Deltares
Enabling Delta Life

Estimating predictive hydrological uncertainty using Quantile Regression

Jan Verkade
Albrecht Weerts
Steven Weijs

June 2011

This presentation is also available on slideshare.net

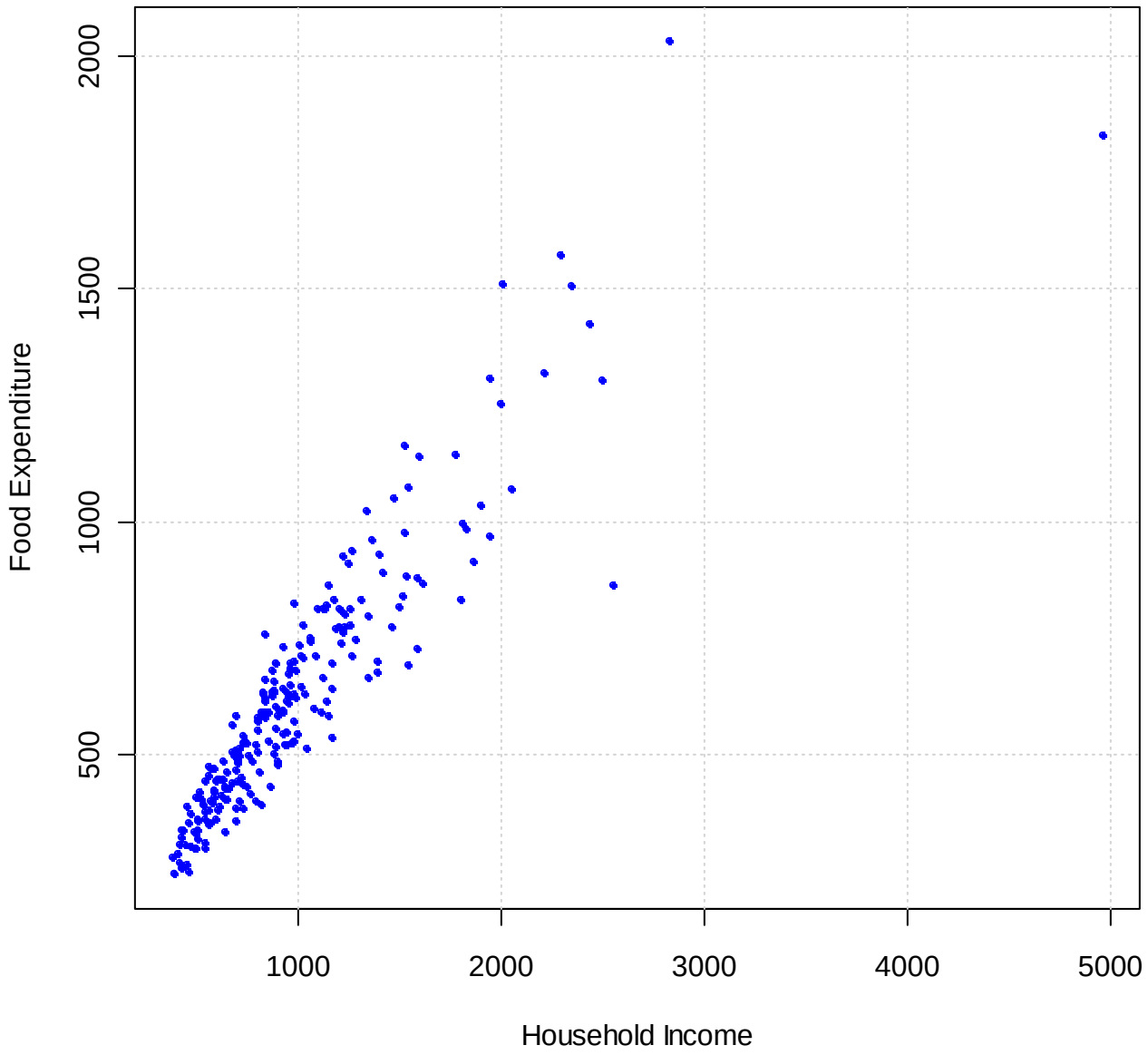
Visit twitter.com/janverkade for the address

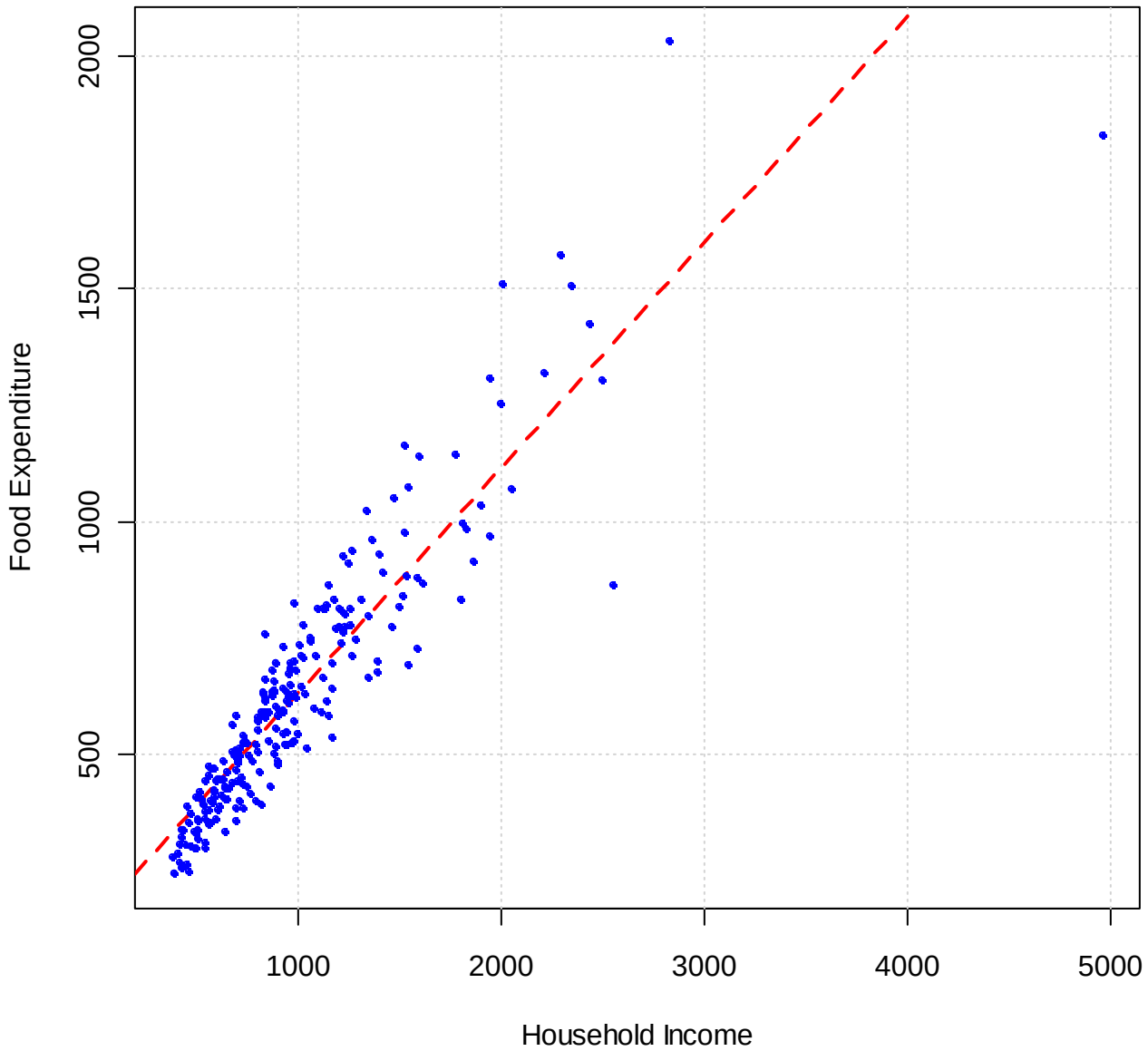

Quantile Regression

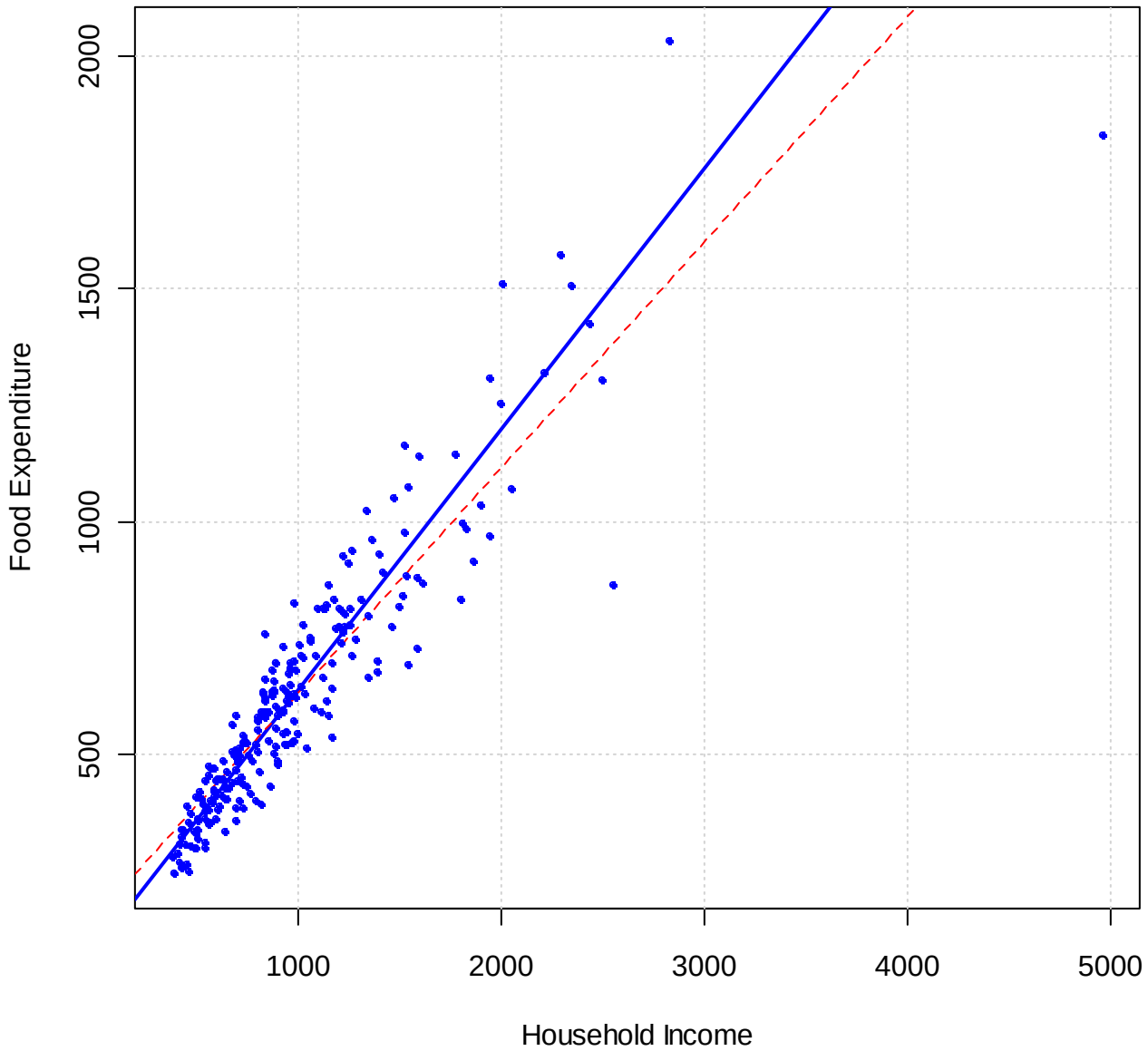
- Origin in econometrics (Koenker et al.)
- Brought to our attention by Andy Woods @ 2008 HEPEX meeting
- First research implementations:
  - ~25 NFFS basins (England and Wales)
  - White Cart (Scotland)
  - Ovens River (Australia)
  - HEPEX-basins (USA) (in progress)
- Now operationally used in National Flood Forecasting System (England and Wales)

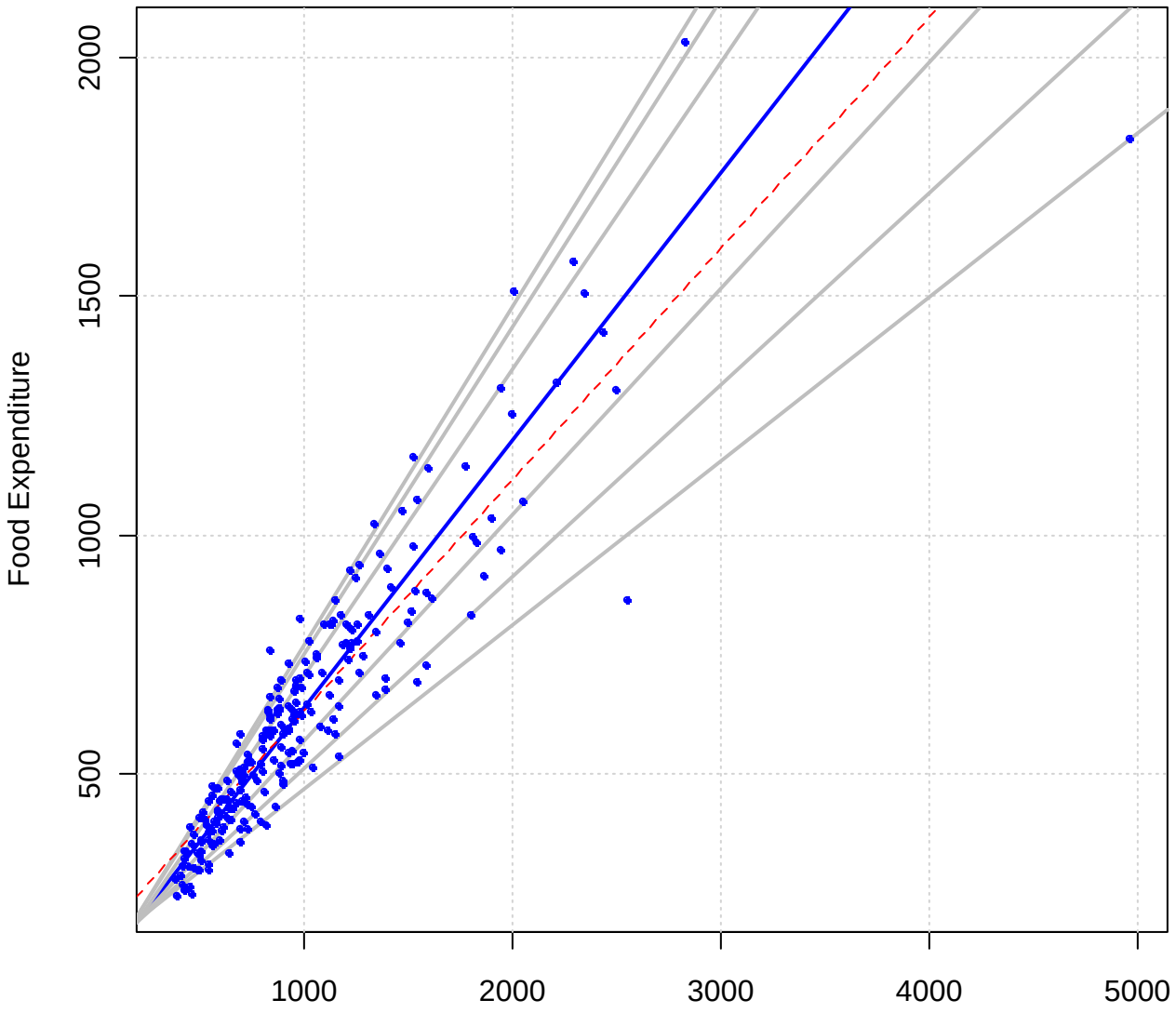

Quantile Regression: principles

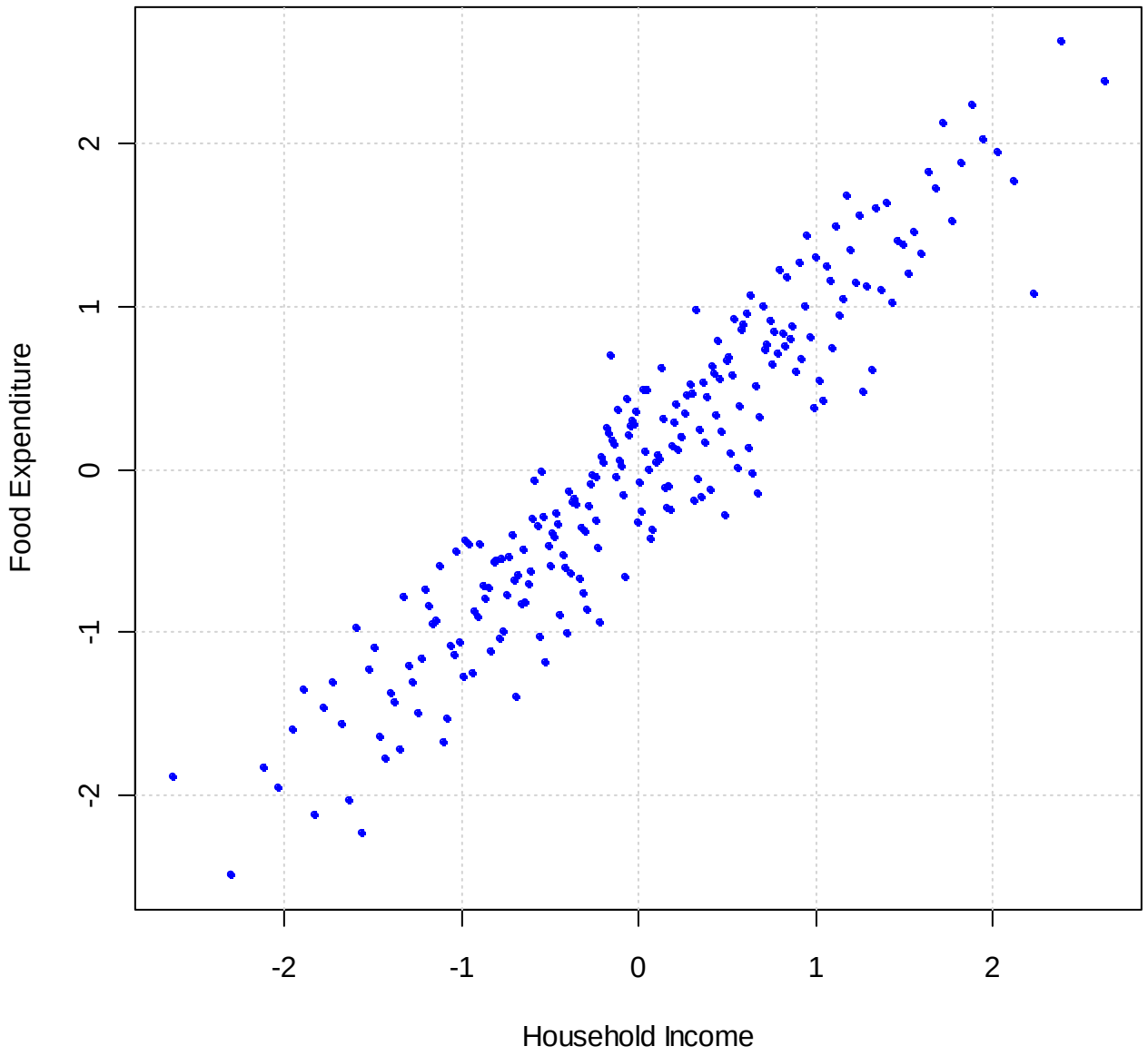
- QR is a method for describing conditional quantiles
- Rather than minimising the mean *squared* error (MSE)
- QR is based on minimising the mean *absolute* error (MAE)
- This yields not the sample *mean* but the sample *median*
- Other quantiles may be derived by adding weights to errors
- E.g. weight = .1 for positive errors and .9 for negative errors
- Fitting models may be done in transformed space to account for heteroscedasticity

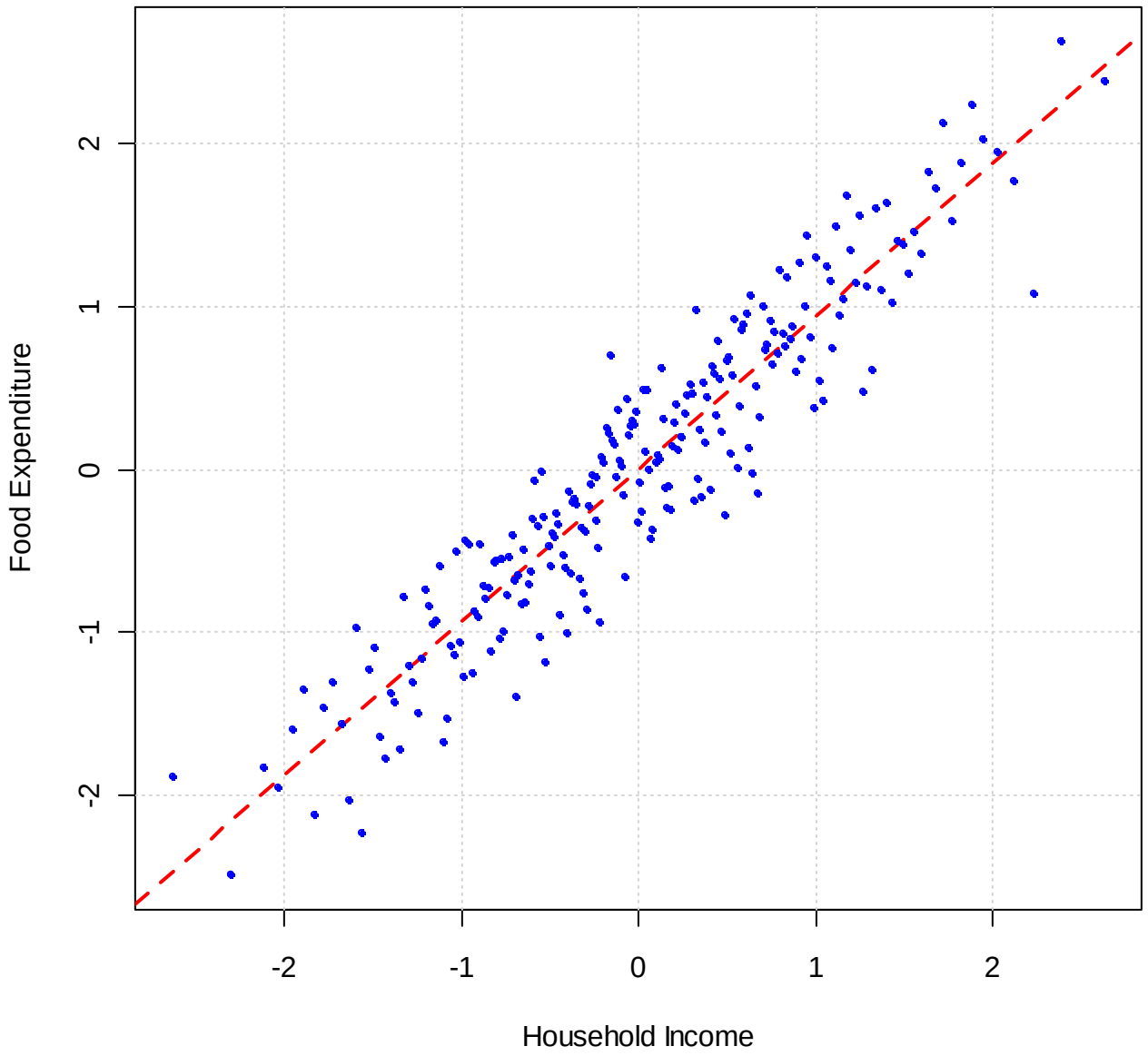

Engel data, 19th century, Belgium
Food Expenditure
500
1000
1500
2000
1000
2000
3000
4000
5000
Household Income


Least Squares Estimate
Food Expenditure
500
1000
1500
2000
1000
2000
3000
4000
5000
Household Income

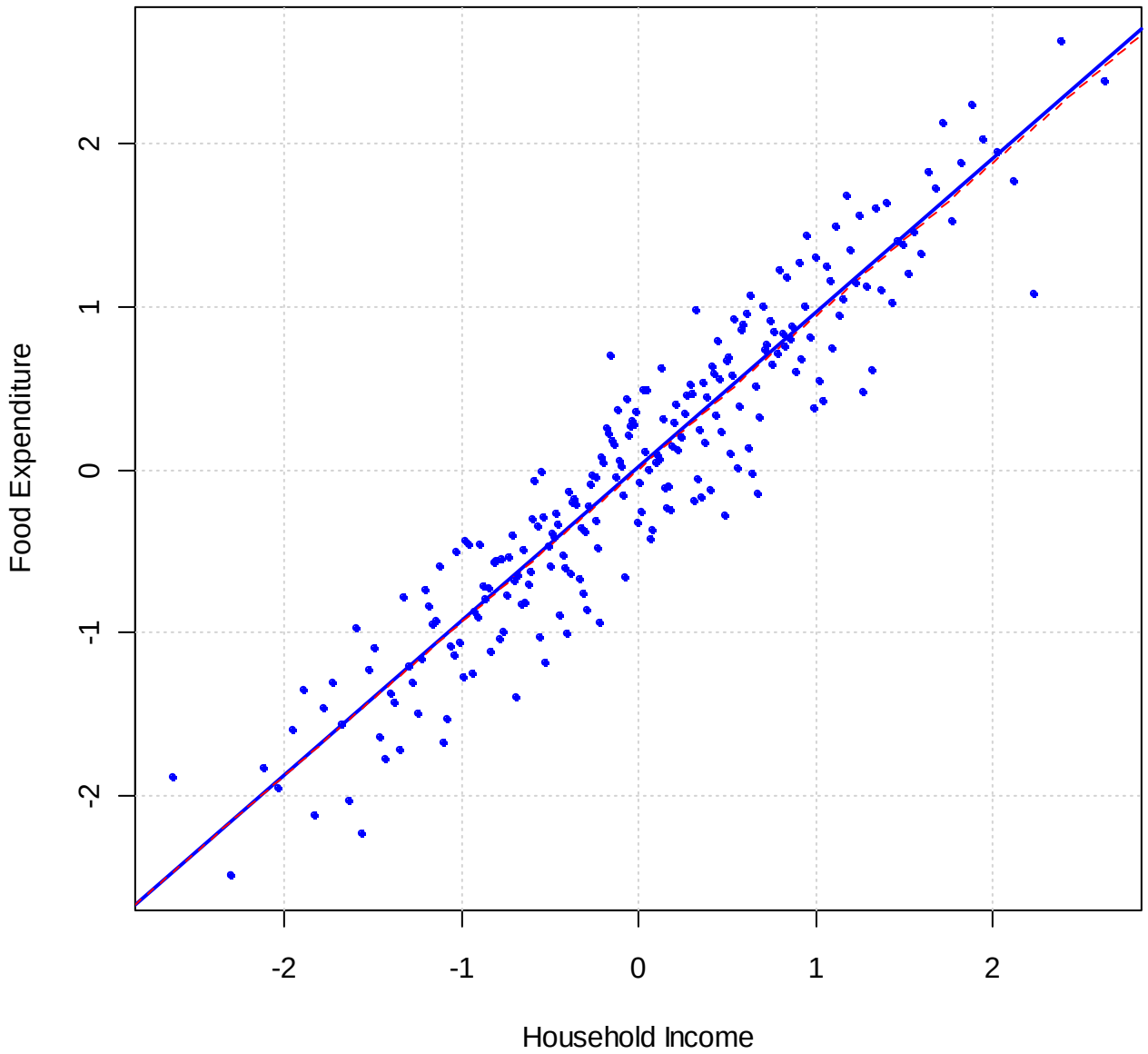

Quantile Regression - median
Food Expenditure
500
1000
1500
2000
1000
2000
3000
4000
5000
Household Income


Quantile Regression, multiple quantiles
Food Expenditure
500
1000
1500
2000
1000
2000
3000
4000
5000
Household Income

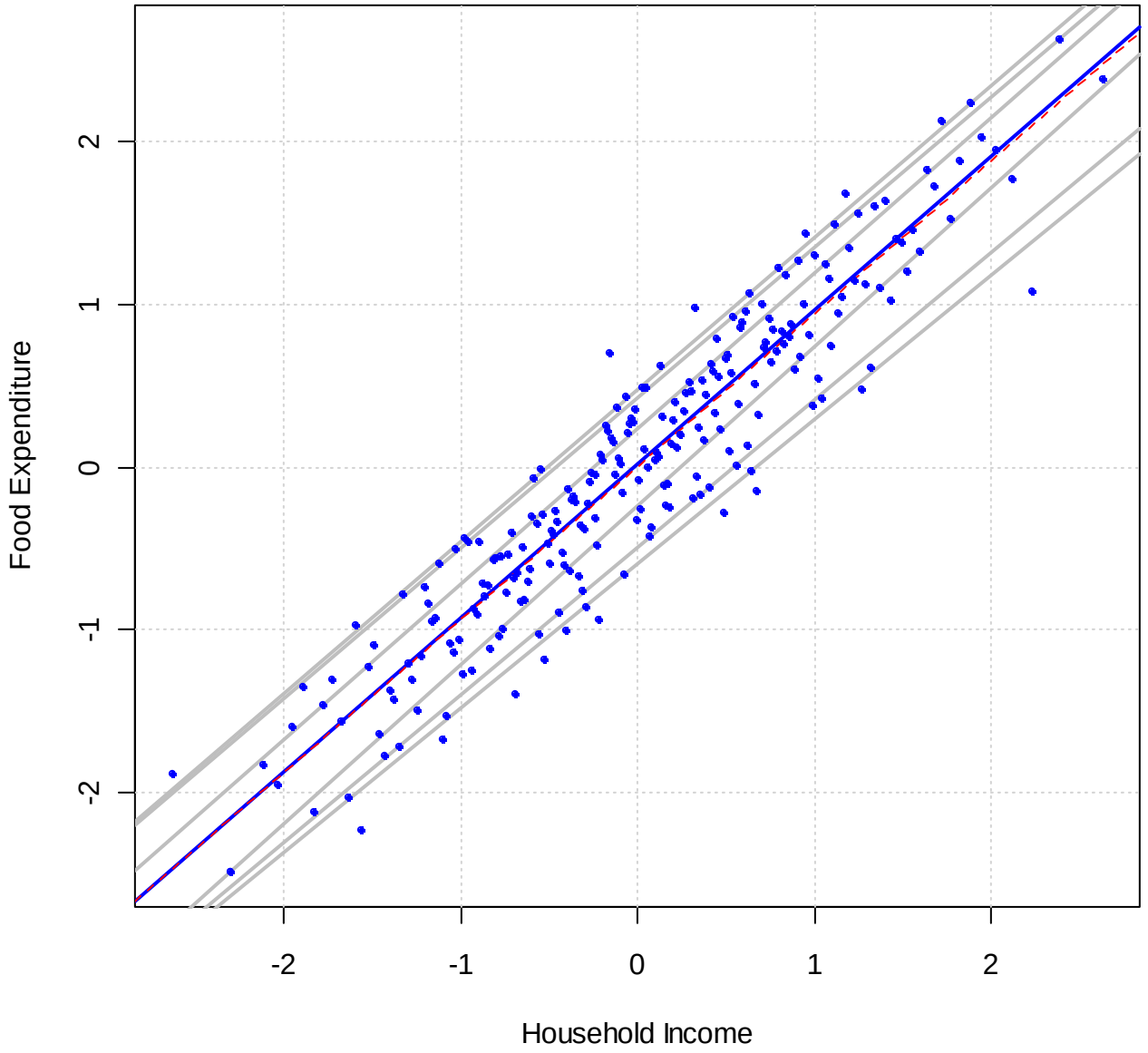

Engel data, 19th century, Belgium
Food Expenditure
Household Income
-2
-1
0
1
2


Least Squares Estimate
Food Expenditure
Household Income
2
1
0
-1
-2
-2
-1
0
1
2

eltares


Quantile Regression - median
Food Expenditure
Household Income
-2
-1
0
1
2
-2
-1
0
1
2

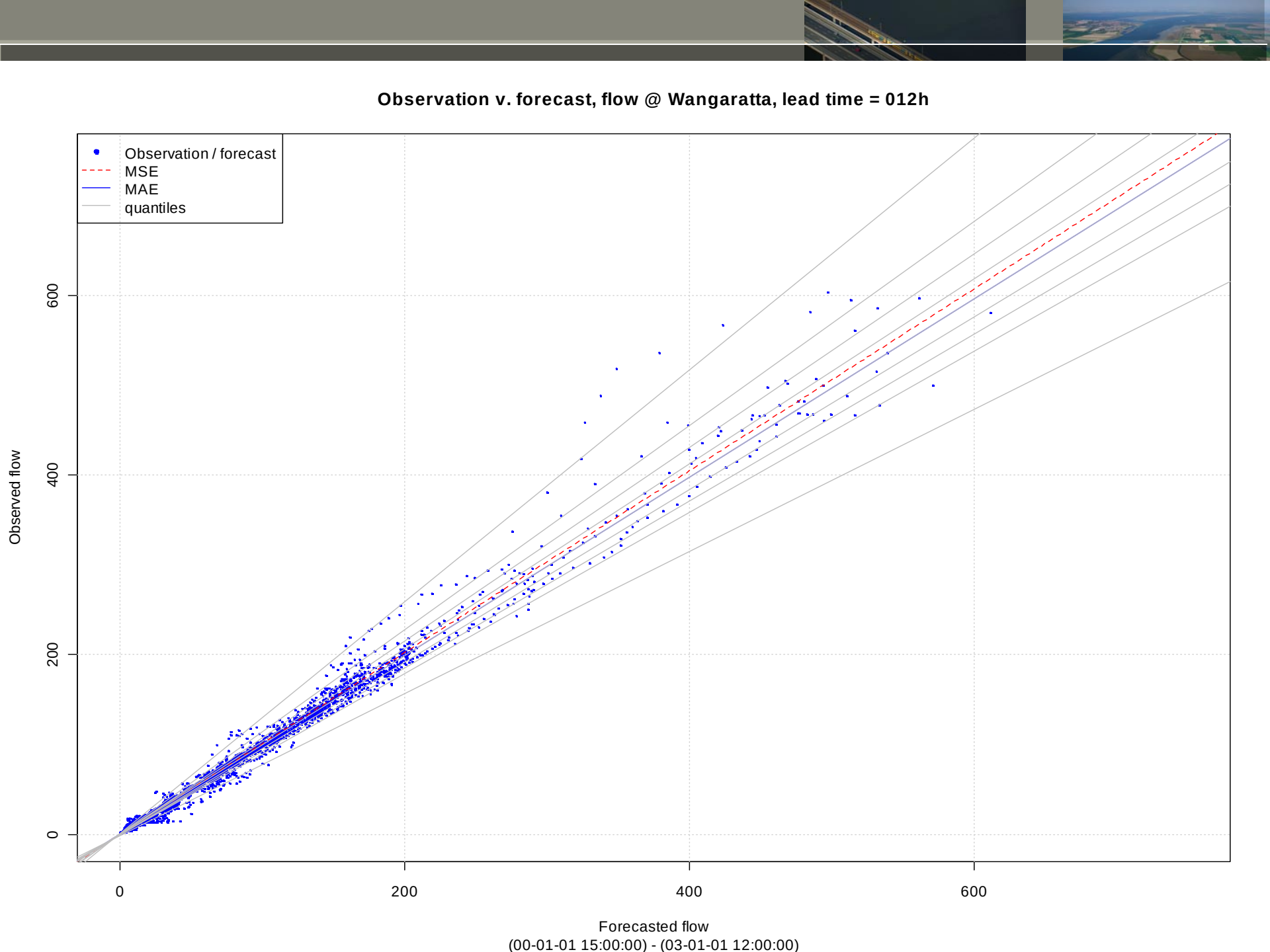
Quantile Regression, multiple quantiles

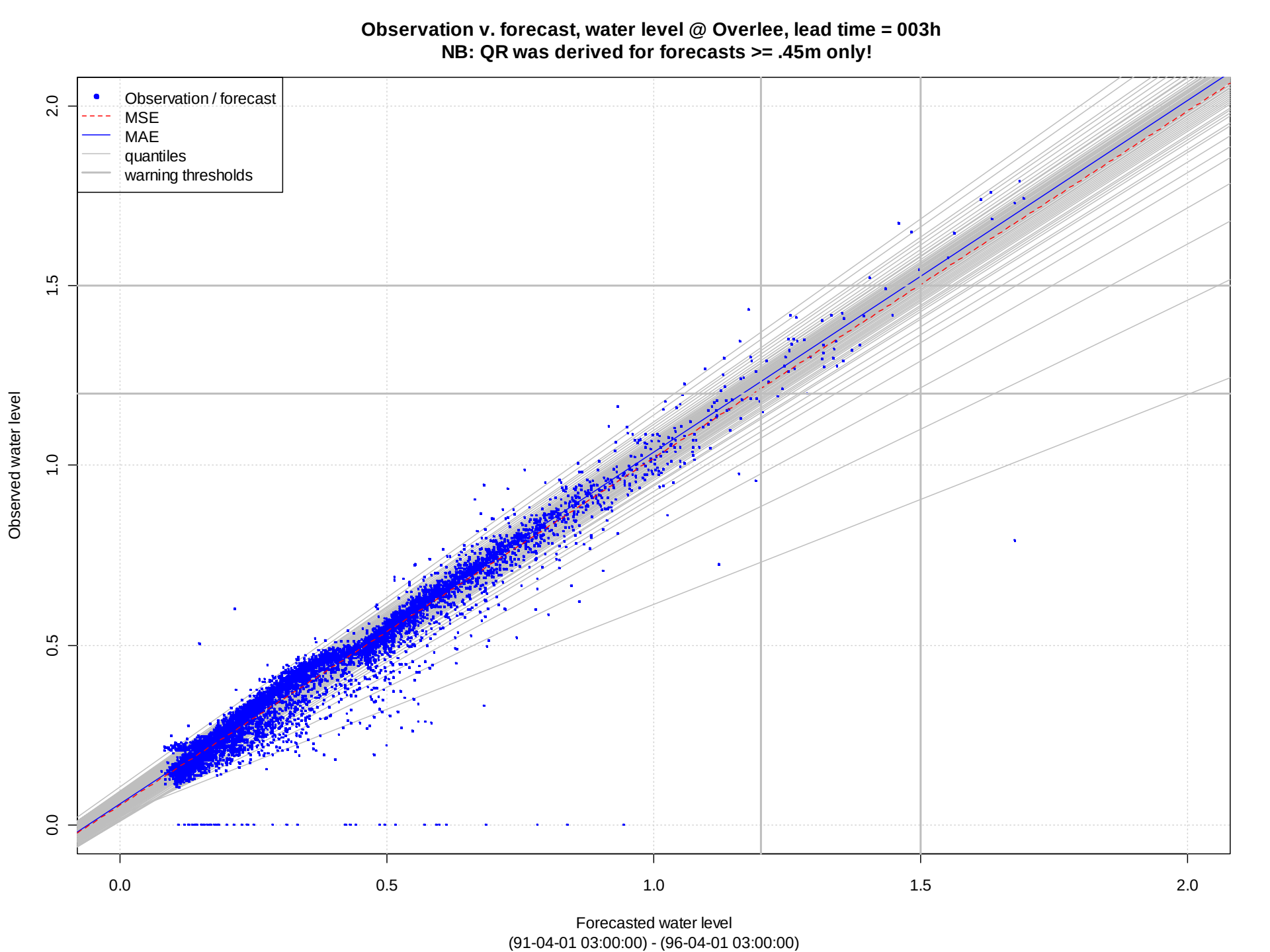


eltares

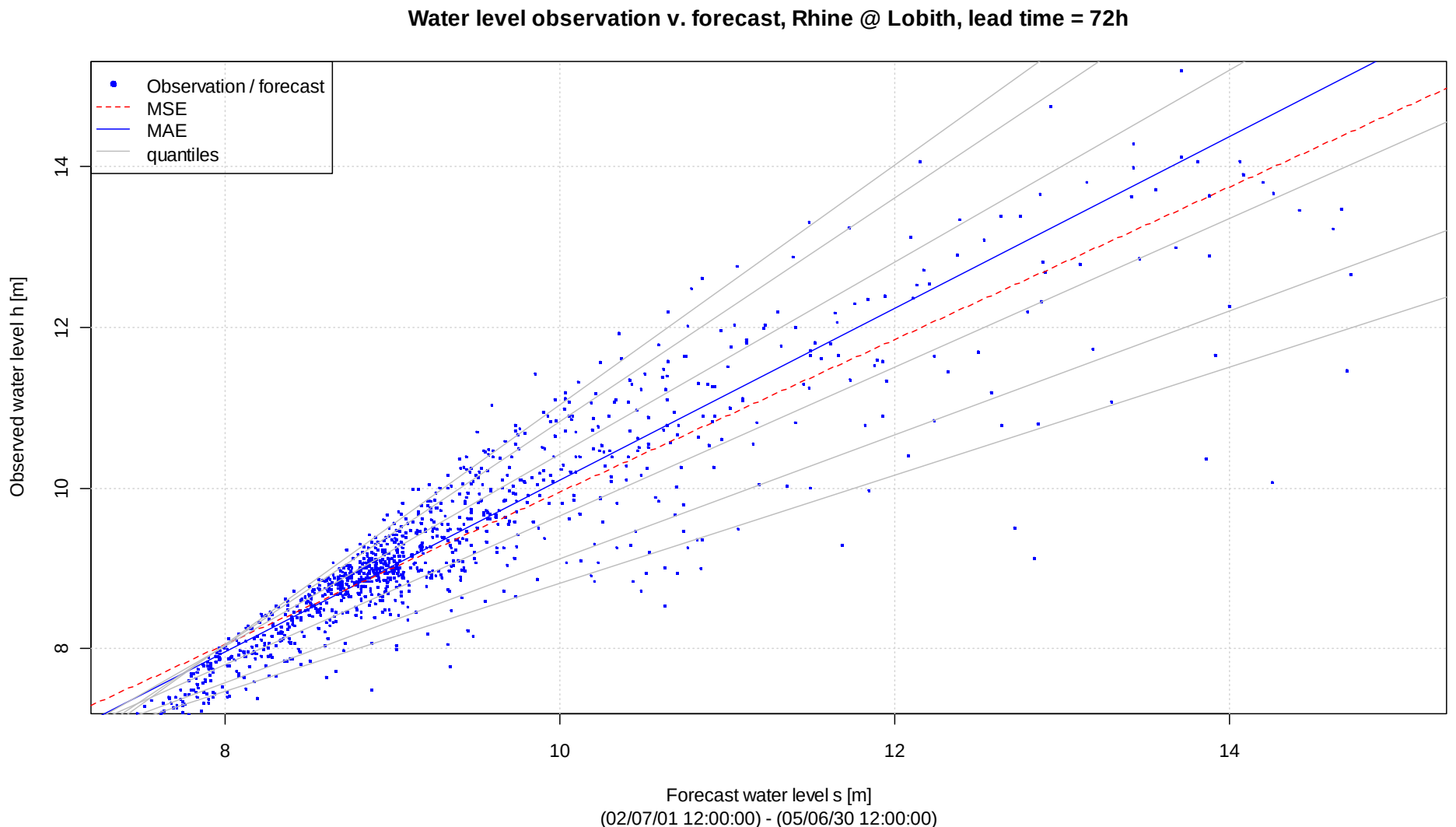

Estimating predictive hydrological uncertainty

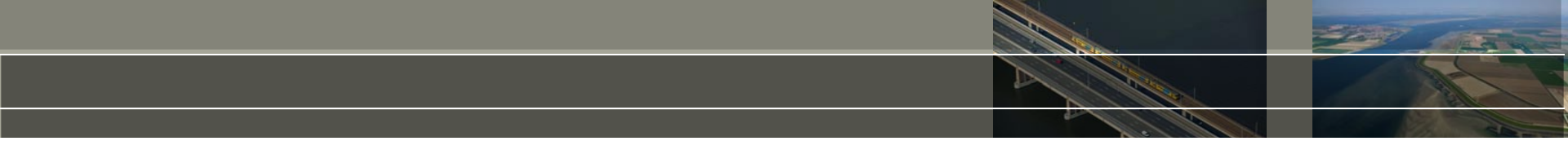
- Fit QR-models where observation is ‘a function of’ a forecast
- Alternative model: *forecast error* as ‘a function of’ a forecast
- Other predictors may be added to further constrain uncertainties
- QR-models can be polynomials, don’t have to be lines
- Derive QR-models on calibration set, then hope that the error structure will remain unchanged

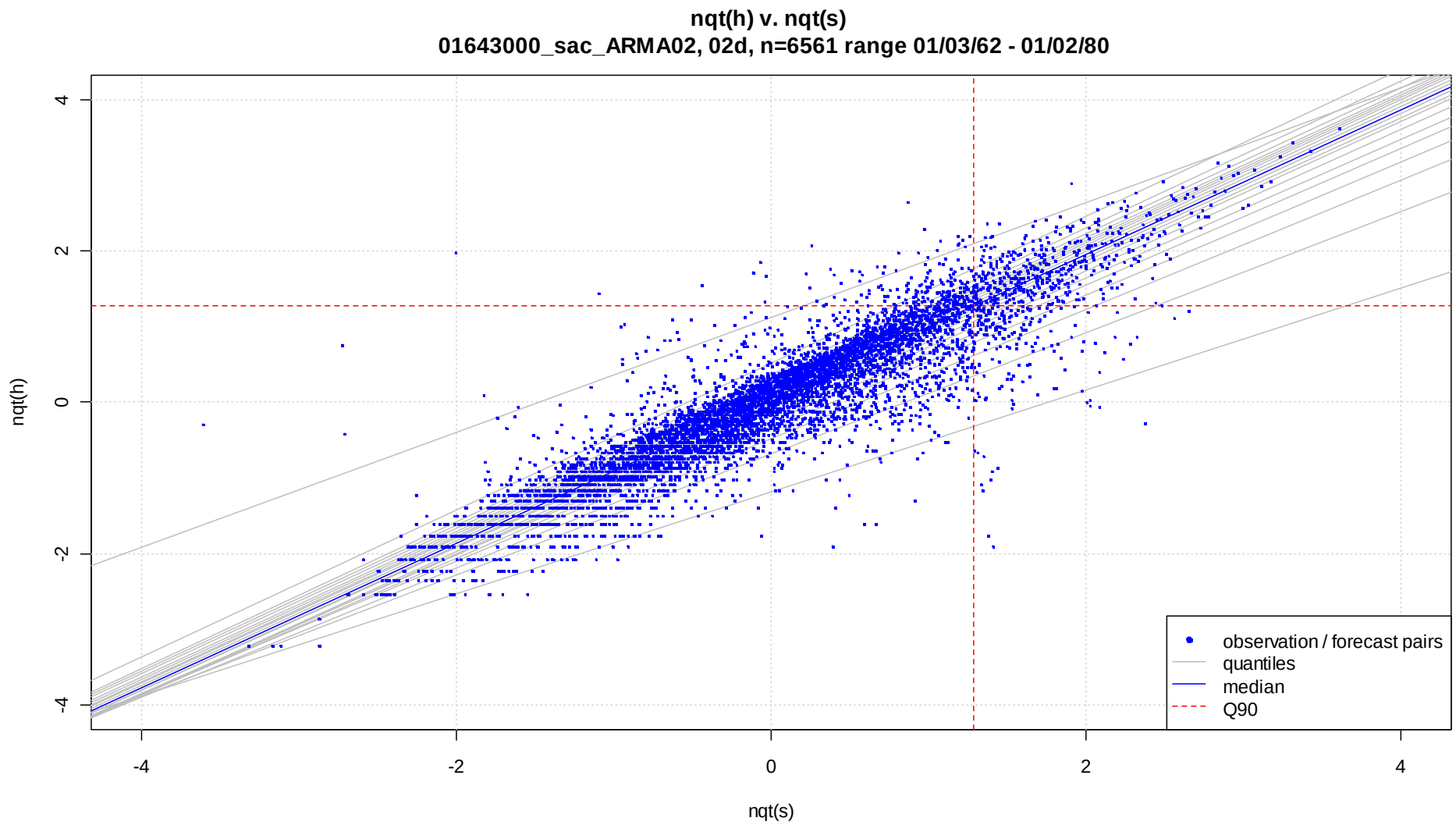

Observation v. forecast, flow @ Wangaratta, lead time = 012h
Observation / forecast
MSE
MAE
quantiles
Observed flow
0
200
400
600
0
200
400
600
Forecasted flow
(00-01-01 15:00:00) - (03-01-01 12:00:00)


Observation v. forecast, water level @ Overlee, lead time = 003h
NB: QR was derived for forecasts >= .45m only!
Observation / forecast
MSE
MAE
quantiles
warning thresholds
Observed water level
Forecasted water level
(91-04-01 03:00:00) - (96-04-01 03:00:00)
0.0
0.5
1.0
1.5
2.0

Watch out for quantile crossing

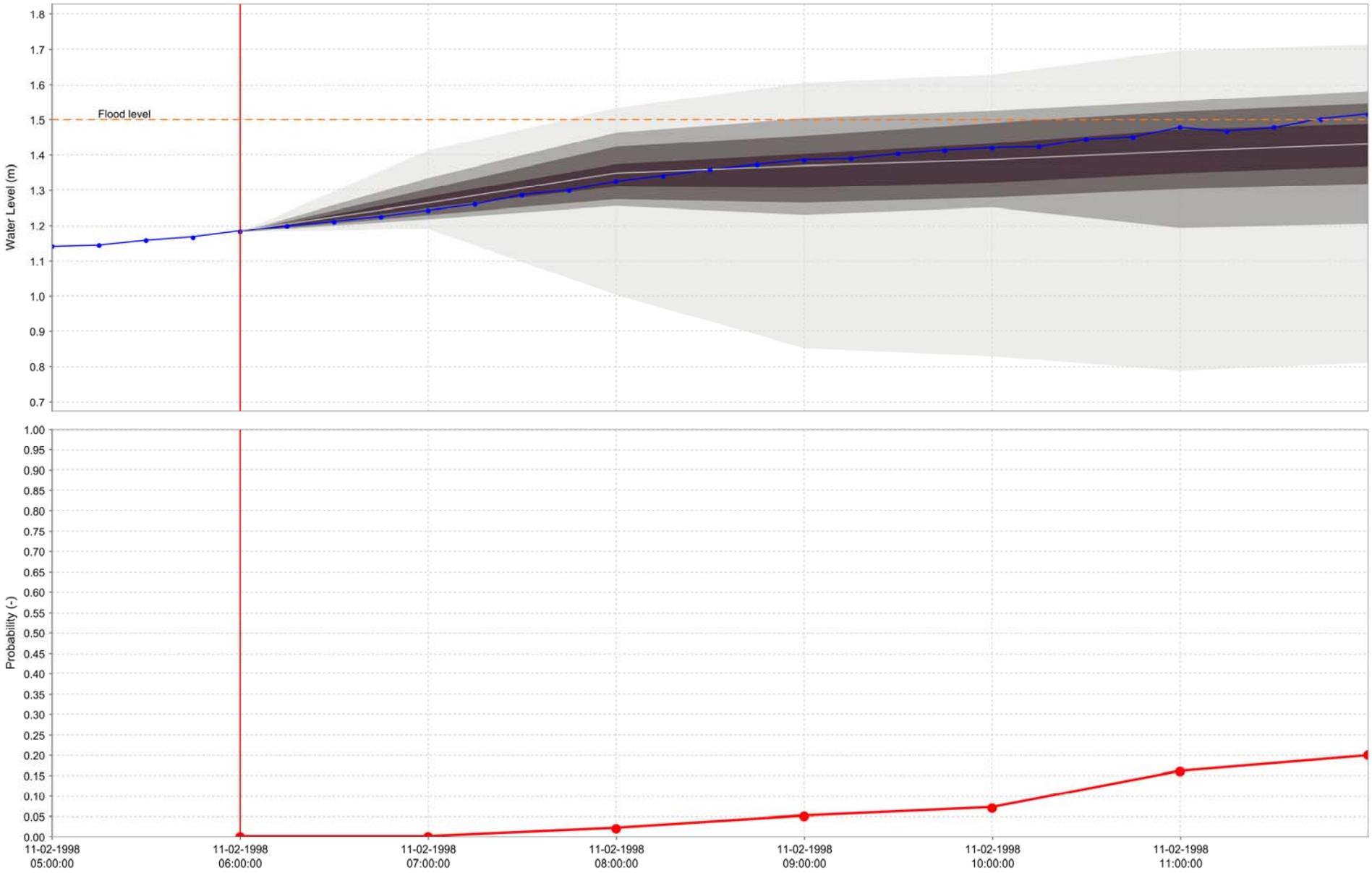





nqt(h) v. nqt(s)
01643000_sac_ARMA02, 02d, n=6561 range 01/03/62 - 01/02/80
nqt(h)
nqt(s)
4
2
0
-2
-4
observation / forecast pairs
quantiles
median
Q90

Deltares

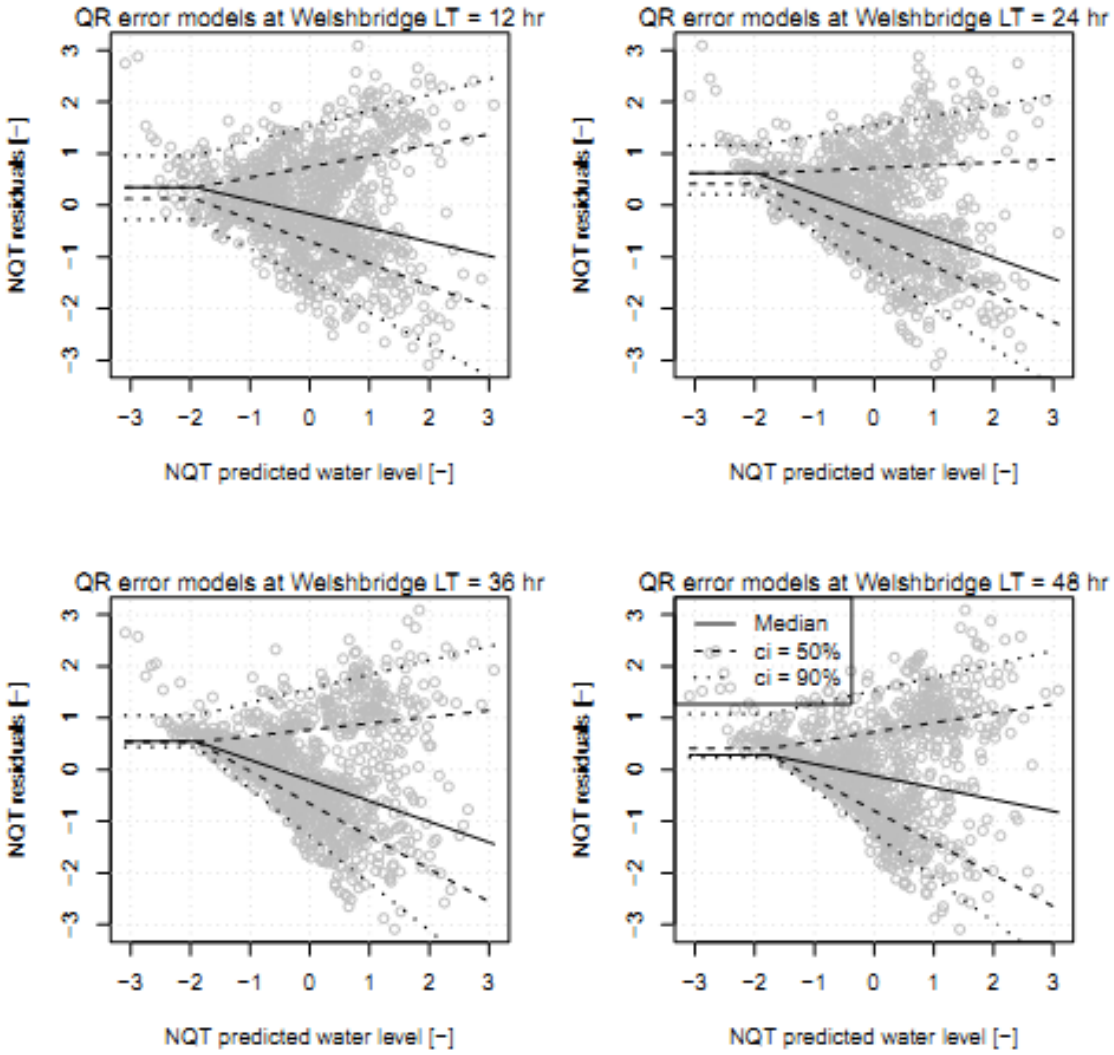
Overlee


Water Level (m)
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Flood level
Probability (-)
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
11-02-1998 05:00:00
11-02-1998 06:00:00
11-02-1998 07:00:00
11-02-1998 08:00:00
11-02-1998 09:00:00
11-02-1998 10:00:00
11-02-1998 11:00:00

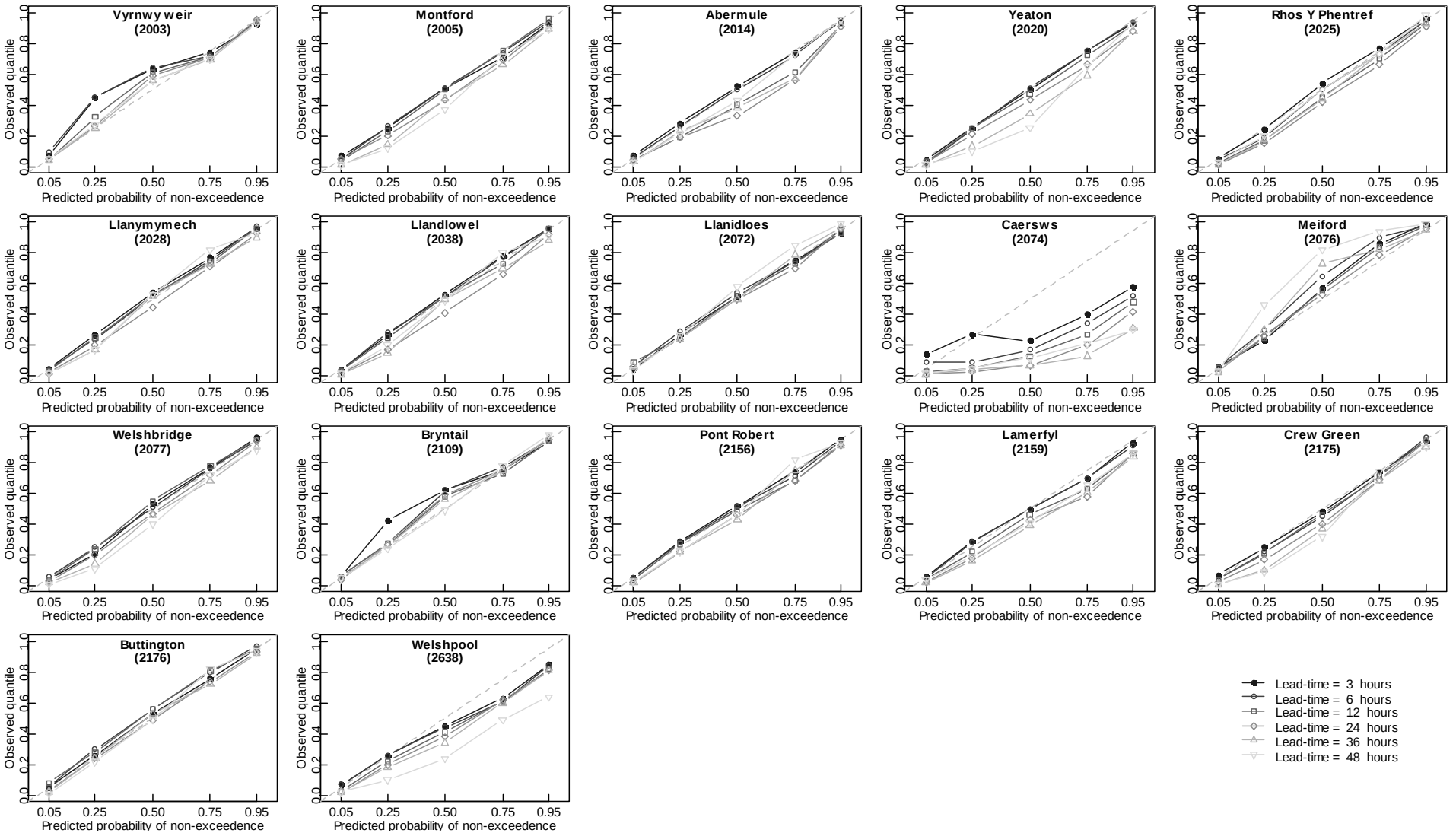

Applications this far

- ~25 NFFS basins (England and Wales)
- White Cart (Scotland)
- Ovens River (Australia)
- HEPEX-basins (USA) (in progress)

NFFS basins: Welshbridge

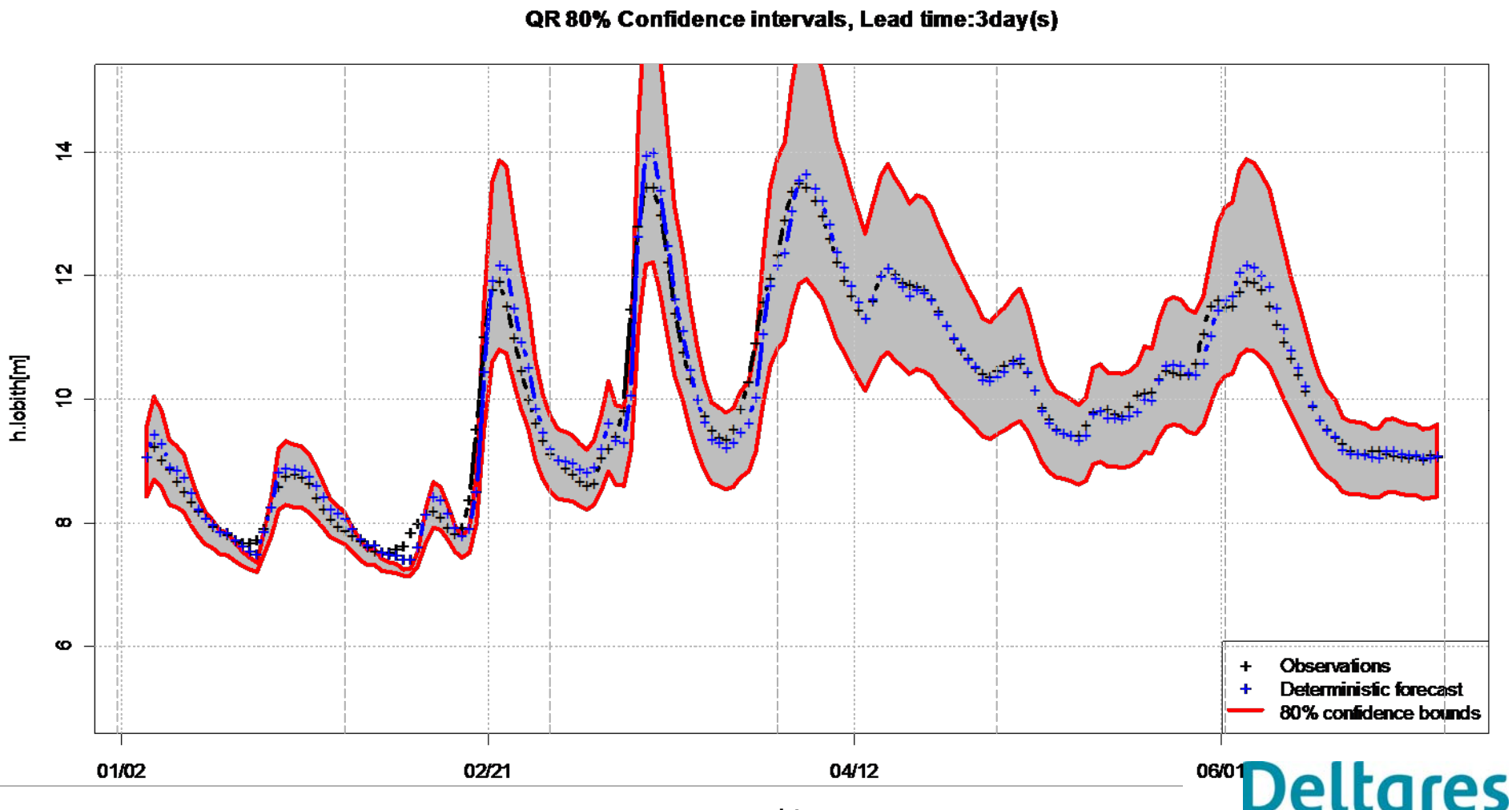


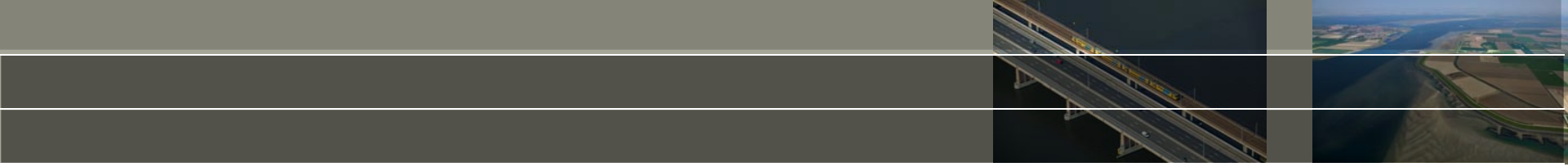
Deltares


Vyrnwy weir (2003)
Montford (2005)
Abermule (2014)
Yeaton (2020)
Rhos Y Phentref (2025)
Llanymymech (2028)
Llandlowel (2038)
Llanidloes (2072)
Caersws (2074)
Meiford (2076)
Welshbridge (2077)
Bryntail (2109)
Pont Robert (2156)
Lamerfyl (2159)
Crew Green (2175)
Buttington (2176)
Welshpool (2638)
Observed quantile
Predicted probability of non-exceedence
0.05 0.25 0.50 0.75 0.95
0.0 0.2 0.4 0.6 0.8 1.0
Lead-time = 3 hours
Lead-time = 6 hours
Lead-time = 12 hours
Lead-time = 24 hours
Lead-time = 36 hours
Lead-time = 48 hours

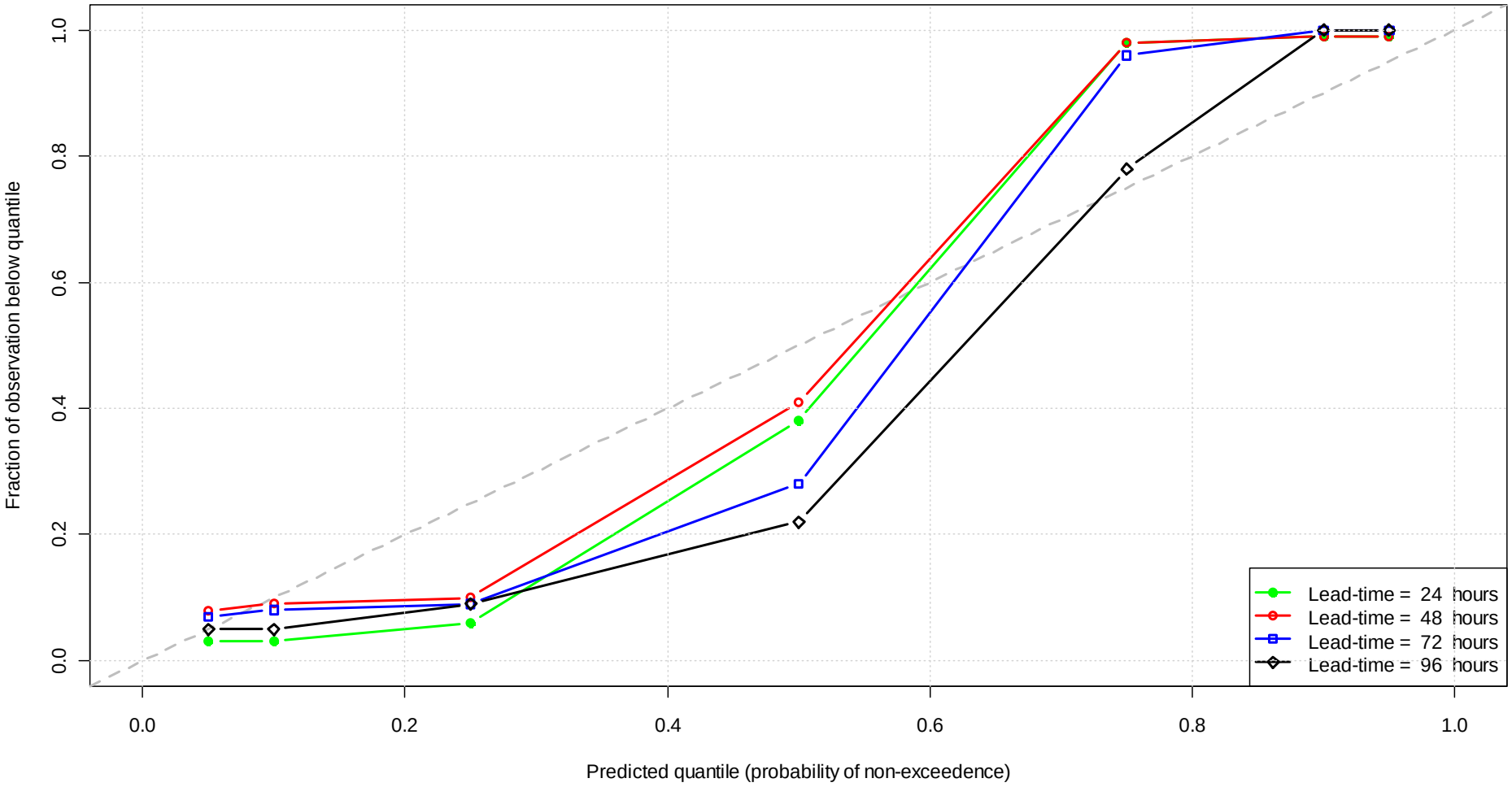
Deltares

Application: Lobith forecasting system

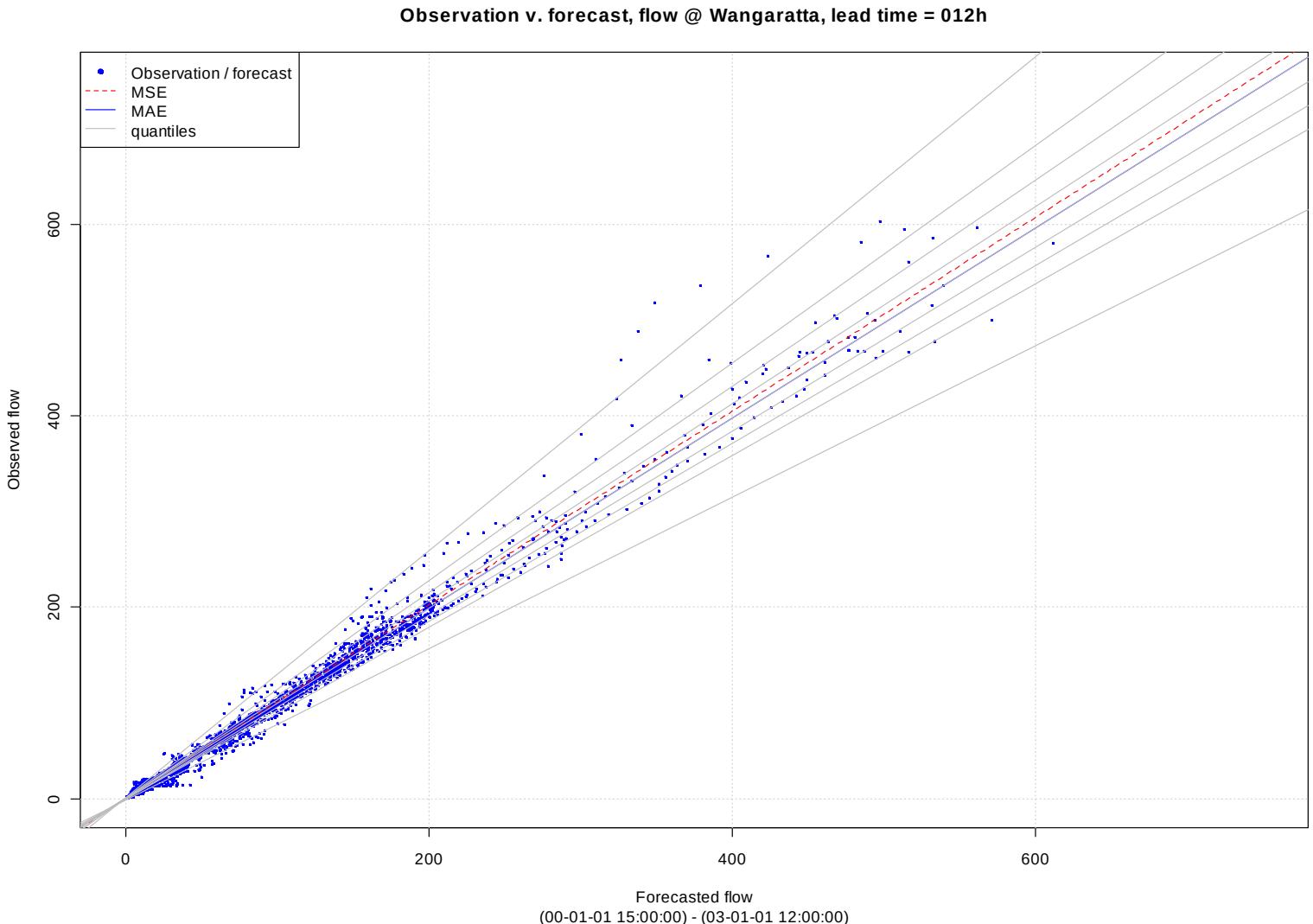


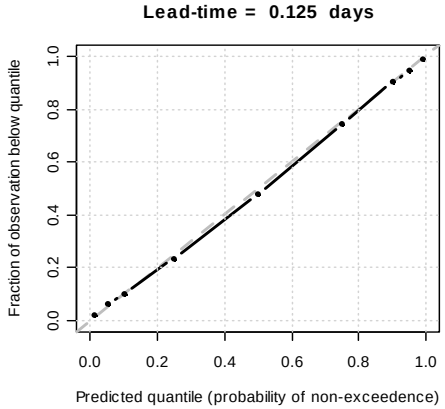


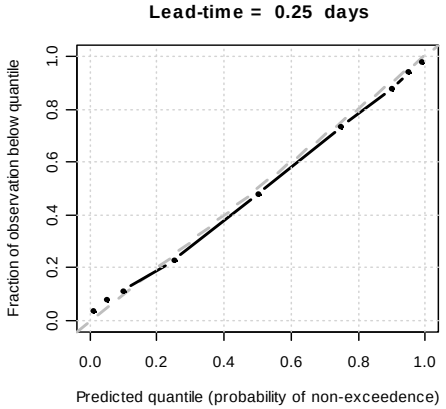
QR Lobith, Quantile plot

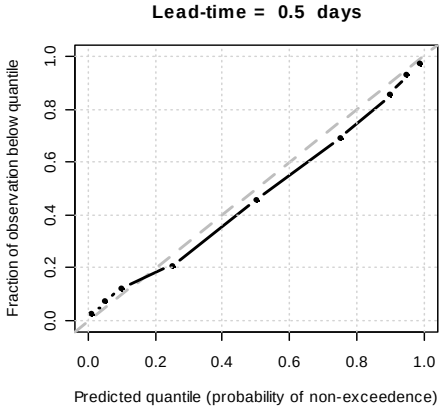


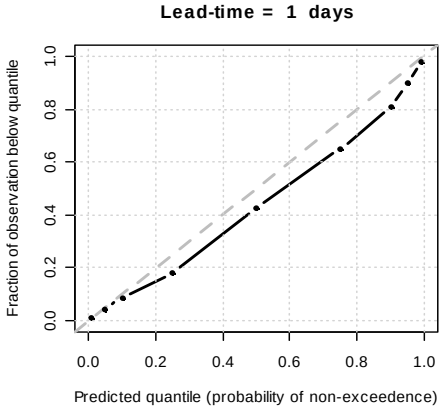
Ovens River @ Wangaratta

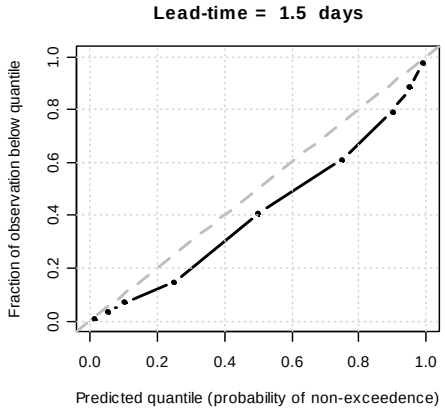


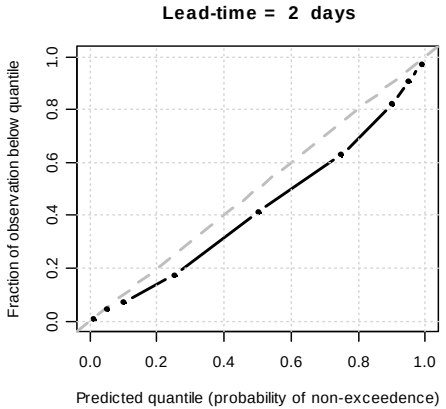

Lead-time = 0.125 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

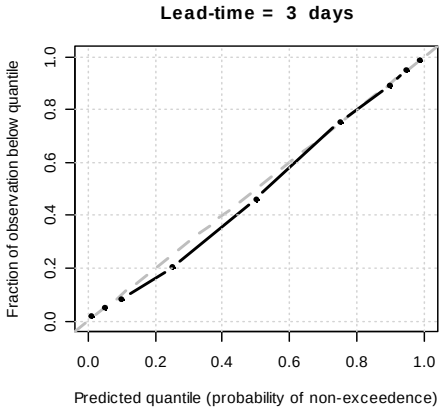

Lead-time = 0.25 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

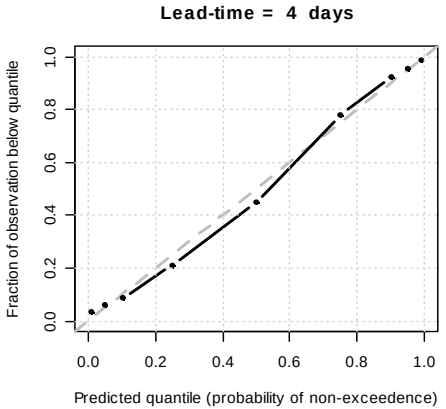

Lead-time = 0.5 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

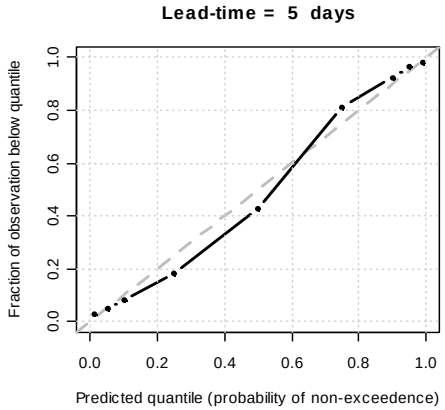

Lead-time = 1 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

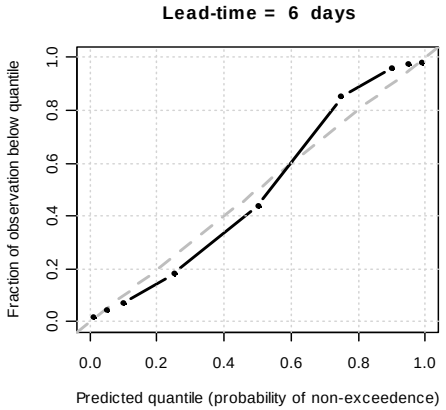

Lead-time = 1.5 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

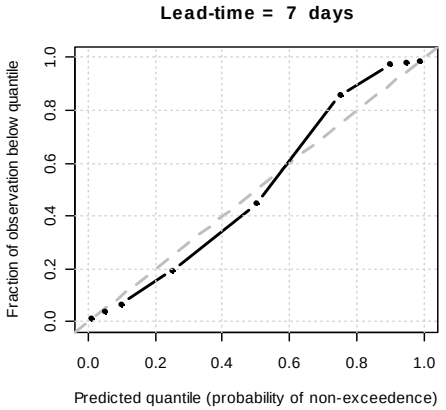

Lead-time = 2 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

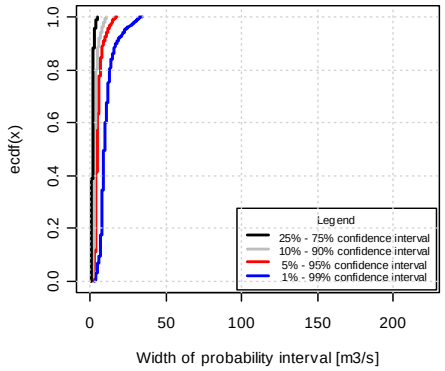

Lead-time = 3 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

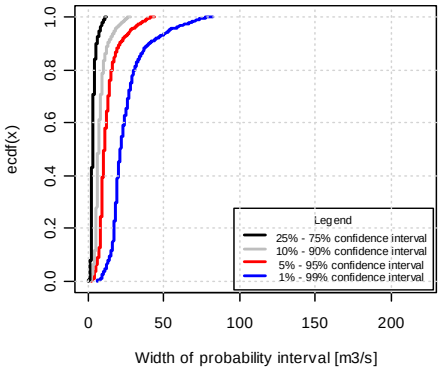

Lead-time = 4 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

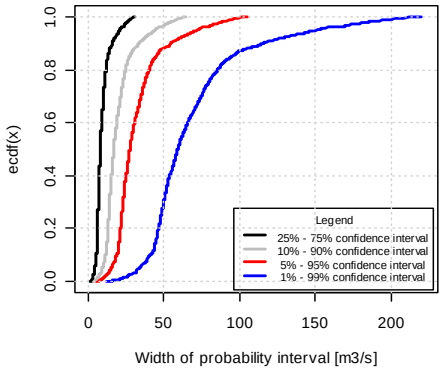

Lead-time = 5 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

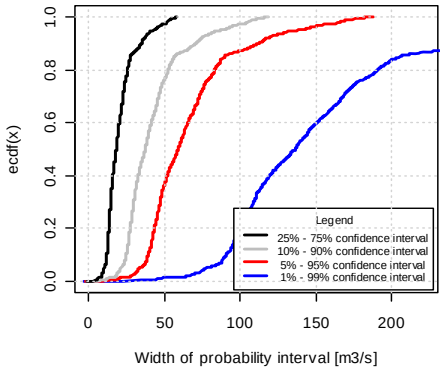

Lead-time = 6 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

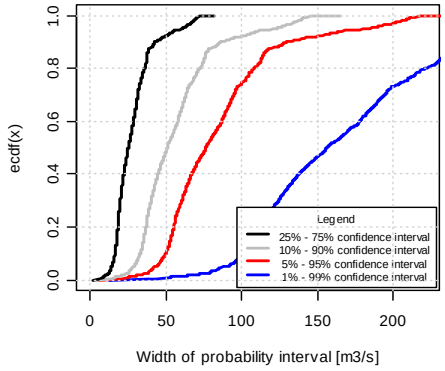

Lead-time = 7 days
Fraction of observation below quantile
Predicted quantile (probability of non-exceedence)

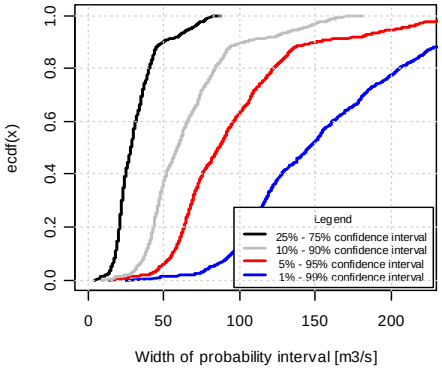

Lead time 0.125 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

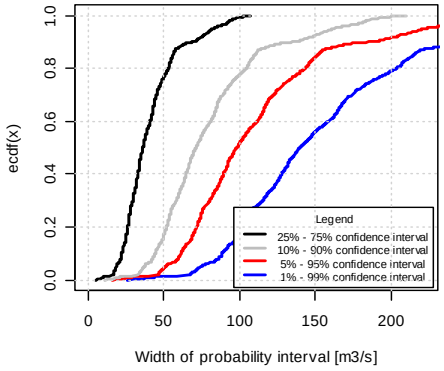

Lead time 0.25 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

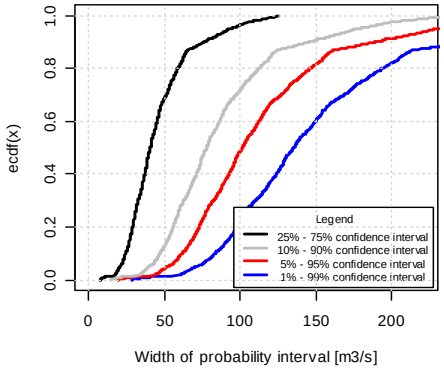

Lead time 0.5 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

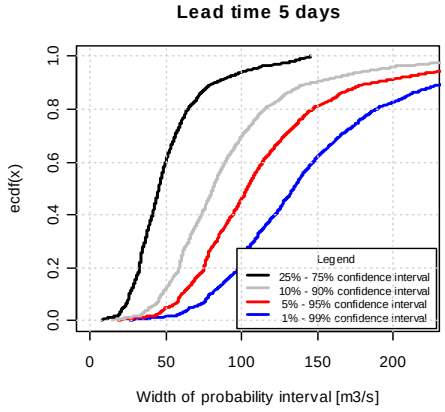

Lead time 1 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

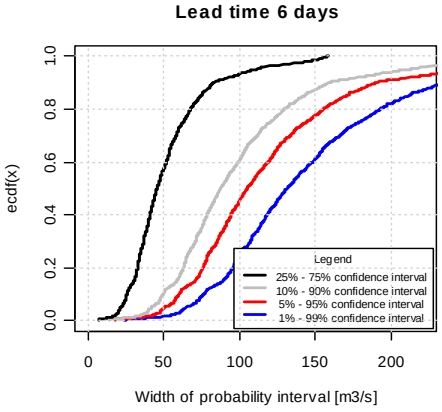

Lead time 1.5 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

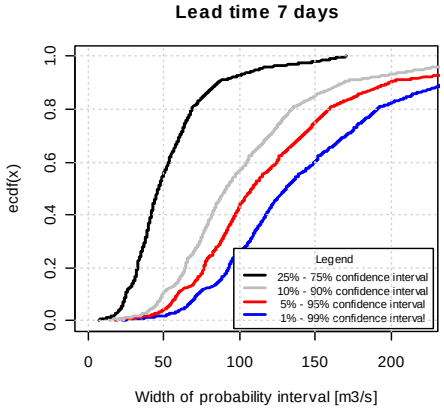

Lead time 2 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

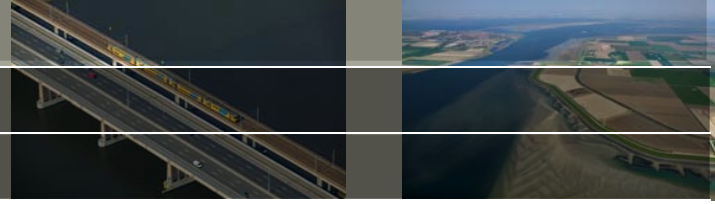

Lead time 3 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval


Lead time 4 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval


Lead time 5 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval


Lead time 6 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval


Lead time 7 days
ecdf(x)
Width of probability interval [m3/s]
Legend
25% - 75% confidence interval
10% - 90% confidence interval
5% - 95% confidence interval
1% - 99% confidence interval

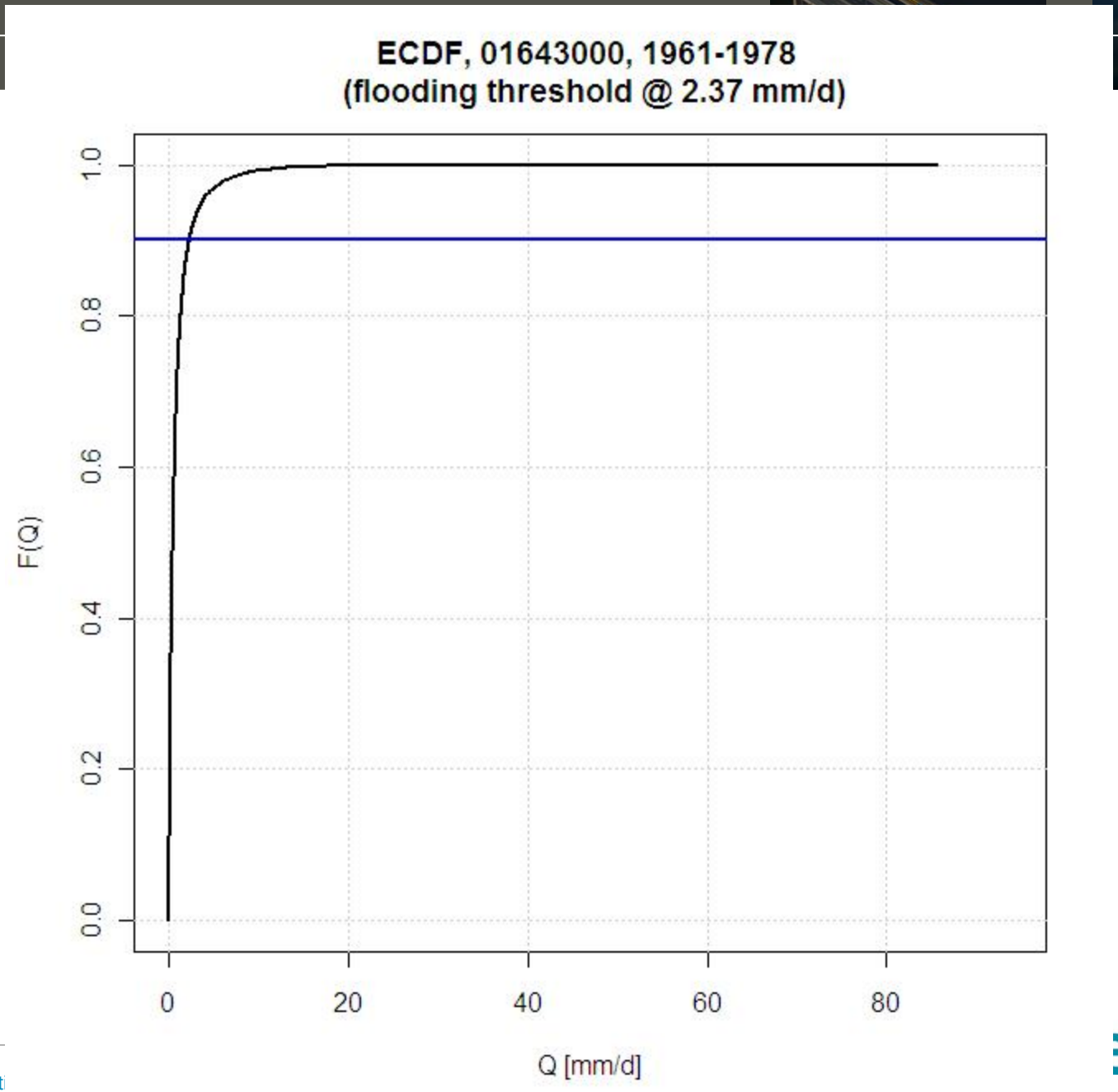


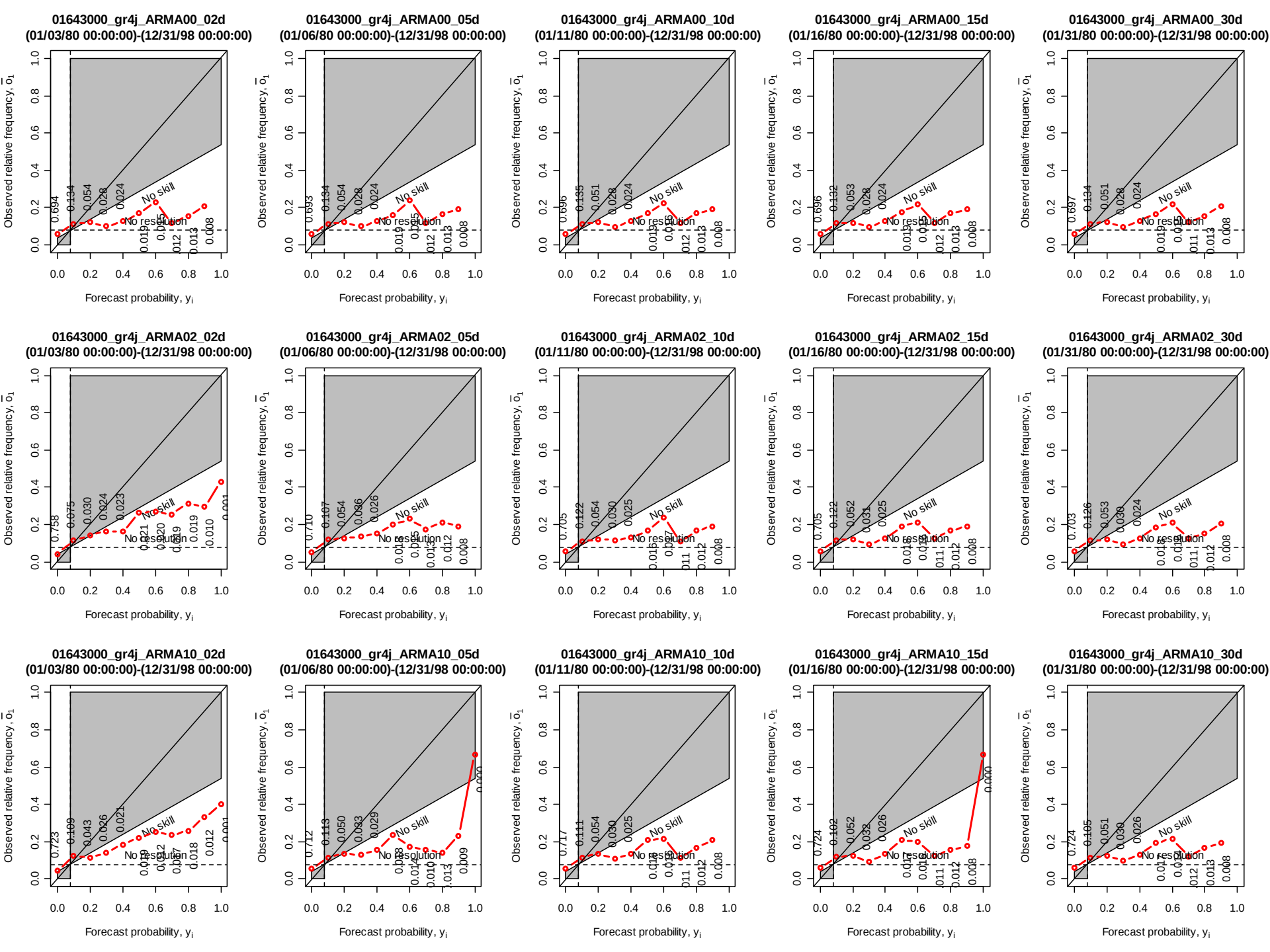
HEPEX scenario 1: hydrol. model simulations

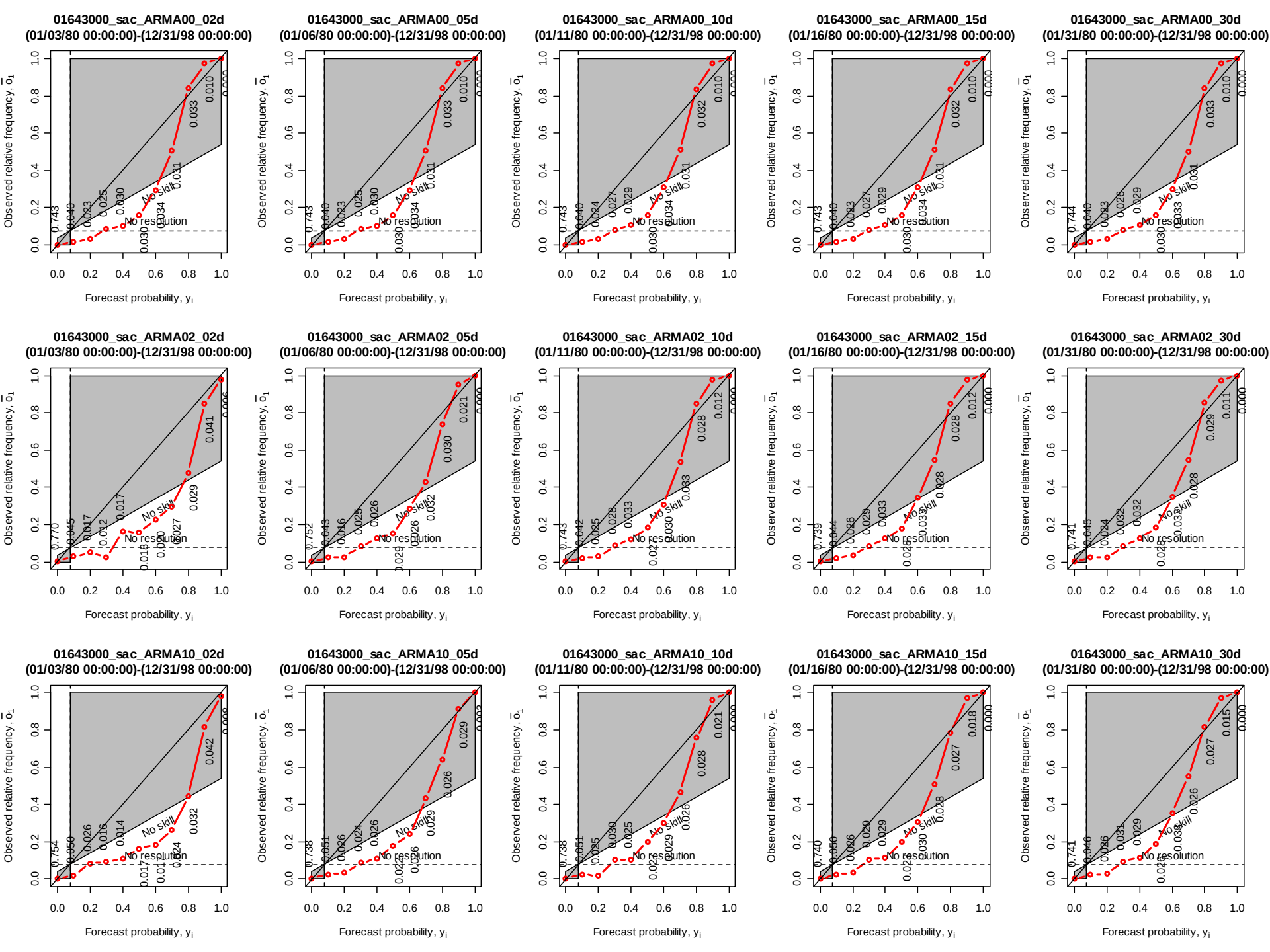
- Monocacy River @ Frederick (0164 3000)
- daily flow, not accumulated
- (calibrated) sac and gr4j models
- lead times considered: 2, 5, 10, 15 and 30 days
- 3 alternative error correction options:
  - 1. no correction: simulation = forecast
  - 2. ARMA-correction, fixed parameters
  - 3. ARMA-correction, dynamically parameterised on past 10d

→ 1 location x 2 models x 3 error corr'n options x 5 leadtimes = 30 cases to be evaluated

→ so far, only evaluation is that of probability of high flow


ECDF, 01643000, 1961-1978
(flooding threshold @ 2.37 mm/d)
F(Q)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Q [mm/d]


01643000_gr4j_ARMA00_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA00_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA00_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA00_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA00_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA02_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA02_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA02_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA02_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA02_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA10_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA10_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA10_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA10_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_gr4j_ARMA10_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
Observed relative frequency, $\bar{o}_1$
Forecast probability, y_i
No skill
No resolution


01643000_sac_ARMA00_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA00_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA00_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA00_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA00_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA02_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA02_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA02_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA02_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA02_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA10_02d
(01/03/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA10_05d
(01/06/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA10_10d
(01/11/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA10_15d
(01/16/80 00:00:00)-(12/31/98 00:00:00)
01643000_sac_ARMA10_30d
(01/31/80 00:00:00)-(12/31/98 00:00:00)
Observed relative frequency, $\bar{o}_1$
Forecast probability, y_i
No skill
No resolution


locId	1643000

			Data					
model	arma	leadtime	bs	bs.baseline	ss	bs.uncert	bs.resol	bs.reliability
sac	ARMA00	2	0.04	0.07	0.38	0.07	0.04	0.01
		5	0.04	0.07	0.38	0.07	0.04	0.01
		10	0.04	0.07	0.38	0.07	0.04	0.01
		15	0.04	0.07	0.39	0.07	0.04	0.01
		30	0.04	0.07	0.39	0.07	0.04	0.01
	ARMA02	2	0.04	0.07	0.36	0.07	0.04	0.01
		5	0.04	0.07	0.38	0.07	0.04	0.01
		10	0.04	0.07	0.4	0.07	0.04	0.01
		15	0.04	0.07	0.41	0.07	0.04	0.01
		30	0.04	0.07	0.41	0.07	0.04	0.01
	ARMA10	2	0.05	0.07	0.33	0.07	0.04	0.02
		5	0.04	0.07	0.38	0.07	0.04	0.01
		10	0.04	0.07	0.41	0.07	0.04	0.01
		15	0.04	0.07	0.41	0.07	0.04	0.01
		30	0.04	0.07	0.42	0.07	0.04	0.01


Next steps in HEPEX experiment

- Thoroughly check intermediate results
- Drop gr4j dataset in favour of vic model simulations
- Try constrain uncertainties by either adding predictors or by stratifying:
  - initial conditions / antecedent precipitation
  - flow at t0
- Extend forecast evaluation
  - full distribution in addition to probability of flooding
- Hopefully compare with other post-processors

Further reading

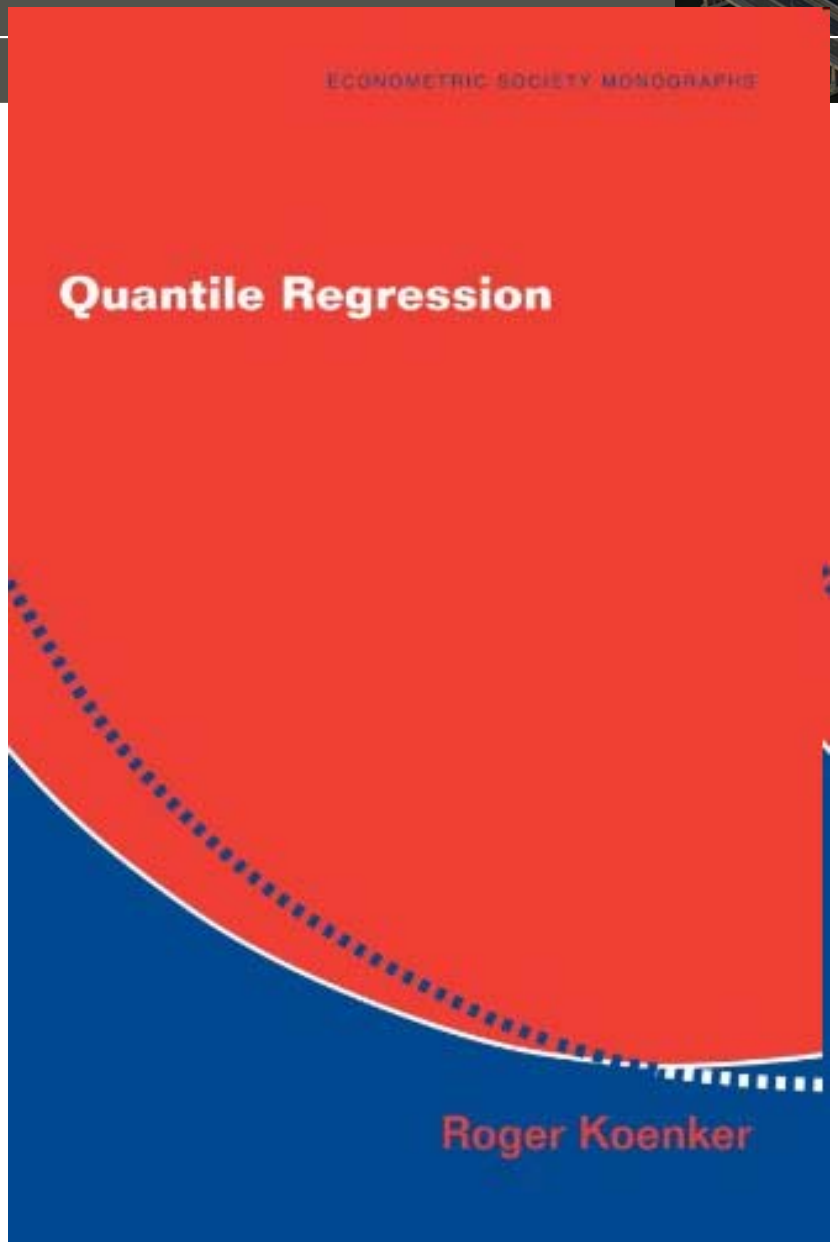
QUANTILE REGRESSION IN R: A VIGNETTE

ROGER KOENKER

ABSTRACT. Quantile regression is an evolving body of statistical methods for estimating and drawing inferences about conditional quantile functions. An implementation of these methods in the R language is available in the package `quantreg`. This vignette offers a brief tutorial introduction to the package. R and the package `quantreg` are open-source software projects and can be freely downloaded from CRAN: `http://cran.r-project.org`.

1. INTRODUCTION

Beran's (2003) provocative definition of statistics as "the study of algorithms for


ECONOMETRIC SOCIETY MONOGRAPHS
Quantile Regression
Roger Koenker

Deltares

Hydrol. Earth Syst. Sci., 15, 255–265, 2011
www.hydrol-earth-syst-sci.net/15/255/2011/
doi:10.5194/hess-15-255-2011

Hydrology and Earth System Sciences

Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales)

A. H. Weerts[1], H. C. Winsemius[1], and J. S. Verkade[1,2]

[1]Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands
[2]Delft University of Technology, Department of Flood Risk Management and Hydraulic Structures, Delft, The Netherlands

Received: 18 July 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 11 August 2010
Revised: 4 January 2011 – Accepted: 5 January 2011 – Published: 21 January 2011

http://www.hydrol-earth-syst-sci.net/15/255/2011/hess-15-255-2011.html

Deltares
Enabling Delta Life

Questions?

Contact details:

about.me/janverkade
jan.verkade@deltares.nl
+31.88.335.8348