Levee protected area detection for improved flood risk assessment in global hydrology models

Masato Ikegawa[1], Tristan Hascoet[1], Victor Pellet[2], Xudong Zhou[3], Tetsuya Takiguchi[1], Dai Yamazaki[3]

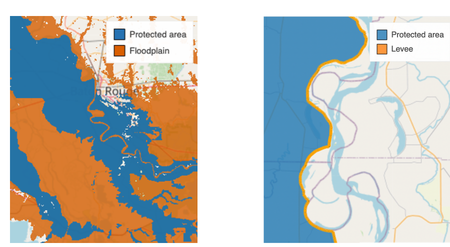
[1]Graduate School of System Informatics, Kobe University, Kobe, Japan
[2]Observatoire de Paris, Paris, France
[3]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Abstract

Precise flood risk assessment is needed to reduce human societies vulnerability as climate change increases hazard risk and exposure related to floods. Levees are built to protect people and goods from flood, which alters river hydrology, but are still not accounted for by global hydrological model. Detecting and integrating levee structures to global hydrological simulations is thus expected to enable more precise flood simulation and risk assessment, with important consequences for flood risk mitigation. In this work, we propose a new formulation to the problem of identifying levee structures: instead of detecting levees themselves, we focus on segmenting the region of the floodplain they protect. This formulation allows to better identify protected areas, to leverage the structure of hydrological data, and to simplify the integration of levee information to global hydrological models.

1 Introduction

Flood risk is defined as the product of hazard (the probability and strength of a flood event), exposure (the population and goods subject to flooding) and vulnerability (the capacity of a society to deal with the event) [1]. Climate change is intensifying hazard risk by rising sea levels and increasing flood frequency, while overall exposure increases with population growth and densification of coastal and fluvial flood-prone regions [2]. Although flooding events occur locally, their impact goes beyond the local tragedies of devastation. With the globally interconnected nature of economic activities, local disasters also disturb supply chains at global scale [3, 4], making flood risk assessment and mitigation a global challenge. To protect people and goods from flood risks, our two main assets are water management infrastructure (i.e.; levees and dams), and early warning systems. In the context of water management, it has been shown that overwhelmed levees are a source of disastrous floods [2] as it comes



(a) Floodplain and protected areas

(b) Levees & protected areas

Figure 1: Illustration of our problem formulation (a) We aim to accurately segment levee protected areas (in blue) of the floodplain (in orange) (b) Different from related work, our approach explicitly models protected areas (blue) instead of levee locations (yellow)

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

with surprise. Better dimensioning of flood defense requires more accurate floods risk map accounting for climate change induced hazard risk (e.g. shift of the event characteristic distribution).

Hydrological simulations contribute to reduce vulnerability as they allow to predict and thus mitigate flood events [5]. Global hydrodynamic models rely on global spaceborn Digital Elevation Model (DEMs) to route river discharge through landscape and represent river/floodplain interactions [6, 7], that do not include man-made river flow control infrastructure such as levees and dams. Thus, flood predictions computed using current global hydrological models do not account for the impact of levees on river discharge. Integrating levees to global hydrology models would allow to generate more accurate and reliable global flood risk assessment. While the simulation of dams impact on river discharge at global scale have been introduced by the hydrology modeling community [8, 9, 10], the representation of levee impact on river discharge has not yet passed from case studies to global modeling [11]. The main bottleneck is the non-availability of a global dataset gathering worldwide river control equipment.

Flood-protection levee datasets are generally confined to territorial boundaries (national datasets) and are scarcely available and poorly documented. A very recent effort has been put to collect, standardize and homogenize flood-protection levee data from various sources into a single open source dataset [12]. But this effort has been limited so far to delta areas that present both fluvial (from river) and coastal (from sea) flood risks. The most-up-to-date river levee dataset is the National Levee Dataset (NLD) from the U.S. Army Corps of Engineers (USACE), which we use in this study. One way to extend our knowledge of levee locations is to use Machine Learning (ML) to learn the relationship between levee locations and a set of globally available variables from available local levee data, so as to infer levee locations on a global scale from these variables. Unfortunately, this problem has proved very challenging: Levees are built to have maximum impact on flood protection at minimum construction cost (and so at minimum size), which makes it notoriously hard to detect. While some local high precision DEMs are precise enough to identify levees from elevation data [13, 14, 15], the precision needed to detect levee structures is far higher than that of global DEMs. Another line of work [16, 17], closer to the present study, have taken an other path and tried to leverage the NLD dataset to infer levee locations from data including economical and hydrological data.

We propose a new formulation of the problem: Instead of identifying the position of levees themselves, we aim to identify the areas that are protected by levees. The benefit of this formulation is threefold: First, we find that it better fits the distribution of discriminative features. Second, it allows us to leverage additional structure to our modeling using hydrological data: We propagate our model classification outputs following the reverse hydrological flow direction path. Finally, this problem formulation allows for seamless integration to existing global catchment level hydrological models: aggregating our model output per hydrological catchment provides with a single parameter representing the catchment's ratio of protected floodplain. This parameter can be readily integrated into hydrological models to limit the growth of river width within the catchment. To detect levee-protected areas, we formulate the hypothesis that protected areas should exist where human activity and flood risk overlap. We compute potential floodplains using hydrological model simulations and quantify human activity using different variables including GDP, population and land use. We use this data as input and train ML models to classify levee protected areas using NLD dataset. Preliminary experiments provide us with encouraging results as we find our porposed problem definition to improve accuracy of levee protected area segmentation. This work focuses on the Mississippi basin area as it is the region most densely annotated with levee locations. However, our ultimate goal is to generalize this analysis globally. Generalizing to a global scale will come with the additional difficulty of assessing accuracy in areas with sparse ground truth and possibly non-stationary levee distributions. These are important challenges to global hydrology and global flood risk assessment, which we leave open fur future work, and welcome the community to join us in tackling.

2 Data

Different from previous works, we use the information of *protected area* provided by the NLD dataset, instead of the levee location themselves, as our classification target. This difference is illustrated in Figure 1. The NLD dataset does not provide an exhaustive collection of U.S. levees, which leads to the existence of false negative ground truth labels, i.e; areas marked as non-protected, while they are in fact protected by an unreferenced levee. One notable exception is the region surrounding the Mississippi river basin, which is densely annotated. In an attempt to minimize the impact of false negative labels on our analysis, we focus on this region only, using the three states of Mississippi, Arkansas and Louisiana as our study area. We rasterized the levee-protected area provided by the NLD dataset to a resolution of 90m, resulting in one binary label maps over each of the three states.

We ran hydrological simulations and estimated 100-year return period flood inundation depth at 90m resolution [18]. We binarized flood inundation depth to segment the hydrological floodplains (i.e.; areas subject to potential inundation), as illustrated in Figure 1a (orange). We also use Global Surface Water Occurrence datasets [19] to quantify flood risk. To quantify human activity, we use land use data from LCCS [20] and GFSAD [21], GDP data provided by [4], and population data from Landscan [22]. Finally, our proposed model propagates classification outputs using the inverse flow direction as illustrated in Figure 3. The flow direction is a hydrological variable indicating the direction in which water flows from every pixel. It is computed using the gradient of hydrological DEMs elevation, i.e.; the slope of the land. In our experiments, we used the flow direction provided by the MERIT Hydro [6] dataset.

3 Model and experiments

To motivate our approach, Figure 2 illustrates the benefits of our proposed problem formulation. Levees are built in locations designed to most efficiently protect a targeted area from floods. In some cases, the ideal location for a levee may be closer to the protected area, while in other cases, levees may be located closer to the river and further away from the targeted area. From a ML perspective, this leads to different distribution of the input features representing levee locations. In addition, levees are thin structures often built on wide plains. This means that only a small fraction of the plain on which they are built is actually covered by the levee structure, despite most of the surrounding pixels having almost exactly the same features. From a ML perspective, this leads to a complex situation in which positive samples share similar input feature distribution with (an even greater number of) negative samples, which complicates the learning task. On the other hand, target protected areas show significantly different and characteristic feature distribution as they represent zones of more intense human activities, as illustrated in Figure 2. This motivates us to classify pixels as levees *protected areas* instead of levees *location*.

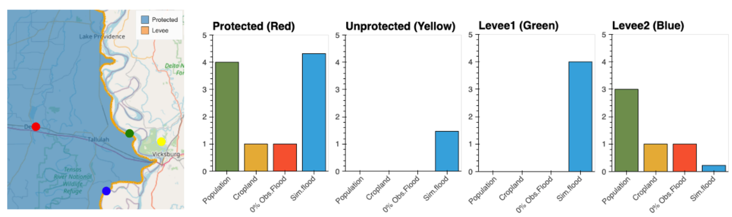
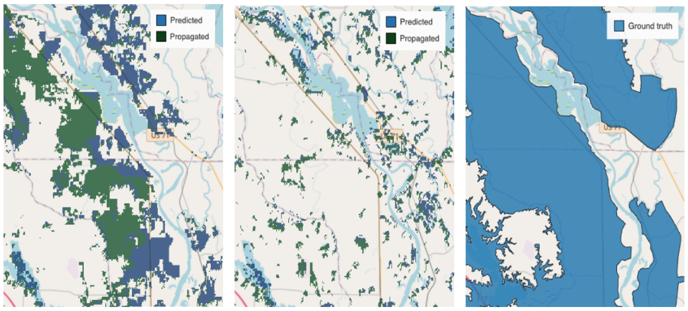


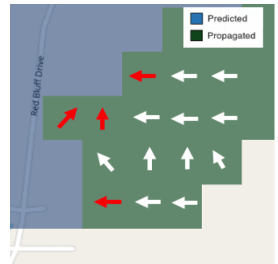
Figure 2: Illustration of the feature expressivity of levees and protected areas. We find protected areas to have more discriminative feature distributions.

We train a linear classifier pixel-wise to classify levee protected areas using a binary cross entropy loss. The model is optimized using stochastic gradient descent with momentum. During inference, we experiment with an original strategy: We post-process the classification output of our model by propagating positive output labels (levee protected areas) along the inverse flow direction. Figure 3(b) schematically illustrates this process. The rationale behind this approach is that if a downstream pixel is to be protected, then the area located upstream of this pixel should also be protected or else it would eventually flood the downstream pixel.

We use a three-fold cross validation strategy in which, for each fold, we use one state as training, one for validation, and one as our test set. We report accuracy of protected in terms of pixel-wise levee protected area segmentation accuracy. We compare our results to that of a model classifying levee existence similar to prior works. We derive the protected areas resulting from levee locations given by this model by propagating positive outputs following the inverse flow direction. This allows us to compare both approaches with a same metric: levee-protected area segmentation accuracy.



(a) Propagate examples



(b) I.P. illustration

Figure 3: Sample model output with inverse flow propagation(I.P.). (a) Modeling levee protected areas with inverse flow propagation better captures (Left) Protected area segmentation. (Middle) Levee location segmentation and their propagated protected area (Rigth) Ground truth protected area (b) Each pixels has one forward flow direction. Protected areas are propagated from the model output (blue) following the inverse flow, starting from the red arrows.

Table 1 shows the levee-protected area segmentation results from both propagated (I.P.) and non-propagated model outputs, and contrasts these results to the accuracy reached by levee propagation (column Levee). Both flow propagation and our proposed problem formulation are shown to improve accuracy. The impact of propagating propagated areas along the inverse flow is illustrated in Figure 3(a). Flow propagation seems to appropriately fill some wholes of the floodplain area to be protected by propagating the positive label from high human activity areas to lower activity regions located upstream. We additionally experimented with more expressive non-linear models, but have found little improvement over the linear baseline. We report their results in Table 1 for completeness.

Table 1: Mean accuracies reached by different models. The Levee column refers to a linear model trained to classify levee locations.

		Models			
I.P.	Levee	Linear	Lgbm	XgB	MLP
No	—	0.6294	0.6170	0.6189	0.6558
Yes	0.5138	0.6620	0.6302	0.6363	0.6616

4 Conclusion and future work

Knowledge of man made structures impacting global surface water processes is critical for flood risk assessment in a time when climate change threatens to increase flood related damages, creating both local catastrophes and global disruptions through interconnected logistic chains and economical activities. In this paper, we proposed a new formulation to the problem of levee detection, which is both more performant and amenable to integration into global hydrological models. Preliminary experiments on a local region densely annotation with levee information have shown the merits of our formulation. Despite encouraging results, this work is still in its early phase; we believe that further optimization of the ML model has good chances of improving detection. In addition, important challenges lie agead towards generalizing the application of our model on a global scale for global flood risk assessment. We encourage the community in joining our effort to identify levee structures globally for better flood risk assessment.

References

[1] Kron, W. (2005). Flood risk= hazard• values• vulnerability. Water international, 30(1), 58-68.

[2] Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., & Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. Nature Reviews Earth & Environment, 2, 592-609.

[3] Haraguchi, M., and Lall, U. (2015). Flood risks and impacts: A case study of Thailand ' s floods in 2011 and research questions for supply chain decision making. International Journal of Disaster Risk Reduction, 14, 256-272.

[4] Taguchi, R., Tanoue, M., Yamazaki, D., and Hirabayashi, Y. (2022). Global-Scale Assessment of Economic Losses Caused by Flood-Related Business Interruption. Water, 14(6), 967.

[5] Lavers, D. A., Ramos, M. H., Magnusson, L., Pechlivanidis, I., Klein, B., Prudhomme, C., ... & Pappenberger, F. (2020). A vision for hydrological prediction. Atmosphere, 11(3), 237.

[6] Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A high - resolution global hydrography map based on latest topography dataset. Water Resources Research, 55(6), 5053-5073.

[7] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... & Alsdorf, D. (2007). The shuttle radar topography mission. Reviews of geophysics, 45(2).

[8] Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. Journal of Hydrology, 327(1-2), 22-41.

[9] Shin, S., Pokhrel, Y., & Miguez - Macho, G. (2019). High - resolution modeling of reservoir release and storage dynamics at the continental scale. Water Resources Research, 55(1), 787-810.

[10] Hanazaki, R., Yamazaki, D., & Yoshimura, K. (2022). Development of a Reservoir Flood Control Scheme for Global Flood Models. Journal of Advances in Modeling Earth Systems, 14(3), e2021MS002944.

[11] Fleischmann, A., Paiva, R., & Collischonn, W. (2019). Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. Journal of Hydrology X, 3, 100027.

[12] O'Dell, J., Nienhuis, J. H., Cox, J. R., Edmonds, D. A., & Scussolini, P. (2021). A global open-source database of flood-protection levees on river deltas (openDELvE). Natural Hazards and Earth System Sciences Discussions, 1-16.

[13] Bishop, M. J., McGill, T. E., & Taylor, S. R. (2004, September). Processing of laser radar data for the extraction of an along-the-levee-crown elevation profile for levee remediation studies. In Laser Radar Technology and Applications IX (Vol. 5412, pp. 354-359). SPIE.

[14] Choung, Y. (2014). Mapping levees using LiDAR data and multispectral orthoimages in the Nakdong river basins, South Korea. Remote Sensing, 6(9), 8696-8717.

[15] Mahrooghy, M., Aanstoos, J., Parasad, S., & Younan, N. H. (2012). On land slide detection using terrasar-x over earthen levees. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39, B3.

[16] Ikegawa, M., Hascoet, T., Pellet, V., Watanabe, M., Zhou, X., Tanaka, Y., Takiguchi, T., Yamazaki, D. (2022). Where do humans build levees? A case study on the contiguous united states. International Geoscience and Remote Sensing Symposium, 2765-2768.

[17] Knox, R. L., Morrison, R. R., & Wohl, E. E. (2022). Identification of artificial levees in the contiguous United States. Water Resources Research, 58(4), e2021WR031308.

[18] Zhou, Xudong, et al. "The uncertainty of flood frequency analyses in hydrodynamic model simulations." Natural Hazards and Earth System Sciences 21.3 (2021): 1071-1085.

[19] Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.

[20] http://www.cec.org/north-american-environmental-atlas/land-cover-2010-landsat-30m/

[21] https://www.usgs.gov/centers/wgsc/science/global-food-security-support-analysis-data-30-m-gfsad

[22] https://landscan.ornl.gov