Empirical Flow Parameters: A Tool for Hydraulic Model Screening

Theodore G. Cleveland
Caroline M. Neale
Cristal C. Tay
George R. Herrmann

Acknowledgements

- Contributions by W.H. Asquith, Hatim Sharif, Xiaofeng Liu, and K.B. Strom
- Funding and direction provided by the Texas Department of Transportation through research contract 0-6654 Empirical Flow Parameters - A Tool for Hydraulic Model Validity Assessment

Background

- Highway crossing in Texas (West of Guadalupe Mountain National Park)

Background

- Flow is uncommon – arid climate, but it does happen from time-to-time

Background

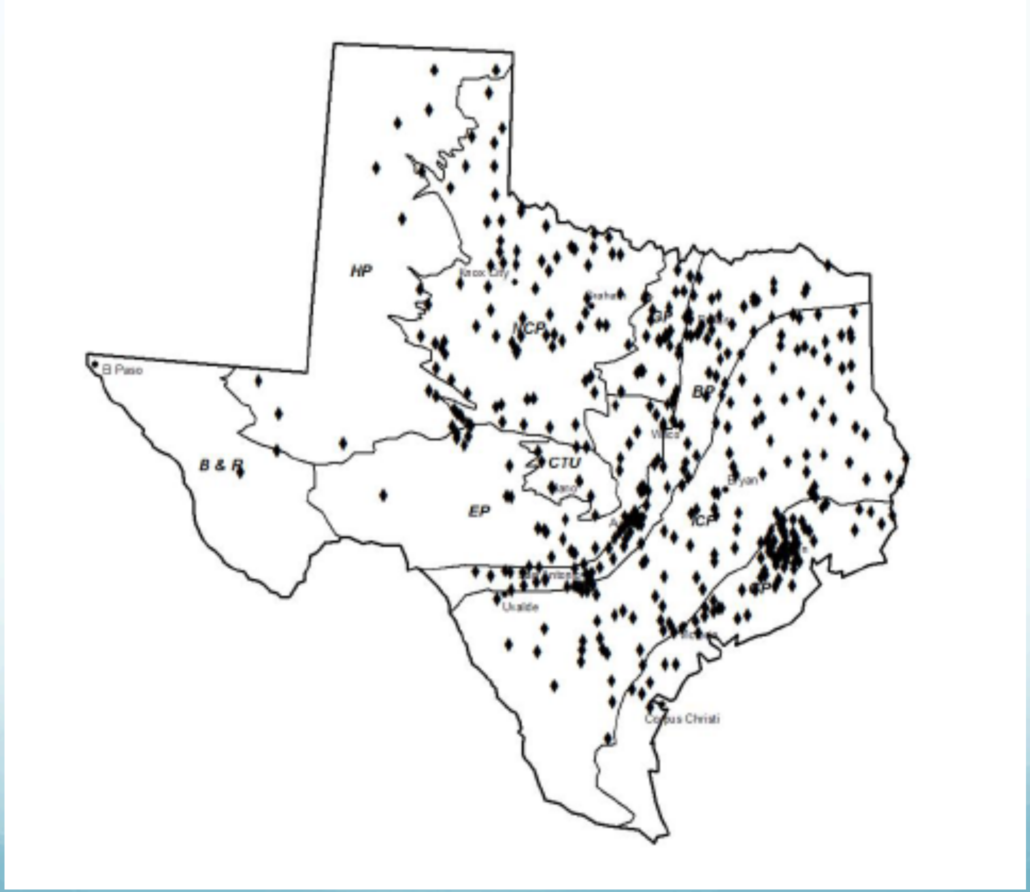
- Flow is uncommon – arid climate, but it does happen from time-to-time

Background

- Good faith to reporting velocities (needed for assessing forces on bridge piers, and assessing erosion and scour potential) that are unusually large and in some instances absurd.
  - What is unusually large (or small)?
  - Are there independent ways to assess computed velocities based on prior *observational* experience?

Background

- NWIS database for Texas, and just plotted Q and V



Background

- NWIS database for Texas, and just plotted Q and V

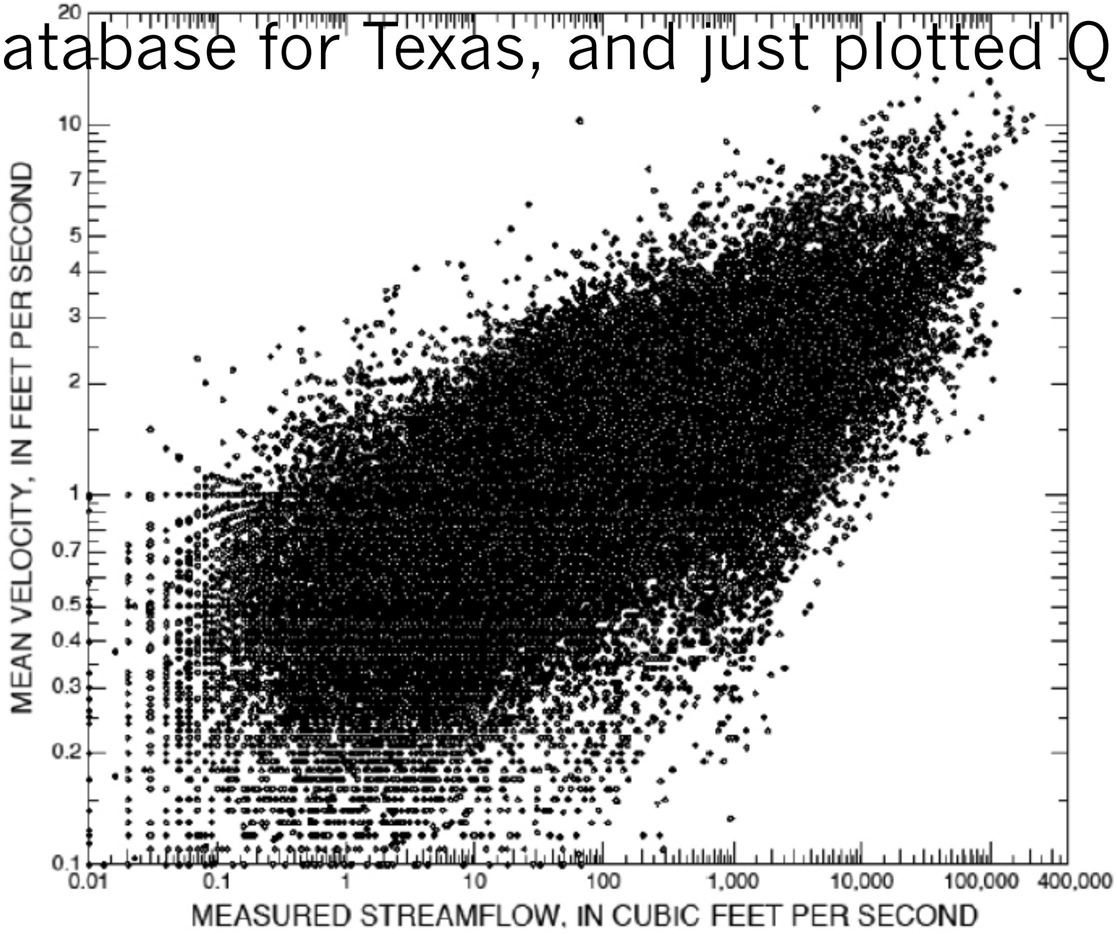


Figure 1: Mean Section Velocity versus Streamflow from U.S. Geological Survey Streamflow Measurement Database in Texas (adapted from Asquith and Herrmann, 2009).

Measurements

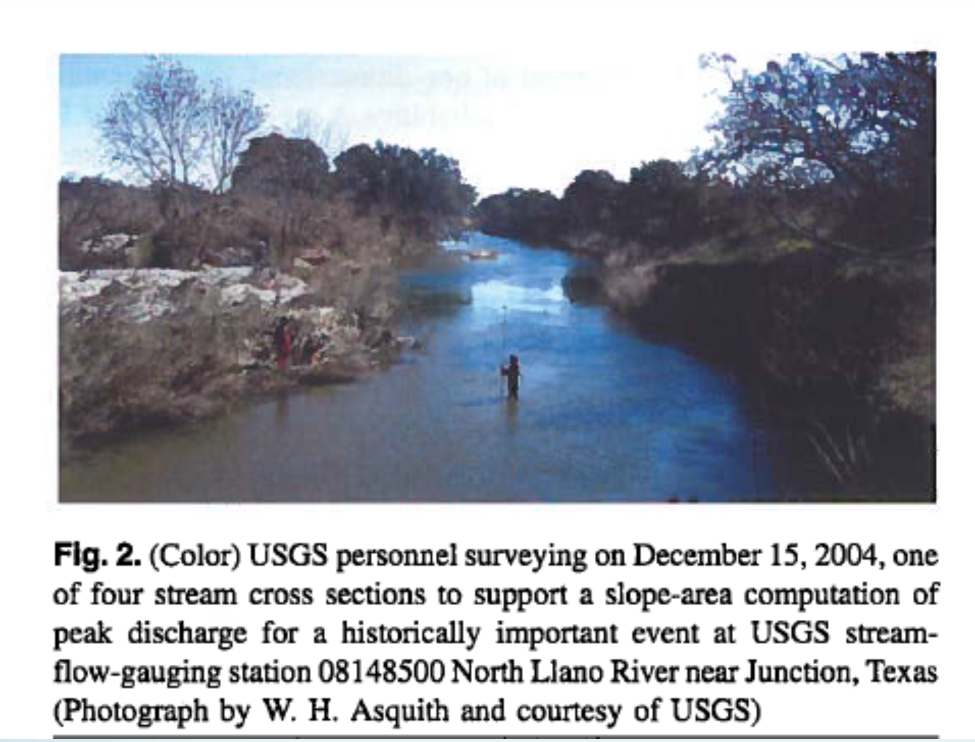


Fig. 2. (Color) USGS personnel surveying on December 15, 2004, one of four stream cross sections to support a slope-area computation of peak discharge for a historically important event at USGS streamflow-gauging station 08148500 North Llano River near Junction, Texas (Photograph by W. H. Asquith and courtesy of USGS)

- Low and high magnitude

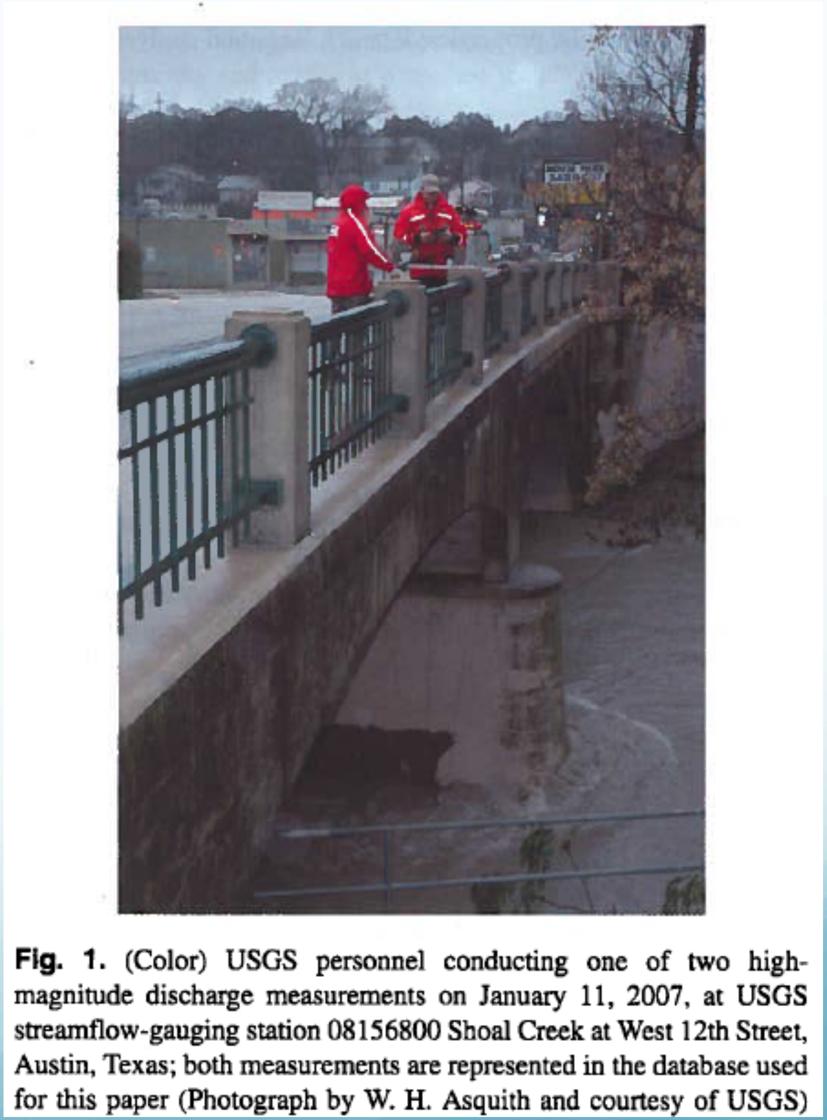
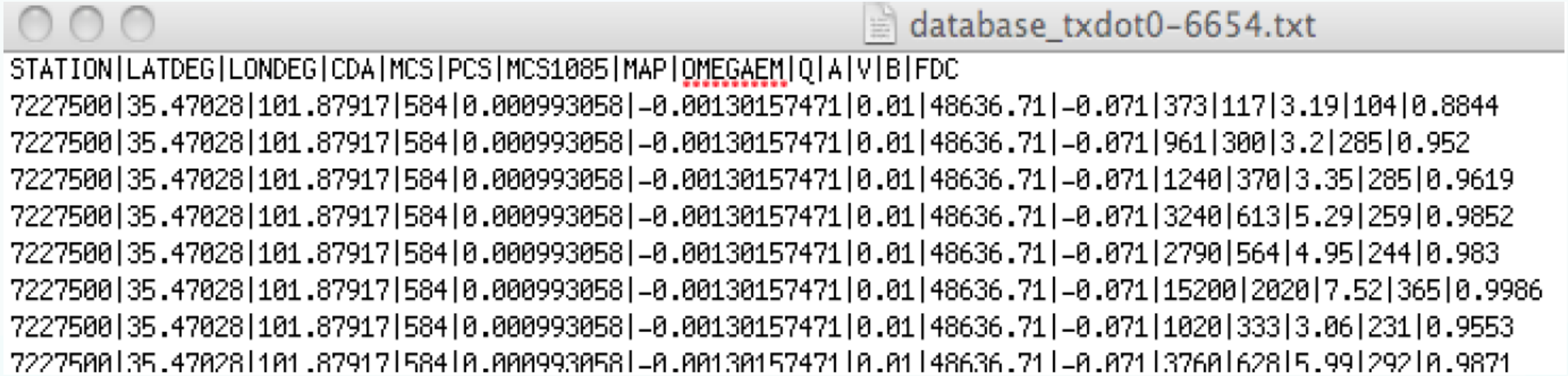


Fig. 1. (Color) USGS personnel conducting one of two high-magnitude discharge measurements on January 11, 2007, at USGS streamflow-gauging station 08156800 Shoal Creek at West 12th Street, Austin, Texas; both measurements are represented in the database used for this paper (Photograph by W. H. Asquith and courtesy of USGS)

Database Description

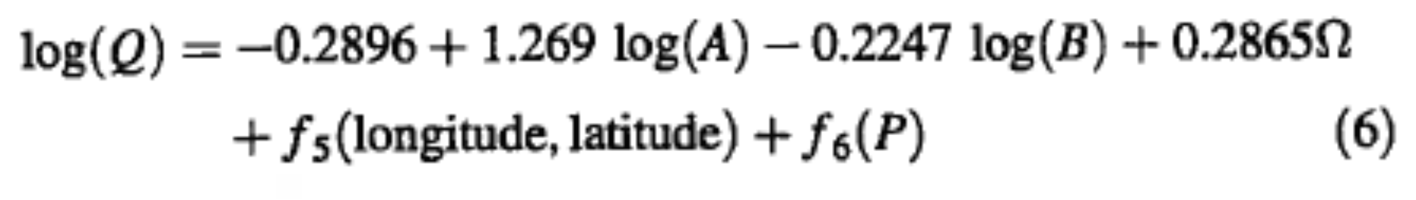


```
STATION|LATDEG|LONDEG|CDA|MCS|PCS|MCS1085|MAP|OMEGAEM|Q|A|V|B|FDC
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|373|117|3.19|104|0.8844
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|961|300|3.2|285|0.952
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|1240|370|3.35|285|0.9619
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|3240|613|5.29|259|0.9852
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|2790|564|4.95|244|0.983
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|15200|2020|7.52|365|0.9986
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|1020|333|3.06|231|0.9553
7227500|35.47028|101.87917|584|0.000993058|-0.00130157471|0.01|48636.71|-0.071|3760|628|5.99|292|0.9871
```

- 17,000 records
- 424 locations

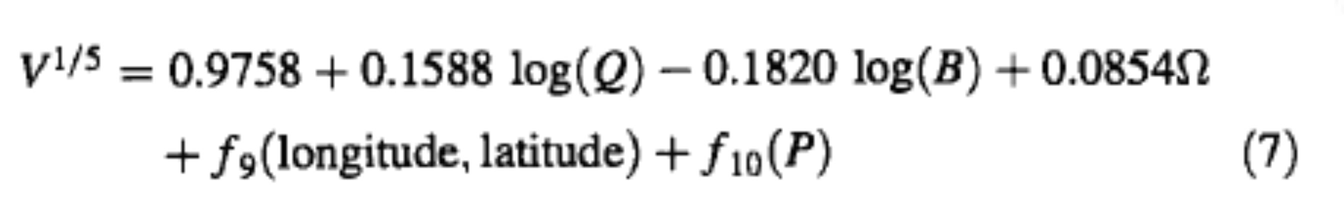
Regression Analysis

- Created a Generalized Additive Model for Q



$$\log(Q) = -0.2896 + 1.269\,\log(A) - 0.2247\,\log(B) + 0.2865\Omega + f_5(\text{longitude, latitude}) + f_6(P) \quad (6)$$

- Created a Generalized Additive Model for V

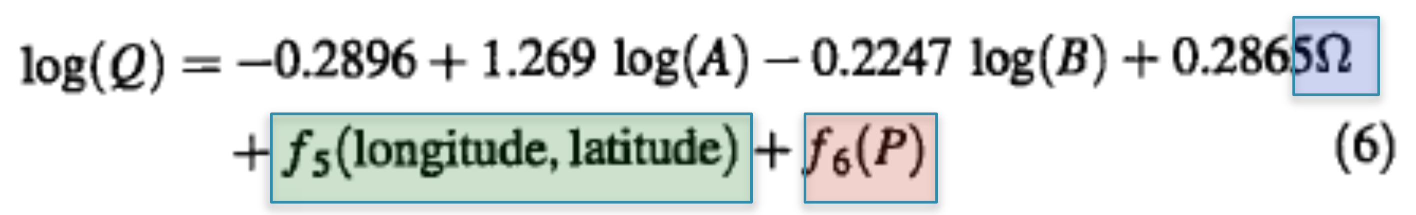


$$V^{1/5} = 0.9758 + 0.1588\,\log(Q) - 0.1820\,\log(B) + 0.0854\Omega + f_9(\text{longitude, latitude}) + f_{10}(P) \quad (7)$$

Asquith, W.H., Herrmann, G.R., and Cleveland, T.G., 2013, “Generalized Additive Regression Models of Discharge and Mean Velocity generally associated with Direct-Runoff Conditions in Texas: The Utility of the U.S. Geological Survey Discharge Measurement Database” American Society of Civil Engineers, Journal of Hydrologic Engineering, Vol. 18, No. 10, pp 1331-1348

Regression Analysis

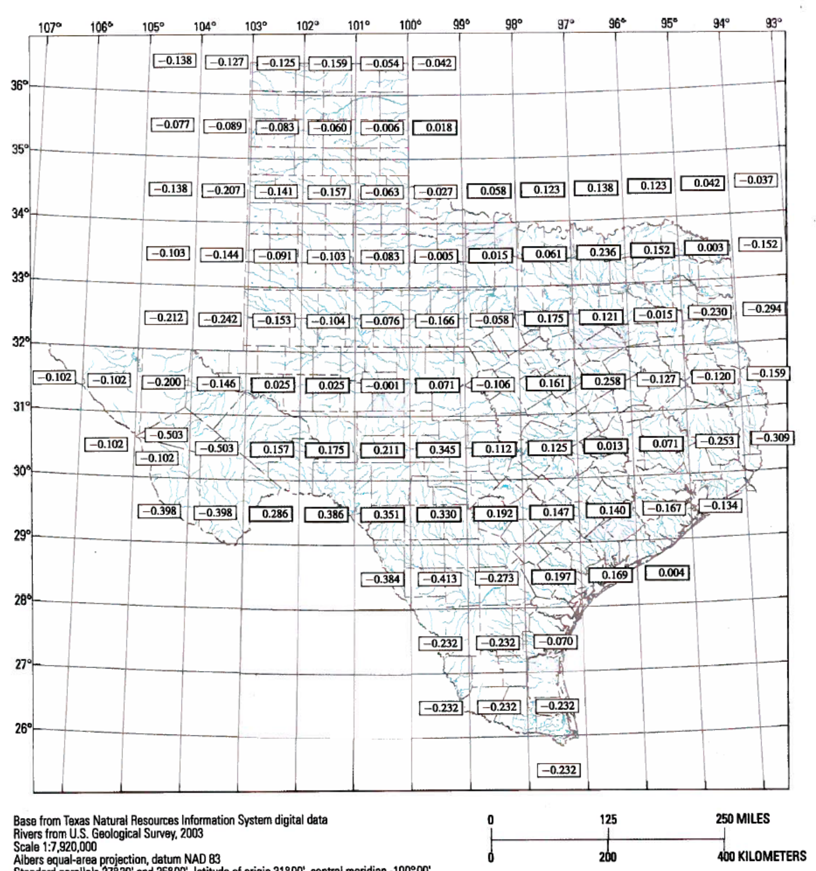
- Created a Generalized Additive Model for Q



$$\log(Q) = -0.2896 + 1.269\log(A) - 0.2247\log(B) + 0.2865\boxed{\Omega} + \boxed{f_5(\text{longitude, latitude})} + \boxed{f_6(P)} \tag{6}$$

Regression Analysis

- “Omega-EM” function



Base from Texas Natural Resources Information System digital data
Rivers from U.S. Geological Survey, 2003
Scale 1:7,920,000
Albers equal-area projection, datum NAD 83
Standard parallels 27°30' and 35°00', latitude of origin 31°00', central meridian -100°00'
Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Fig. 3. (Color) OmegaEM parameter of Asquith and Roussel (2009) to be used in generalized additive model (GAM) of discharge (QGAM) shown in Fig. 4 and Eq. (6) and generalized additive model of mean velocity (VGAM) shown in Fig. 8 and Eq. (7); the OmegaEM parameter represents a generalized terrain and climate index expressing relative differences in peak-streamflow potential across Texas [reproduced from Asquith and Roussel (2009)]

Regression Analysis

- Location adjustment function

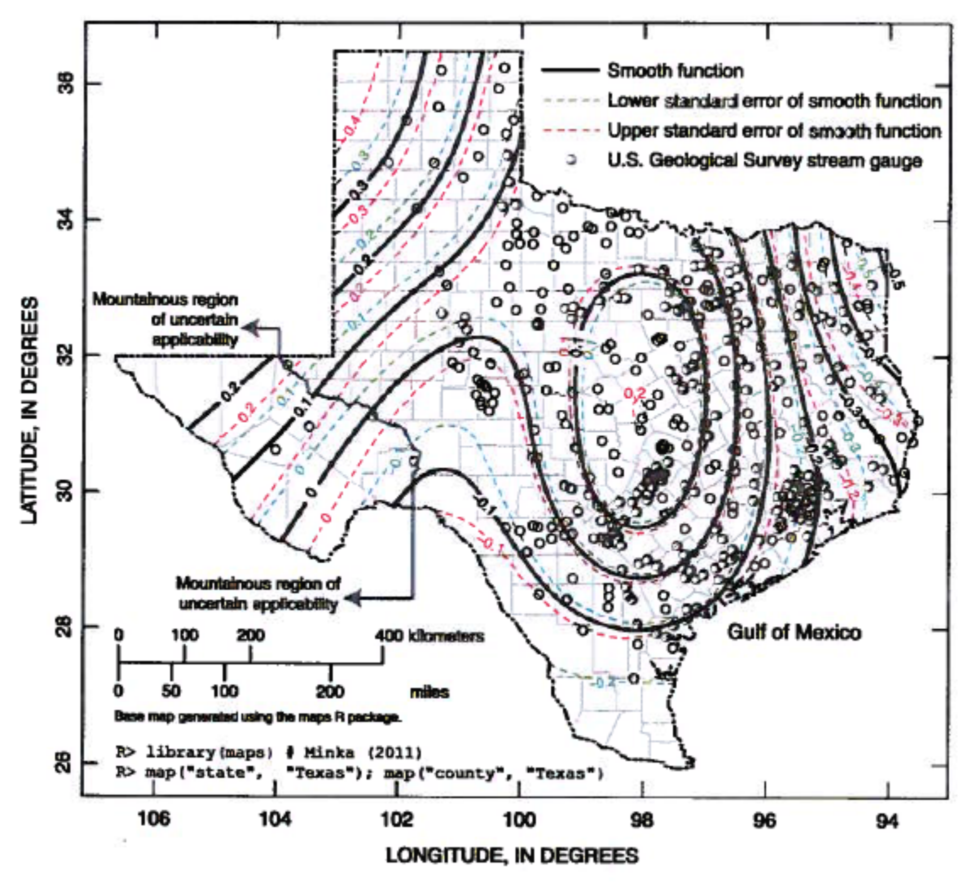


Fig. 5. (Color) Smooth function $f_5(l, k)$ of location in Texas for the discharge model shown in Fig. 4

Regression Analysis

- Precipitation adjustment function

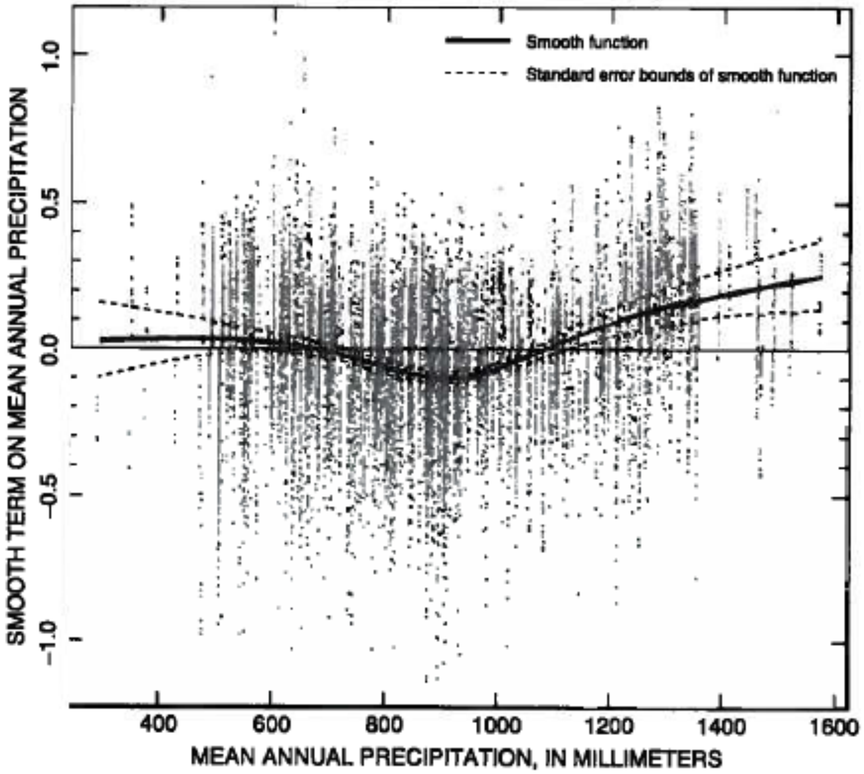


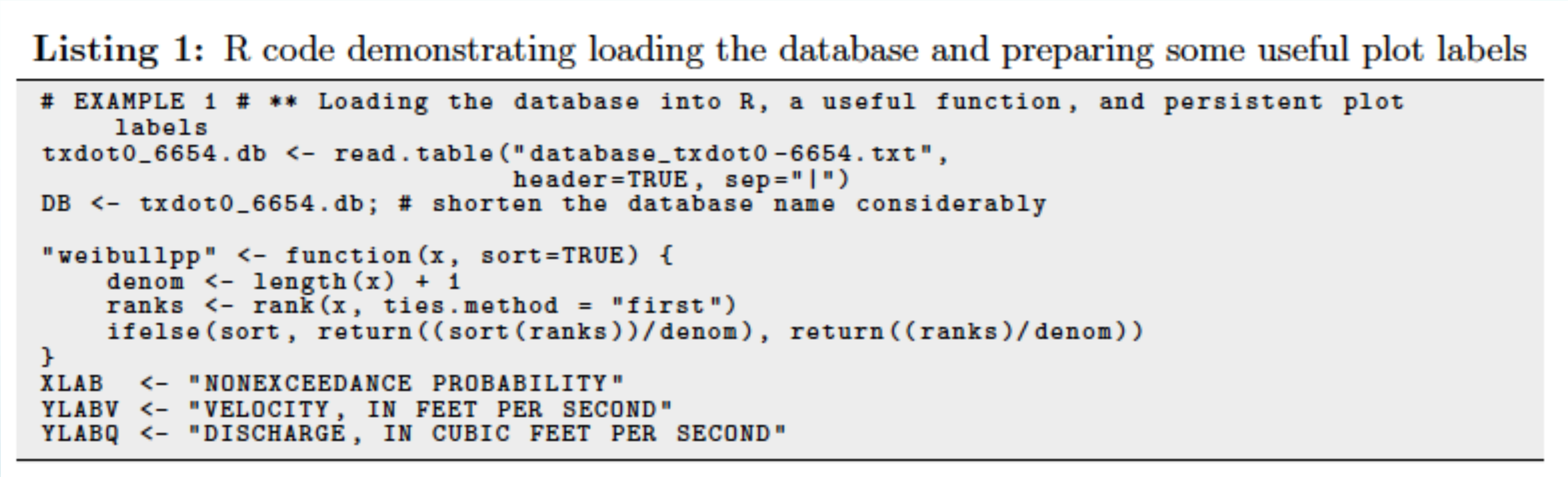
Fig. 6. Smooth function $f_6(P)$ of mean annual precipitation for the discharge model shown in Fig. 4

Idea

- Why not just directly access the database?
  - Use conditioning to restrict the range of explanatory variables (i.e. all values for topwidth between 10 and 50 feet)
  - From the conditioned results, the empirical distributions should be a decent estimator tool.

Script to Access the Database

- Database access is pretty straightforward – we used R, but could use MatLab, SAS, S or even Excel (17000 records is kind of a nuisance in Excel)



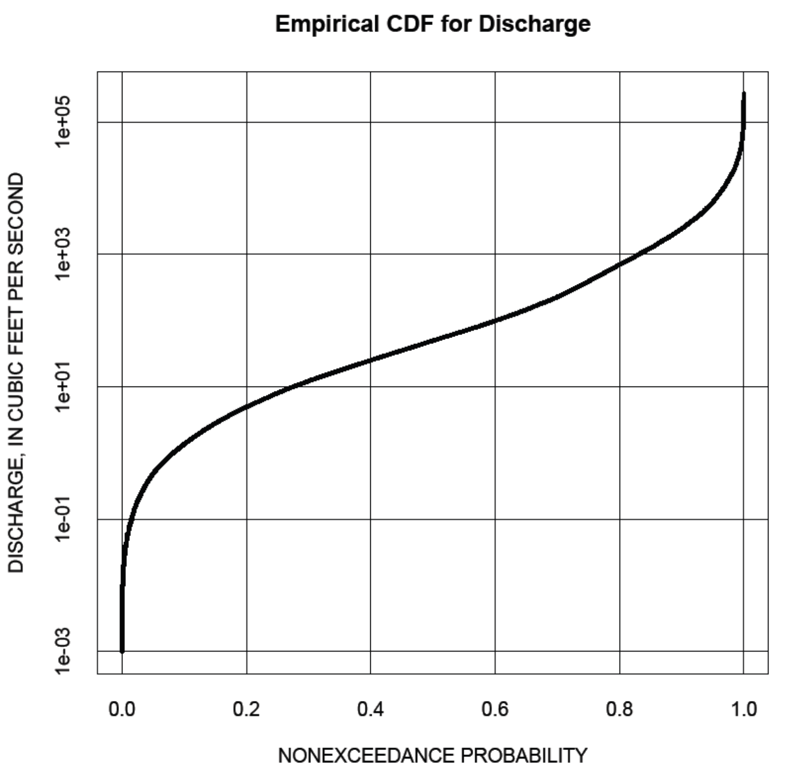
Listing 1: R code demonstrating loading the database and preparing some useful plot labels

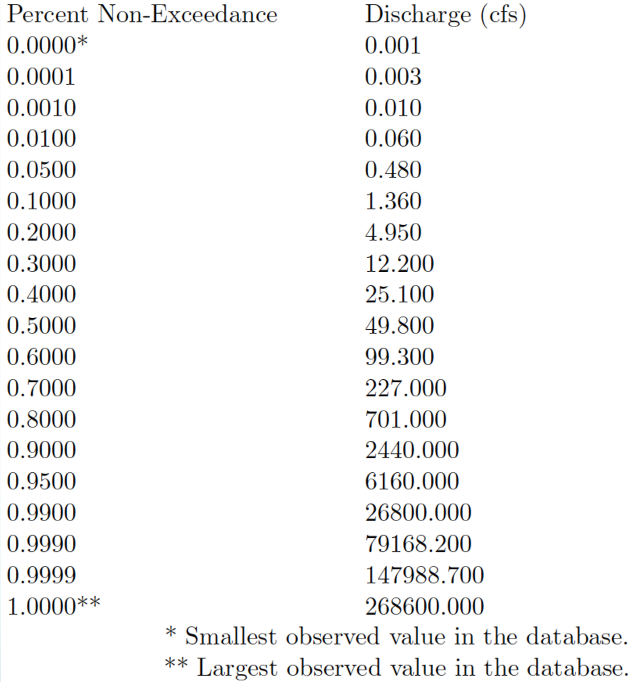
```
# EXAMPLE 1 # ** Loading the database into R, a useful function, and persistent plot
     labels
txdot0_6654.db <- read.table("database_txdot0-6654.txt",
                             header=TRUE, sep="|")
DB <- txdot0_6654.db; # shorten the database name considerably

"weibullpp" <- function(x, sort=TRUE) {
    denom <- length(x) + 1
    ranks <- rank(x, ties.method = "first")
    ifelse(sort, return((sort(ranks))/denom), return((ranks)/denom))
}
XLAB  <- "NONEXCEEDANCE PROBABILITY"
YLABV <- "VELOCITY, IN FEET PER SECOND"
YLABQ <- "DISCHARGE, IN CUBIC FEET PER SECOND"
```

Empirical Distributions

- Cumulative distribution – all values


Empirical CDF for Discharge



Percent Non-Exceedance	Discharge (cfs)
0.0000*	0.001
0.0001	0.003
0.0010	0.010
0.0100	0.060
0.0500	0.480
0.1000	1.360
0.2000	4.950
0.3000	12.200
0.4000	25.100
0.5000	49.800
0.6000	99.300
0.7000	227.000
0.8000	701.000
0.9000	2440.000
0.9500	6160.000
0.9900	26800.000
0.9990	79168.200
0.9999	147988.700
1.0000**	268600.000

* Smallest observed value in the database.

** Largest observed value in the database.

Empirical Distributions

- Cumulative distribution – all values

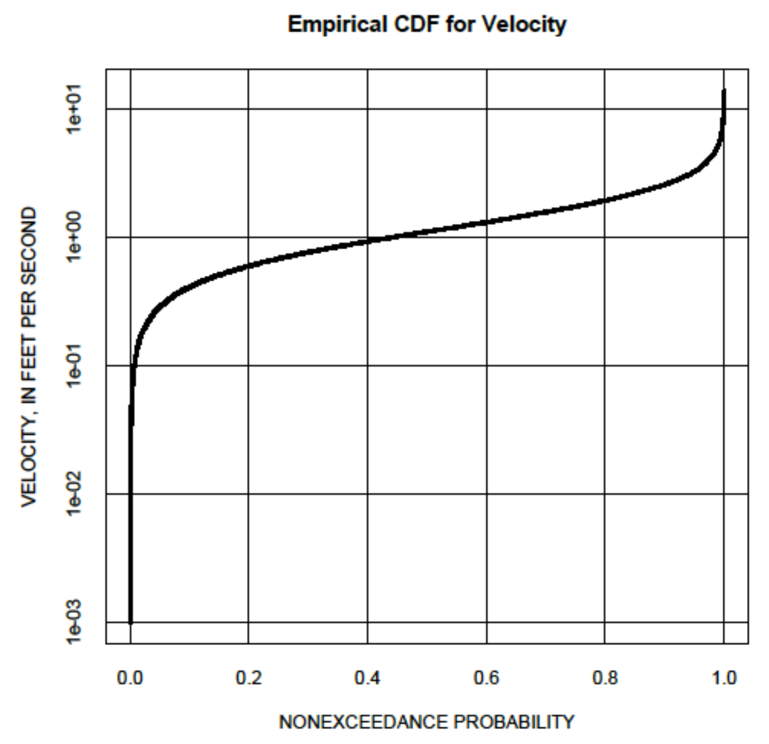

Empirical CDF for Velocity

Figure 34: Empirical cumulative distribution for velocity in Texas

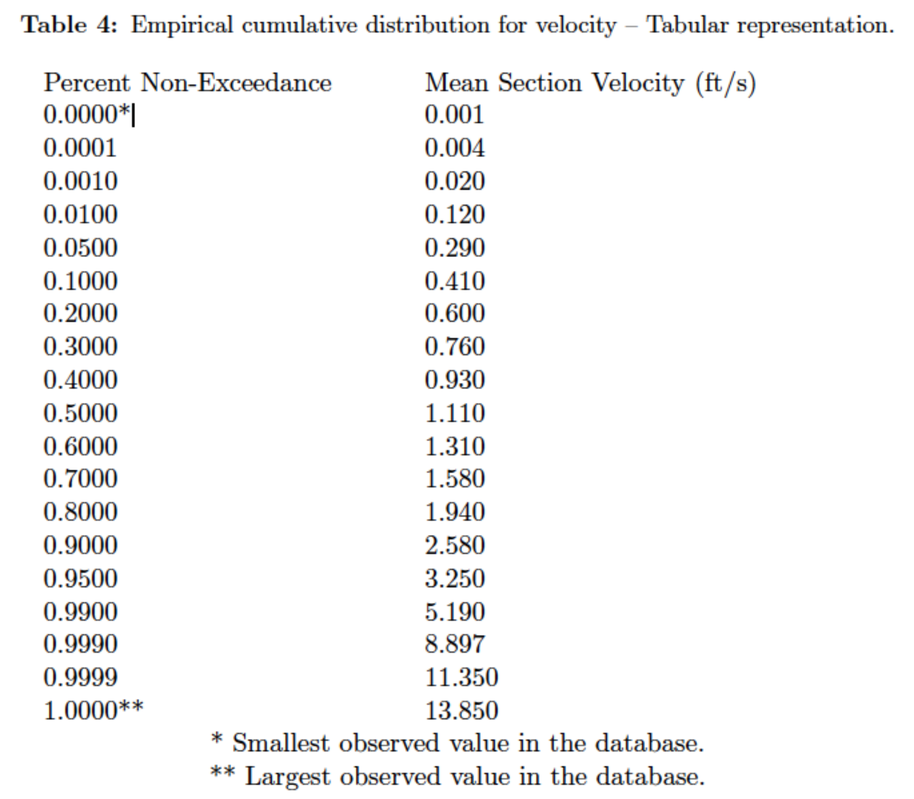


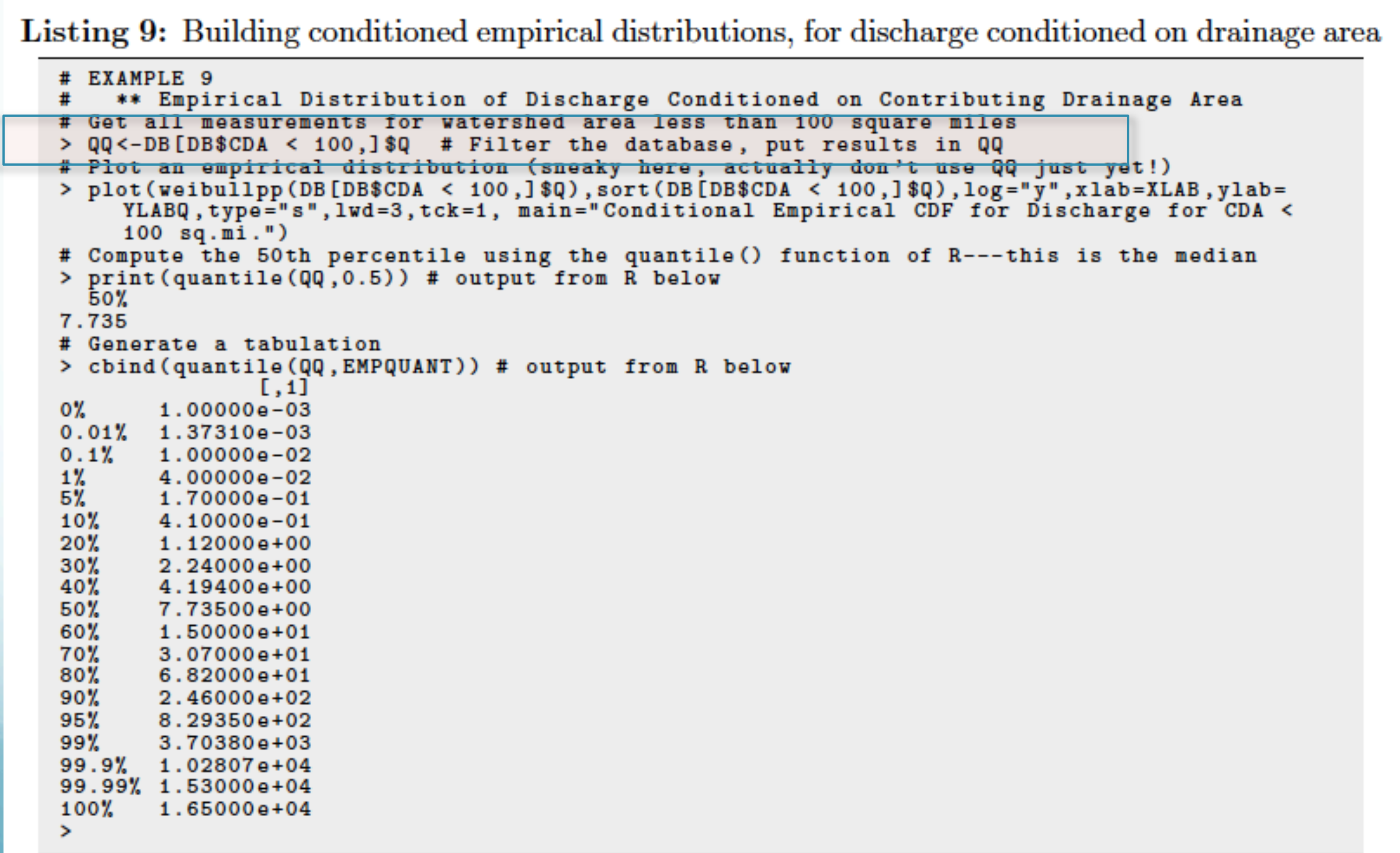
Table 4: Empirical cumulative distribution for velocity – Tabular representation.

Percent Non-Exceedance	Mean Section Velocity (ft/s)
0.0000*	0.001
0.0001	0.004
0.0010	0.020
0.0100	0.120
0.0500	0.290
0.1000	0.410
0.2000	0.600
0.3000	0.760
0.4000	0.930
0.5000	1.110
0.6000	1.310
0.7000	1.580
0.8000	1.940
0.9000	2.580
0.9500	3.250
0.9900	5.190
0.9990	8.897
0.9999	11.350
1.0000**	13.850

* Smallest observed value in the database.
** Largest observed value in the database.

Conditional Distributions

- Condition to use what we know about a location

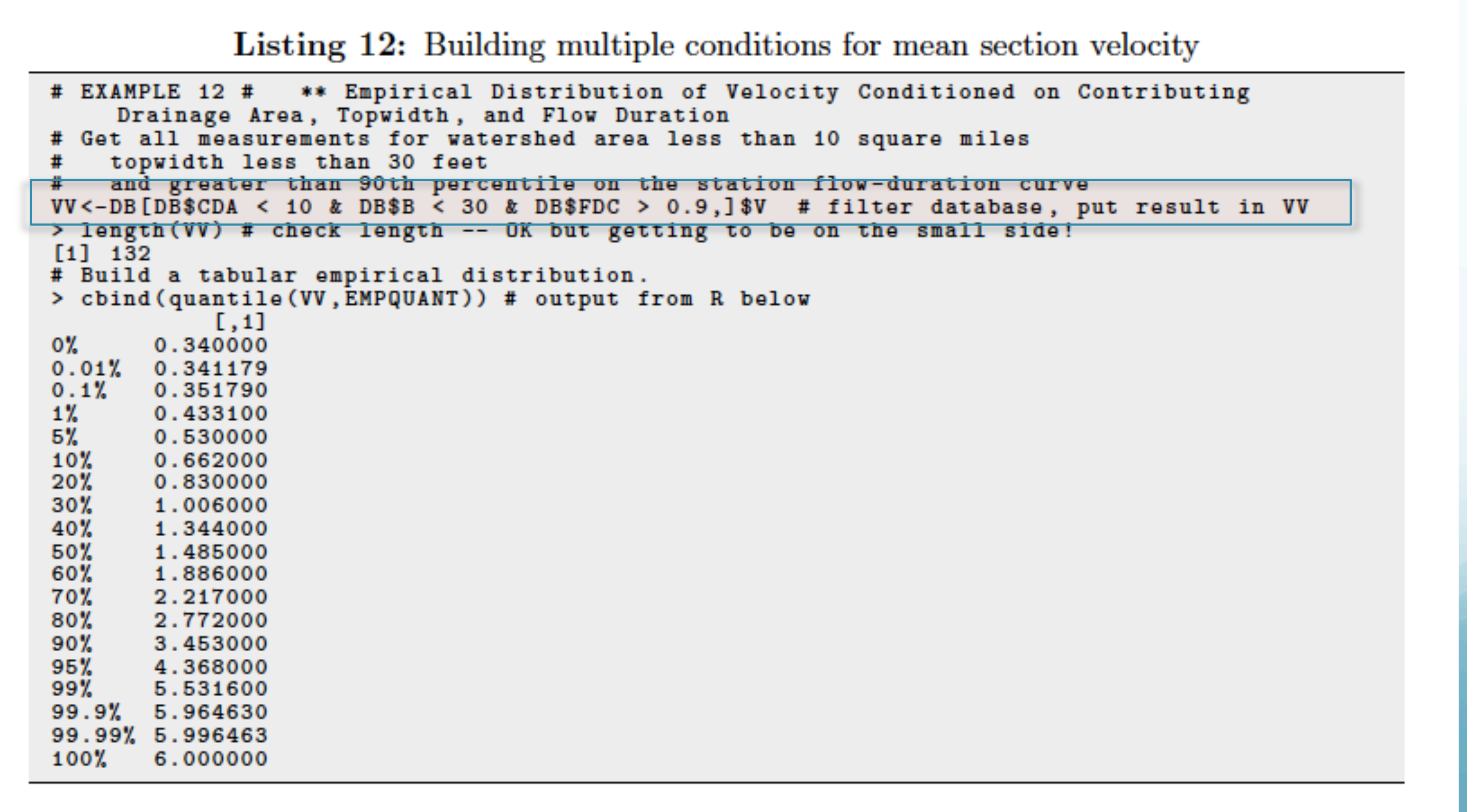


Listing 9: Building conditioned empirical distributions, for discharge conditioned on drainage area

```
# EXAMPLE 9
#   ** Empirical Distribution of Discharge Conditioned on Contributing Drainage Area
# Get all measurements for watershed area less than 100 square miles
> QQ<-DB[DB$CDA < 100,]$Q  # Filter the database, put results in QQ
# Plot an empirical distribution (sneaky here, actually don’t use QQ just yet!)
> plot(weibullpp(DB[DB$CDA < 100,]$Q),sort(DB[DB$CDA < 100,]$Q),log="y",xlab=XLAB,ylab=
    YLABQ,type="s",lwd=3,tck=1, main="Conditional Empirical CDF for Discharge for CDA <
    100 sq.mi.")
# Compute the 50th percentile using the quantile() function of R---this is the median
> print(quantile(QQ,0.5)) # output from R below
  50%
7.735
# Generate a tabulation
> cbind(quantile(QQ,EMPQUANT)) # output from R below
              [,1]
0%     1.00000e-03
0.01%  1.37310e-03
0.1%   1.00000e-02
1%     4.00000e-02
5%     1.70000e-01
10%    4.10000e-01
20%    1.12000e+00
30%    2.24000e+00
40%    4.19400e+00
50%    7.73500e+00
60%    1.50000e+01
70%    3.07000e+01
80%    6.82000e+01
90%    2.46000e+02
95%    8.29350e+02
99%    3.70380e+03
99.9%  1.02807e+04
99.99% 1.53000e+04
100%   1.65000e+04
>
```

Multiple Conditioning

- Condition to use what we know about a location



Listing 12: Building multiple conditions for mean section velocity

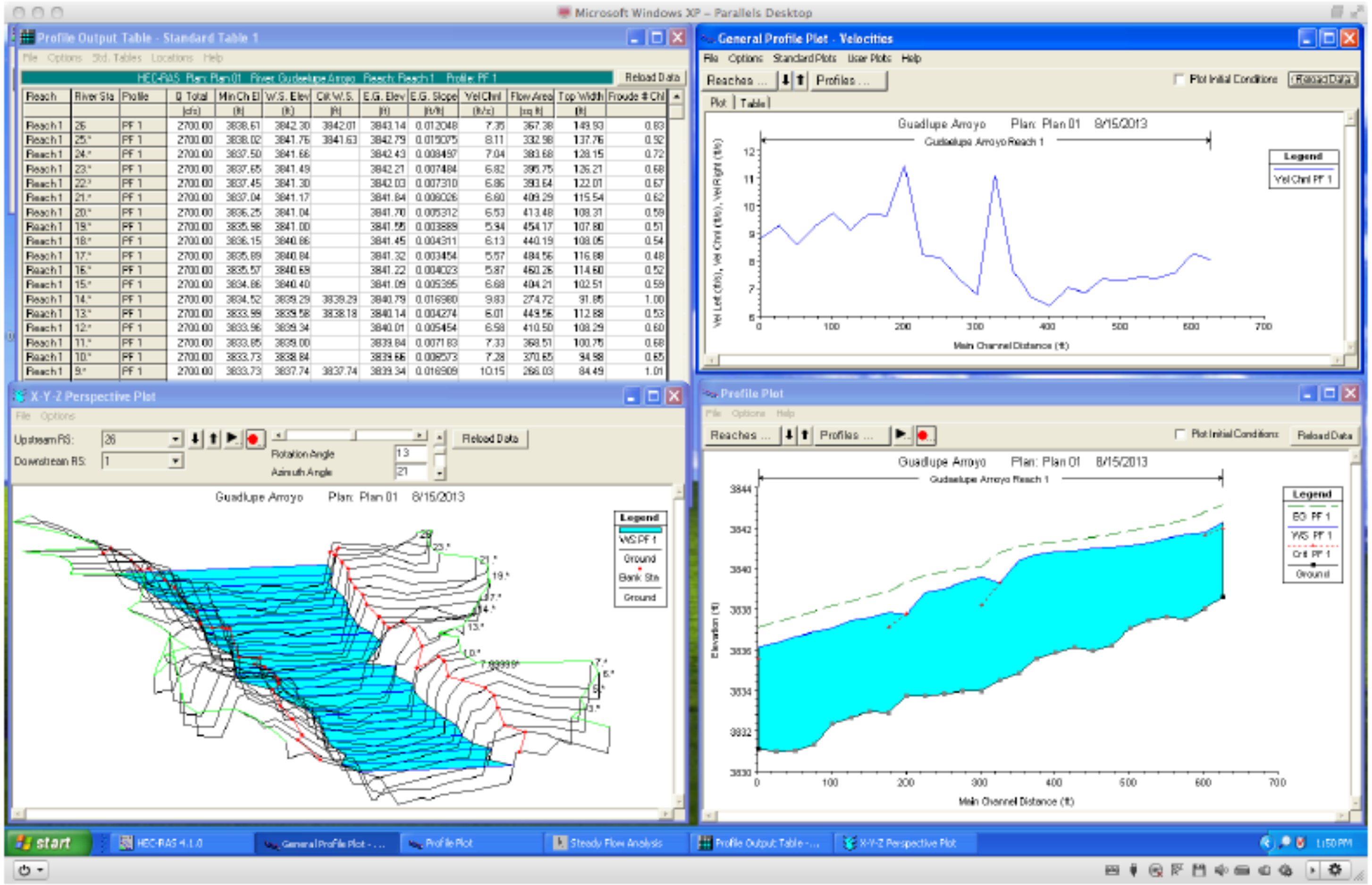
```
# EXAMPLE 12 #   ** Empirical Distribution of Velocity Conditioned on Contributing
     Drainage Area, Topwidth, and Flow Duration
# Get all measurements for watershed area less than 10 square miles
#   topwidth less than 30 feet
#   and greater than 90th percentile on the station flow-duration curve
VV<-DB[DB$CDA < 10 & DB$B < 30 & DB$FDC > 0.9,]$V  # filter database, put result in VV
> length(VV) # check length -- OK but getting to be on the small side!
[1] 132
# Build a tabular empirical distribution.
> cbind(quantile(VV,EMPQUANT)) # output from R below
           [,1]
0%       0.340000
0.01%    0.341179
0.1%     0.351790
1%       0.433100
5%       0.530000
10%      0.662000
20%      0.830000
30%      1.006000
40%      1.344000
50%      1.485000
60%      1.886000
70%      2.217000
80%      2.772000
90%      3.453000
95%      4.368000
99%      5.531600
99.9%    5.964630
99.99%   5.996463
100%     6.000000
```

Application for Guadelupe Arroyo Example

- HEC-RAS Model (reconstructed)

Figure 53: Google Earth aerial photograph of US 62/180 bridge over Guadalupe Arroyo, Culberson County, Texas

HEC-RAS Model



HEC-RAS Plan: Plan 01 River: Gudaelupe Arroyo Reach: Reach 1 Profile: PF 1

Reach	River Sta	Profile	Q Total (cfs)	Min Ch El (ft)	W.S. Elev (ft)	Crit W.S. (ft)	E.G. Elev (ft)	E.G. Slope (ft/ft)	Vel Chnl (ft/s)	Flow Area (sq ft)	Top Width (ft)	Froude # Chl
Reach 1	26	PF 1	2700.00	3838.61	3842.30	3842.01	3843.14	0.012048	7.35	367.38	149.93	0.83
Reach 1	25.*	PF 1	2700.00	3838.02	3841.76	3841.63	3842.79	0.015075	8.11	332.98	137.76	0.92
Reach 1	24.*	PF 1	2700.00	3837.50	3841.66		3842.43	0.008497	7.04	383.68	128.15	0.72
Reach 1	23.*	PF 1	2700.00	3837.65	3841.49		3842.21	0.007484	6.82	395.75	126.21	0.68
Reach 1	22.*	PF 1	2700.00	3837.45	3841.30		3842.03	0.007310	6.86	393.64	122.01	0.67
Reach 1	21.*	PF 1	2700.00	3837.04	3841.17		3841.84	0.006026	6.60	409.29	115.54	0.62
Reach 1	20.*	PF 1	2700.00	3836.25	3841.04		3841.70	0.005312	6.53	413.48	108.31	0.59
Reach 1	19.*	PF 1	2700.00	3835.98	3841.00		3841.55	0.003889	5.94	454.17	107.80	0.51
Reach 1	18.*	PF 1	2700.00	3836.15	3840.86		3841.45	0.004311	6.13	440.19	108.05	0.54
Reach 1	17.*	PF 1	2700.00	3835.89	3840.84		3841.32	0.003454	5.57	484.56	116.88	0.48
Reach 1	16.*	PF 1	2700.00	3835.57	3840.69		3841.22	0.004023	5.87	460.26	114.60	0.52
Reach 1	15.*	PF 1	2700.00	3834.86	3840.40		3841.09	0.005395	6.68	404.21	102.51	0.59
Reach 1	14.*	PF 1	2700.00	3834.52	3839.29	3839.29	3840.79	0.016980	9.83	274.72	91.85	1.00
Reach 1	13.*	PF 1	2700.00	3833.99	3839.58	3838.18	3840.14	0.004274	6.01	449.56	112.88	0.53
Reach 1	12.*	PF 1	2700.00	3833.96	3839.34		3840.01	0.005454	6.58	410.50	108.29	0.60
Reach 1	11.*	PF 1	2700.00	3833.85	3839.00		3839.84	0.007183	7.33	368.51	100.76	0.68
Reach 1	10.*	PF 1	2700.00	3833.73	3838.84		3839.66	0.008573	7.28	370.65	94.98	0.65
Reach 1	9.*	PF 1	2700.00	3833.73	3837.74	3837.74	3839.34	0.016909	10.15	266.03	84.49	1.01

Figure 54: HEC-RAS results from example conditions

HEC-RAS Model

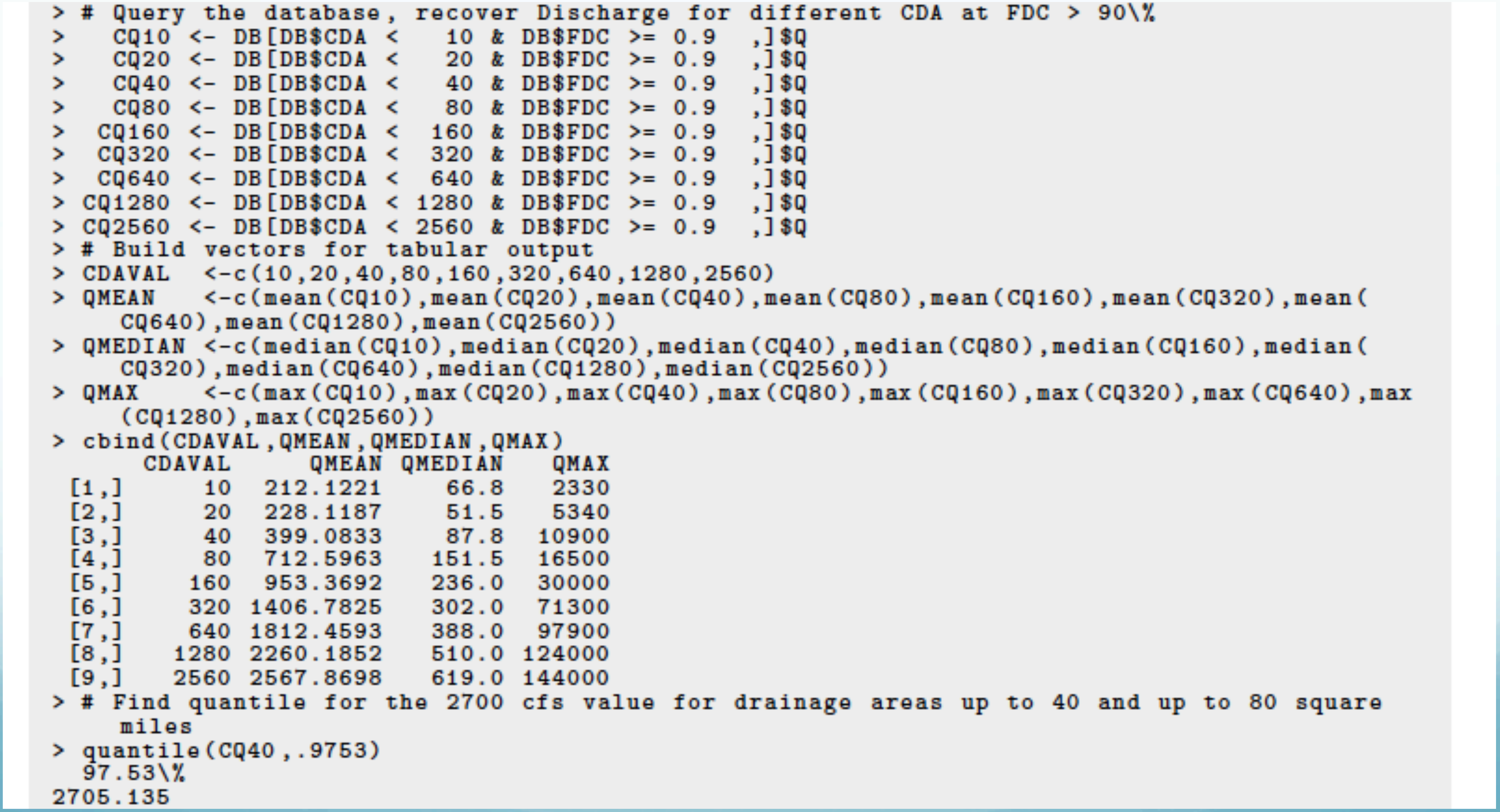
- Two excursions into supercritical flow at sections 14 and 9.
- Velocity being in excess of 6 ft/second everywhere
- Two velocity spikes of about 11 feet per second.

HEC-RAS Model

- For drainage areas up to the 40 square mile range, with individual low duration curve values greater than the 90-th percentile:

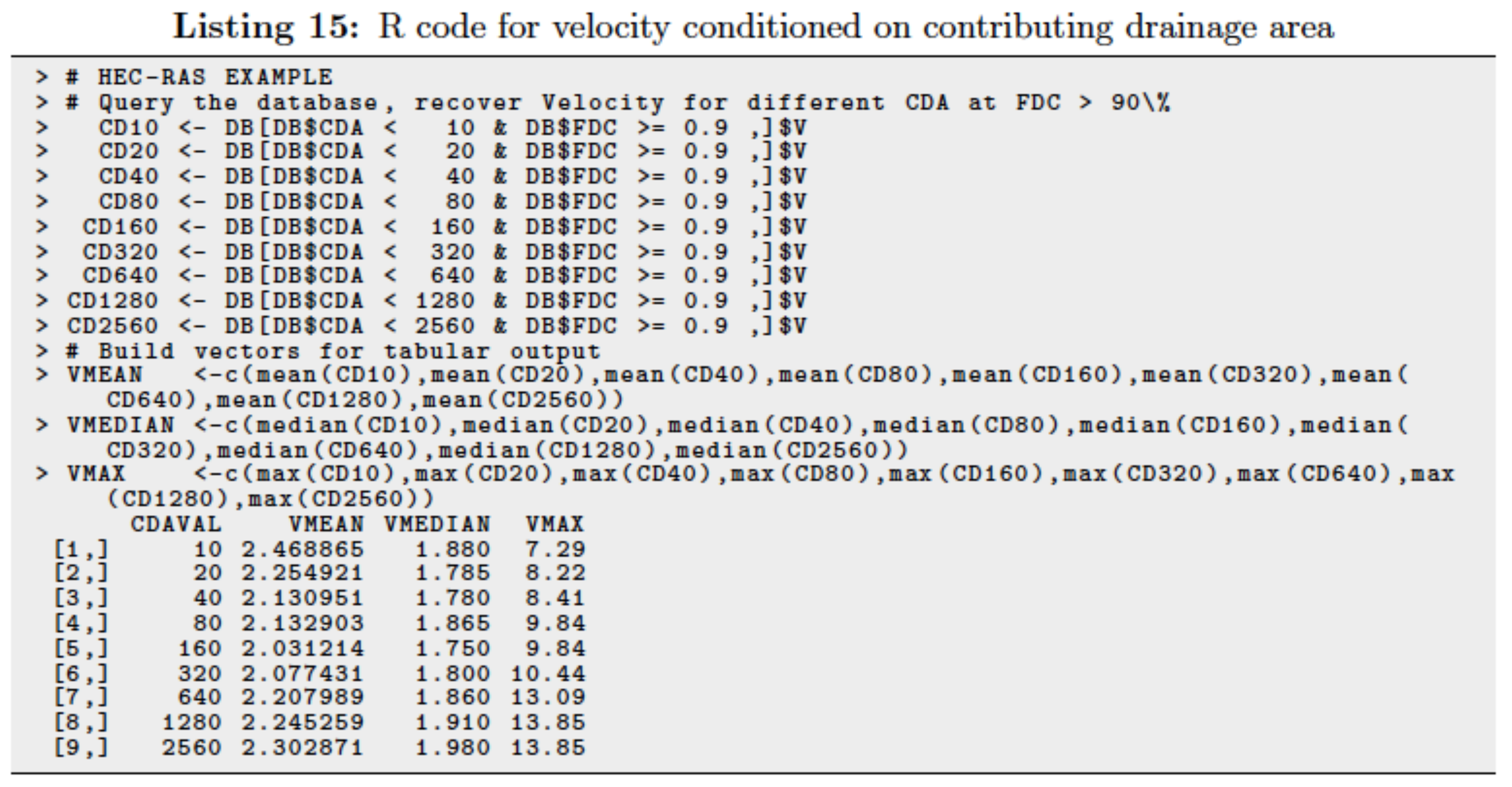
1) Is 2700 cfs an accessible discharge?
2) What is the anticipated water velocity?
3) What is a typical topwidth?

HEC-RAS Model



```
> # Query the database, recover Discharge for different CDA at FDC > 90\%
>    CQ10 <- DB[DB$CDA <    10 & DB$FDC >= 0.9  ,]$Q
>    CQ20 <- DB[DB$CDA <    20 & DB$FDC >= 0.9  ,]$Q
>    CQ40 <- DB[DB$CDA <    40 & DB$FDC >= 0.9  ,]$Q
>    CQ80 <- DB[DB$CDA <    80 & DB$FDC >= 0.9  ,]$Q
>   CQ160 <- DB[DB$CDA <   160 & DB$FDC >= 0.9  ,]$Q
>   CQ320 <- DB[DB$CDA <   320 & DB$FDC >= 0.9  ,]$Q
>   CQ640 <- DB[DB$CDA <   640 & DB$FDC >= 0.9  ,]$Q
>  CQ1280 <- DB[DB$CDA <  1280 & DB$FDC >= 0.9  ,]$Q
>  CQ2560 <- DB[DB$CDA <  2560 & DB$FDC >= 0.9  ,]$Q
> # Build vectors for tabular output
> CDAVAL  <-c(10,20,40,80,160,320,640,1280,2560)
> QMEAN   <-c(mean(CQ10),mean(CQ20),mean(CQ40),mean(CQ80),mean(CQ160),mean(CQ320),mean(
    CQ640),mean(CQ1280),mean(CQ2560))
> QMEDIAN <-c(median(CQ10),median(CQ20),median(CQ40),median(CQ80),median(CQ160),median(
    CQ320),median(CQ640),median(CQ1280),median(CQ2560))
> QMAX    <-c(max(CQ10),max(CQ20),max(CQ40),max(CQ80),max(CQ160),max(CQ320),max(CQ640),max
    (CQ1280),max(CQ2560))
> cbind(CDAVAL,QMEAN,QMEDIAN,QMAX)
      CDAVAL     QMEAN QMEDIAN   QMAX
 [1,]     10  212.1221    66.8   2330
 [2,]     20  228.1187    51.5   5340
 [3,]     40  399.0833    87.8  10900
 [4,]     80  712.5963   151.5  16500
 [5,]    160  953.3692   236.0  30000
 [6,]    320 1406.7825   302.0  71300
 [7,]    640 1812.4593   388.0  97900
 [8,]   1280 2260.1852   510.0 124000
 [9,]   2560 2567.8698   619.0 144000
> # Find quantile for the 2700 cfs value for drainage areas up to 40 and up to 80 square
    miles
> quantile(CQ40,.9753)
  97.53\%
2705.135
```

HEC-RAS Model



Listing 15: R code for velocity conditioned on contributing drainage area

```
> # HEC-RAS EXAMPLE
> # Query the database, recover Velocity for different CDA at FDC > 90\%
>   CD10 <- DB[DB$CDA <   10 & DB$FDC >= 0.9 ,]$V
>   CD20 <- DB[DB$CDA <   20 & DB$FDC >= 0.9 ,]$V
>   CD40 <- DB[DB$CDA <   40 & DB$FDC >= 0.9 ,]$V
>   CD80 <- DB[DB$CDA <   80 & DB$FDC >= 0.9 ,]$V
>  CD160 <- DB[DB$CDA <  160 & DB$FDC >= 0.9 ,]$V
>  CD320 <- DB[DB$CDA <  320 & DB$FDC >= 0.9 ,]$V
>  CD640 <- DB[DB$CDA <  640 & DB$FDC >= 0.9 ,]$V
> CD1280 <- DB[DB$CDA < 1280 & DB$FDC >= 0.9 ,]$V
> CD2560 <- DB[DB$CDA < 2560 & DB$FDC >= 0.9 ,]$V
> # Build vectors for tabular output
> VMEAN   <-c(mean(CD10),mean(CD20),mean(CD40),mean(CD80),mean(CD160),mean(CD320),mean(
     CD640),mean(CD1280),mean(CD2560))
> VMEDIAN <-c(median(CD10),median(CD20),median(CD40),median(CD80),median(CD160),median(
     CD320),median(CD640),median(CD1280),median(CD2560))
> VMAX    <-c(max(CD10),max(CD20),max(CD40),max(CD80),max(CD160),max(CD320),max(CD640),max
     (CD1280),max(CD2560))
      CDAVAL    VMEAN VMEDIAN  VMAX
 [1,]     10 2.468865   1.880  7.29
 [2,]     20 2.254921   1.785  8.22
 [3,]     40 2.130951   1.780  8.41
 [4,]     80 2.132903   1.865  9.84
 [5,]    160 2.031214   1.750  9.84
 [6,]    320 2.077431   1.800 10.44
 [7,]    640 2.207989   1.860 13.09
 [8,]   1280 2.245259   1.910 13.85
 [9,]   2560 2.302871   1.980 13.85
```

HEC-RAS Model

- 1. Is 2700 cfs an accessible discharge?
  - Yes, the value has been observed for drainage areas about the same as the area at the point of interest, however the value is comparatively rare and not anticipated for a “median" event.
- 2. What is the anticipated water velocity?
  - Somewhere in the 2-8 feet per second range. The HEC-RAS values are high, suggesting either too much discharge (hydrology) or model boundary conditions are inappropriate.
- 3. What is a typical topwidth?
  - For the location of interest, based on contributing drainage area, 58 feet is the anticipated value. The HEC-RAS values are larger, but not by too much.

Conclusions

- An approach for assessing modeling validity from observational experience is presented.
- Empirical distributions are an alternative to regression-like approach -- the authors think these are simpler to actually use
- Prediction limits are analogous to empirical distributions presented herein.

Going Forward

- The R script is reasonably simple, but for practical use an interface is needed
  - Perhaps RExcel (an Excel add-in that allows direct access to R scripts)
  - Need to build a compendium of examples to illustrate various uses rather than a single case study in an admittedly remote part of Texas
- Attractiveness is that because the conditioning is really a database look-up, as new data is available can simply add it to the database, whereas the Regression Models are unable to incorporate new data (without reanalysis)