Response to the first reviewers' comments (RC1) on the paper "*Integrated hydrodynamic and machine larning models for compound flooding prediction in a data-scarce estuarine delta*".

General comments

This study develops a scheme to integrate physically-based and data-driven approaches as proposed in similar studies with either conventional machine learning algorithms (Hosseiny et al., 2020; French et al., 2017; Kabir et al., 2020b) or more advanced deep learning techniques (Kabir et al., 2020a; Muñoz et al., 2021). Yet, the scheme has critical technical flaws and is not well-aligned with the main goal of this study that consists of reducing computation burden for water level prediction in data-scarce regions. At this point, I believe that the manuscript requires major improvements to be considered for publication. Nevertheless, please find below a list of suggestions that can help improve the proposed scheme as well as my concerns that should be addressed and/or clarified.

Response:

We want to thank the reviewer for taking the time to review our paper. Their comments are beneficial and helped us to improve the article. In what follows, we addressed the comments, where the reviewer's comments are presented in italic type and my response in roman type.

Specific comments

Title: 'compound flooding prediction' certainly draws the reader's attention, but unfortunately there is no formal analysis in this study suggesting that Pontianak experienced such compound events in the past. Although the introduction elaborates on the mechanisms triggering compound flooding, the authors should conduct preliminary process-based or statistical analyses to confirm this. See for example the works of Kumbier et al., (2018), Valle-Levinson (2020), Ward et al., (2018), Ghanbari et al., (2021) among others.

Response:

We did the assessment of compound flooding over the area using a 2D hydrodynamic model in the previous work (Sampurno et al., 2021). The paper related to the work was accepted for publication in the journal Biogeosciences and is expected to be available online in due course. So, this work is to continue what we have done in the previous work. As a complement, we added these sentences in the introduction session:

This city experienced a compound flooding event on 29 December 2018 (Sampurno et al., 2021), and the impact was severe (Madrosid, 2018). At that moment, the water level dynamic is about to go down after passing its peak elevation, when suddenly a strong force pushes it to go up again for a short moment. The interaction between tides, storm surges, and discharges along the tidal river in the Kapuas River delta is responsible for a 30 cm increase in the water level during the event.

Abstract: *'Compound flood' scenarios should be derived from preliminary statistical analysis that account for the dependence among flood drivers. See for example Serafin et al., (2019), Moftakhari et al., (2019), etc. The abstract should be revised to clearly report the findings of this study. Random forest is used to predict water levels only, yet 'flooding hazards' and 'compound flooding' are mentioned repeatedly without presenting any flood maps in Pontianak. A very preliminary assessment could be achieved by projecting water level to adjacent areas of the Kapuas-Kecil River in case 2D flood modeling is computationally demanding.*

Response:

We followed the suggestion and modified the abstract. We only focus on flood prediction but not yet on a flood hazard assessment. Regarding the "compound flooding" term, the study area has the possibility of experiencing three possible flooding types (coastal, urban/flash, and riverine). Since these events generally coincide or occur in the near time in the area, therefore, compound flooding term is chosen to represent the issues.

Here is the updated abstract:

"Flood forecasting based on hydrodynamic modeling is an essential non-structural measure against compound flooding over the globe. With the risk increasing under climate change, all coastal areas are now in need of flood risk management strategies. Unfortunately, for local water management agencies in developing countries, building such a model is challenging due to the limited computational resources and the scarcity of observational data. We attempt to solve this issue by proposing an integrated hydrodynamic and machine learning approach to predict water level dynamics as a proxy of compound flooding risk in a data-scarce delta. As a case study, this integrated approach is implemented in Pontianak, the densest coastal urban area over the Kapuas River delta, Indonesia. Firstly, we built a hydrodynamic model to simulate several compound flooding scenarios. The outputs are then used to train the machine learning (ML) model. To obtain a robust machine learning model, we consider three machine learning algorithms, i.e., Random Forest, Multi Linear Regression, and Support Vector Machine. Our results show that the integrated scheme works well. The Random Forest (RF) is the most accurate algorithm to model water level dynamics in the study area. Meanwhile, the machine-learning model with the RF algorithm can predict eleven out of seventeen compound flooding events during the implementation phase. It could be concluded that RF is the most appropriate algorithm to build a reliable ML model capable of estimating the river water level dynamics within Pontianak, whose output can be used as a proxy for predicting compound flooding events in the city."

Introduction: *Literature review falls short in content and cohesion. The introduction should mention the state-of-the-art techniques for compound flood hazard modeling and assessment. (Bevacqua et al., 2019; Couasnon et al., 2020; Ye et al., 2021; Muñoz et al., 2022).*

Response:
The suggestion has been followed. We mention these references in the introduction. Then, we also add a new paragraph as follows:
A new paradigm that combines deterministic and machine learning components has been proposed and implemented to tackle data and computational limitations in environmental modeling (Krasnopolsky and Fox-Rabinovitz, 2006; Goldstein and Coco, 2015). However, to the best of our knowledge, no previous modeling frameworks have developed a deterministic model to train a machine learning model for compound flooding studies. As a common practice, compound flood modeling typically uses the coupling of two or more hydrodynamic, hydraulic, or hydrological models (Hsiao et al., 2021; Santiago-Collazo et al., 2021; Ikeuchi et al., 2017). The coupling could be one-way, two-way, or dynamic coupling. Another approach is deep learning and data fusion (Muñoz et al., 2021), and data assimilation (Muñoz et al., 2022).

Material and methods: *This section is of major concern since the methodology presented in this study is not technically sound. I agree that developed countries might face challenges to implement hydrodynamic models due to data scarcity and computational resource limitations. Nevertheless, assuming that those countries develop such a model and want to implement the proposed scheme. Are 6000 simulations really necessary to train machine learning algorithms? Would not it be better to wisely sample a small set of realistic forcing conditions that effectively lead to compound flooding? Although machine learning is a 'data-hungry' technique, I consider thousand of hydrodynamic simulations a bit exaggerated. I suggest the authors train the models with a small sample size (e.g., hundred of simulations) and report the results. This can help reduce computation time associated with hydrodynamic simulations. How long does it take to run the 2D-model of the Kapuas-Kecil in a regular desktop computer?*
Response:
We are sorry because there is a misunderstanding between what we did and what the reviewer thinks we did regarding the term "scenarios." Actually, we only ran ten months of hydrodynamic simulations, and then we extracted only 6000 pairs of data points (hourly water level in Pontianak vs. its associated input variables). To run the 2D model of the Kapuas-Kecil on a regular desktop computer (with 16 GB memory) took about 12 hours.

Section 2.2: *There is no information regarding model calibration. This is critical as the authors rely on hydrodynamic simulations to train the machine learning algorithms. Referring to a pre-print/unpublished work (Sampurno et al., 2021) for additional details of the model is not acceptable. Please describe the model calibration process in detail.*

Response:

Actually, the previous work we mentioned has been accepted for publication and will be available online soon. However, the suggestion has been followed. We added a new section dedicated to model setup and calibration:

2.3 Hydrodynamic model setup and calibration

In order to run the hydrodynamic model, we defined a computational domain that covers both the river and the ocean parts. Next, we generated an unstructured mesh to cover the domain, with a resolution of 50 m over the riverbanks, 400 m over the coast near the river mouth, 1 km over the rest of the coastline, and 5 km over the offshore (Fig. 3). The multi-scale mesh was generated using an algorithm developed by Remacle and Lambrechts (2018). Next, we set the bathymetry constructed from two data sets: first, the river and estuary bathymetry maps, obtained from the Indonesian Navy (Kästner, 2019), and second, the Karimata Strait bathymetry, obtained from BATNAS (BATimetri NASional, 2021). Furthermore, we set the bulk bottom drag coefficients, which are 2.5×10^{-3} over the ocean (which corresponds to a sandy seabed) and 1.9×10^{-2} over the river bed (Kästner et al., 2018). Lastly, we imposed the rainfall, as observed by the Pontianak Maritime Meteorological Station (PMMS).

The hydrodynamic model simulation is forced by wind and atmospheric pressure from ECMWF (Hersbach et al., 2020), and tides from TPXO (Egbert and Erofeeva, 2002). As upstream boundary conditions, we imposed discharge from the Kapuas River and the Landak River. The discharge data were retrieved from the Global Flood Monitoring System (GFMS) (Wu et al., 2014).

We also imposed runoff, which was obtained by converting rainfall over the Kapuas Kecil River catchment area as an inlet water flux at some channels entering the domain. The runoff of every channel was calculated from rainfall data using SWAT+ (Bieger et al., 2017), which considered the pressure, the humidity, and other weather parameter input. Unfortunately, during the tuning of the SWAT model, the correlation between the output of the model (runoff) and the observation data is still low (0.32). However, we decided to use the output as the channels' inlet boundary condition in the hydrodynamic model because the channel runoff volume is much less than the river discharge. Therefore, we assumed that it does not significantly affect the hydrodynamics of the river.

To evaluate the SLIM model performance, we ran a simulation for January 2019 and compared the simulated water elevation with the observations in Pontianak. The model errors correspond to an NSE of 0.87 and an RMSE of 0.12 m (Fig. 4). This RMSE is deemed sufficiently small to consider model outputs as a good proxy of the real system (Moriasi et al., 2015).

We simulated the hydrodynamics with oceanic, atmospheric, and river forcings to forecast flood events based on the water levels. Based on the Pontianak Maritime Meteorological Station report, the city is flooded when the water level exceeds 2.5 m. We, therefore, set this value as the threshold of a flood event. We ran the hydrodynamic model for ten months and extracted the output hourly to produce the scenarios (see Table 1). Then, we selected 6,000 sample points of the predicted water levels at Pontianak with their associated input dataset. We merged the data as a single dataset to train the machine learning model, encompassing all possible flood events resulting from the combination of the external forcings. The dataset shows that several flooding occurred within the simulations, indicated by sample points with water elevations greater than 2.5 m (Fig. 5).

Section 2.4. *Another point of major concern is the calibration of machine learning algorithms. It is no clear whether the authors tuned random forest and support vector machine in the training phase or not. In that regard, the training dataset (e.g., 6000 model outputs with the associated input variables) should have been split into training/validation datasets to conduct hyperparameter tuning and so prevent overfitting issues. Using all model outputs to train the algorithms (as reported here) and relying on default parameter-values is not a wise use of machine learning (e.g., Random requires tuning of the number of trees, sample leaf, sample split, etc.). The authors should conduct a thorough 'hyperparameter' tuning as it substantially improves the performance of machine learning algorithms.*

Response:
The suggestion has been followed. We did the calibration for both algorithms. We searched for the optimal value of the number of variables randomly sampled as candidates at each split for Random Forest and got 16 as the optimal one. For the SVM algorithm, we tuned it to select the correct choice of kernel parameters, which is crucial for obtaining good results. We tested four kernel algorithms, i.e., linear, polynomial, radial basis, and sigmoid. We found that the radial basis kernel gave the best performance for the SVM algorithm.

<u>We added these sentences on section 2.4.2, for RF:</u>

To obtain the optimal parameter for the RF, we first tune the algorithm by searching for the optimal value of the number of variables randomly sampled as candidates at each split (mtry). As a result, the optimal number is 16 (Fig. 7).

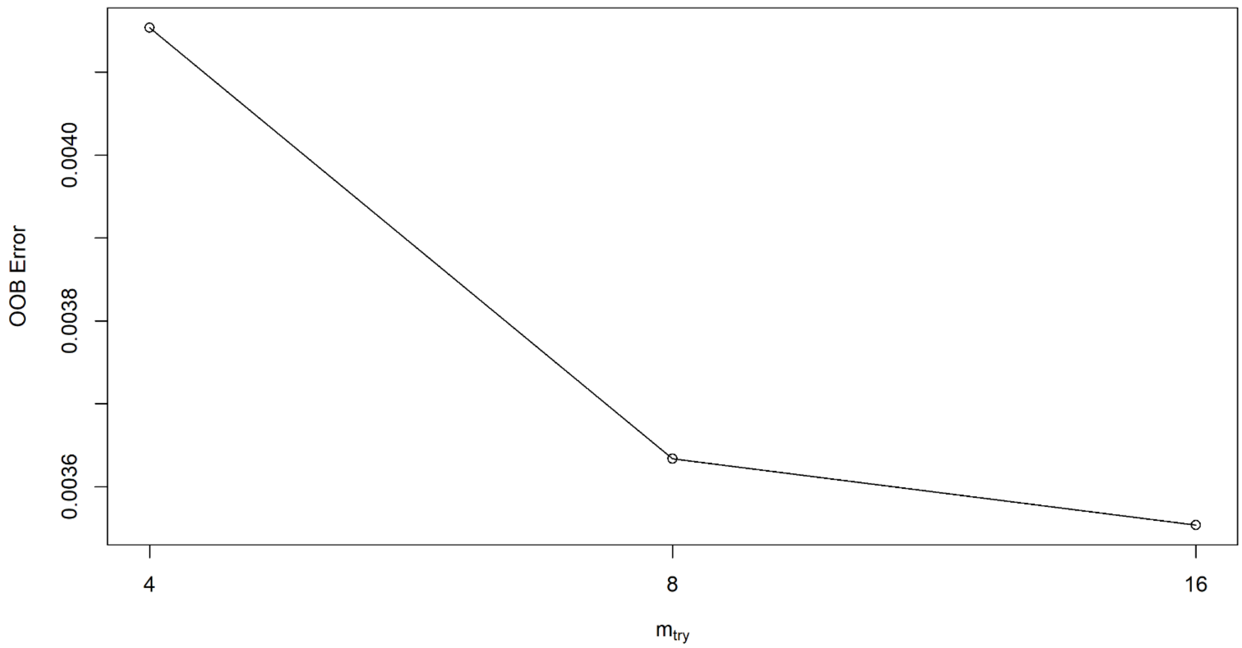

Figure: Tuned randomForest algorithm for the optimal number of variables randomly sampled as candidates at each split (mtry) parameter.

<u>And we added this paragraph for MLR:</u>

To obtain the best performance of the MLR algorithm, we did a statistical analysis to evaluate the multicollinearity among the predictor variables using the Variance Inflation Ratio (VIF). Since multicollinearity negatively affects the performance of the MLR model, VIF can help reduce the number of predictors (Alipour et al., 2020). Here, we found that some variables have VIF more significant than 5, which indicates a potentially severe correlation between these variables in the model (Fig. 8). Therefore, combined with the output of *MI* analysis, we removed some variables which have low *MI* and high VIF.

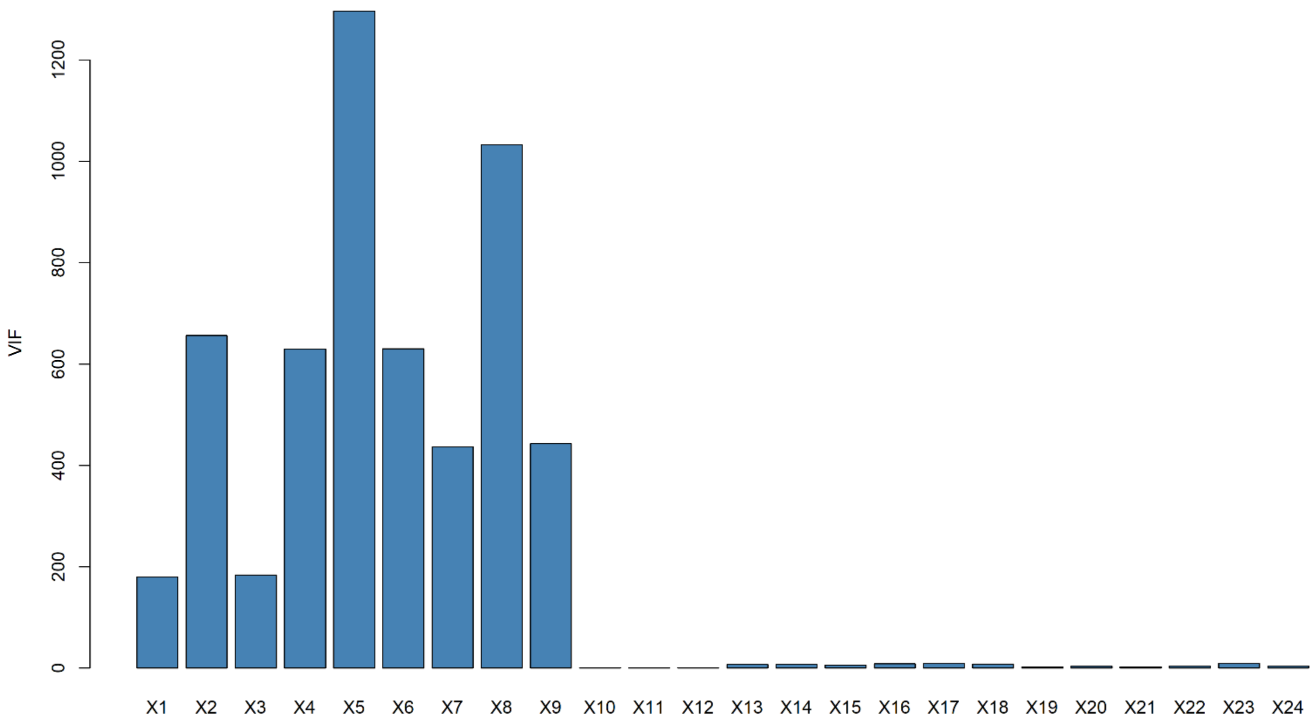

Figure: Variance Inflation Factor values of all predictor's variables in 3 months of observational data.

<u>And this paragraph for SVM:</u>

Since kernel function is critical in SVM, we tuned the SVM algorithm to obtain good results by selecting the most appropriate kernel parameter. We tested four kernels, i.e., linear, polynomial, radial basis, and sigmoid, as the candidates. We found that the radial basis kernel performed the best for the SVM algorithm.

Results. *"Even though all algorithms perform very well during the training phase, the performances are different during the testing phases". This is known as overfitting (Ying, 2019) and occurs because random forest and support vector machine are not calibrated/tuned in the training phase.*

Response:
As mentioned above, we are following the suggestion. We did the tuning in the training phase for all algorithms to avoid overfitting.

Technical corrections

L10 *and thorough the text: There are odd terms that should be corrected like hydrodynamic modeling instead of 'water level modeling'.*
Response: The suggestion has been followed.

L28*: What is the growth rate in the last decade?*
Response: In the Low Elevated Coastal Zone, the population will increase from 638 million in 2000 by 58% to 71% by 2050 (Merkens et al., 2016).

L38*: Please elaborate more on non-structural measures. This sentence is not clear.*
Response: The suggestion has been followed. We added more sentences in the paragraph:
Non-structural measures mean any actions to manage the risk of compound flooding without involving a physical construction (UNDRR, 2022), including land-use regulations, flood forecasting, warning systems, floodproofing and disaster prevention, and preparedness and response mechanisms.

L40*: Which issue? Please explain clearly.*
Response: We replaced word "the issue" with "the water-level prediction issue"

L45-47*: 'Machine learning can enable us…' How? Please, elaborate more on this. More references are needed discussing the benefits of machine learning for water level prediction and/or flood forecasts.*
Response: The suggestion has been followed. We added more details after this sentence:
For instance, by assuming that flood events are stochastic, machine learning can predict major flood events based on certain probability distributions from the historical discharge data (Mosavi et al., 2018). In some cases, their performance is even more accurate than traditional statistical models (Xu and Li, 2002). In other words, we can prepare strategies to mitigate the flood risks using a machine learning model.

L96*: Please locate Pontianak in Figure 2.*
Response: Figure 2 updated. We added the perimeter of Pontianak on the map.

L97*: Section 2.3 is very short (4 lines) and should be included in the previous section.*
Response: The suggestion has been followed.

L65*: More details of the study area are needed. What is the catchment size, average river flow, tidal regime, rate of local sea level rise at the Kapuas River?*
Response: The suggestion has been followed. We added these sentences on the paragraph:
Its water catchment area spreads over about 93000 km2 (about 12.5% of the Borneo Island area, Fig. 1), with about 66.7% of it consisting of forests (Wahyu et al., 2010). Its topography comprises hills over its upstream, covered mainly by Acrisol soils (Fig. 2). In contrast, its downstream comprises plains with more heterogeneous soil types (Fig. 2), such as *Humic Gleysols* (derived from grass or forest vegetation) and *Dystric Fluvisols* (young soil in alluvial deposits).
And, create a new paragraph in this section:
As a tidal river, the tidal regime within the Kapuas River downstream area is mixed but mainly diurnal (Kästner, 2019). The dominant tidal constituent is K1, O1, P1, M2, and S2 (Pauta, 2018). The average tidal amplitude downstream is set in a microtidal regime, with a mean spring range of 1.45 m at its river mouth (Kästner, 2019).

L115*: I suggest a more robust statistical analysis to evaluate multicollinearity among the variables (e.g., Variance Inflation Ratio (VIF), see Alipour (2020)). Multicollinearity negatively affects the performance of support vector machine and multilinear regression models. VIF can help reduce the number of predictors.*

Response: The suggestion has been followed. We now test the multicollinearity among the predictor variables and found some variables having VIF greater than 5, which indicates a potentially severe correlation between these variables in the model. Therefore, we removed some variables based on this VIF values combine with MI coefficients, without reducing the model's performance.

L155-156*: NSE and RMSE might improve after a thorough hyperparameter tuning of the machine learning algorithms.*

Response: We did the tuning and reduced the number of predictors. The NSE and RMSE are better in the training and the testing phases but still not too different in the implementation phase.

L158*: There are no such inundation scenarios (no flood maps). This should be clarified and better replaced for water level scenarios.*

Response:

The suggestion has been followed. We replace this sentence:

We then simulated several inundation scenarios to produce datasets used to train the machine learning model.

<u>With the following:</u>

We simulated the hydrodynamics with different oceanic, atmospheric, and river forcings to forecast flood events based on the water levels. Based on the Pontianak Maritime Meteorological Station report, the city is flooded when the water level exceeds 2.5 m. We, therefore, set this value as the threshold of a flood event. We ran the hydrodynamic model for ten months and extracted the output hourly to produce the scenarios (see Table 1).

We also moved this sentence from the result section to the Material and Method section, as suggested by Reviewer2.

Table 1*. What are the criteria to come up with those range of values?*

Response:

We set those ranges based on minimum and maximum observation data from 2016 to 2021.

Figure 2. *Scale bar and north arrow are missing.*

Response:

The figure has been updated. We added scale bare and north arrow. We also added grid-coordinates and combined it with the bathymetry map (as suggested by Reviewer 2). We changed Figure 2 to Figure 3.

Figure 6. *Comparison of predicted and 'simulated' hourly water levels of training data. There are no observed water levels in the training phase.*
Response: The suggestion has been followed. We updated the Figure.

Figure 8. *X-axis is not observation but hydrodynamic simulation.*
Response: The suggestion has been followed. We updated the Figure.

L295. *There are references not included in the main text. See for example Rozum et al., 2020.*
Response: Actually, there is no **Rozum et al., 2020**. It is still the part of **Hersbach, et al, 2020**. We will put more space between references to make them more easily checked.
The full reference is:
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., **Rozum**, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020

<u>Reference:</u>

Alipour, A., Ahmadalipour, A., Abbaszadeh, P., and Moradkhani, H.: Leveraging machine learning for predicting flash flood damage in the Southeast US, Environ. Res. Lett., 15, 024011, https://doi.org/10.1088/1748-9326/AB6EDD, 2020.

BATimetri NASional: https://tanahair.indonesia.go.id/demnas/#/batnas, last access: 14 July 2021.

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M., and Srinivasan, R.: Introduction to SWAT+, a Completely Restructured Version of the Soil and Water Assessment Tool, J. Am. Water Resour. Assoc., 53, 115–130, https://doi.org/10.1111/1752-1688.12482, 2017.

Egbert, G. D. and Erofeeva, S. Y.: Efficient inverse modeling of barotropic ocean tides, J. Atmos. Ocean. Technol., 19, 183–204, 2002.

Goldstein, E. B. and Coco, G.: Machine learning components in deterministic models: Hybrid synergy in the age of data, Front. Environ. Sci., 3, 33, https://doi.org/10.3389/FENVS.2015.00033/BIBTEX, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hsiao, S. C., Chiang, W. S., Jang, J. H., Wu, H. L., Lu, W. S., Chen, W. B., and Wu, Y. T.: Flood risk influenced by the compound effect of storm surge and rainfall under climate change for low-lying coastal areas, Sci. Total Environ., 764, 144439, https://doi.org/10.1016/J.SCITOTENV.2020.144439, 2021.

Ikeuchi, H., Hirabayashi, Y., Yamazaki, D., Muis, S., Ward, P. J., Winsemius, H. C., Verlaan, M., and Kanae, S.: Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 Cyclone Sidr in Bangladesh, J. Adv. Model. Earth Syst., 9, 1847–1862, https://doi.org/10.1002/2017MS000943, 2017.

Kästner, K.: Multi-scale monitoring and modelling of the Kapuas River Delta, Wageningen University, https://doi.org/10.18174/468716, 2019.

Krasnopolsky, V. M. and Fox-Rabinovitz, M. S.: A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components, Ecol. Modell., 191, 5–18, https://doi.org/10.1016/J.ECOLMODEL.2005.08.009, 2006.

Madrosid: Cerita Warga, Detik-detik Banjir Rob Melanda Kota Pontianak, Trib. Pontianak, 2018.

Merkens, J. L., Reimann, L., Hinkel, J., and Vafeidis, A. T.: Gridded population projections for the coastal zone under the Shared Socioeconomic Pathways, Glob. Planet. Change, 145, 57–66, https://doi.org/10.1016/J.GLOPLACHA.2016.08.009, 2016.

Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P.: Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria, Trans. ASABE, 58, 1763–1785, https://doi.org/10.13031/TRANS.58.10715, 2015.

Mosavi, A., Ozturk, P., and Chau, K.: Flood Prediction Using Machine Learning Models: Literature Review, Water 2018, Vol. 10, Page 1536, 10, 1536, https://doi.org/10.3390/W10111536, 2018.

Muñoz, D. F., Muñoz, P., Moftakhari, H., and Moradkhani, H.: From local to regional compound flood mapping with deep learning and data fusion techniques, Sci. Total Environ., 782, 146927, https://doi.org/10.1016/J.SCITOTENV.2021.146927, 2021.

Muñoz, D. F., Abbaszadeh, P., Moftakhari, H., and Moradkhani, H.: Accounting for uncertainties in compound flood hazard assessment: The value of data assimilation, Coast. Eng., 171, 104057, https://doi.org/10.1016/J.COASTALENG.2021.104057, 2022.

Pauta, D. F. M.: Tidal influence on the discharge distribution at two junctions of the Kapuas River (West Kalimantan, Indonesia), 2018.

Sampurno, J., Vallaeys, V., Ardianto, R., and Hanert, E.: Modeling interactions between tides, storm surges, and river discharges in the Kapuas River delta, Biogeosciences Discuss. [preprint], https//doi.org/10.5194/bg-2021-273, Rev. 2021., https://doi.org/10.5194/BG-2021-273, 2021.

Santiago-Collazo, F. L., Bilskie, M. V., Bacopoulos, P., and Hagen, S. C.: An Examination of Compound Flood Hazard Zones for Past, Present, and Future Low-Gradient Coastal Land-Margins, Front. Clim., 3, 76, https://doi.org/10.3389/FCLIM.2021.684035/BIBTEX, 2021.

UNDRR: Structural and non-structural measures: https://www.undrr.org/terminology/structural-and-non-structural-measures, last access: 12 May 2022.

Wahyu, A., Kuntoro, A., and Yamashita, T.: Annual and Seasonal Discharge Responses to Forest/Land Cover Changes and Climate Variations in Kapuas River Basin, Indonesia, J. Int. Dev. Coop., 16, 81–100, https://doi.org/10.15027/29807, 2010.

Xu, Z. X. and Li, J. Y.: Short-term inflow forecasting using an artificial neural network model, Hydrol. Process., 16, 2423–2439, https://doi.org/10.1002/HYP.1013, 2002.