AI-Improved Resolution Projections of Population Characteristics and Imperviousness Can Improve Resolution and Accuracy of Urban Flood Predictions

Christa Brelsford[1], Chris Krapu[1], Joe Tuccillo[1], Matt McCarthy[1], Jake McKee[1], and Nagendra Singh[1]

[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Focal Area

(2) Predictive modeling through the use of AI techniques and AI-derived model components. Specifically, we call for the use of deep generative models from ML for providing high-resolution projections of impervious surface area, as well as neural network solvers for fast approximation of urban hydrodynamics to provide greatly enhanced forecasts of future urban flood dynamics.

Science Challenge

Interactions between urban populations and the water cycle occur across a range of spatial and temporal scales. Cities are heavy water users and, in aggregate, have a dramatic influence on basin-scale water systems by moving large volumes of water. However, urban influences on the water cycle are much more distributed than the settlements themselves. Hydraulic infrastructure, combined with agriculture functioning as a buffer, enables cities to maintain highly stable and robust water deliveries, even in the face of extreme inter-annual variation in natural water flows. This infrastructure enables localized shortages to be transferred across regions, and even to different hydraulic basins. Urban heat island effects may influence regional-scale precipitation extremes (Singh et al. 2020). These feedbacks influence local and basin-scale extreme events, which may in turn influence future settlement patterns.

Better prediction of population growth will enable water cycle predictions that explicitly include the role of human settlements and their infrastructure in shaping water cycle changes. Current projections of built infrastructure are at a resolution that is too coarse to enable accurate modeling of intra-city hydrological processes like infiltration, runoff, and evapotranspiration. The physical details of urban form—including the configuration of impervious surfaces, the extent of infiltration infrastructure like green roofs and retention basins, and the scale of drainage infrastructure—govern the magnitude and extent of urban flood risk, and shape interactions between surface water and groundwater. Similarly, human experiences of extreme flooding shape subsequent investment choices and so influence the future urban infrastructure.

Assessment of the interactions between population settlements and the water cycle is a transformational science challenge. Despite widespread acknowledgement that anthropogenic processes have a large influence on water cycles and ecosystems, there has been little research focused on identifying interactions between these sectors. Accurate understanding of the mechanisms and consequences through which population growth influences the water cycle will improve predictability.

Rationale

Current land use land cover (LULC) data products that provide information about spatial distribution of land cover lack characteristics for fine-scale flood modeling. Most regional land cover data sets have medium to coarse spatial resolution (30 m –1000 m) and have 5 year upgrade cycles. In addition, users are constrained by fixed class definitions that may not translate well over diverse geographic areas and data-model integration. With high-resolution satellite imagery and

DOI: https://doi.org/10.2172/1769673
DOE BER, AI4ESP 2021 #AI4ESP1016 (https://www.ai4esp.org)

computing resources combined with emerging AI techniques, there is potential to produce high-resolution land cover data at very high temporal frequency. Computer vision along with classification techniques using neural networks offers the capability to deliver information without the need for expensive sampling data. This has been one of the biggest hindrances to producing high-resolution LULC data at scale at very high spatial and temporal resolution. In addition, the flexibility to produce user-defined classifications removes a major hindrance for data-model integration.

AI techniques such as deep-learning and neural networks provide capabilities to project future distributions of population not captured by the conventional trend analysis method; this fails to capture anomalous trends driven by climate-related stressors and socioeconomic inequality. The availability of data captured from a plethora of sources, including social media and smart devices, provides a rich data source to model and map human migration and settlement patterns.

The opportunity to accurately map LULC at scale at high temporal resolution, along with the capability to predict future population distribution under different potential environmental and socioeconomic scenarios, provides us with an unprecedented opportunity to model and predict the location and intensity of future floods. Changes in LULC driven by population changes will drive substantial changes in the location and intensity of future floods. Understanding the effects of human settlement patterns on the water cycle is a significant scientific challenge.

Narrative

Synthetic Populations

Synthetic populations are realistic recreations of people and households in an area based on known aggregate statistics (Harland et al. 2012) and are generated using agent-based models and simulation. They provide fine-scale, probabilistic representations of the characteristics of individuals who live in a particular location; and they preserve annual total population, fertility, and survival ratio statistics in the National Population Projections, subject to constraints on sex, age, race, and ethnicity. To derive occupancy preferences, a sub-model can be trained to learn associations among demographics, socioeconomic status, migration, and home sales from publicly available data, creating a spatially explicit representation of projected population density and characteristics.

These projections can then be used to parameterize projections of the fine-scale extent and arrangement of impervious surfaces, based on demographic and socioeconomic occupancy distributions drawn from synthetic populations.

Impervious Surface Area Projections

The multispectral sensors aboard commercial satellite platforms facilitate accurate classification of impervious surfaces. Supervised ML algorithms such as neural networks and support vector machine can further distinguish among buildings, roads, and unpaved hard surfaces (Lottering et al. 2019; Shao and Lunetta 2012). Open-source algorithms now exist to efficiently process large volumes of imagery; and when paired with high-performance computing resources, they reduce computation time from days to seconds. However, additional training data sets are still needed to enable these algorithms to generalize across different impervious surfaces and address other empirical challenges, such as quantifying noise in the data and identifying calibrated confidence scores.

Modeling the growth of impermeable surfaces in an urban context is challenging because of the close interconnections between physical and social characteristics which determine the fine-scale details of the built environment (Reilly et al. 2004). Posing this as an image-to-image translation problem (Santhanam et al. 2017; Zhu et al. 2018), in which the original image represents the pre-development or current status of the landscape and the target image represents post-development land cover, will facilitate more representative projections of urban growth. These projections of urban impervious surface area, when paired with synthetic population projections, can provide a rich foundation for modeling urban flood risk, extent, and consequences.

Urban Flood Risk Projections

ML frameworks can address the most significant computational tasks associated with risk assessment for urban floods. Constructing scenarios of future urban landscapes, which are finely resolved and spatially explicit with regard to both impervious surfaces and drainage channels, will enable impact forecasts of future runoff, assuming current practices in runoff management. Existing optical imagery data sets and LiDAR-derived digital elevation models could be used to provide imperviousness and channel placement training data for an ML model aimed at generating these future urban landscapes. Relying on high-resolution projections of urban imperviousness and other hydraulic infrastructure, neural network-based solvers for partial differential equations (PDE), such as the Navier-Stokes equations, can then be used to simulate floodwater flow in urban and semi-urban settings with much lower computational cost than fully physics-based models.

However, even with high-resolution maps of impervious area and elevation, a major impediment to flood risk assessment on the scale of a metropolitan area is the high computational burden of hydrodynamic simulations (Xing et al. 2019). Exploring the replacement of a standard PDE numerical integration scheme with a neural operator that is trained to approximate hydrodynamics at an appropriate spatial scale (Li et al. 2020) may reduce the time required for approximate numerical integration by several orders of magnitude.

These methods suggest that it is possible to dramatically improve imperviousness projections; and doing so would result in valuable insights on urban flood risk, surface runoff rates, and subsequent impacts to infiltration/groundwater recharge.

Flood Consequences

Cities respond to floods in consistent ways: floods have a large effect on the value of recently flooded areas, but this effect fades as time passes (Di Baldassarre et al. 2013). The effect also depends on the characteristics of the affected population. By using synthetic populations to estimate the characteristics of populations exposed to flood events, it is possible to model the subsequent urban disinvestment from flooded areas.

Recent studies of long-term population change in communities with repeat exposure to climate disasters like Gulf Coast hurricanes suggest that people with more social and economic resources are able to migrate away from hazards, while marginalized populations remain in more exposed areas (Logan et al. 2016). In Hurricane Harvey, minority and low-income residents experienced disproportionately high impacts (Flores et al. 2020), while increasing urbanization influenced Harvey's severity (Chakraborty et al. 2019). Thus, simultaneous assessment of changes in demographic characteristics and imperviousness can anticipate convergence between locational and social vulnerability—something that is not possible with current population projections.

References

Chakraborty, J., Collins, T. W., and Grineski, S. E. 2019. Exploring the environmental justice implications of Hurricane Harvey flooding in Greater Houston, Texas. *American Journal of Public Health*, 109(2), 244-250.

Di Baldassarre, G., Kooy, M., Kemerink, J. S., and Brandimarte, L. 2013. Towards understanding the dynamic behaviour of floodplains as human-water systems. *Hydrology and Earth System Science*, *17*(8), 3235–3244.

Flores, A. B., Collins, T. W., Grineski, S. E., and Chakraborty, J. 2020. Social vulnerability to Hurricane Harvey: unmet needs and adverse event experiences in Greater Houston, Texas. *International Journal of Disaster Risk Reduction*, 46, 101521.

Harland, K., Heppenstall, A., Smith, D., and Birkin, M. H. 2012. Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. *Journal of Artificial Societies and Social Simulation*, 15(1).

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A., 2020. Fourier Neural Operator for Parametric Partial Differential Equations. ArXiv201008895 Cs Math.

Logan, J. R., Issar, S., and Xu, Z. 2016. Trapped in place? Segmented resilience to hurricanes in the Gulf Coast, 1970–2005. *Demography*, 53(5), 1511-1534.

Lottering, R., Mutanga, O., Peerbhay, K., and Ismail, R. 2019. Detecting and mapping Gonipterus scutellatus induced vegetation defoliation using WorldView-2 pan-sharpened image texture combinations and an artificial neural network. *Journal of Applied Remote Sensing*, *13*(1), 014513.

Reilly, J., Maggio, P., and Karp, S. 2004. A model to predict impervious surface for regional and municipal land use planning purposes. *Environmental Impact Assessment Review* 24, 363–382.

Santhanam, V., Morariu, V.I., and Davis, L.S. 2017. Generalized Deep Image to Image Regression. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5609–5619.

Shao, Y., and Lunetta, R. S. 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 78–87.

Singh, J., Karmakar, S., Pai Mazumder, D., Ghosh, S., and Niyogi, D. 2020. Urbanization alters rainfall extremes over the contiguous United States. *Environmental Research Letters*, *15*(7), 074033.

Xing, Y., Liang, Q., Wang, G., Ming, X., and Xia, X. 2019. City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Natural Hazards* 96, 473–496.

Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., and Shechtman, E. 2018. Toward Multimodal Image-to-Image Translation. ArXiv171111586 Cs Stat.