Predicting Flash Floods in the Dallas-Fort Worth Metroplex Using Workflows and Cloud Computing

Eric Lyons[§], Dong-Jun Seo[*], Sunghee Kim[*], Hamideh Habibi[¶]
George Papadimitriou[‡], Ryan Tanaka[‡], Ewa Deelman[‡], Michael Zink[§], Anirban Mandal[†]

[*]University of Texas Arlington, TX, USA
[†]RENCI, University of North Carolina at Chapel Hill, NC, USA
[‡]Information Sciences Institute, University of Southern California, CA, USA
[§]Electrical and Computer Engineering Department, University of Massachusetts at Amherst, MA, USA
[¶]Walter P. Moore and Associates Inc. Houston, TX, USA

***Abstract*—Accurate and timely prediction of flash flooding events can be a very useful tool for stormwater officials and first responders. Having lead time with which to issue evacuation directives, to close flood prone roadways, to deploy rescue gear and personnel, and to fortify areas against flooding is essential to minimize property damage and risk of casualties. In this poster, we are presenting a flash flooding prediction workflow based on the Hydrology Lab-Research Distributed Hydrologic Model (HL-RDHM). This workflow leverages cloud computing and the Pegasus Workflow Management System to provide continuous high resolution flood predictions for the Dallas-Fort Worth Metroplex area in North Texas, and can be easily expanded to other regions.**

I. Introduction

The Research Distributed Hydrologic Model (RDHM) [1] [2] was the byproduct of a collaborative effort between several National Oceanic and Atmospheric Administration (NOAA) [3] and National Weather Service (NWS) [4] hydrology research laboratories. This model was developed to improve streamflow predictions in streams and rivers and to improve flash flood forecasting by incorporating evolving estimates of relevant parameters, such as soil moisture, soil temperature, surface permeability, and vegetation, in addition to the primary forcing mechanism, rainfall. Since its inception in 2008, local domain knowledge has improved. In the Dallas-Fort Worth (DFW) metroplex in North Texas, high resolution Lidar based surveys have enabled the creation of accurate topology mappings of riverbeds and man-made drainage systems. Additionally, urban surface changes have been well documented thanks to such mappings. Improvements of gridded rainfall estimates [5] have the added to the model accuracy. The RDHM model has traditionally been run on a single, powerful dedicated server. In an effort to improve scalability as resolutions and domain sizes increase, and to tailor the output for stakeholder decision support, we have optimized the source code, containerized the model for use in public and educational clouds, and adopted the Pegasus Workflow Management System [6] to manage the workflow executions.

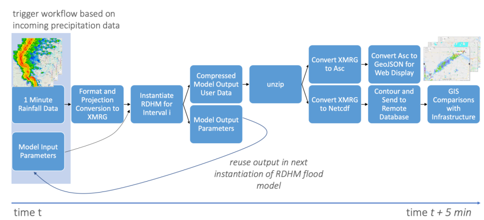
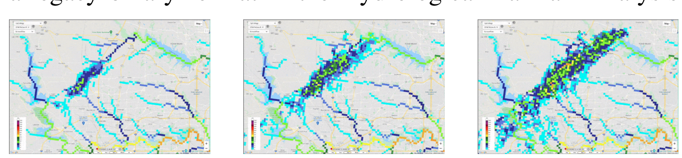


Fig. 1: RDHM Workflow Diagram

II. RDHM Workflow

The RDHM workflow (depicted in Figure 1) was developed using the Pegasus 5.0 Python API. The workflow is executed in an ExoGENI [7] cloud deployment orchestrated by the Mobius multi-cloud provisioning system [8] and Pegasus on a continual basis using the latest precipitation data provided by two sources: the Next Generation Weather Radar (NEXRAD) system and a network of Doppler radars, operated by the Collaborative Adaptive Sensing of the Atmosphere (CASA) [9] partnership. This data is combined and fed into the RDHM flood model, along with a multitude of input parameters derived from high-resolution topography, land cover, land use, and soil data of the DFW region, and other hydrologic and climatological data. To ensure that the RDHM flood model has access to best estimates of the current hydrologic conditions, its output from the previous run of the workflow is also passed in as an input, forming a feedback loop over all consecutive runs of this workflow. Once complete, the RDHM model outputs streamflow, runoff, water depth, and return period estimates valuable for flash flood forecasting, along with the updated parameterizations to be used for the next instantiation of the model. This output data is formatted as XMRG [10], a legacy binary format in the Hydrological Rainfall Analysis



(a) T-1 (b) T-2 (c) T-3

Fig. 2: RDHM generated streamflow

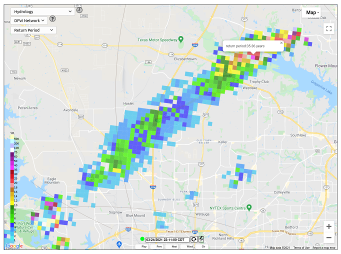


Fig. 3: RDHM generated flooding return period

Projection (HRAP) coordinate system, used by the NWS. The next steps in this workflow involve converting the XMRG output file(s) into the modern GIS format GeoJSON, and the packed binary grid format NetCDF, with standard WGS84 map projections. The output GeoJSON files can be ingested by mapping services such as Google maps for web display (Figures 2, 3, 4), while the NetCDF files are contoured into geofenced polygons representing areas where the hydrologic values exceed standard thresholds for alerting. These contours are sent to a remote database in near real time, triggering targeted flash flood alerts sent to appholders currently within in the risk area, as well as city emergency managers and stormwater management personnel.

To ensure portability across various clouds, this workflow utilizes Pegasus' Singularity container support. The RDHM model and all the downstream processes are containerized with the latest model parameters stored locally for continuity across runs and also to persist offline should the compute cloud architecture go down.

III. Visualizing RDHM Data

As GeoJSON files are generated from each workflow run on a minute by minute basis, they can be made available for viewing on the web. Streamflow, return period, and runoff rates are some of the features which can be visualized and interact with on the map.

Streamflow. Utilizing the rainfall data provided by the radar systems and layers of hydrologic parameters, the RDHM model can simulate stream flows. This is illustrated in Figure 2 where the colored lines represent stream flows through river basins. Typically, as seen on the lower right portion, the streams flow within narrow, well defined channels. But in the case of floods, as in the center of the image, the streams exceed its banks and can flow over much larger areas.

Flooding Return Period. Return period is the time interval that you would expect a flood of some size to occur on average at a given location. Return periods are very often used in engineering and zoning codes, dictating the type of events for which infrastructure must be designed to withstand. Indications that an event may exceed a 50 year return period for example, would be an indication that property damage may be likely to occur. The colored regions in Figure 3 show the estimated return period of flooding in the area, and based on these areas automated alerts can be sent to stormwater managers.

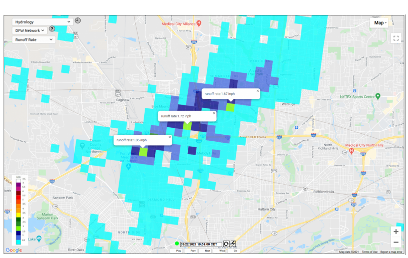


Fig. 4: RDHM generated runoff rate

Runoff rate and Water Depth. Depending on how wet or dry the soil conditions are, a certain amount of rainfall infiltrates into the ground. Runoff rate is the rainfall rate subtracted by the infiltration rate. When the soil is saturated, runoff rate is as high as rainfall rate itself which is more likely to produce flooding than in dryer soil conditions. This is depicted in Figure 4. Runoff is a primary indicator for high water, particularly in low lying areas, where the water will accumulate with a lag from the rainfall itself. In conjunction with topography, the model accumulates runoff and calculates water depth explicitly. This is very important for the protection of motorists who may become stuck should they attempt to cross these areas.

IV. Conclusion - Community Impact

The results of the RDHM pipeline in the cloud using Pegasus are displayed live to over 1000 emergency managers and first responders in the DFW metroplex, providing visual clues and mobile app alerts when observed values suggest that flash flooding is occurring over customizable regions of interest. Stakeholders can help validate model output by sending descriptions and pictures of flood affected areas within the app, as shown in Figure 5. The CASA website also displays live values from a network of road overtop sensors that measure the water level of adjacent streams in relation to the road. This can enhance confidence in the accuracy of the spatially continuous RDHM predictions across the region, and may eventually be able to be fed back into the model parameterizations. Additionally, experiments are ongoing to use streamflow variations from RDHM as a trigger to perform water quality sampling by the Environmental Affairs Department (EAD) at DFW International Airport to ensure pollutants are not being dispersed into the watersheds. While the current implementation of this workflow is for the North Texas region only, proposals have been submitted to expand its use to the Connecticut River Valley in MA and beyond.

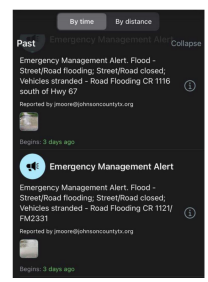


Fig. 5: Mobile Alert

Acknowledgment

The development of the RDHM Pegasus workflow is supported by the National Science Foundation through DyNamo (Award #1826997) and FlyNet (Award #2018074). Pegasus is supported by the NSF Award #1664162.

References

[1] V. Koren, S. Reed, M. Smith, Z. Zhang, and D.-J. Seo, "Hydrology laboratory research modeling system (hl-rms) of the us national weather service," *Journal of Hydrology*, vol. 291, no. 3, pp. 297–318, 2004, catchment modelling: Towards an improved representation of the hydrological processes in real-world model applications. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002216940400037X

[2] S. Reed, J. Schaake, and Z. Zhang, "A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations," *Journal of Hydrology*, vol. 337, no. 3, pp. 402–420, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169407000923

[3] National Oceanic and Atmospheric Administration (NOAA), https://oceanservice.noaa.gov/.

[4] National Weather Service (NWS), https://www.weather.gov/.

[5] H. Chen and V. Chandrasekar, "The quantitative precipitation estimation system for dallas-fort worth (dfw) urban remote sensing network," *Journal of Hydrology*, vol. 531,, pp. 259–271, 05 2015.

[6] E. Deelman, K. Vahi, G. Juve, M. Rynge, S. Callaghan, P. J. Maechling, R. Mayani, W. Chen, R. Ferreira da Silva, M. Livny, and K. Wenger, "Pegasus: a workflow management system for science automation," *Future Generation Computer Systems*, vol. 46, pp. 17–35, 2015, funding Acknowledgements: NSF ACI SDCI 0722019, NSF ACI SI2-SSI 1148515 and NSF OCI-1053575. [Online]. Available: http://pegasus.isi.edu/publications/2014/2014-fgcs-deelman.pdf

[7] I. Baldin, J. Chase, Y. Xin, A. Mandal, P. Ruth, C. Castillo, V. Orlikowski, C. Heermann, and J. Mills, "Exogeni: A multi-domain infrastructure-as-a-service testbed," in *The GENI Book*, R. McGeer, M. Berman, C. Elliott, and R. Ricci, Eds. Springer International Publishing, 2016, pp. 279–315.

[8] E. Lyons, G. Papadimitriou, C. Wang, K. Thareja, P. Ruth, J. Villalobos, I. Rodero, E. Deelman, M. Zink, and A. Mandal, "Toward a dynamic network-centric distributed cloud platform for scientific workflows: A case study for adaptive weather sensing," in *15th International Conference on eScience (eScience)*, 2019, pp. 67–76, funding Acknowledgments: NSF 1826997.

[9] D. McLaughlin, D. Pepyne, V. Chandrasekar, B. Philips, J. Kurose, M. Zink, K. Droegemeier, S. Cruz-Pol, F. Junyent, J. Brotzge, D. Westbrook, N. Bharadwaj, Y. Wang, E. Lyons, K. Hondl, Y. Liu, E. Knapp, M. Xue, A. Hopf, K. Kloesel, A. DeFonzo, P. Kollias, K. Brewster, R. Contreras, B. Dolan, T. Djaferis, E. Insanic, S. Frasier, and F. Carr, "Short-wavelength technology and the potential for distributed networks of small radar systems," *Bulletin of the American Meteorological Society*, vol. 90, no. 12, pp. 1797–1818, 2009. [Online]. Available: https://doi.org/10.1175/2009BAMS2507.1

[10] NOAA, "XMRG File Format," https://www.nws.noaa.gov/ohd/hrl/dmip/2/xmrgformat.html.