

# EVALUATION OF OPERATIONAL HYDROLOGICAL ENSEMBLE FORECASTS IN SWEDEN



by

Jonas Olsson, Göran Lindström, Anna Johnell, Karin Jacobsson

Swedish Meteorological and Hydrological Institute (SMHI)

Norrköping, Sweden

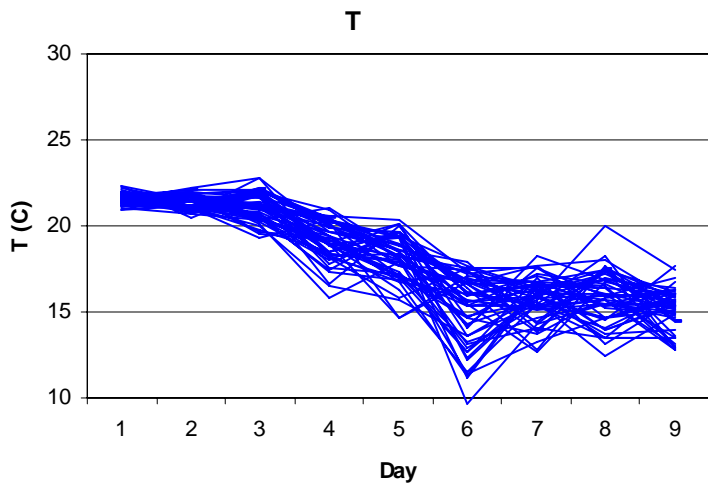
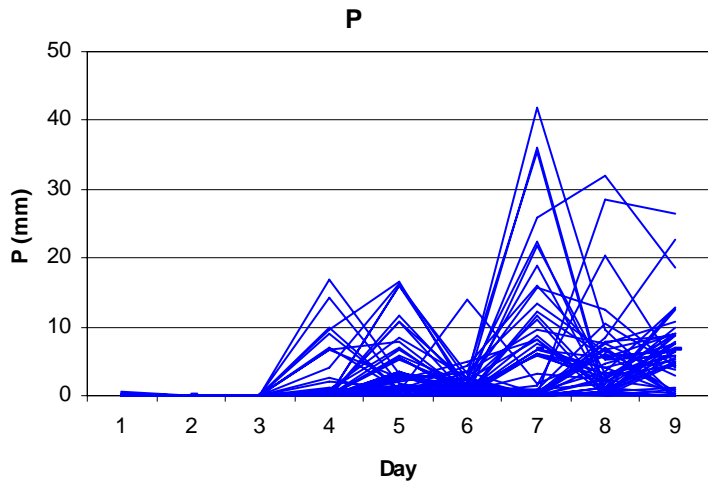


# CONTENTS

- **Overview** of the hydrological ensemble prediction system (EPS) at SMHI
- **Evaluation** of 18 months of data from 45 catchments in Sweden
- **Discussion** on the interpretation and presentation of hydrological EPS forecasts

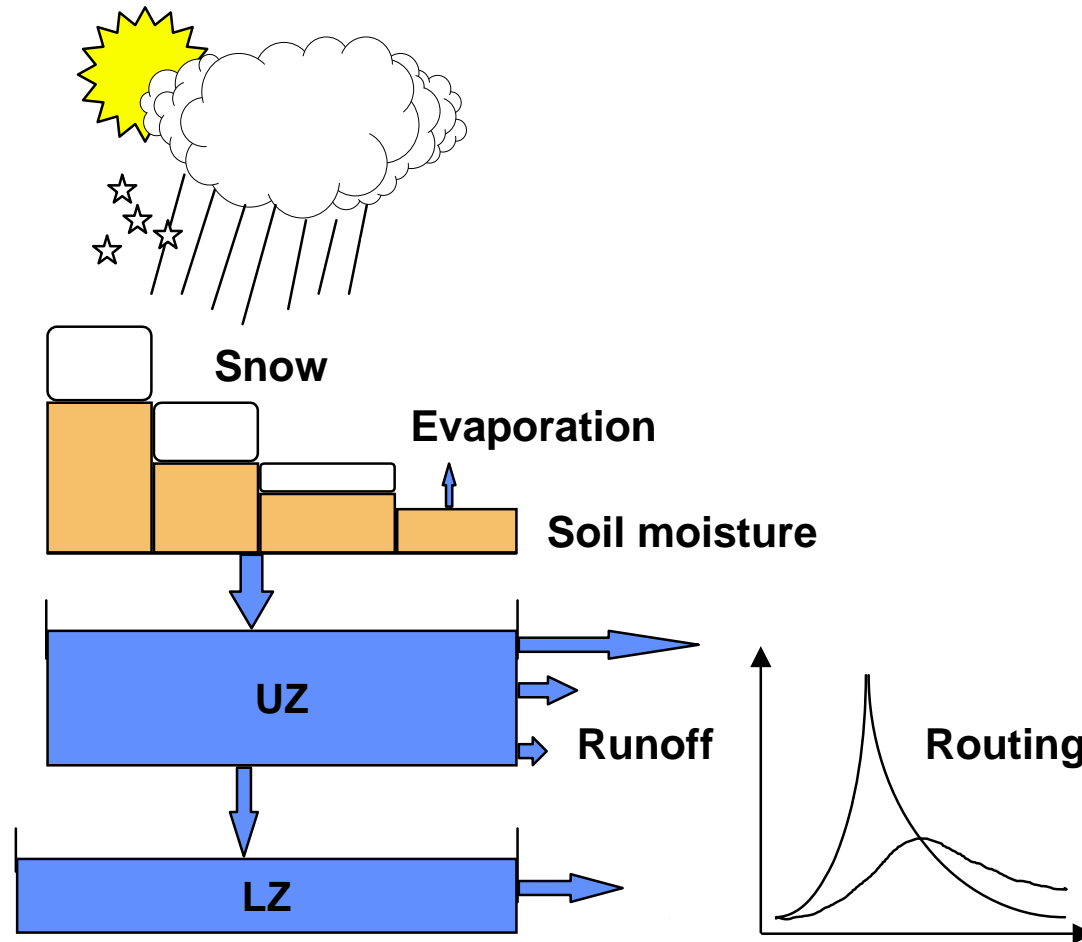
# HYDROLOGICAL EPS FORECASTS By the HBV model

## P and T from ECMWF



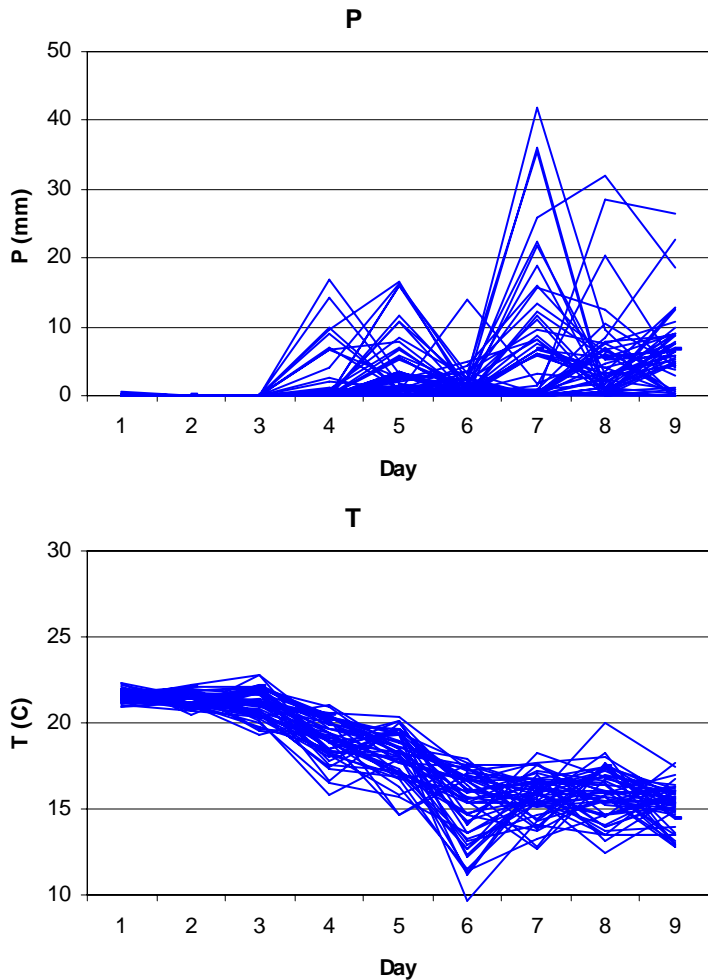
— HBV →

# THE HBV MODEL



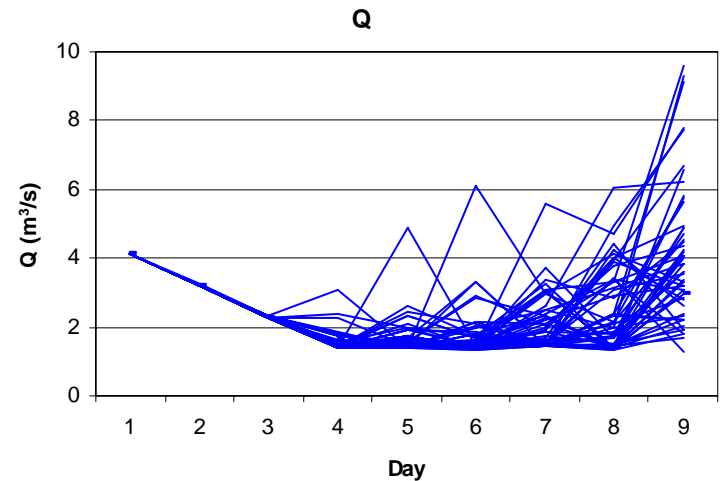
# HYDROLOGICAL EPS FORECASTS By the HBV model

## P and T from ECMWF



— HBV →

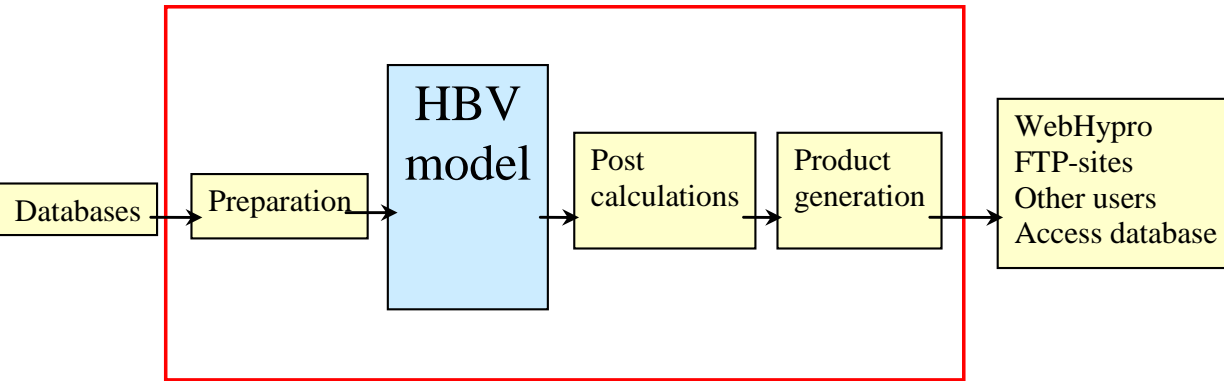
## Discharge Q



# OPERATIONAL HYDRO-EPS AT SMHI

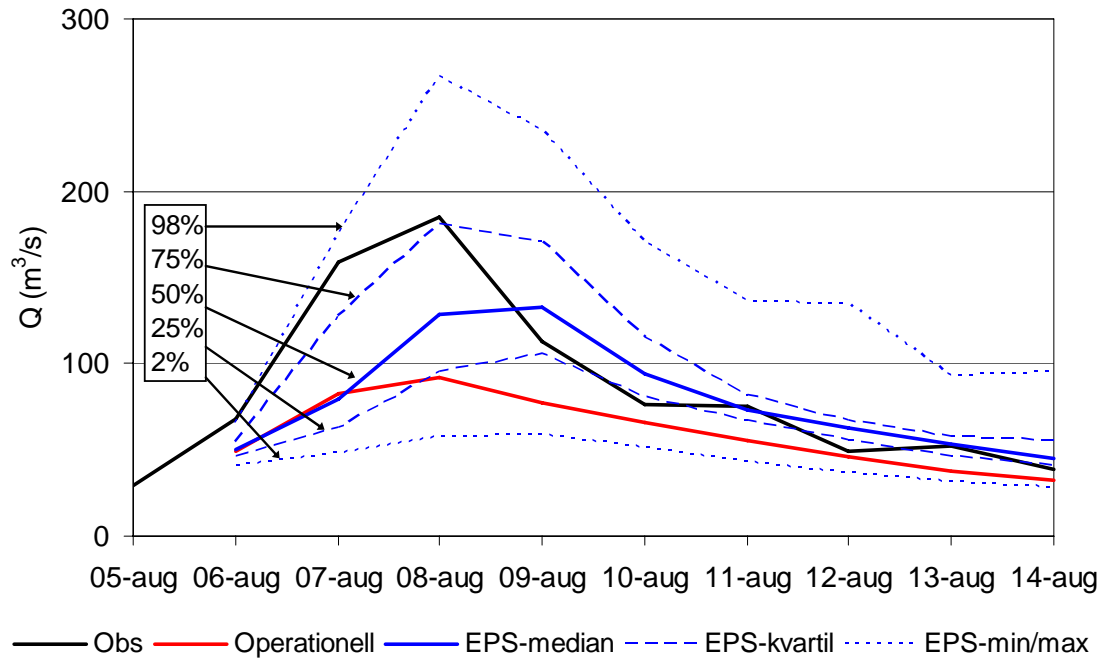
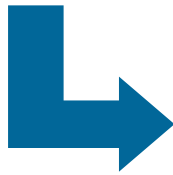
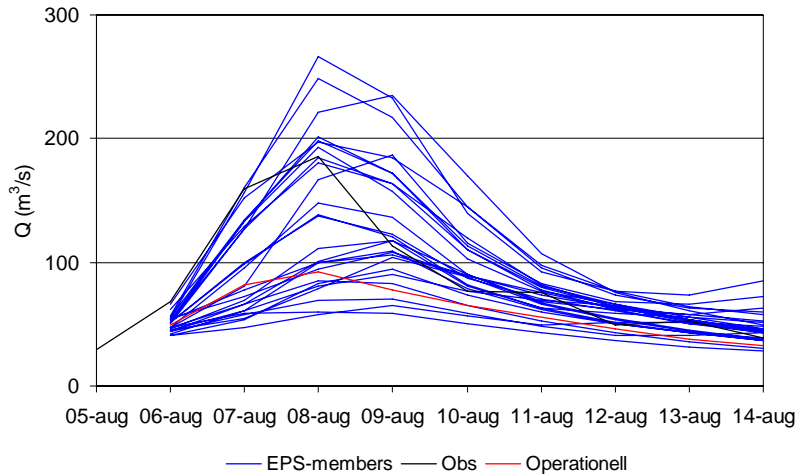
## Production system and presentation (July 2004)

Aegir



# HYDROLOGICAL EPS FORECASTS

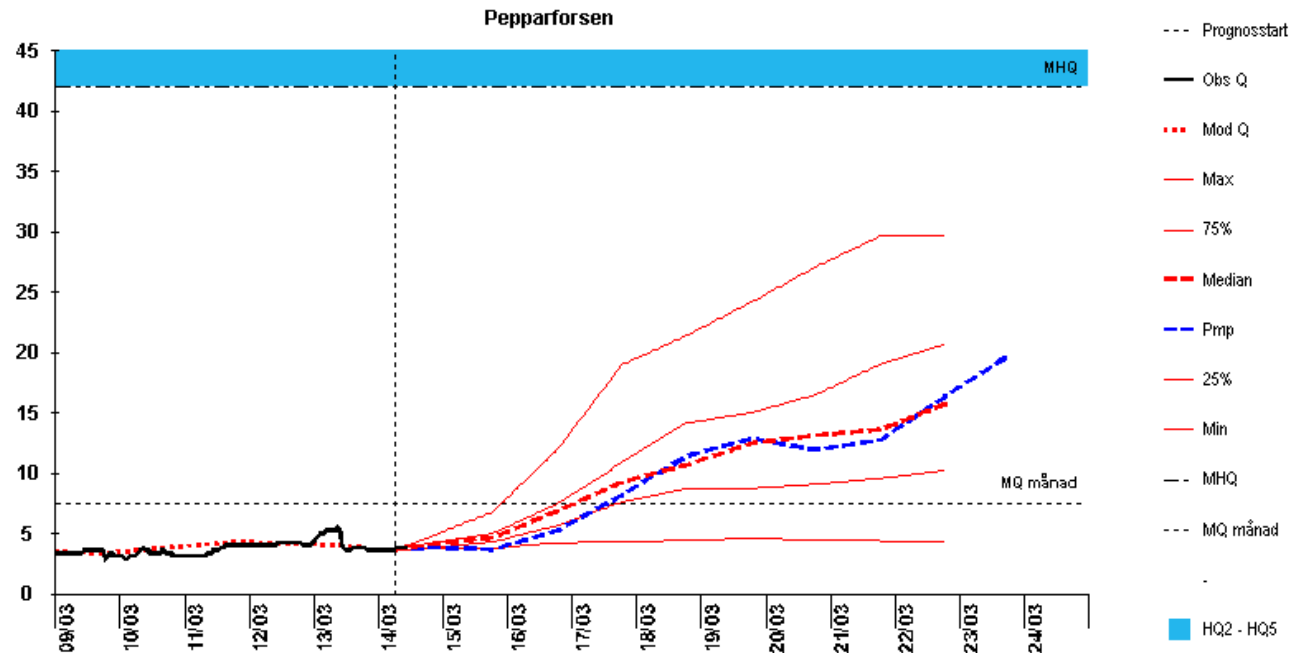
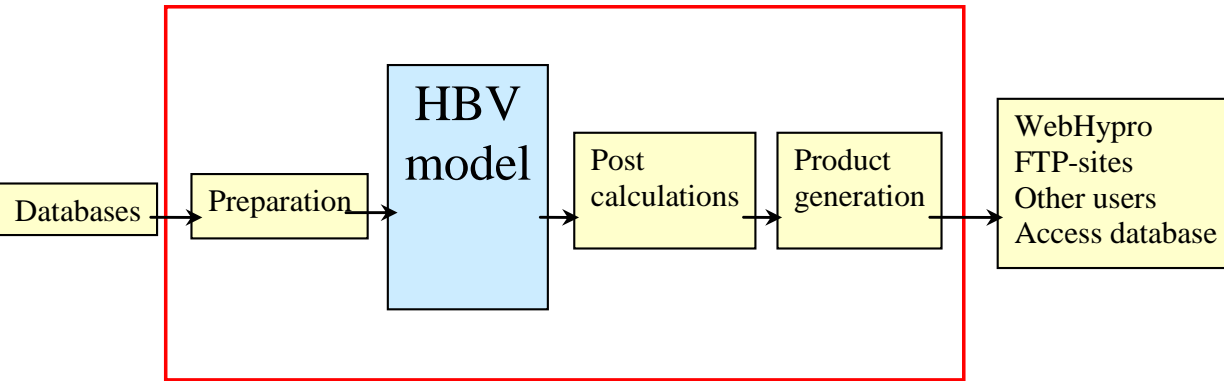
## From members to statistics



# OPERATIONAL HYDRO-EPS AT SMHI

## Production system and presentation (July 2004)

Aegir



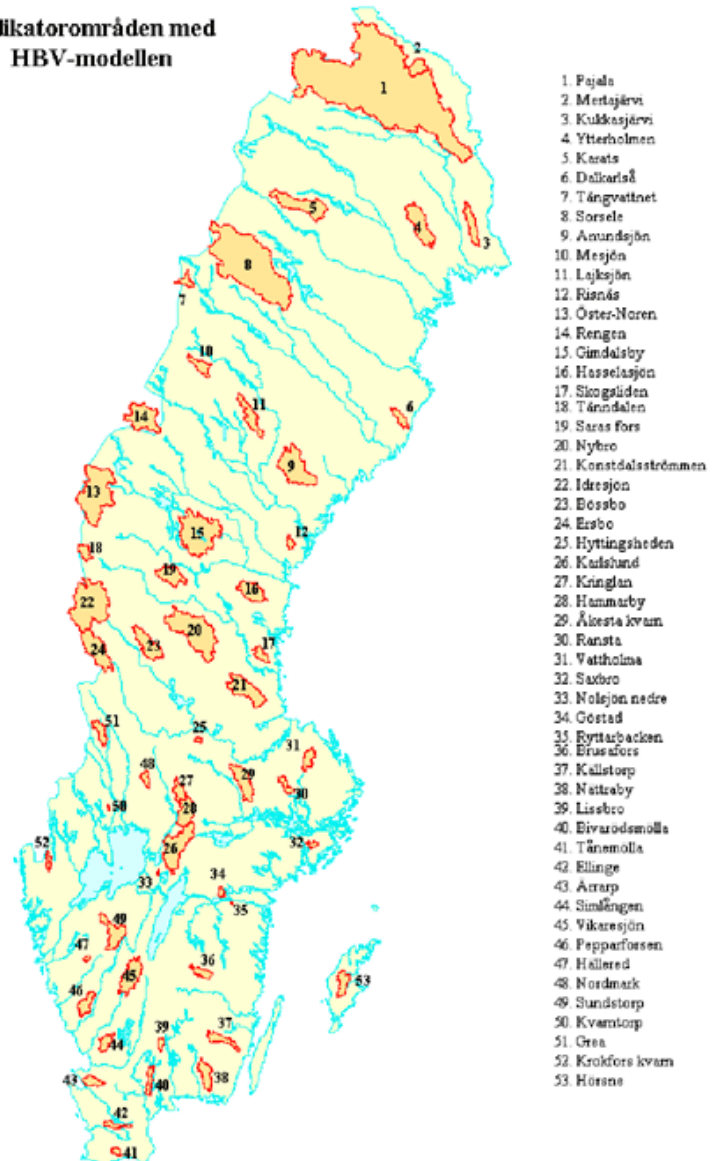


# OPERATIONAL HYDRO-EPS AT SMHI

## Indicator catchments and available data

- July 2004 – December 2005 (18 months)
- After quality screening, 45 catchments

Indikatorområden med  
HBV-modellen



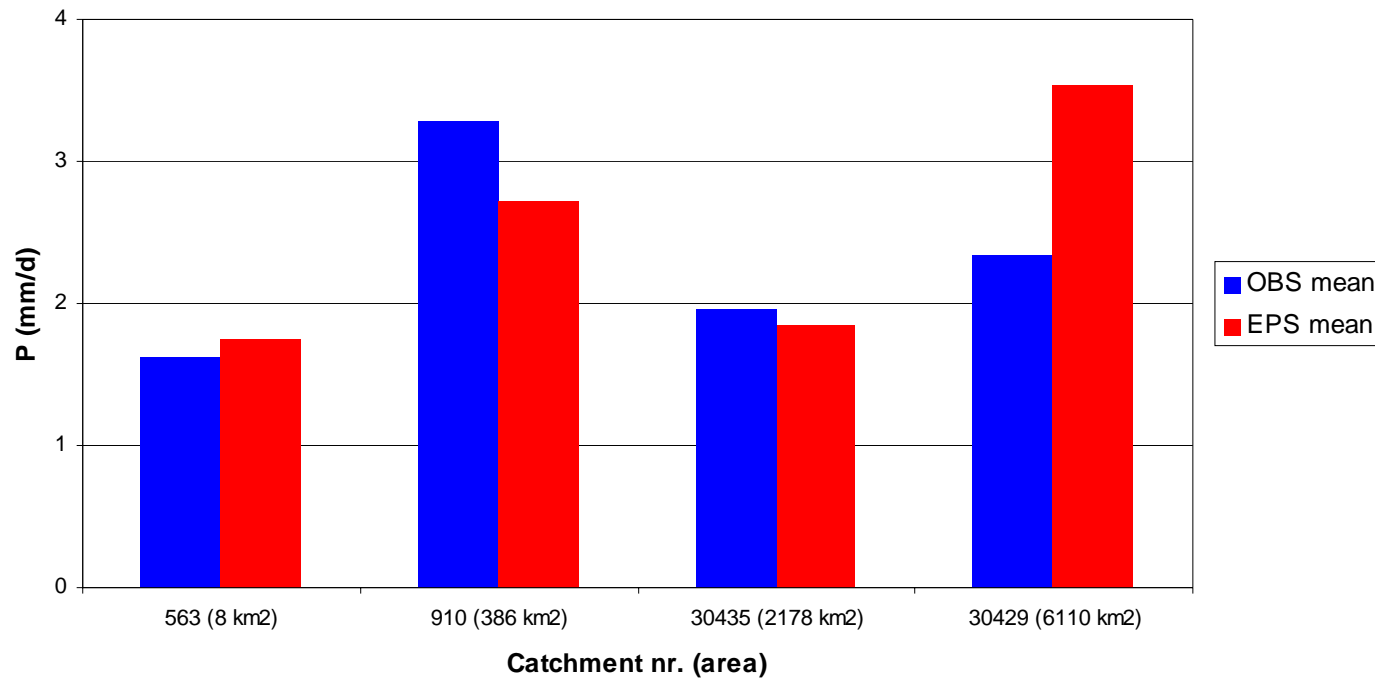
# EVALUATION

- **P, T:** general comparisons between EPS and observations (incomplete)
- **Q, deterministic:** comparison of EPS median with operational, categorical forecast (PMP) in terms of e.g. BIAS and RMSE
- **Q, probabalistic:** evaluation of EPS spread in terms of EPS percentiles and threshold exceedances

# EVALUATION

## P

- Mean of observations and EPS forecasts in four catchments



- Dry days(%):

**39.9**

**29.7**

**16.3**

**26.0**

**28.0**

**25.4**

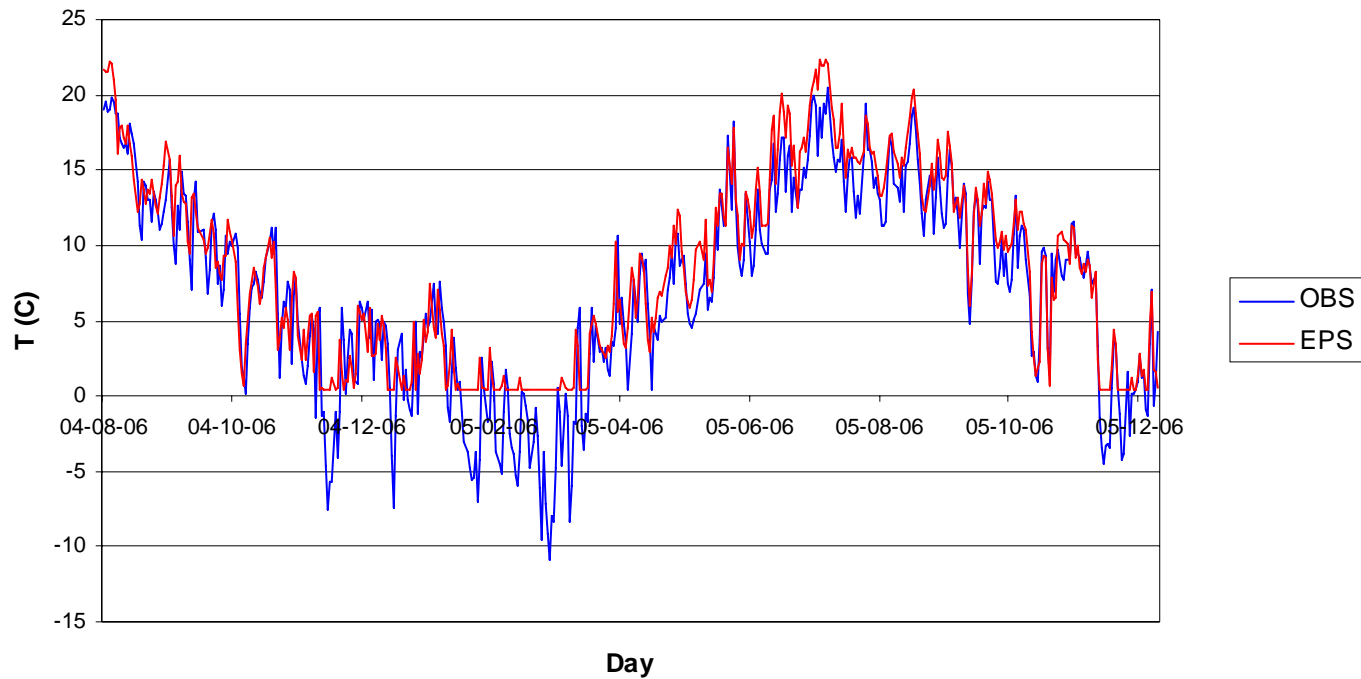
**24.3**

**10.3**

# EVALUATION

## T

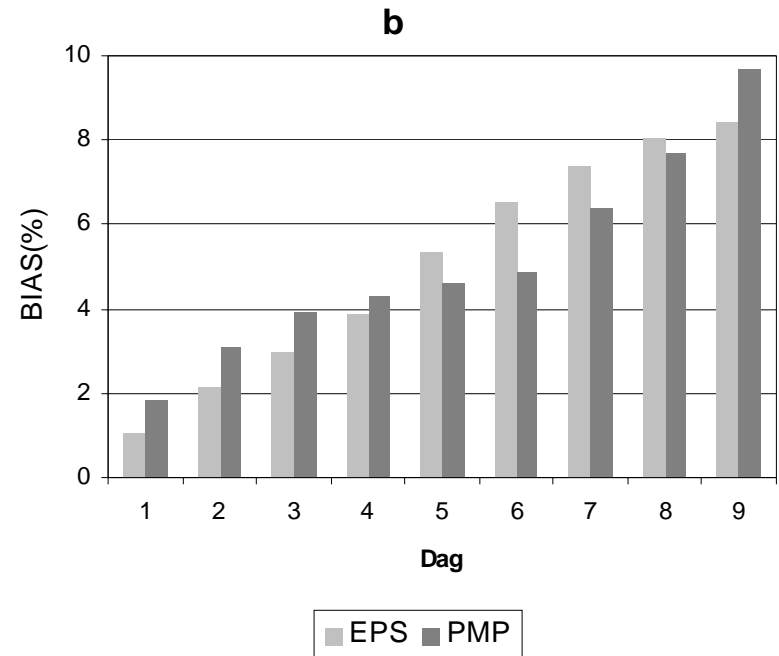
- Observation and EPS mean 1-day forecast in one catchment



# DETERMINISTIC EVALUATION OF Q

- EPS median vs. existing operational deterministic forecast (PMP)

## All catchments



# PROBABALISTIC EVALUATION OF Q Methods

- **Spread-skill:** how does the general relation between EPS spread and forecast error look?
- **Percentile-based:** how often does the observation fall between different EPS percentiles?  
Talagrand type diagram.
- **Threshold-based:** how well do estimated and observed probabilities of exceeding fixed discharge levels agree? Reliability diagram.
- **Two types of reference discharge:**
  - (1) observed Q (**OBS**) – error includes both uncertainty in the meteorological forecast and in the hydrological model
  - (2) HBV-simulated Q using a perfect meteorological forecast (**HBV**) – error includes only uncertainty in the meteorological forecast

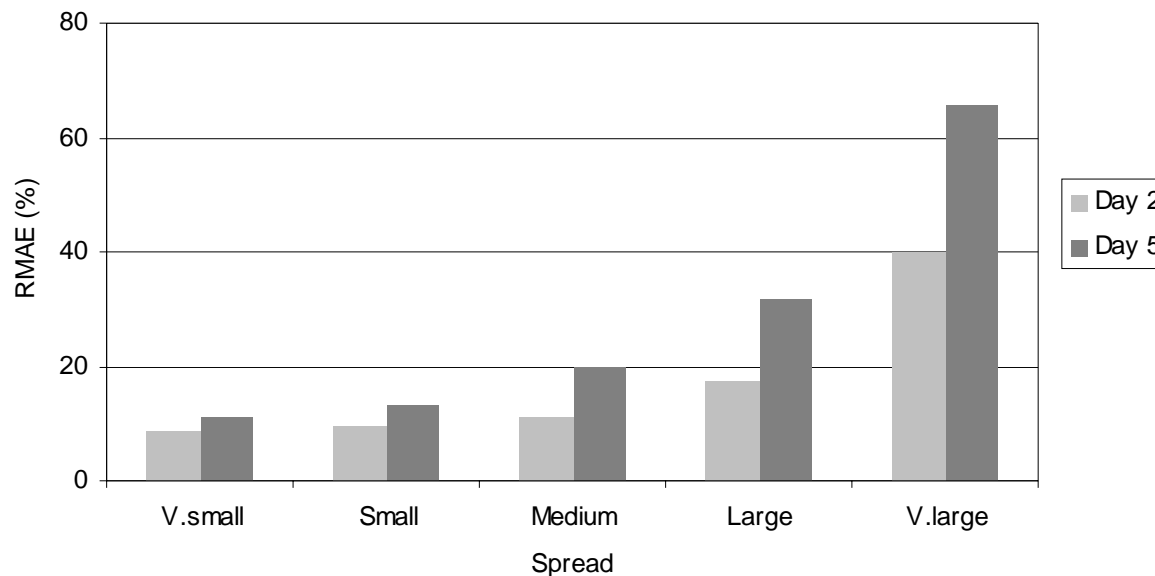
# PROBABALISTIC EVALUATION OF Q

## Spread-skill: results OBS

**Spread:** distance between EPS quartiles 25% and 75%

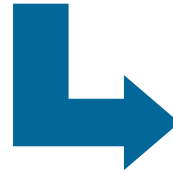
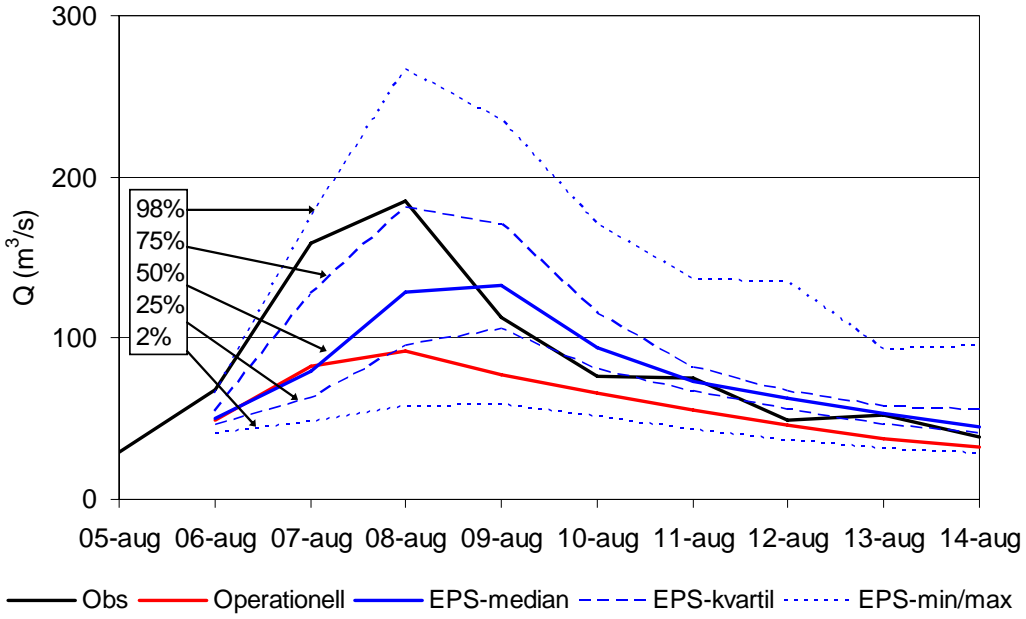
**Skill:** relative mean absolute error, with error = EPS median – observation

### All catchments

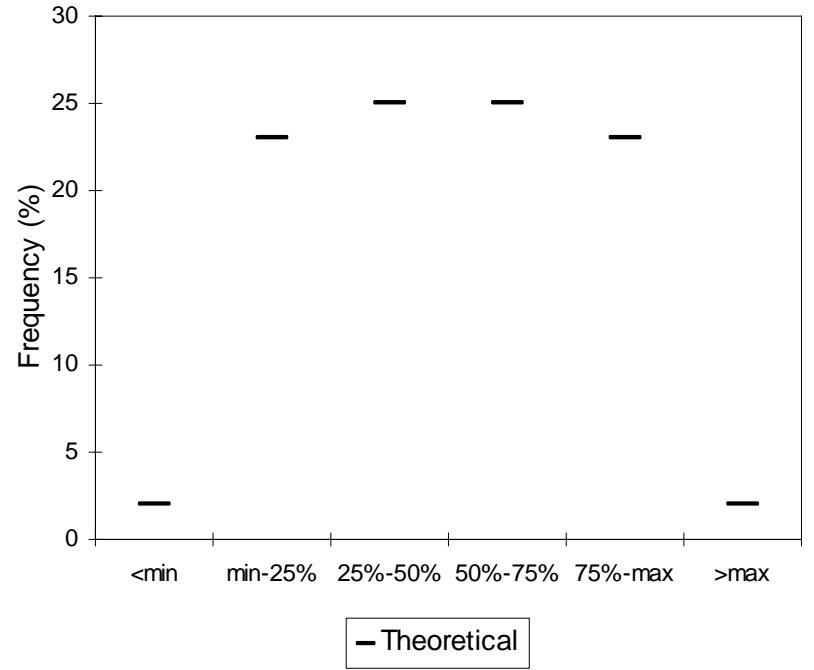


# PROBABALISTIC EVALUATION OF Q

## Percentile-based: methodology



frequency of observations  
if EPS spread is accurate





# PROBABALISTIC EVALUATION OF Q

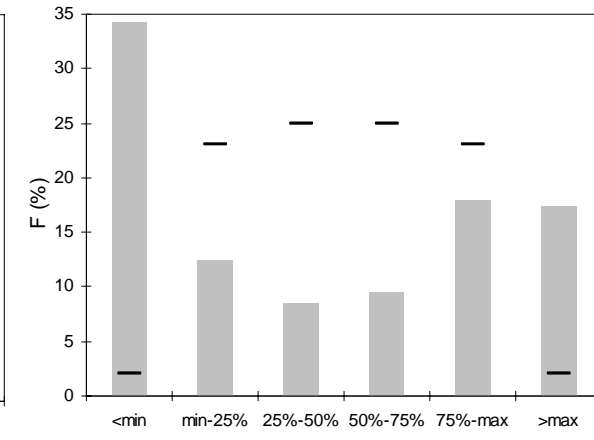
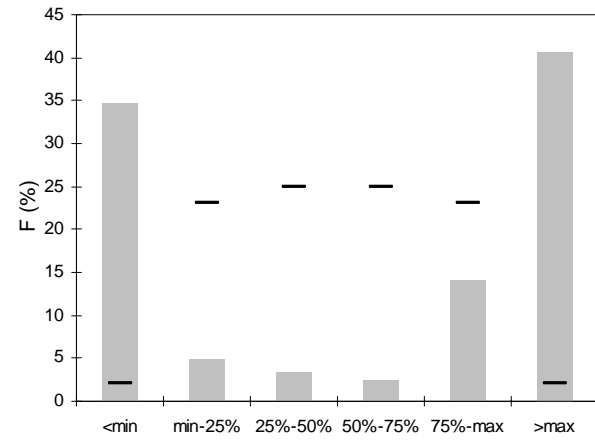
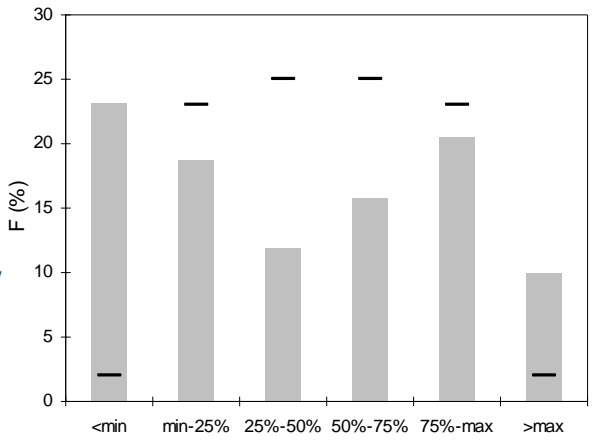
## Percentile-based: results day 2

Good catchment

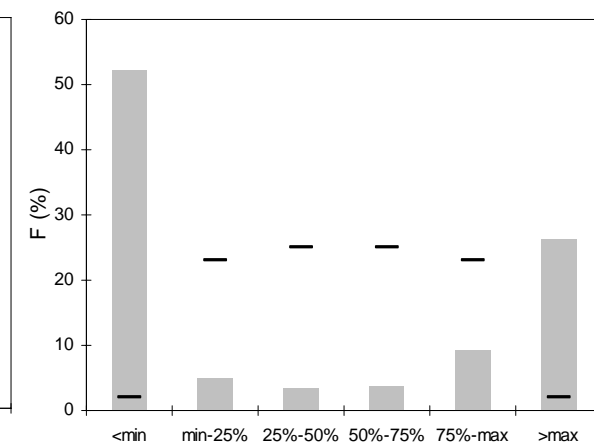
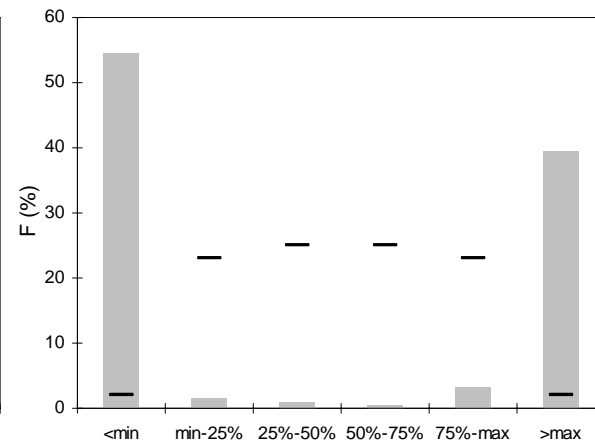
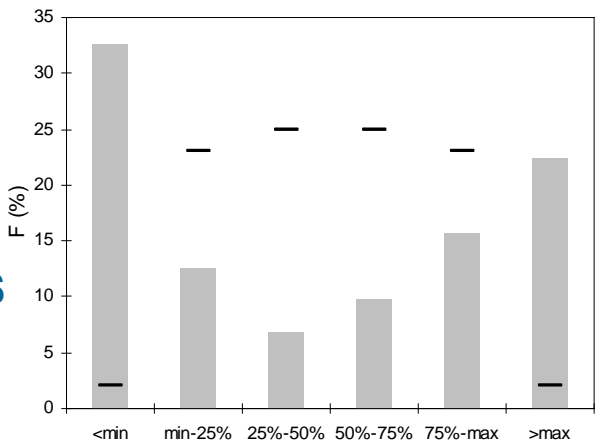
Bad catchment

All catchments

Ref.  
HBV



Ref.  
OBS



# PROBABALISTIC EVALUATION OF Q

## Percentile-based: results day 1-9

All catchments

Day 1

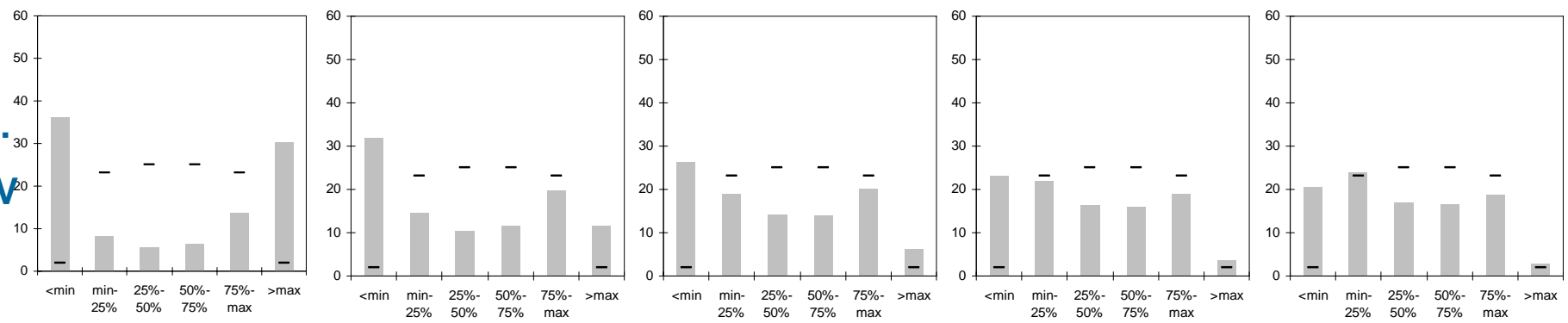
Day 3

Day 5

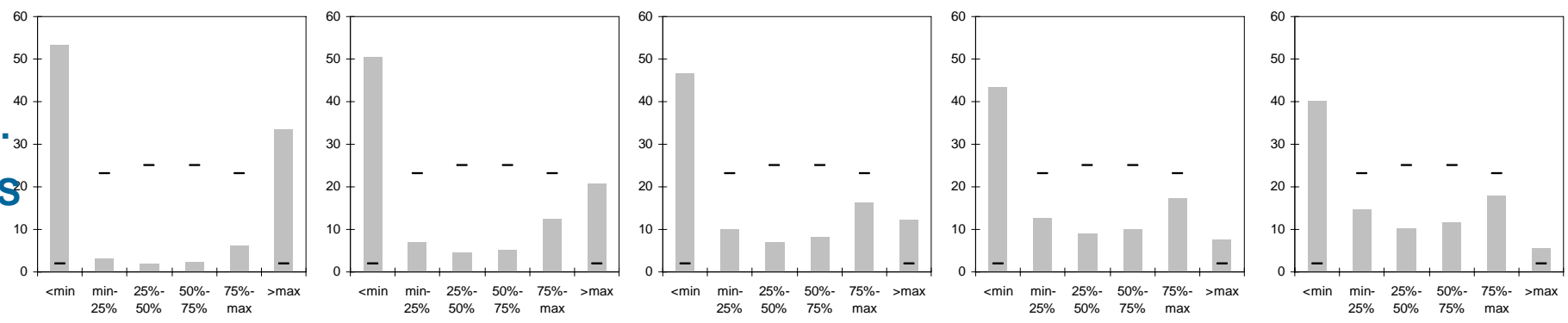
Day 7

Day 9

Ref.  
HBV



Ref.  
OBS



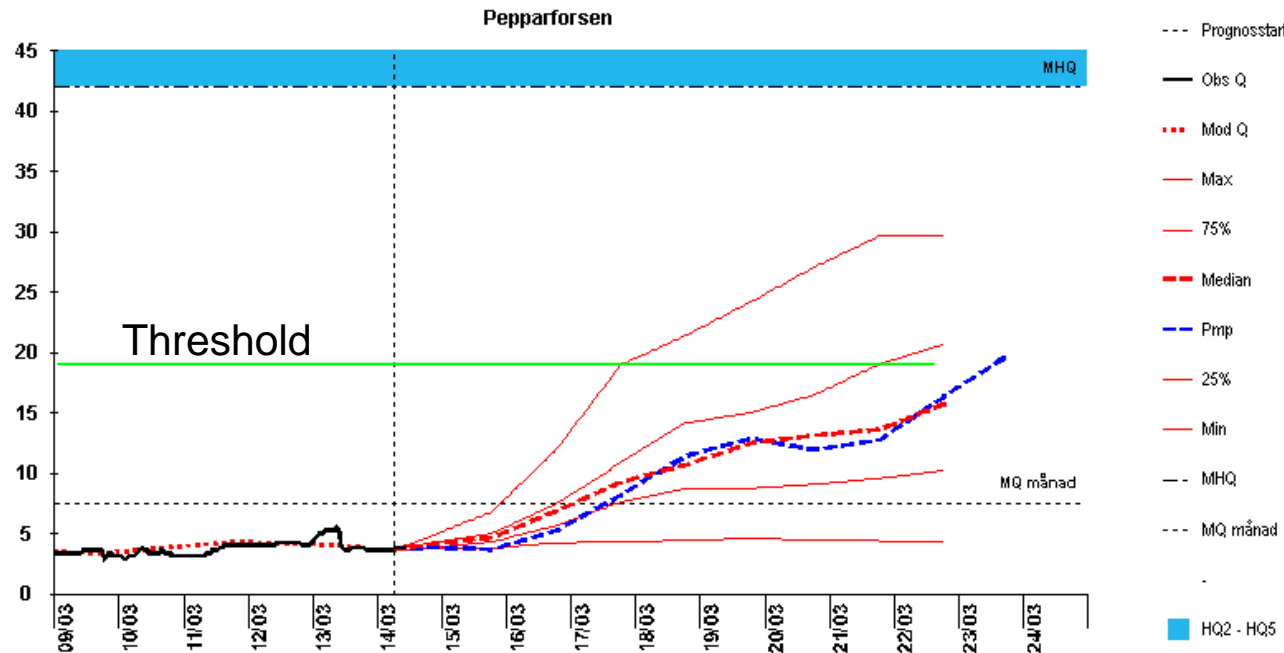
# PROBABALISTIC EVALUATION

## Threshold-based: methodology

**Discharge threshold levels:** 'high' (exceeded 30% of the evaluation period) and 'very high' (10%)

**Included:** only cases when discharge below threshold at the time of forecast

**Evaluation:** comparison of estimated exceedance probabilities and corresponding observed frequency in categorical terms (false alarms, total misses) and as reliability diagrams



# PROBABALISTIC EVALUATION

## Threshold-based: categorical - results

**False alarm (FA):** EPS min above threshold (i.e. all members) but in reality not exceedance

**Total miss (TM):** EPS max below threshold (i.e. all members) but in reality exceedance

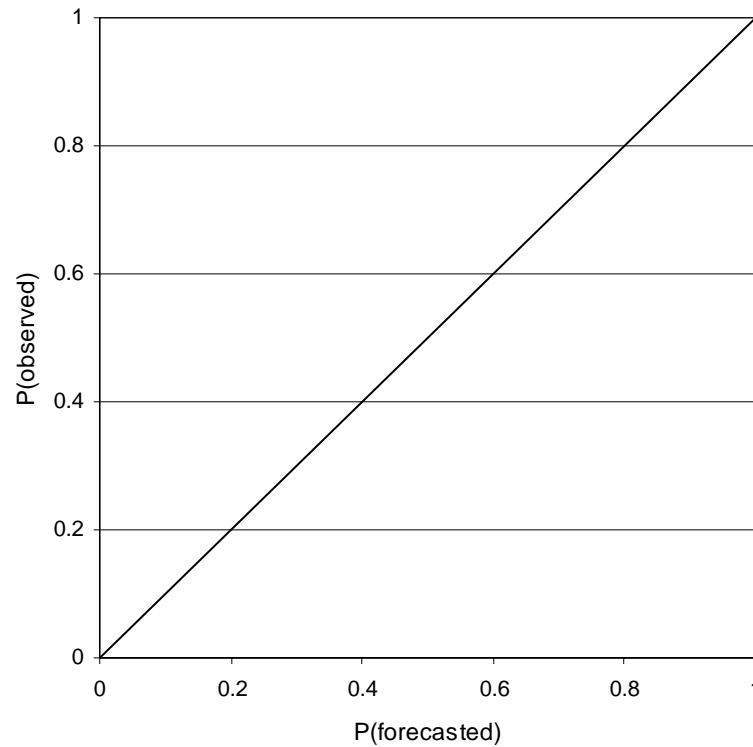
### All catchments

		Day 2		Day 5	
		FA (%)	TM (%)	FA (%)	TM (%)
High Q	HBV	22.5	0.5	25.9	0.6
	OBS	60.4	1.4	54.4	1.4
Very high Q	HBV	20.7	0.2	19.7	0.3
	OBS	50.4	0.5	40.5	0.5

# PROBABALISTIC EVALUATION

## Threshold-based: reliability - methodology

**Reliability diagram:** plot forecasted exceedance probabilities vs. corresponding observed frequencies

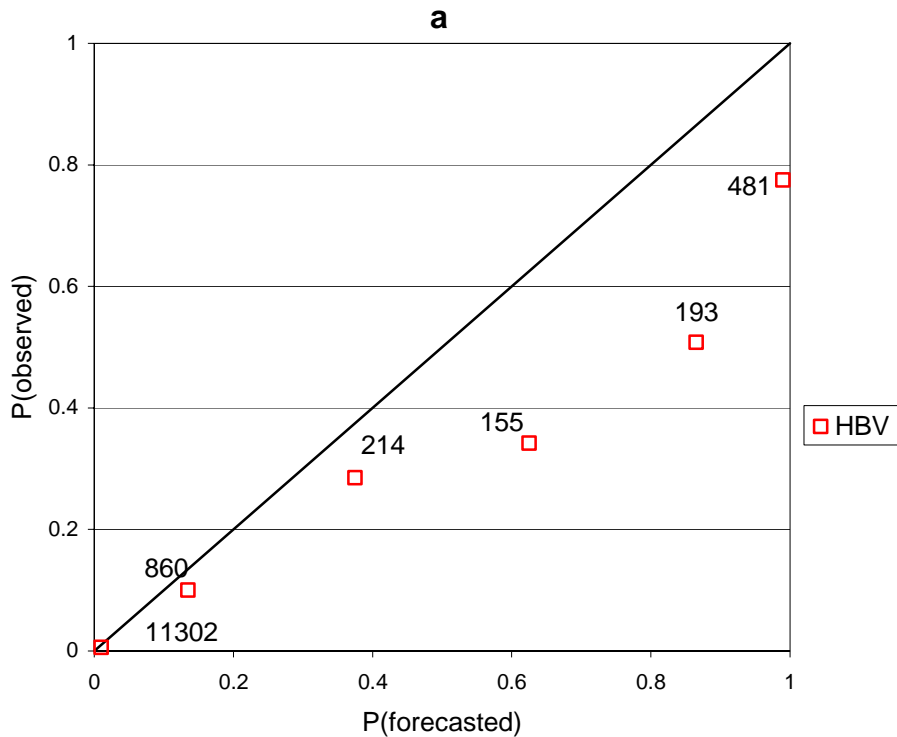


# PROBABALISTIC EVALUATION

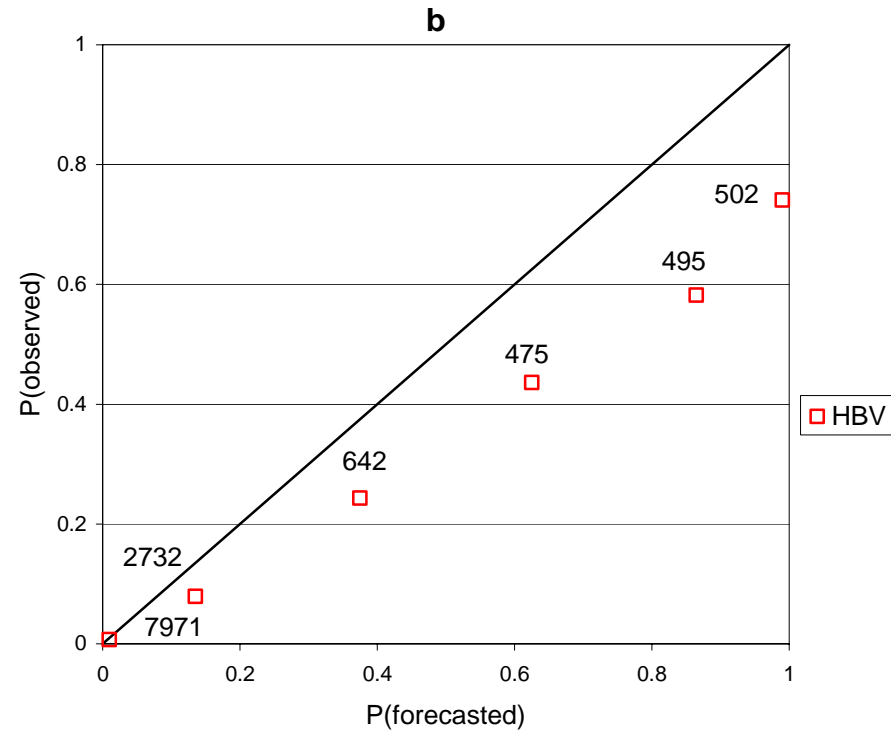
## Threshold-based: reliability – results HBV

All catchments, threshold 'high'

Day 2



Day 5



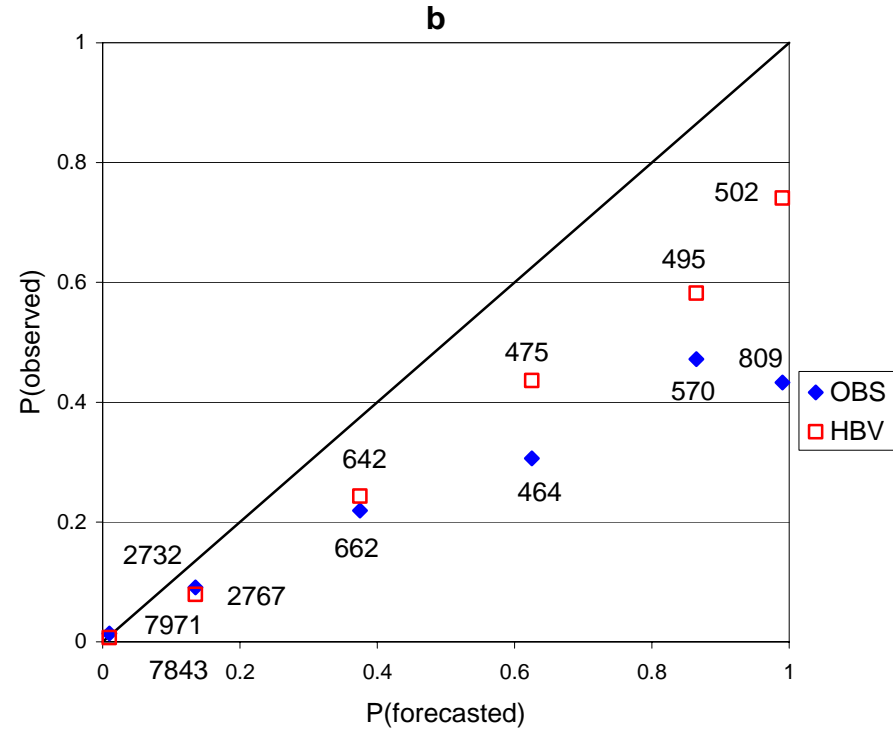
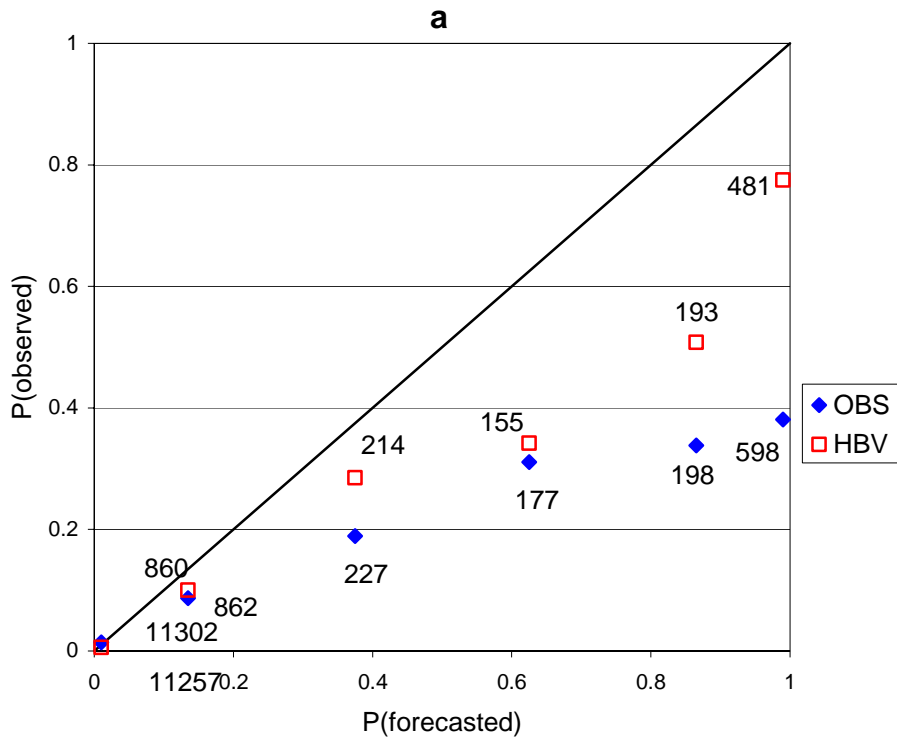
# PROBABALISTIC EVALUATION

## Threshold-based: reliability – results OBS

All catchments, threshold 'high'

Day 2

Day 5



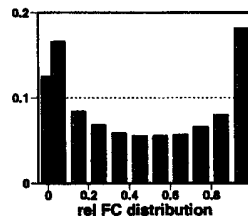
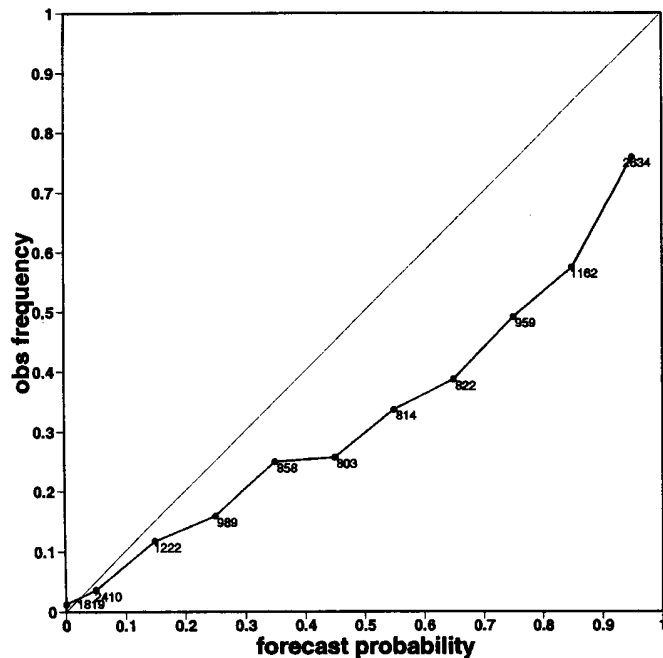
# PROBABALISTIC EVALUATION

## Threshold-based: two aspects

How good?

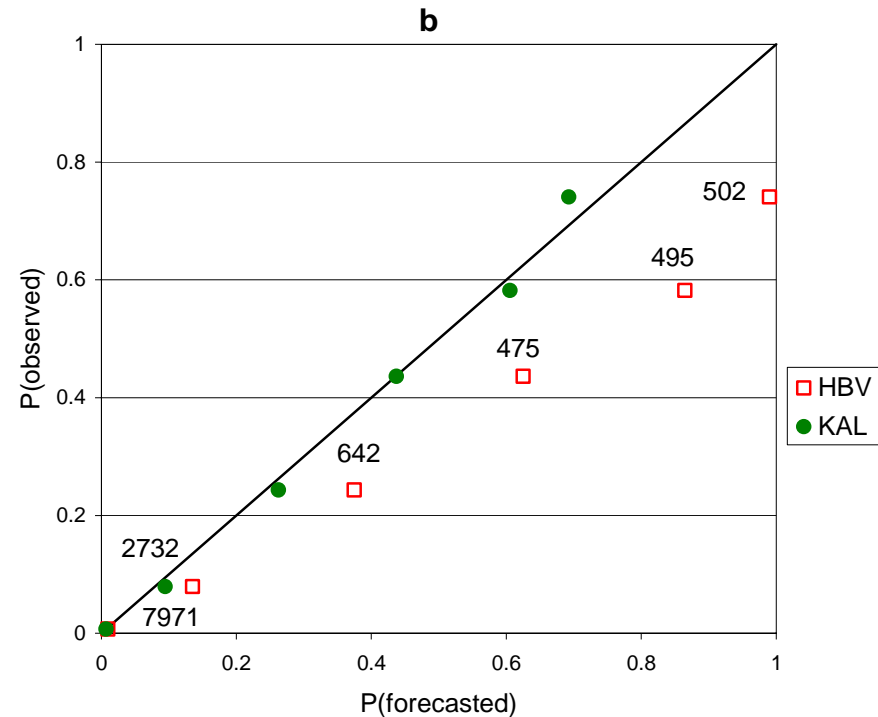
Comparison with ECMWF precipitation

Dec00-Feb01 t + 96 Europe obs 24h-precip gt 1 mm



How to improve?

Calibration (multiply by 0.7)

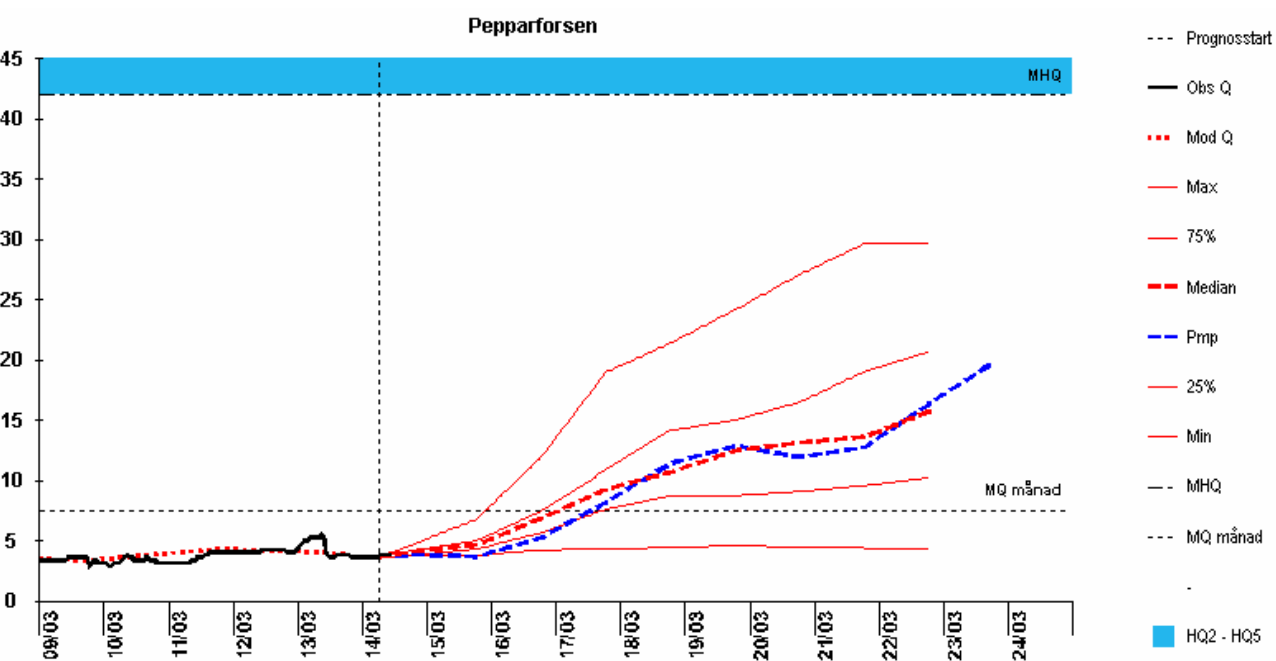




# INTERPRETATION AND PRESENTATION

## Presentation: today

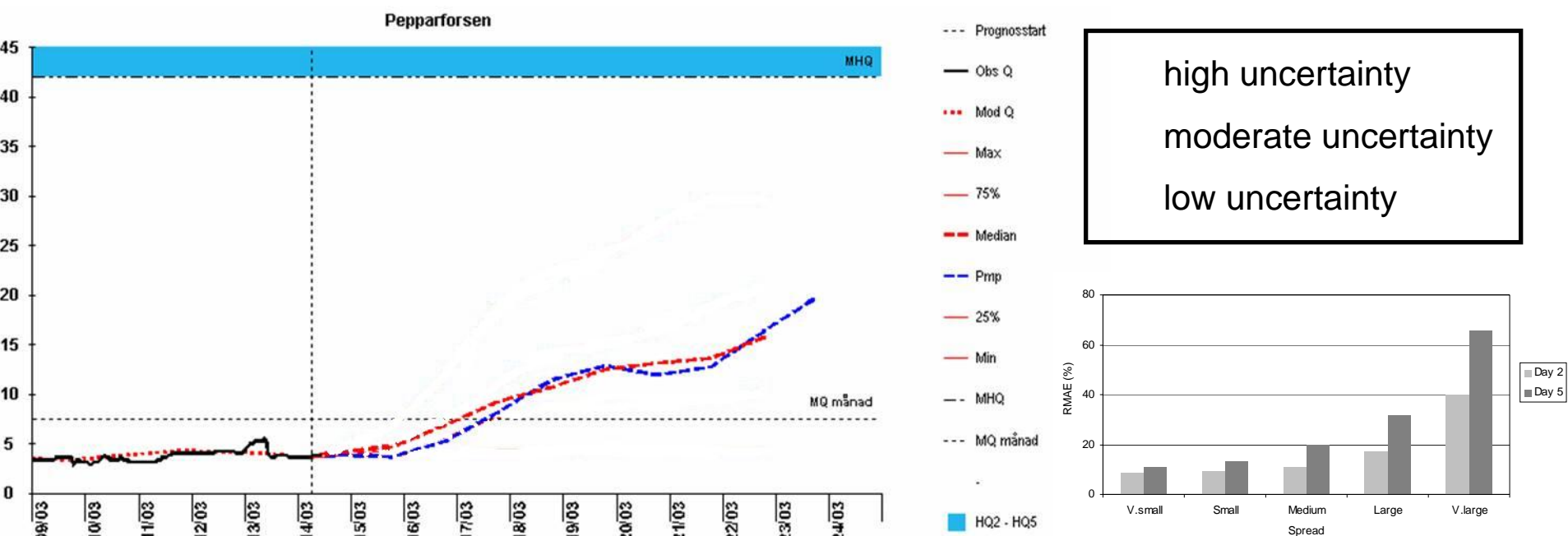
### Percentile plot



# INTERPRETATION AND PRESENTATION

## Presentation: qualitative

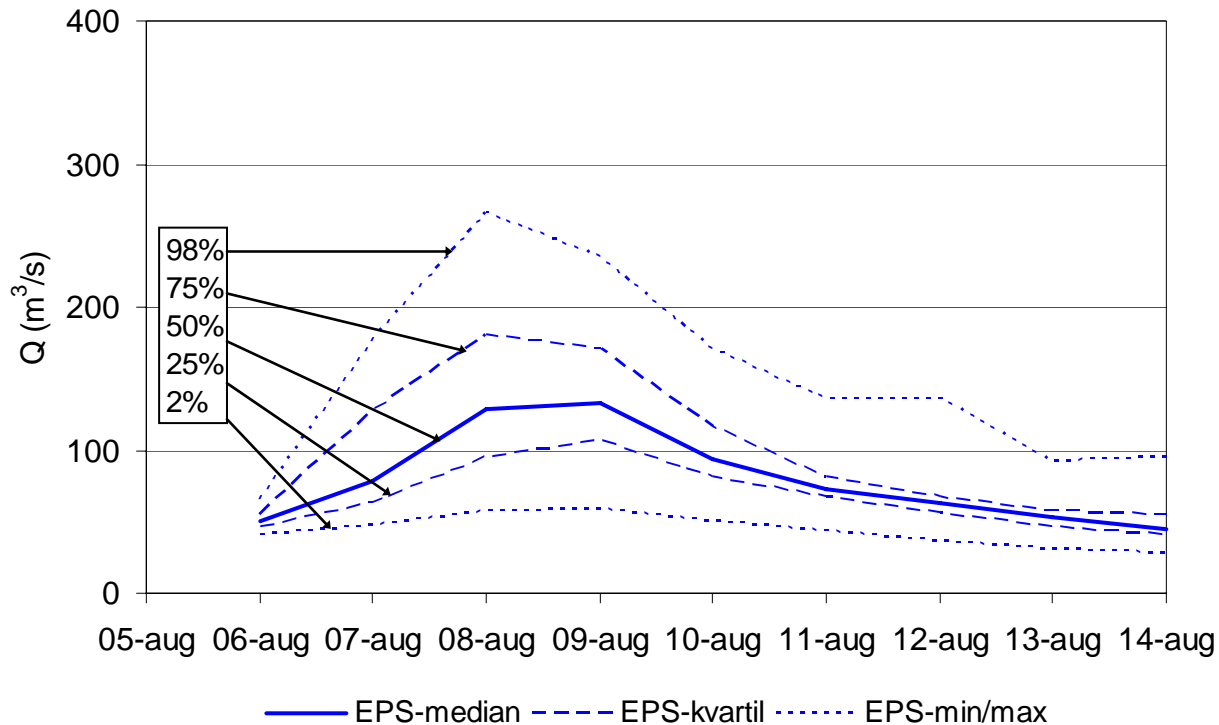
Categorical forecast (PMP or EPS median) + 'forecast uncertainty class'



# INTERPRETATION AND PRESENTATION

## Adjustment: methodology

What type of adjustment would produce an accurate EPS spread?

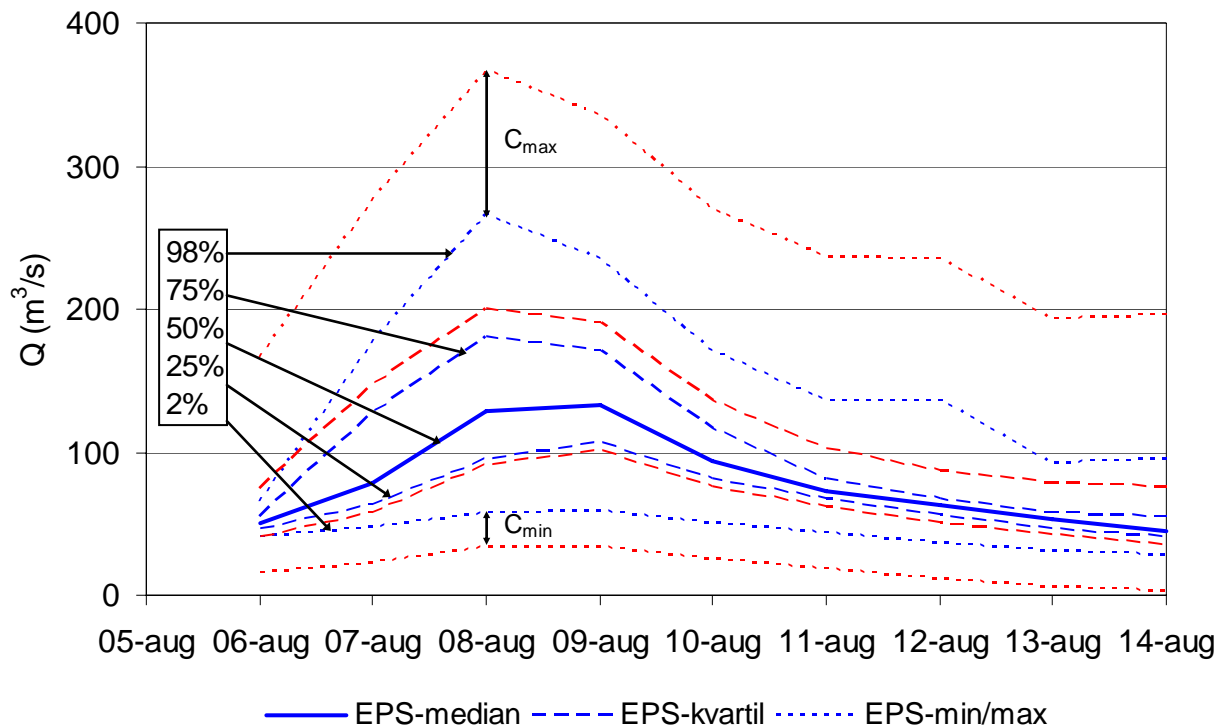


# INTERPRETATION AND PRESENTATION

## Adjustment: methodology

What type of adjustment would produce an accurate EPS spread?

Additive constants  $C_{\max}$  and  $C_{\min}$  applied on EPS max and min ( $\sim 0.1C_{\max}/C_{\min}$  on quartiles), independent of forecast day.



# INTERPRETATION AND PRESENTATION

## Adjustment: results day 1-9

### Catchment Sundstorp

Day 1

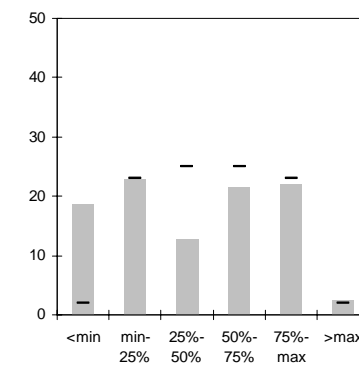
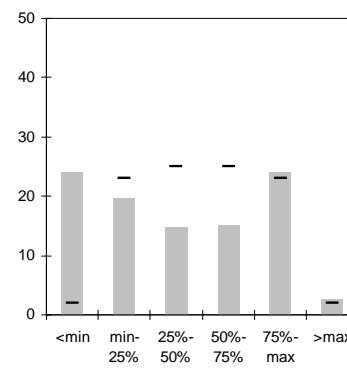
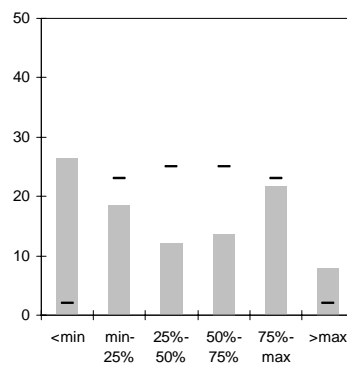
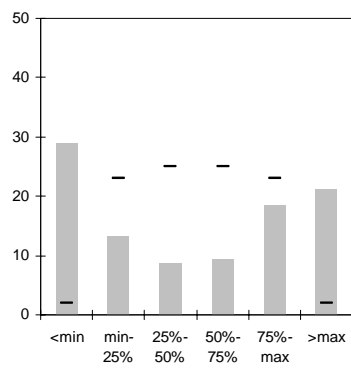
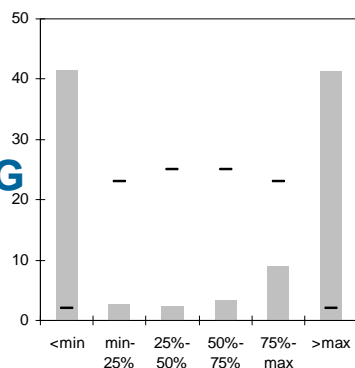
Day 3

Day 5

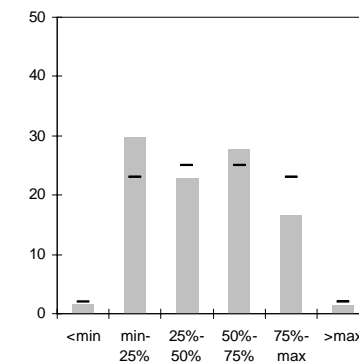
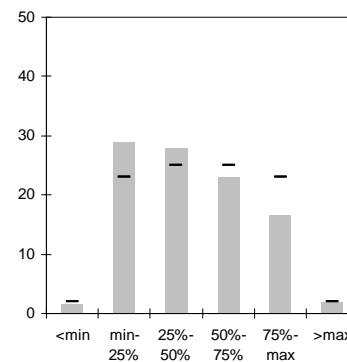
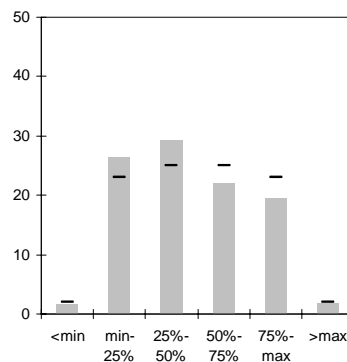
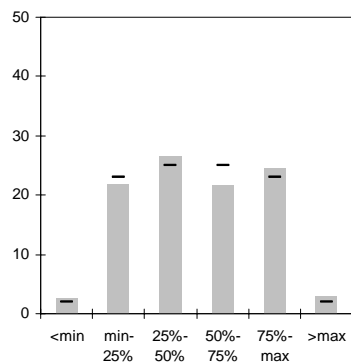
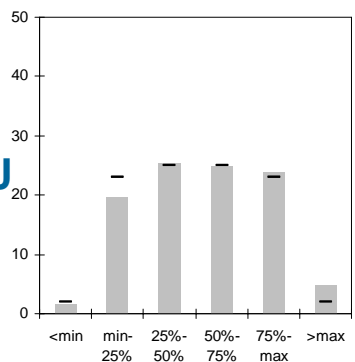
Day 7

Day 9

ORG



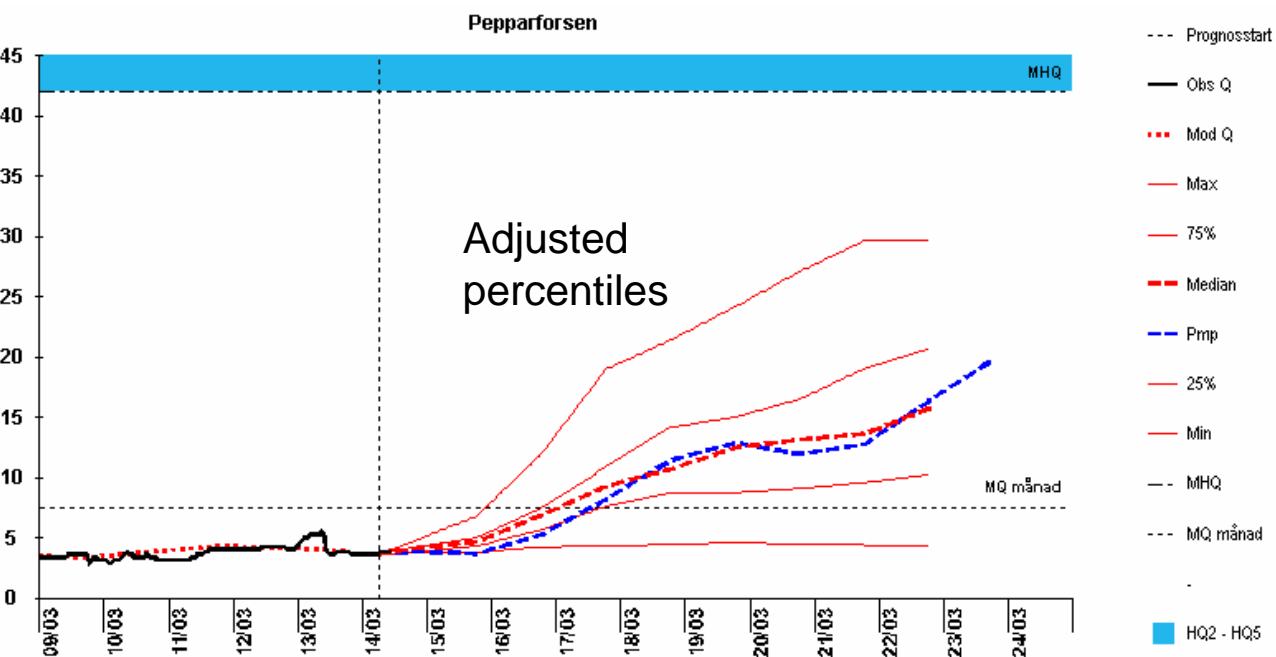
ADJ



# INTERPRETATION AND PRESENTATION

## Presentation: quantitative

EPS median and adjusted percentiles + table with (calibrated) exceedance probabilities (or probability classes) of warning levels



high probability			
moderate probability			
low probability			
	HQ2	HQ10	HQ25
15/03	Green	Green	Green
16/03	Yellow	Green	Green
17/03	Red	Yellow	Green
...	...	...	...

# SUMMARY

**Deterministic evaluation:** EPS median equal to or slightly better than the PMP forecast

**Probabilistic evaluation:**

- qualitative use of EPS by e.g. 'forecast uncertainty classes' is possible
- quantitative use requires adjustment of EPS spread
- constant shift of EPS percentiles can adjust the spread

**But...** technical problems and many remaining open issues