What is RFFA – regression for RFFA

Mulitple flood durations and why we need multiple estimates

How to incorporate duration into design values: event-based vs aggregation

Regression models at different durations require special considerations

There is a gap – we build models independently for each duration. We must be able to compare to existing models

What is a GAM – advantages, description

Where have GAMs been used before

Why predictor selection is difficult in GAMs…shrinkage based methods

Need to select a small, uncorrelated set of predictors

Why we describe our predictor selection method here

Objectives and research questions

Flow-duration curves

Moving window analyses

Copulas and multivariate analyses

Synthetic design hydrographs

Moving windows – average out the time spent above a threshold.

A common challenge in engineering design for retention-based systems is the need for duration-specific design values at ungauged locations. These applications often require frequency estimates of average streamflow for pre-determined durations. Moving-window analyses (alternatively: n-day flood) enable frequency analysis of the period with the highest average streamflow over n consecutive hours; see, for example . Frequency analyses are usually performed separately for each pre-determined duration, although some approaches (Barna, Cunderlik, Ourada, Javelle) attempt to scale estimates across durations by fitting an “average” distribution under strict assumptions.

In the typical case, where frequency analyses are performed separately, regression equations are typically built individually for each duration to regionalize the estimates…(or built jointly, re-estimated for each duration?)

This introduces some assumptions…

Our focus is the structure of the regionalization

*Durations at ungauged locations*

Developing regional regression equations for duration-specific methods

Nonlinear and data-driven approaches

Thorough (practically motivated) reliability and predictive performance assessment

the period of n consecutive hours with the highest average streamflow, where the streamflow is sometimes averaged over multiple flood events.

“Flood-duration flows are reported in dimensions of volume per time and units of cubic feet per second; to convert to total volume the flow rate is simply multiplied by the length of the duration interval considered.»

**Random thoughts:**

We average discharge across duration (average flowrate)

Regional.

Hydrologically: The processes that produce a high average flowrate over 1 hour may be different than the processes that produce a high average flowrate over 24 hours. These differences may be so fundamentally different that we may have to adapt the functional form of the relationship, not just the model coefficients. Practically: a model would have to adapt to different relationships at different durations, or we risk imposing an artificial mathematical relationship on a changing hydrological process.

This is acknowledged by developing, or if it is, separate parametric regression models are developed for different average flowrates (burden to practitioners). Examples: NIFS v RFFA2018

Recent developments leave us well-positioned to develop more accurate and more reliable regression models. Data-driven approaches. And examine potential watershed processes governing different durations.

A challenge when working with some of the newer statistical approaches is available performance metrics

Practically-oriented

We did not find evidence for different covariates at different durations, just different functional relationships…maybe because we use annual maxima.

A common challenge in engineering design for retention-based applications is the need for duration-specific design values at ungauged locations. Since these applications focus on total storage capacity, they often require frequency estimates of average streamflow over pre-determined durations. Moving-window analyses (alternatively termed sustained flood flow, n-day flood or flood-duration-frequency analyses) enable frequency analysis of the period with the highest average streamflow over n consecutive hours; see, for example . Frequency analyses are typically performed separately for each pre-determined duration, although some approaches (see, e.g. Barna, Cunderlik, Ourada, Javelle) attempt to scale estimates across durations by fitting an “average” distribution under strict assumptions.

In the typical case, where frequency analyses are performed separately for each pre-determined duration, regression models are usually used to extend estimates to ungauged locations.

What is done, what is done.

What they do – regression

Why this is problematic. This approach is not without its limitations. What are those?

* Impose an artificial structure on the relationship
* Burden to practitioners

Why is this a burden to practitioners? Always challenging to select covariates and model for structure (general RFFA problem). So, although there is some evidence we shouldn’t be using the same model on all durations, building and analyzing these models is time consuming and it’s generally not practical to construct a separate one for each duration.

Building a rffa model is time consuming (predictor selection, structure selection, estimation, validation, validation constraining sometimes what models we use). It is usually not practical to rebuild this relationship for every pre-determined duration. For this reason, a single regression relationship is typically built, and then the parameters are re-estimated for each pre-determined durations (lamontagne, lind, kennedy, nve).

Development of a regional regression relationship is a significant burden.

This approach has its limitations. For starters, even though model coefficients re-estimated….assumes same functional form…