A Statistical-Distributed Modeling Approach for Flash Flood Prediction: Phase 1 Results

To be presented Wed., June 9, 2004 at the DOH/RDM 2004 (Session 17)

Seann Reed*
John Schaake
Victor Koren
Michael Smith
Dong-Jun Seo

HL/OHD/NWS/NOAA

Brief History

2002 DOH Conference: John Schaake proposes approach

2002 – Jan 2003:

- General description of approach included in FFGIT Report
- Continued HL-RMS/DMS development makes testing approach feasible

June 2003 – June 2004:

Proof of concept with available data and tools

June 2004 – June 2005

Proposed: Validate science at smaller scales and prepare for focused testing

Definition

Statistical: Use statistical characteristics of modeled historical events as comparative indices to assess the threat level of forecasted flood events. **Addresses large modeling errors over small areas.**

Distributed: Use HL-RMS **distributed parameter estimation** and modeling techniques to provide high resolution information useful for WFO scale modeling.

Characteristics of Current and Proposed Approaches

Traditional FFG

- Difficult to implement consistently across space.
- Lumped rainfall-runoff models (300-5,000 km^2)
- Gridded threshold runoff awkwardly attempts to address scale issues
- Does not address large uncertainties for small, ungauged basins

Statistical-distributed (FFG)

- Easier to implement consistently.
- Distributed rainfall-runoff calculations
- Distributed routing calculations
- Addresses simulation uncertainty through frequency-based rather than flow-based comparisons

Goals

- **Phase 1: Proof of concept using readily available data and parameterization schemes (StageIII/MPE and RFC headwater basins)**

- **Phase 2: Scientific validation and testing at scales more commensurate with anticipated application (WFO scales)**

  (use improvement upon the current lumped FFG approach as the standard for success)

Outline

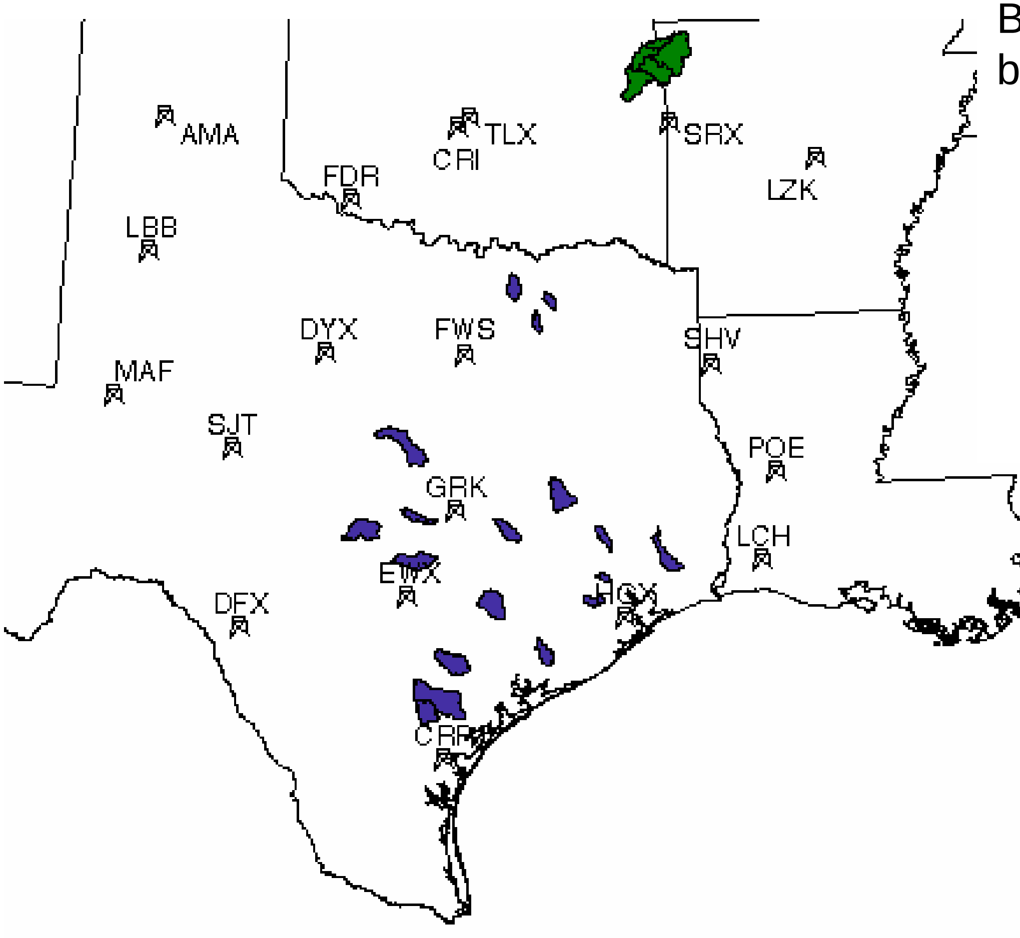
Phase I

- **Results at basin outlets**
- **Verification using the critical success index (CSI)**
- **Gridded results**
- **Which frequency to choose as the flooding frequency?**
- **Conclusions**

Planning for Phase II

- **Modeling/data requirements for different scales**
- **Proposed work**

Study Basin and Radar Locations

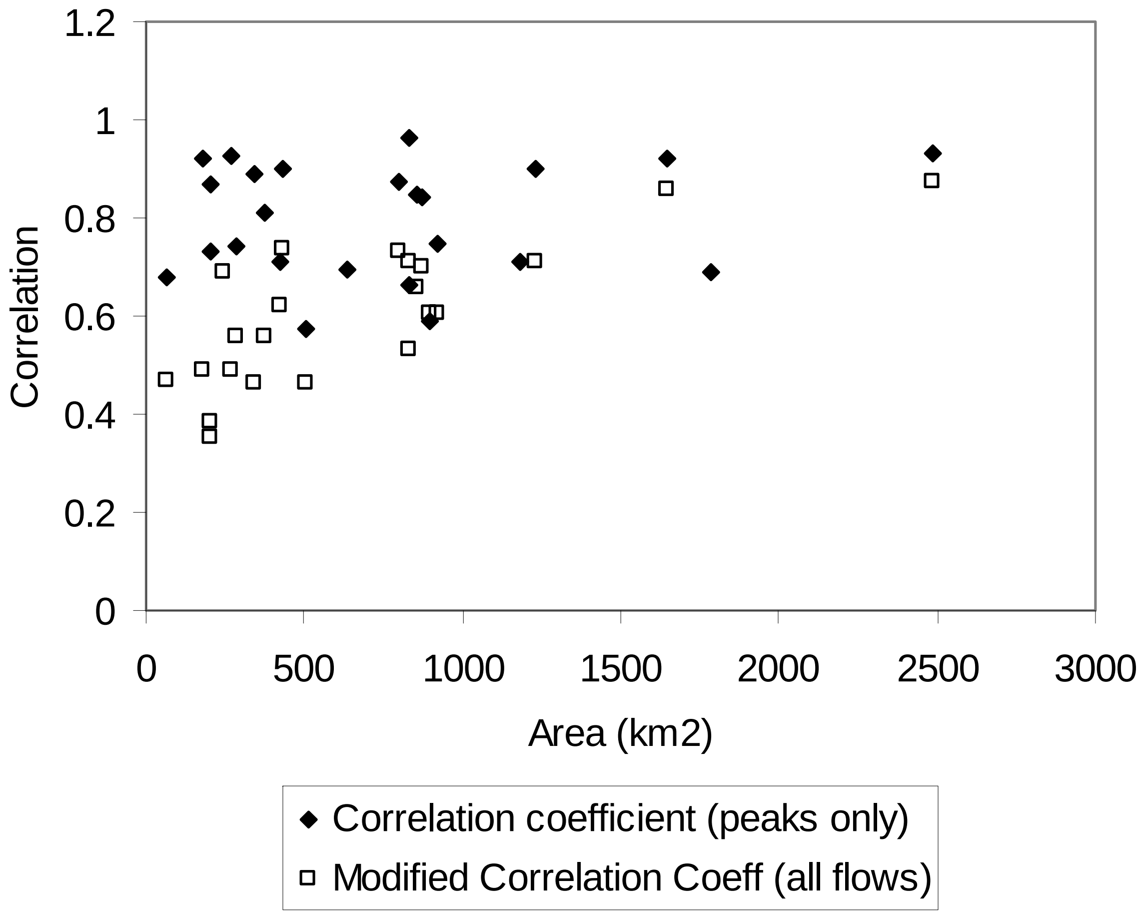


Basins where distributed model has been run:

Green: ABRFC basins (7)
Size Range: 65 – 2484 km2
Smallest tpeak: ~ 7 - 8 hrs

Blue: WGRFC basins (19)
Size Range: 178 – 1790 km^2
Smalles tpeak: ~ 10 hrs

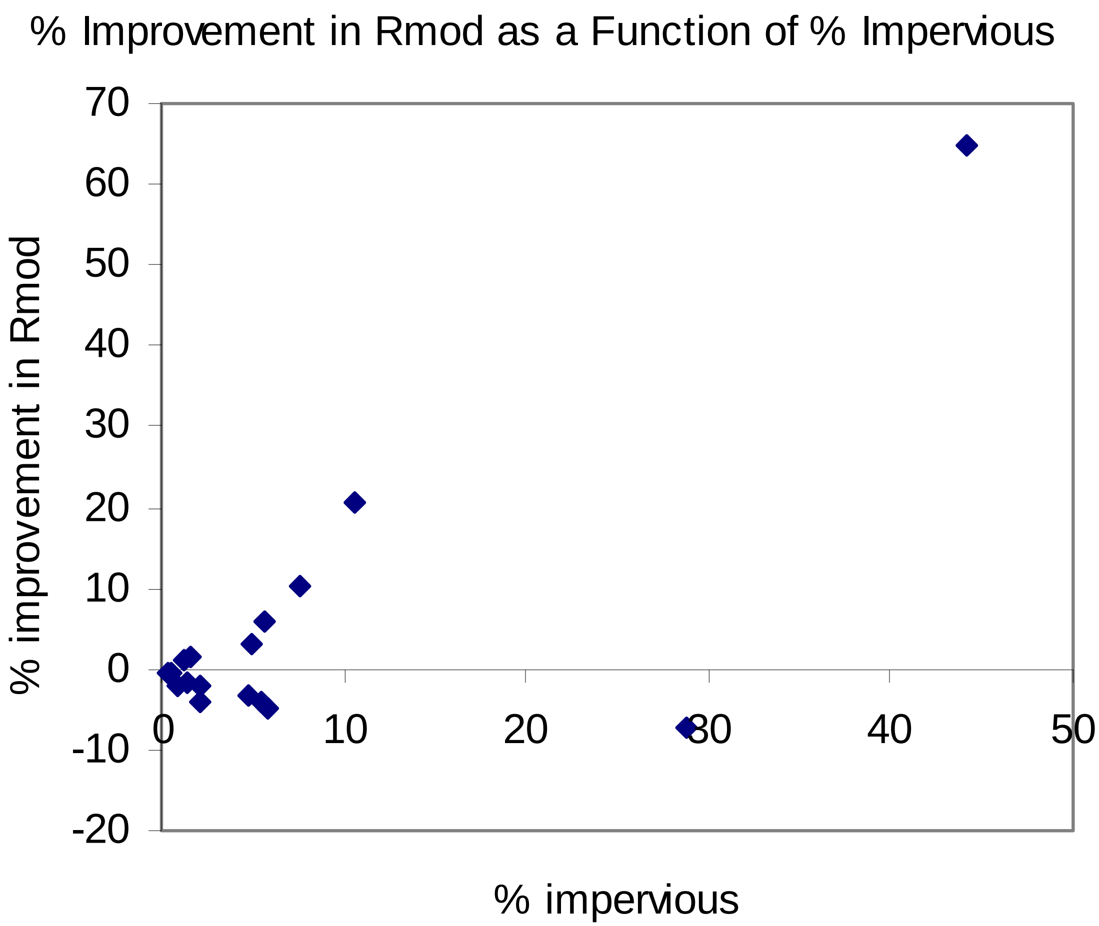
Simulation Accuracy as a Function of Scale for Uncalibrated HL-RMS Model Runs



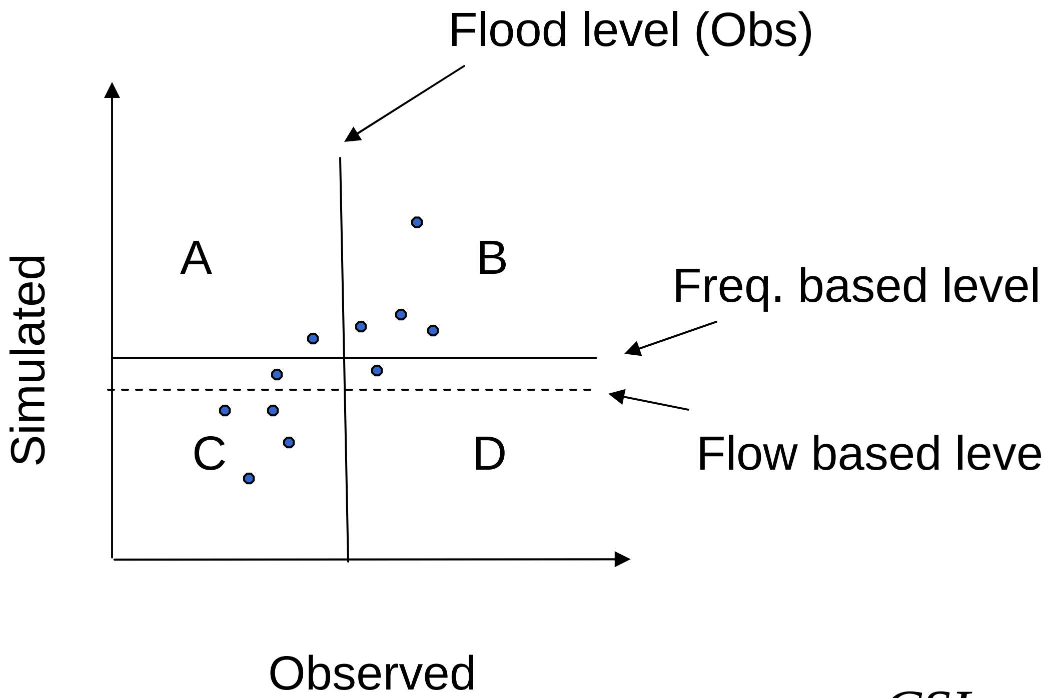
% Improvement from Uncalibrated to Calibrated (DMIP Results)

	Area (km2)	%Improvement Event Peakflow	
Christie	65	-26.7	
Kansas	285	-0.7	
Savoy	433	0.1	
Eldon	795	19.5	
Blue	1,233	15.4	
Watts	1,645	3.9	
Tiff City	2251	11.8	
Tahlequah	2,484	-1.3	

Improvement by Using GIS-based Impervious % Estimates



POD, FAR, CSI Definitions



$$POD = \frac{B}{B+D}$$

$$FAR = \frac{A}{A+B}$$

$$CSI = \frac{1}{1/(1-FAR)+1/POD-1}$$

Notes:

- The statistical-distributed approach forces POD = 1- FAR
- CSI is equally sensitive to changes in POD or FAR when POD = 1 – FAR

(Gerapetritis and Pelissier, 2004)

Distributed vs. Statistical-distributed results for Different Flood Frequency Thresholds

Aggregate Results for 7 ABRFC Basins (Uncalibrated Model)

Ret. Period (yrs)	**0.25**	**0.5**	**0.7**	**1.0**	**1.5**
# of floods per basin	22	10	7	5	3
POD	0.94	0.77	0.78	0.69	0.57
PODS	0.86	0.78	0.86	0.71	0.67
FAR	0.36	0.24	0.19	0.25	0.56
FARS	0.14	0.22	0.14	0.29	0.33
CSI	0.614	0.615	0.655	0.558	0.333
CSIS	0.760	0.638	0.750	0.556	0.500
% CSI Improvement	24	4	15	-0.4	50

Aggregate Results for 7 ABRFC Basins (Calibrated Model)

Ret. Period (yrs)	**0.25**	**0.5**	**0.7**	**1.0**	1.5
# of floods per basin	22	10	7	5	3
CSI	0.571	0.547	0.591	0.674	0.467
CSIS	0.760	0.656	0.690	0.707	0.448
% CSI Improvement	33	20	17	4.9	-4

Total Simulation Period is 5.5 years

Distributed vs. Statistical-distributed results for Different Flood Frequency Thresholds

Aggregate Results for 17 WGRFC Basins (Uncalibrated Model)

Ret. Period (yrs)	**0.25**	**0.5**	**0.7**	**1.0**	1.5
# of floods per basin	22	10	7	5	3
CSI	0.436	0.432	0.389	0.407	0.419
CSIS	0.483	0.442	0.494	0.418	0.425
% CSI Improvement	10.7	2.3	27.0	2.7	1.4

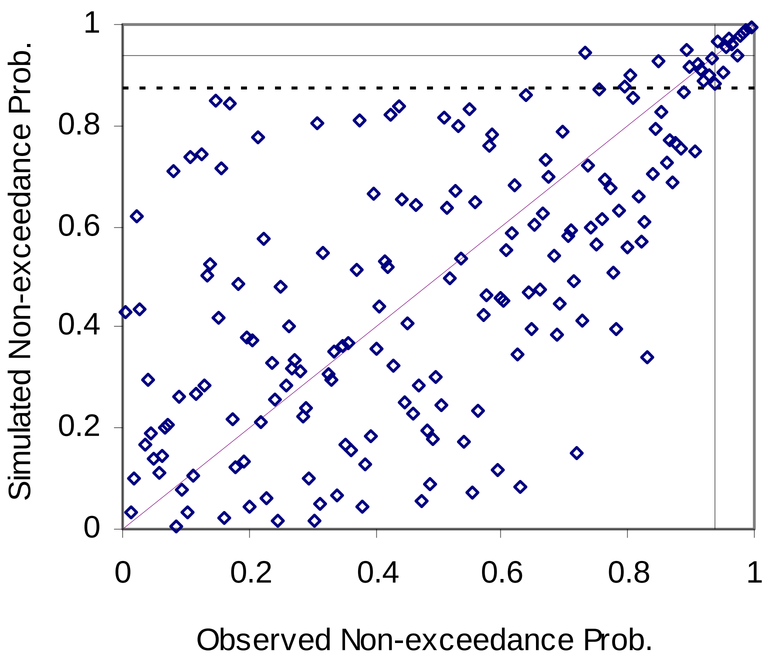
Total Simulation Period is 5.5 years

Example Individual Basin Results

ATIT2

Ret Per: 0.5 yrs (11 floods)

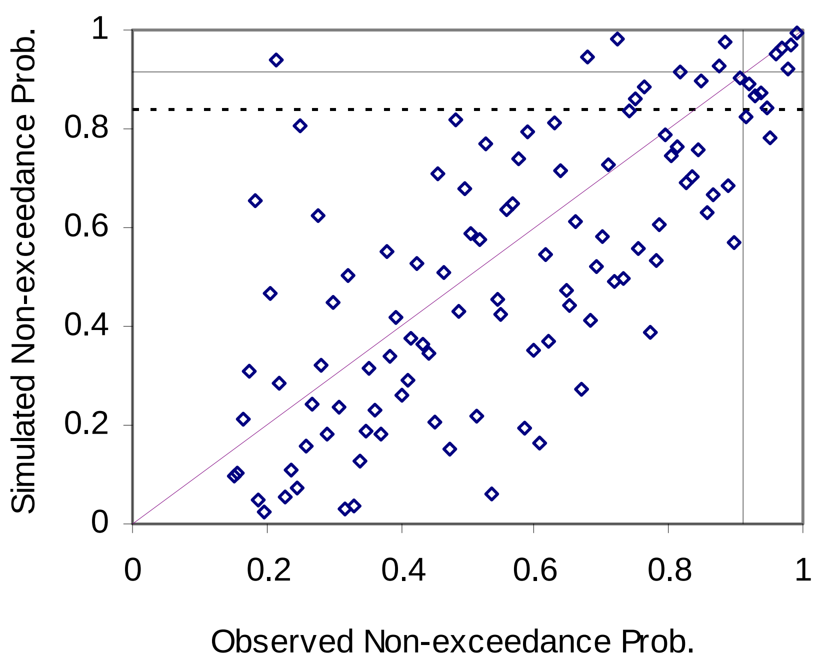
CSI Imp., 10.7%



KNLT2

Ret Per: 0.5 yrs (11 floods)

CSI Imp., -11%

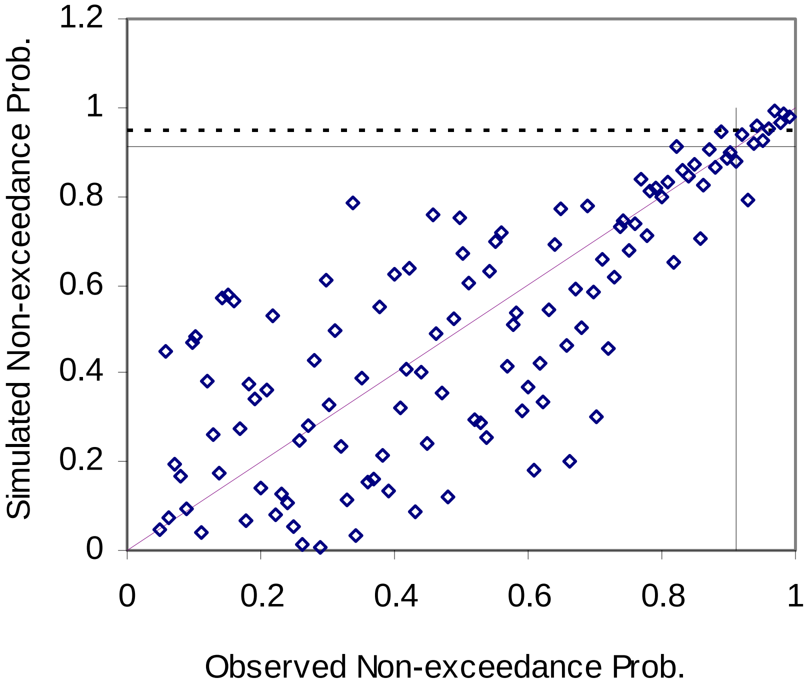


Example Individual Basin Results

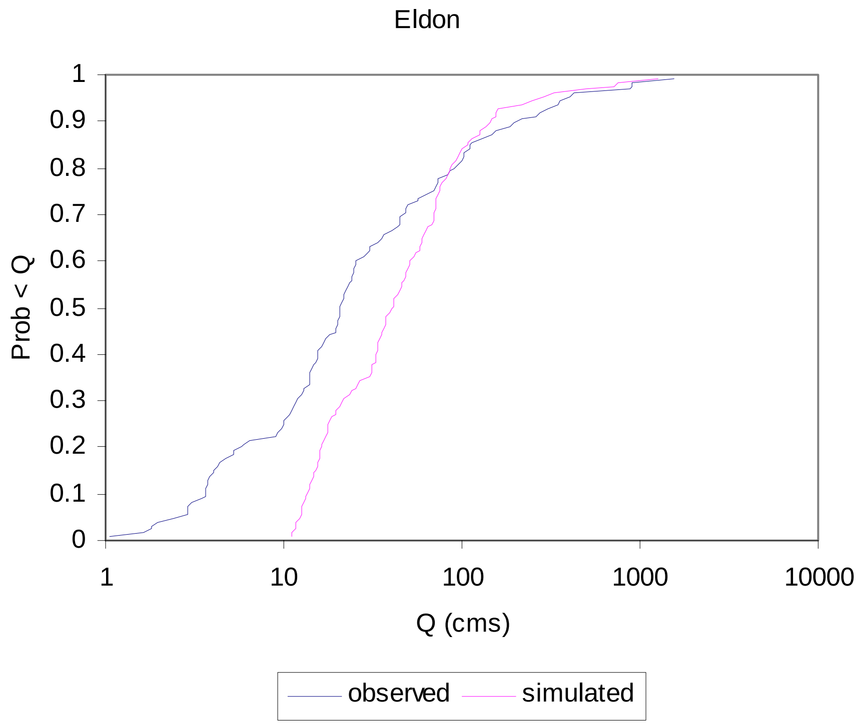
Eldon

RetPer: 0.5 yrs (11 floods)

CSI Imp., 27%



Distribution Functions for Eldon



Comparison to Archived FFG to in Basins with Known Flood Stage (WGRFC)

- Obtained archived gridded FFG data from National Precipitation Verification Unit (NPVU)

- Generated time series of basin mean FFG

- Visually compared 1 hr, 3 hr, 6 hr time series with 1, 3, and 6 hr cumulative MAPX; tabulated hits, misses, and false alarms

- Tabulated hits and misses using distributed modeling and statistical distributed approaches by comparing with known flooding flow.

* Results from distributed model and statistical distributed model were at least as good as the FFG approach.

Functions Added to the Object-oriented Research Modeling System (ORMS)

1. **Calculate basin average FFG for a historical run**

2. **Calculate hits based on lumped FFG from 1 (using lumped or gridded rainfall)**

3. **Identify/store simulated flood peaks for all pixels**

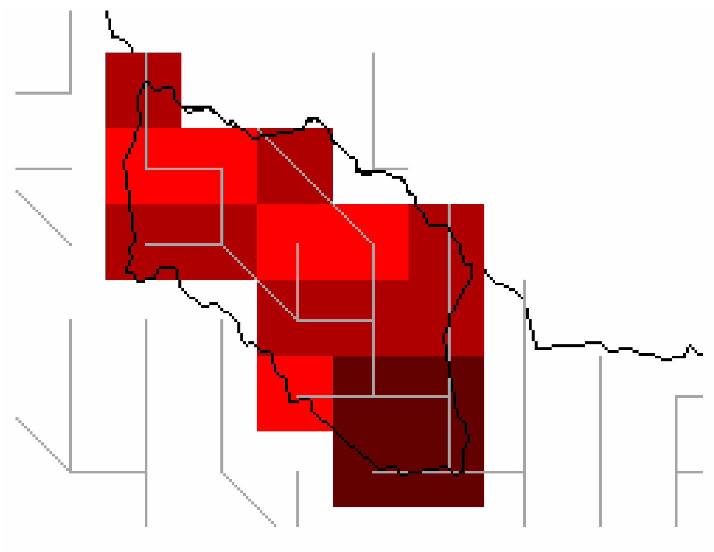
4. **Compute frequency associated with current flow based on information from 3.**

5. **Calculate "hits" for each pixel based on a prescribed flood frequency (output a grid or frequency time series)**

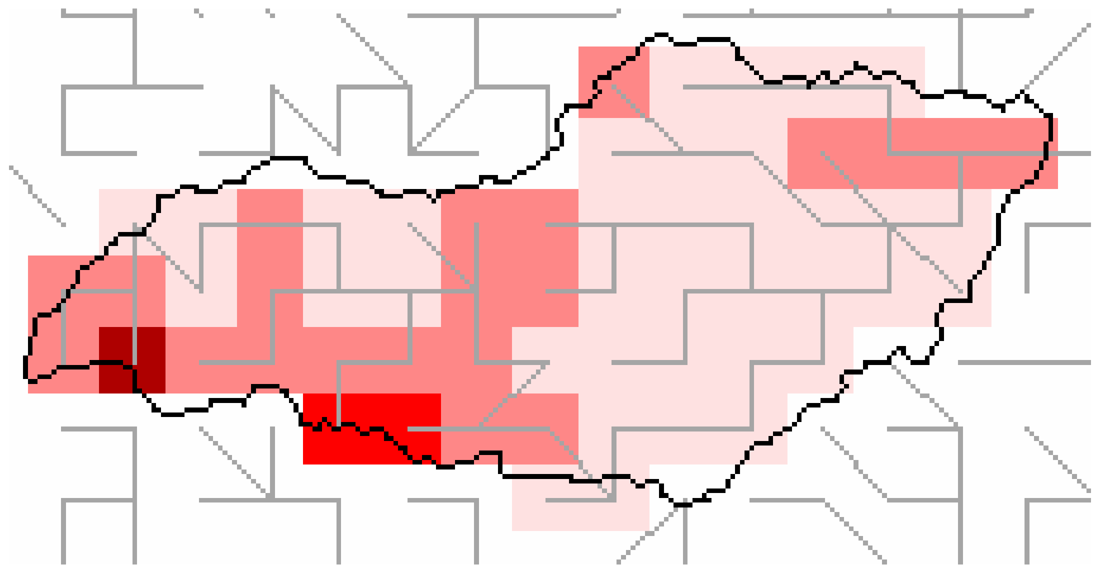
6. **Pre-processing procedure for estimating impervious percent: complements other, more extensive, HLRMS parameter estimation research.**

FFG and Spatial Scale Issues

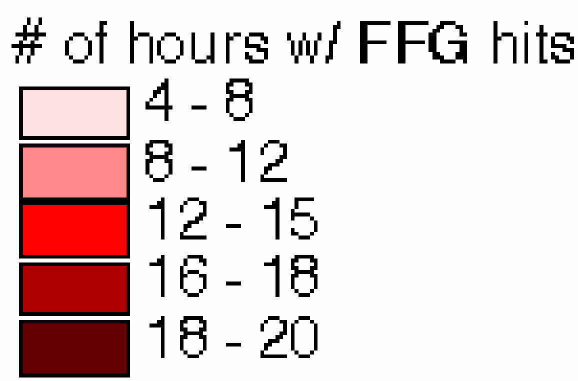
Cowleech Fork at Greenville, TX (GNVT2) (201 km2)



Onion Creek (ATIT2) (831 km2)

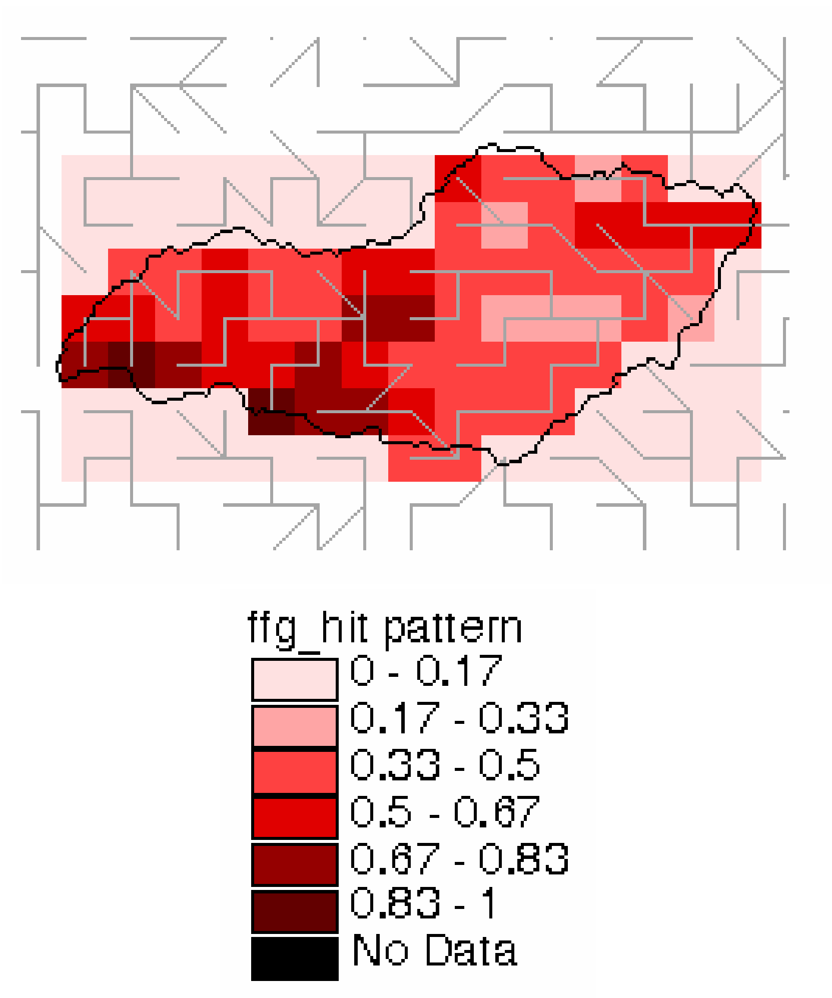


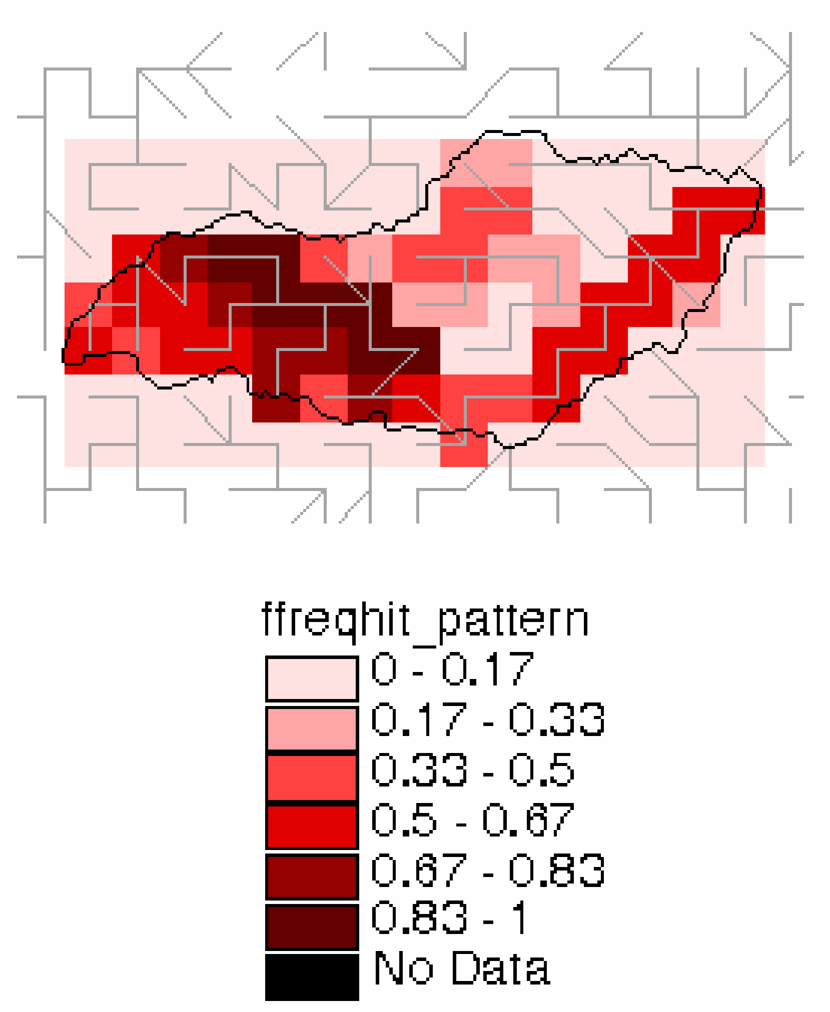
# of hours with lumped FFG hits: 16



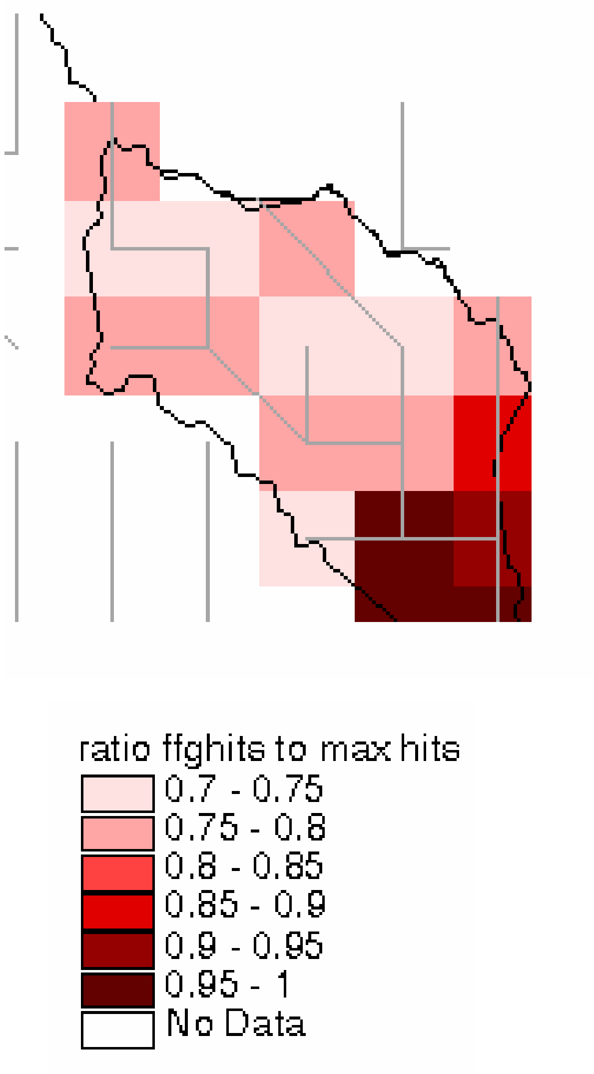
# of hours with umped FFG hits: 5

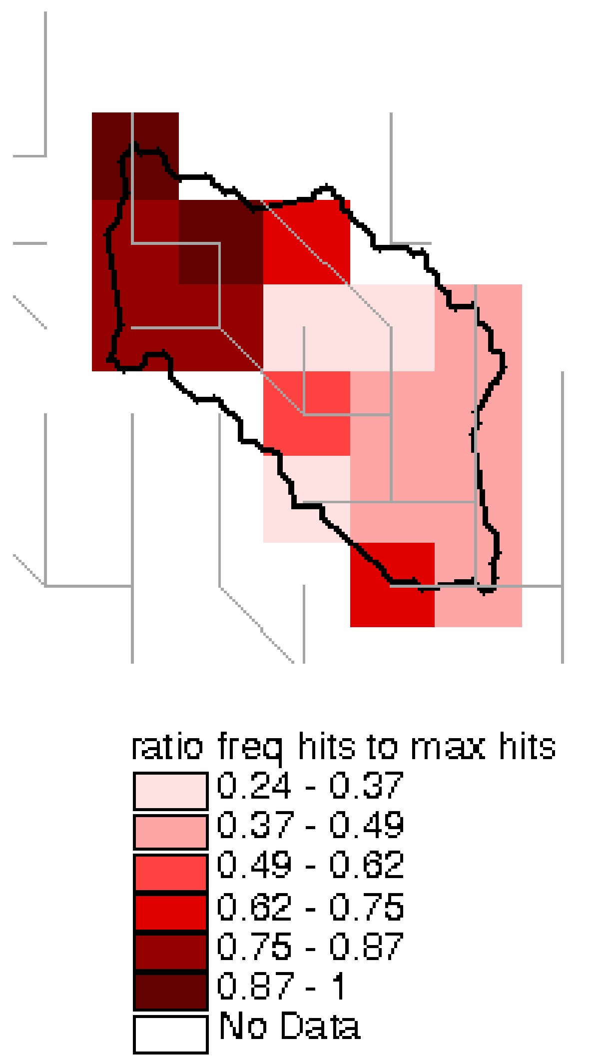
Spatial Pattern of Stat-dist Hits versus FFG Hits (ATIT2)



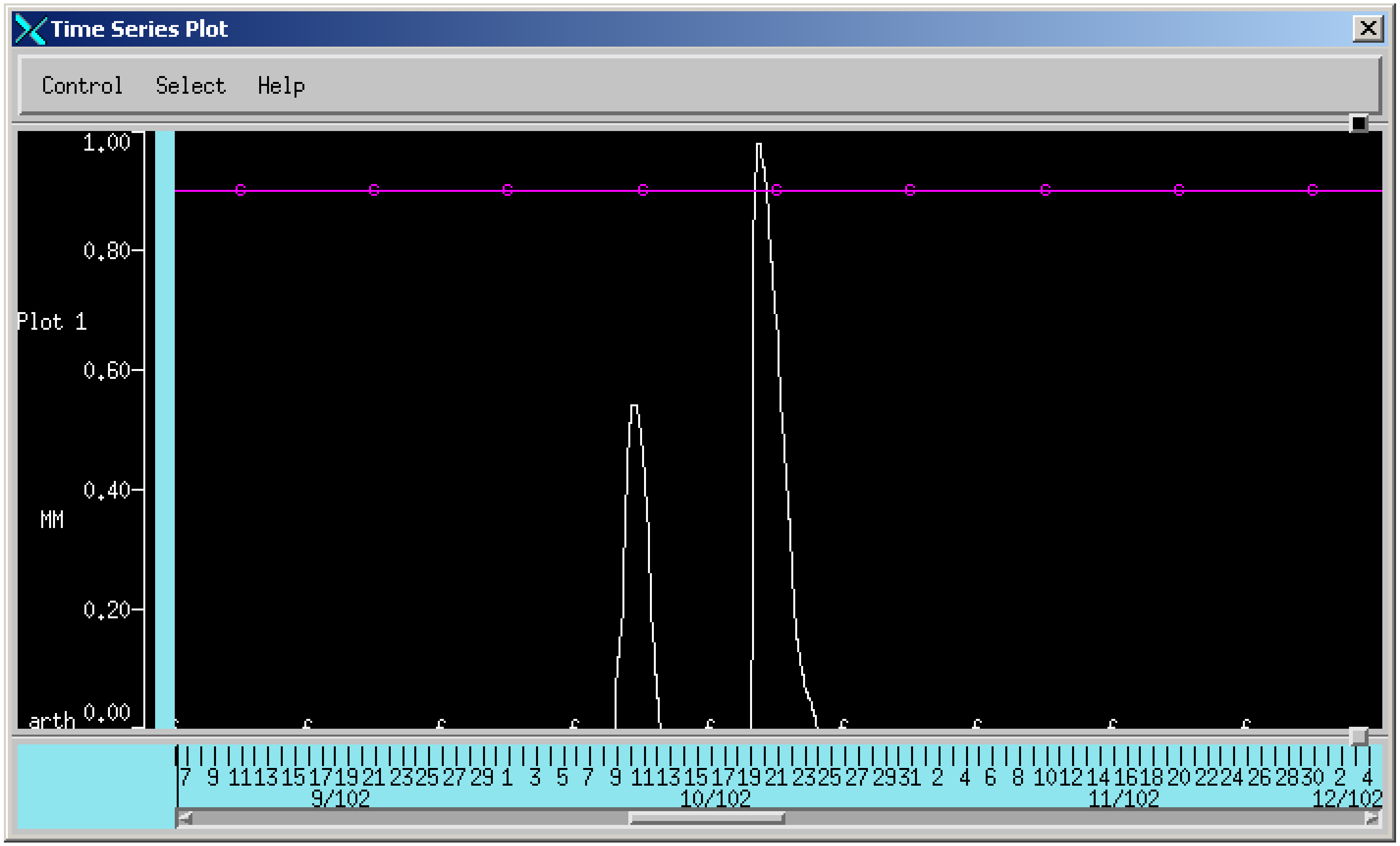


Spatial Pattern of Stat-dist Hits versus FFG Hits (GNVT2)

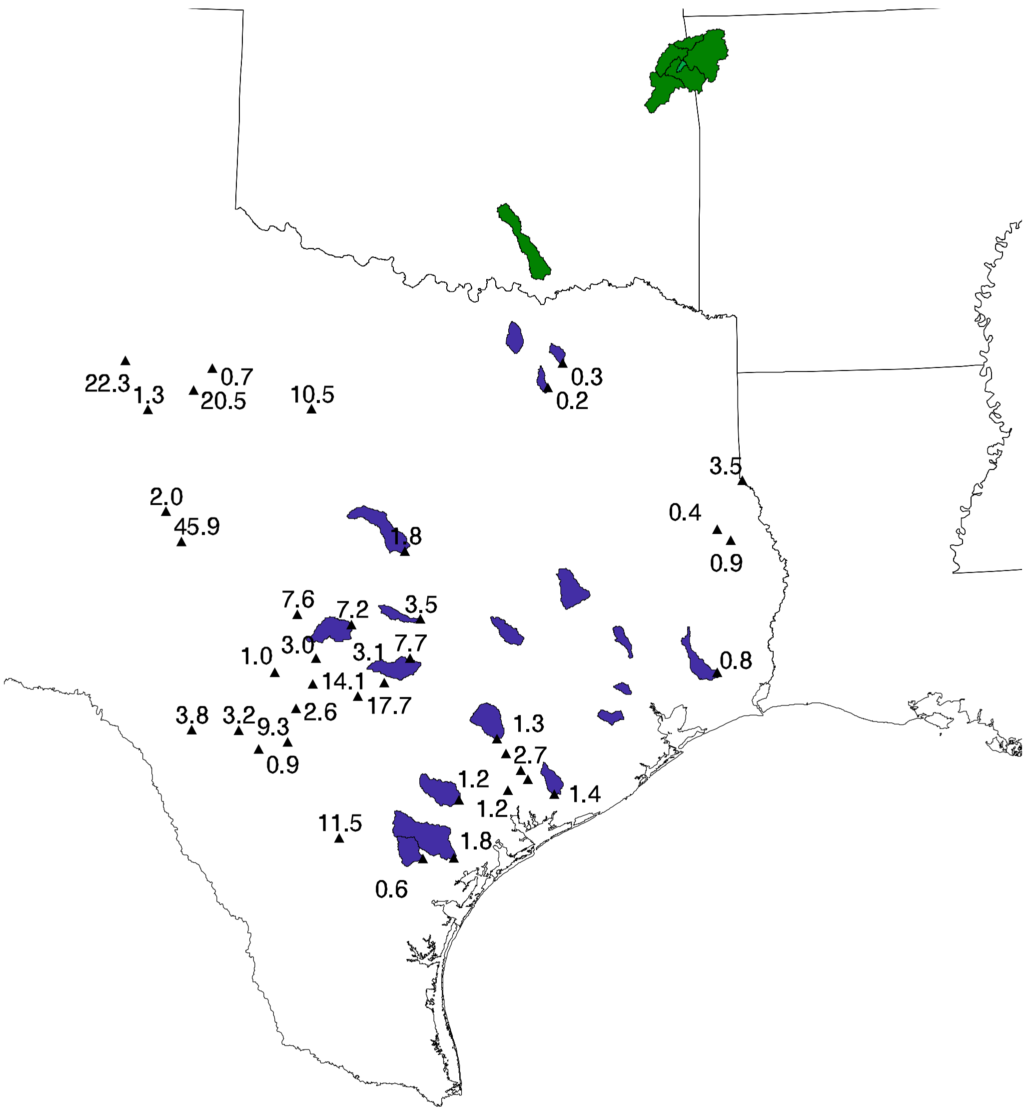




Simulated Frequency Time Series for GNVT2



Which frequency to choose as the flooding frequency?

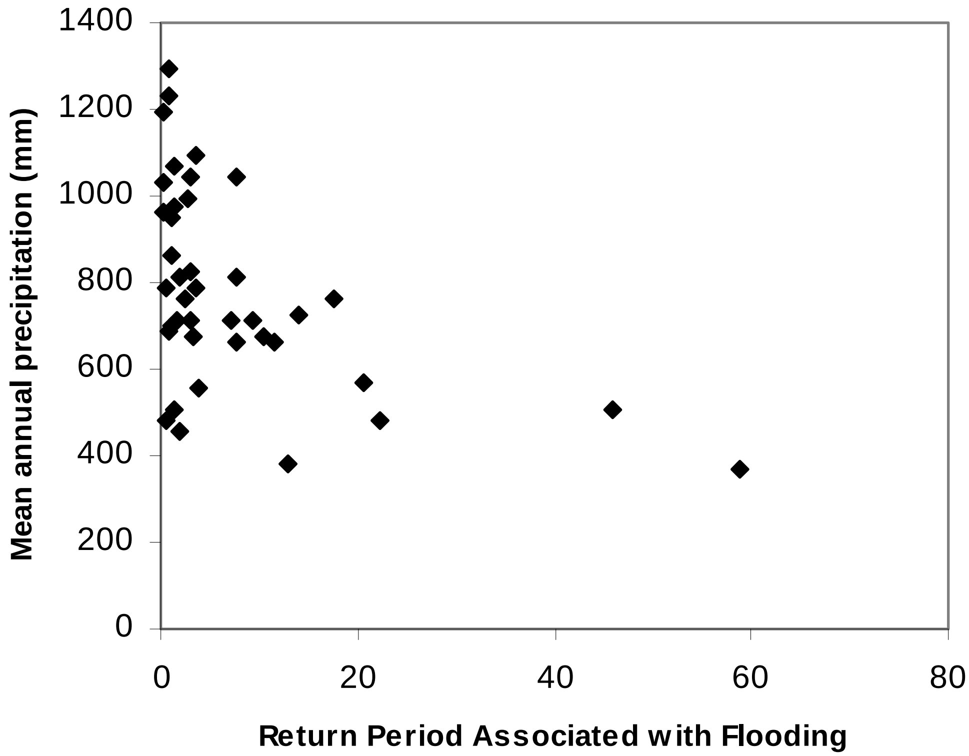


▲ Locations where **return period associated with flooding flow** has been estimated

- flooding flows from RFC database
- corresponding frequencies estimated using basin-specific data from Asquith and Slade, WRIR 99-4172.
- only considered basins with no undesirable USGS data flags (e.g. regulation or diversion)

How does frequency associated with flooding flow vary in space? With scale?

Mean annual precipitation vs. flooding flow return period for 40 basins in Texas



- Geomorphology literature suggests a relationship between bankfull flow and scale

- This data set is not adequate to examine the scale issue because climate differences dominate the flood frequency variations.

CorrCoef(RetPeriod, BasArea) = -0.06

CorrCoef(RetPeriod, AnnPrec) = -0.5

Phase 1: Summary and Conclusions

- Statistical-distributed approach accounts for spatial variation in hydrologic properties and increasing errors at small scales

- An analysis of FFG archives shows that At RFC outlets, the performance of a distributed modeling approach is at least as good as FFG

- At RFC outlets, a statistical-distributed approach improves upon distributed approach

- Real benefits from statistical-distributed approach are expected at interior points
    - scale effects were demonstrated
    - Phase II will focus on smaller scale validation

- Statistical-distributed approach leverages and complements other work (HL-RMS development, parameter estimation, radar rainfall re-analysis)

Modeling Approaches for Different Scales

	Areal Flooding	**Small Basins**	**Intermediate Basins**
Basin Areas:	None	5 km^2 – 50 km^2 scales	50 km^2 – RFC scales
Required spatial discretization:	Same as rainfall	1x1 – 2x2 km	2 x 2 – 4 x 4 km
Est. Times to peak:	NA	1 – 2.6 hours	2.6 – 12 hours
Precip time scales required.	**15 min**	**15 min – 30 min**	**1 hour**
Historical data:	Consider using 15 minute gauge data (use nearby “representative” gauge)	High res. (sub-hourly) radar archives (quality and record length suspect)	StageIII/MPE archives
Real-time “event” data:	DHR or MPE 15 min.	DHR or MPE 15 minute	MPE multi-sensor hourly totals
Initial states:	Maintained in large-area gridded SMA model using hourly MPE data.	same	
Major assumptions:	• Uniform rainfall and runoff even for short durations • No routing	Spatial variability of rainfall becomes important for all but the smallest basins	
Limitations/ Unknowns:	More difficult to verify	Limited archived radar data at this spatial-temporal resolution. Quality unknown.	Archived StageIII/MPE data have non-stationary biases (re-analysis)
Short/Medium Term Outlook:	Good	Uncertain. May require statistical disaggregation of rainfall data.	Good

Proposed Phase II Tasks

- Validate approach using small basin streamflow data

    - Identify study areas/collaborators

    - Gather and process precipitation data needed for historical analysis

    - Collect and pre-process spatial data to allow for higher resolution application

    - Make any necessary HL-RMS enhancements for higher resolution data ingest and smaller element modeling