 13(4), 1349–1380. https://doi.org/10.5194/tc-13-1349-2019

Chapman, C.R., Morrison, D. (1994). Impacts on the Earth by asteroids and comets: Assessing the hazard. Nature, 367, 33–40. https://doi.org/10.1038/367033a0

Choi, B.H., Pelinovsky, E., Kim, K.O., Lee, J.S. (2003). Simulation of the trans-oceanic tsunami propagation due to the 1883 Krakatau volcanic eruption. Nat. Hazards Earth Syst. Sci., 3, 321–332. https://doi.org/10.5194/nhess-3-321-2003

Chyba, C., Thomas, P., Zahnle, K. (1993). The 1908 Tunguska explosion: Atmospheric disruption of a stony asteroid. Nature, 361, 40–44. https://doi.org/10.1038/361040a0

Denamiel, C., Šepić, J., Huan, X., Bolzer, C., Vilibić, I. (2019). Stochastic surrogate model for meteotsunami early warning system in the eastern Adriatic Sea. J. Geophys. Res. Oceans 124, 8485-8499. https://doi.org/10.1029/2019JC015574

Denamiel, C., Huan, X., Šepić, J., Vilibić, I. (2020). Uncertainty propagation using polynomial chaos expansions for extreme sea level hazard assessment: The case of the eastern Adriatic meteotsunamis. J. Phys. Oceanogr. 50(4), 1005–1021. https://doi.org/10.1175/JPO-D-19-0147.1D

Denamiel, C., Huan, X., Vilibić, I. (2021). Conceptual design of extreme sea-level early warning systems based on uncertainty quantification and engineering optimization methods. Front. Mar. Sci. 8, 650279. https://doi.org/10.3389/fmars.2021.650279

Denamiel, C., Vasylkevych, S., Žagar, N., Zemunik, P., & Vilibić, I. (2023). Destructive Potential of Planetary Meteotsunami Waves beyond the Hunga Tonga–Hunga Ha‘apai Volcano Eruption, Bull. Amer. Meteor. Soc., 104(1), E178-E191. https://doi.org/10.1175/BAMS-D-22-0164.1

Dragoni, M., Santoro, D. (2020). A model for the atmospheric shock wave produced by a strong volcanic explosion. Geophys. J. Int., 222, 735–742. https://doi.org/10.1093/gji/ggaa205

European Severe Storms Laboratory (2022). The derecho and hailstorms of 18 August 2022. Posted on 25 Aug 2022 by Tomas Pucik. https://www.essl.org/cms/the-derecho-and-hailstorms-of-18-august-2022/

Formaggia, L., Guadagnini, A., Imperiali, I., Lever, V., Porta, G., Riva, M., et al. (2013). Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. Computational Geosciences, 17(1), 25–42. https://doi.org/10.1007/ s10596-012-9311-5

Giraldi, L., Le Maître, O. P., Mandli, K. T., Dawson, C. N., Hoteit, I., & Knio, O. M. (2017). Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Computational Geosciences, 21(4), 683–699. https://doi.org/10.1007/s10596- 017-9646-z

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning. Cambridge, MA, MIT Press.

Harrison, G. (2022). Pressure anomalies from the January 2022 Hunga Tonga–Hunga Ha‘apai eruption. Weather, 77, 87–90. https://doi.org/10.1002/wea.4170

Hersbach, H., et al. (2020). The ERA5 Global Reanalysis. Q. J. Roy. Meteorol. Soc., 146, 1999–2049. https://doi.org/10.1002/qj.3803

Lamb, H., (1911). On atmospheric oscillations. Proc. Roy. Soc. London, A84, 551–572. https://doi.org/10.1098/rspa.1911.0008

Luettich, R.A., Birkhahn, R.H., Westerink, J.J. (1991). Application of ADCIRC-2DDI to Masonboro Inlet. A brief numerical modeling study. U.S.Army Engineer Waterways Experiment Station Rep., 27 pp., https://adcirc.org/wp-content/uploads/sites/2255/2018/11/1991_Luettich02.pdf

Ming, D., Guillas, S. (2021). Linked Gaussian Process Emulation for Systems of Computer Models using Matérn Kernels and Adaptive Design. SIAM/ASA Journal on Uncertainty Quantification, 9(4). https://doi.org/10.1137/20M1323771

Morgan, J.V., Bralower, T.J., Brugger, J., Wünnemann, K. (2022). The Chicxulub impact and its environmental consequences. Nat. Rev. Earth Environ., 3, 338–354. https://doi.org/10.1038/s43017-022-00283-y

Omira, R., et al. (2022). Global Tonga tsunami explained by a fast-moving atmospheric source. Nature, 609(7928), 734-740. https://doi.org/10.1038/s41586-022-04926-4

Poulain, P., Le Friant, A., Pedreros, R., Mangeney, A., Filippini, A.G., Grandjean, G., Lemoine, A., Fernández-Nieto, E.D., Castro Díaz, M.J., Peruzzetto, M. (2022). Numerical simulation of submarine landslides and generated tsunamis: application to the on-going Mayotte seismo-volcanic crisis. Comptes Rendus. Géoscience, 354 (S2), 361-390. https://doi.org/10.5802/crgeos.138

Press, F., Harkrider, D. (1966). Air-sea waves from the explosion of Krakatoa. Science, 154, 1325–1327. https://doi.org/10.1126/science.154.3754.1325

Rasmussen, C.E., Williams, C.K.I. (2006). Gaussian Process for Machine Learning. MIT press, Cambridge, MA.

Salmanidou, DM; Beck, J; Pazak, P; Guillas, S. (2021). Probabilistic, high-resolution tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation. Natural Hazards and Earth System Sciences, 21 (12), 3789-3807. https://doig.org/10.5194/nhess-21-3789-2021

Shchepetkin, A.F., McWilliams, J.C. (2009). Computational kernel algorithms for fine-scale, multiprocess, longtime oceanic simulations. Handb. Numerical Anal., 14, 121–183. https://doi.org/10.1016/S1570-8659(08)01202-0

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D.M., Wang, W., and Powers, J. G. (2005). A description of the Advanced Research WRF Version 2, NCAR Technical Note NCAR/TN468+STR. https://doi.org/10.5065/D6DZ069T

Sraj, I., Mandli, K., Knio, O., Dawson, C. N., & Hoteit, I. (2014). Uncertainty quantification and inference of Manning's friction coefficients using DART buoy during the Tohoku tsunami. Ocean Modelling, 83, 82–97. https://doi.org/10.1016/j.ocemod.2014.09.001

Tojčić, I., Denamiel, C., Vilibić, I. (2021). Performance of the Adriatic early warning system during the multi-meteotsunami event of 11–19 May 2020: an assessment using energy banners. Nat. Hazards Earth Syst. Sci. 21, 2427–2446. https://doi.org/10.5194/nhess-21-2427-2021

Vasylkevych, S., Žagar, N. (2021). A high-accuracy global prognostic model for the simulation of Rossby and gravity wave dynamics. Quart. J. Roy. Meteor. Soc.,147, 1989–2007. https://doi.org/10.1002/qj.4006

Veeramony, J., Condon, A., Hebert, D. (2012). Effect of coupling wave and flow dynamics on hurricane surge and inundation. Proc. Twenty-second ISOPE, Rhodes, Greece, 1503–1507.

Wang, M., Wan, Z., & Huang, Q. (2016). A new uncertain analysis method for the prediction of acoustic field with random and interval parameters. Shock and Vibration, 16 pp, 2016, 1–16. https://doi.org/10.1155/2016/3693262

Warner, J.C., Armstrong, B., He, R., Zambon, J. (2010). Development of a coupled ocean-atmosphere-wave-sediment transport (COAWST) modeling system. Ocean Model., 35, 230–244. https://doi.org/10.1016/j.ocemod.2010.07.010

Winn, S.D., Sarmiento, A.F., Alferez, N., Touber, E. (2023). Two-way coupled long-wave isentropic ocean-atmosphere dynamics. J. Fluid Mech., 959, A22. https://doi.org/10.1017/jfm.2023.131

Wright, C.J., et al. (2022). Surface-to-space atmospheric waves from Hunga Tonga-Hunga Ha'apai eruption. Nature, 609, 741–746. https://doi.org/10.1038/s41586-022-05012-5

Xiu, D., Karniadakis, G.E. (2002). The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM Journal on Scientific Computing, 24(2), 619–644. https://doi.org/10.1137/S1064827501387826

Yeh, K. C., Liu, C.H. (1974). Acoustic-gravity waves in the upper atmosphere. Rev. Geophys., 12, 193–216. https://doi.org/10.1029/RG012i002