A Simple Data-Driven Model for Streamflow Prediction

Towards streamflow prediction in ungaged basins

Valliappa Lakshmanan[1,2] Jonathan J. Gourley[2] Zac Flamig[1,2] Scott Giangrande[3]

[1]Cooperative Institute of Mesoscale Meteorological Studies
University of Oklahoma

[2]Radar Research and Development Division
National Severe Storms Laboratory

[3]McGill University

AMS Annual Meeting, Phoenix, Jan. 12-15, 2009

Outline

1. Introduction
2. FIR Model
3. Results
4. Conclusions

Why a Data-Driven Model?

Justify complexity required in a hydrologic model

- Independent of physical descriptions of basin
- Independent of initial states such as soil moisture
- Can add variables as needed, to reach desired accuracy

Create streamflow predictions for ungaged basins

- Need only satellite/radar rainfall estimates
- Useful in areas without streamflow gages

Overview

Develop a data-driven streamflow prediction model

- Use observations of rainfall and runoff over the heavily instrumented Ft. Cobb basin in western Oklahoma
- Statistical model developed for a dataset of ten hydrologic events in which there was complete coverage by the KOUN polarimetric research radar
- Streamflow observations on three subcatchments in the basin.
- Data-driven model was evaluated for each event by considering the rainfall/runoff observations from the other nine, independent events.
- Will focus on model results from Tropical Storm Erin which produced streamflow having a return period greater than 100 years.

Ft. Cobb Watershed

- 813 km^2 in area and features a Micronet, a network of 15 stations that measure air temperature, rainfall, relative humidity, solar radiation, soil temperature at four depths, and soil water content at three depths.
- majority of the basin is within 100 km of the KOUN (polarimetric) radar
- duration of rainfall for these events ranged from 6-61 hours
- intense convective supercells with severe hail, squall lines with trailing stratiform rain, and tropical rain

Preprocessing

- divided into three bands according to terrain elevation
- rainfall within each band was estimated using polarimetric radar using several techniques
- Average rainfall within each band used to simulate streamflow at three basin subcatchments

Need for FIR Model

Can not ignore time

- simple input-to-output mapping model won't work
- because the streamflow is based on rainfall over a time interval.
- not enough to use rainfall estimates at just the current time.

Use finite impulse response (FIR) model

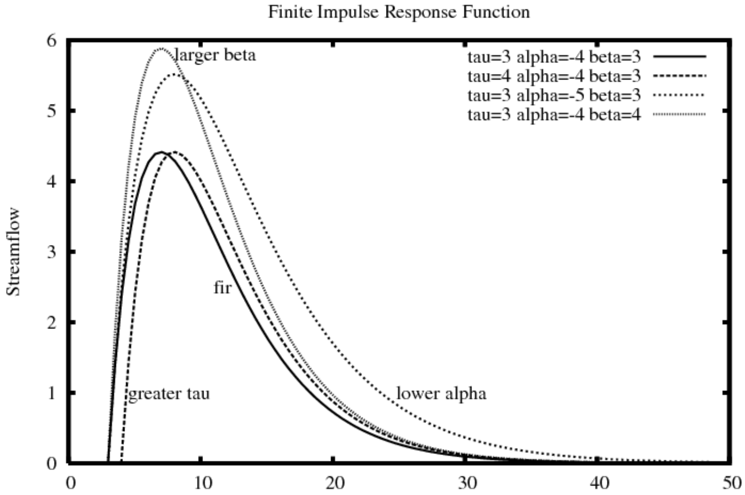
The FIR Model

$$\hat{S}(t) = \sum_{k=0}^{n} \sum_{j=0}^{t-\tau_k} (t - j - \tau_k)\beta_k e^{\frac{t-j-\tau_k}{-\alpha_k}} E_k(j) \tag{1}$$

where $\hat{S}(t)$ is the estimate of the streamflow at time t and E_k is a time series of measurements of rainfall over the k^{th} band. There are three parameters for each of n bands: β_k, α_k and τ_k. The impact of a band is given by β_k, the time delay between rainfall at the k^{th} band and the time that water from it reaches the stream is given by τ_k while the decay rate is given by α_k.

What does FIR mean?

Suppose there were to be rainfall of E_0 at time $t = 0$ and no rainfall after that.



What does FIR mean? (contd.)

Streamflow that would result because of this would begin at $t = \tau_0$, reach a peak and then decay exponentially. Since any digital series of measurements can be considered a sum of such $E_0(j)$, the resulting streamflow is the sum of curves of the form shown. The smaller the α, the slower the curve decays. The larger the β, the faster and higher it ramps up. The larger the τ, the greater the delay before the effect of rainfall shows up in the streamflow.

How to find parameters?

Minimize mean-square-error between observed and predicted streamflow ...

- Use Fourier Transforms
- The Fourier transform of a convolution of two data sets is the product of the Fourier transforms of the two data sets.
- The Fourier transform of $\hat{S}(t)$ follows the form of:

$$F(e^{-t/\alpha}u(t)) = \frac{1}{1/\alpha + j2f\pi} \tag{2}$$

- Generalized non-linear optimization of a complex dataset!

Genetic Algorithm

A more tractable approach is to work within the time domain itself.

- The equation is:

$$\hat{S}(t) = \sum_{k=0}^{n} \sum_{j=0}^{t-\tau_k} (t - j - \tau_k)\beta_k e^{\frac{t-j-\tau_k}{-\alpha_k}} E_k(j) \tag{3}$$

- If α_k and τ_k are known, the parameters b_k can be estimated using linear regression.
- A genetic algorithm was used to create reasonable values of α_k and τ_k.
- Then, linear regression was used to compute the best estimate of β_k.
- The resulting square error was used to determine the fitness of α_k and τ_k, for future generations of the genetic algorithm.

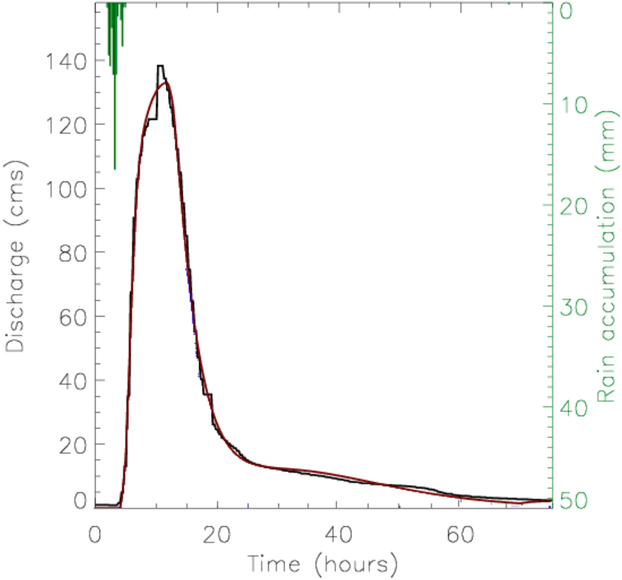
Training

Training of the model was carried out separately on nine events:

- 12-16 Jun 2005: Max rainfall accumulation of 96 mm.
- 30 Sep - 3 Oct 2005: 60 mm, several severe hail reports
- 14-17 Jun 2007: 201 mm, several severe hail reports
- 20-23 Jun 2007: 72 mm.
- 27 Jun - 2 Jul 2007: 120 mm.
- 2-4 Mar 2008: 32 mm, severe hail reports and tornadic storms.
- 31 Mar - 7 Apr 2008: 74 mm.
- 7-13 May 2008: 81 mm.
- 9-13 Jun 2008: 78 mm.

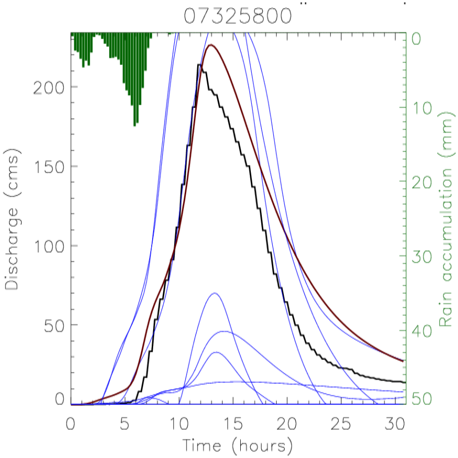
Fitting

12-16 June 2005 calibration case.



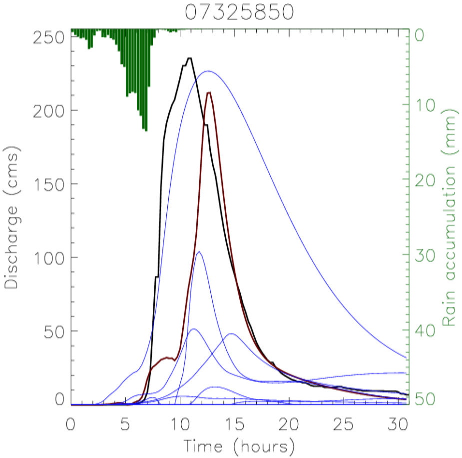
Ensemble

The empirical hydrologic model is an ensemble of nine FIR functions. The ensemble was independently evaluated on the 18-20 Aug 2007 case (Tropical Storm Erin: 307 mm in 24 hr)



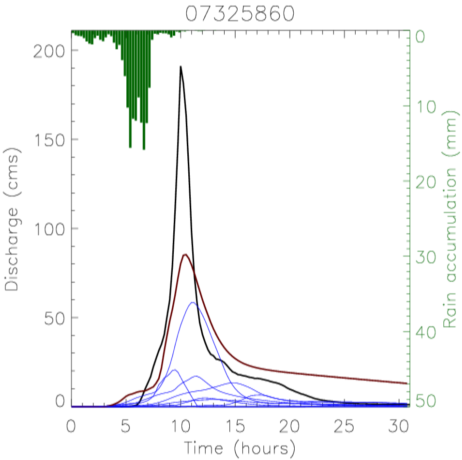
Ensemble

Sub-catchment 2



Ensemble

Sub-catchment 3



Conclusions

- set of fitted functions based on the individual FIR functions with rainfall inputs from the synthetic dual-pol rainfall algorithm encompassed the extreme event quite well for the 342 km^2 catchment
- Impressive considering no events of this extreme magnitude were included in the training data set.
- Skill in simulating runoff apparently decreases with smaller basin sizes.

Additional complexity in hydrologic modeling e.g. inclusion of initial soil moisture conditions, distributed soil types and land cover roughness values, etc. may be warranted at the flash flood scale whereas simplistic, parsimonious approaches are sufficient for larger basins.

Acknowledgments

Funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Severe Storms Laboratory (NSSL) or the U.S. Department of Commerce.