Comparison of point forecasts and interval forecasts in hydrologic applications

Cees Diks[1] Jasper Vrugt[2]

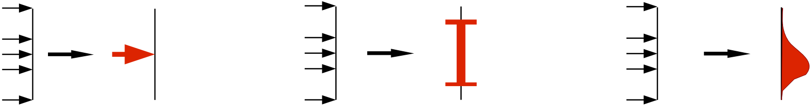
[1]University of Amsterdam

[2]University of California, Irvine

HEPEX Workshop, June 9, 2011

Evaluation of combined hydrologic forecasts

Model averaging: constructing combined hydrologic forecasts from an ensemble of point forecasts



point forecasts   interval forecasts   density forecasts

Preliminary results:

- *Point forecasts*: linear regression performs better than more sophisticated model averaging methods
- *Interval forecasts*: quantile regression does not always perform better than simpler methods
- *Density forecasts*: future research . . .

Model averaging: point forecasts

Y_t, $t = 1, \ldots, T$: hydrological variable of interest
(streamflow, pressure head, ...)

$X_{k,t}$, $k = 1, \ldots, K$: ensemble of point forecasts for Y_t

Combined point forecasts

$$\hat{Y}_t^{\boldsymbol{\beta}} = \boldsymbol{X}_t^{\mathsf{T}} \boldsymbol{\beta} = \sum_{k=1}^{K} \beta_k X_{k,t}$$

Can be associated with forecasts of a linear model

$$Y_t = \boldsymbol{X}_t^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_t = \sum_{k=1}^{K} \beta_k X_{k,t} + \varepsilon_t$$

or finite mixture model, etc...

Choices for β compared (Diks & Vrugt, 2010, SERRA)

- Equal weights averaging: $\hat{\boldsymbol{\beta}}_{\mathrm{EWA}} = \left(\frac{1}{k}, \ldots, \ldots, \frac{1}{k}\right)$
- Bates-Granger averaging: $\hat{\beta}_{\mathrm{BGA},k} = \frac{1/\hat{\sigma}_k^2}{\sum_{j=1}^{K} 1/\hat{\sigma}_j^2}$
- AIC/BIC averaging: $\hat{\beta}_k = \frac{\exp(-I_k/2)}{\sum_{j=1}^{K} \exp\left(-I_j/2\right)}$
- Granger-Ramanathan (GRA) averaging (OLS regression weights) $(*)$:

$$\hat{\boldsymbol{\beta}}_{\mathrm{GRA}} = \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}} Y$$

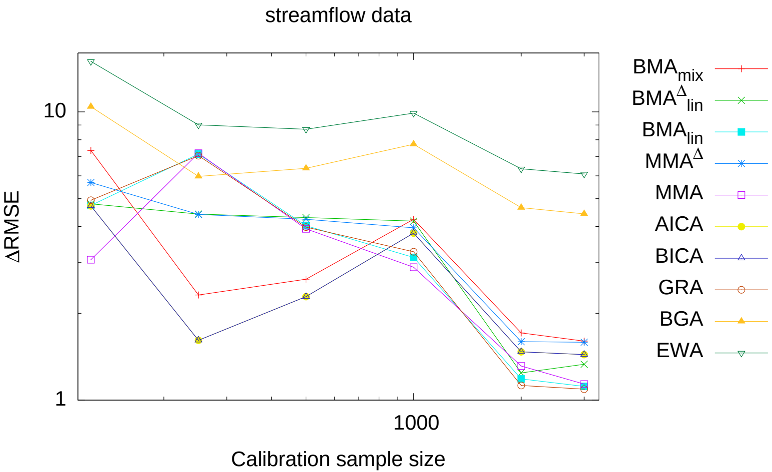
- Bayesian model averaging (BMA) weights
  - BMA in the finite mixture model (Raftery *et al.*, 2005)
  - BMA in the linear regression model (Raftery *et al.*, 1997) $(*)$
- Mallows model averaging (MMA) weights (sum of squared prediction errors + penalty for model complexity) $(*)$

Point forecast accuracy

Diks and Vrugt (2010) examined the RMSPE of point forecasts in two case studies

- Case 1:
  - Daily streamflow through a 1950 km^2 watershed
  - Leaf River, Mississippi, USA
  - 36-year historical record + ensemble forecasts
  - Ensemble members: 8 conceptual hydrologic models (Vrugt & Robinson, 2007)
  - Fixed evaluation period 1961–1988 (10,500 obs.)
  - Varying calibration sample size (up to 3,000 obs.)
- Case 2:
  - Tensiometric pressure head in 5 m deep layered vadose zone
  - Vadose zone of volcanic origin, New Zealand
  - 9,070 hourly observations + ensemble forecasts
  - Ensemble members: 7 soil-hydraulic models (Vrugt & Robinson, 2007)

Streamflow excess prediction error ΔRMSE



Main results (point forecasts)

- Methods with weights restricted to the simplex ($\beta_k > 0$, $\sum_{k=1}^{K} \beta_k = 1$) perform worse than unrestricted methods
- Three best performing methods: GRA weights (linear regression β), BMA in linear regression model and MMA weights
- These best methods are asymptotically equivalent, but GRA weights by far simplest to calculate

$\Rightarrow$ use GRA weights

(For further details, see Diks and Vrugt (2010), SERRA)

Note: not necessary to apply a *bias correction* to the ensemble members. Just add a constant forecast to $\boldsymbol{X}$

Interval forecasts

Aim: compare interval forecasts from

- linear regression model
- quantile regression

Case:

- HEPEX streamflow data + ensemble forecasts
- corresponding ensemble of 8 forecasts + constant

Streamflow data are log-transformed

removes most of the asymmetry in the distribution

Interval forecasts from linear regression model

$$Y_t = \boldsymbol{X}_t^\mathsf{T}\boldsymbol{\beta} + \varepsilon_t = \sum_{k=1}^{K} \beta_k X_{k,t} + \varepsilon_t$$

$\boldsymbol{\beta}$ estimated by OLS

$$\hat{\boldsymbol{\beta}}_{\mathrm{GRA}} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{n} \left(\boldsymbol{X}_t^\mathsf{T}\boldsymbol{\beta} - Y_t\right)^2 = \left(X^\mathsf{T}X\right)^{-1} X^\mathsf{T} Y$$

Assume (non-)parametric distribution on ε_t (e.g. $N(0, \sigma^2)$)

$\Rightarrow$ 95% interval forecasts (equal tails)

$$\left(\boldsymbol{X}_t^\mathsf{T}\boldsymbol{\beta} - 1.96\hat{\sigma}, \boldsymbol{X}_t^\mathsf{T}\boldsymbol{\beta} + 1.96\hat{\sigma}\right)$$

Interval forecasts from quantile regression

Conditional quantile $Q_p(\boldsymbol{x})$ ($p \in (0,1)$) defined through

$$P(Y_t \leq Q_p(\boldsymbol{x}) | \boldsymbol{X}_t = \boldsymbol{x}) = p$$

Linear quantile regression model:

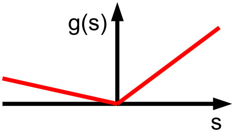
$$Q_p(\boldsymbol{X}_t) = \boldsymbol{X}_t^\mathsf{T} \gamma_p + \varepsilon_t = \sum_{k=1}^{K} \gamma_{p,k} X_{k,t} + \varepsilon_t$$

Given $p \in (0,1)$, γ_p is estimated as

$$\hat{\gamma}_p = \underset{\gamma_p}{\arg\min} \sum_{t=1}^{n} g\left(\boldsymbol{X}_t^\mathsf{T} \gamma_p - Y_t\right)$$

where $g(\cdot)$ is the so-called *tick loss function*

$$g(s) = -(1-p)\, s\, \mathrm{I}(s < 0) + p\, s\, \mathrm{I}(s \geq 0)$$



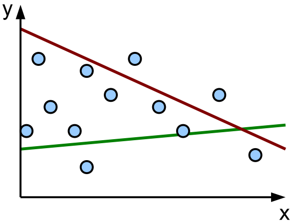
Relative advantages/disadvantages

Linear regression

- more parsimonious $\Rightarrow$ potentially worse fit but few parameters to estimate
- fixed conditional distribution (apart from location)
- hence fixed predictive interval lenghts

Quantile regression

- more flexible $\Rightarrow$ potentially better fits but more parameters to estimate
- allows for local changes in distribution
- hence intervals can change length (heteroskedasticity)
- but beware: estimated quantiles may cross!!!



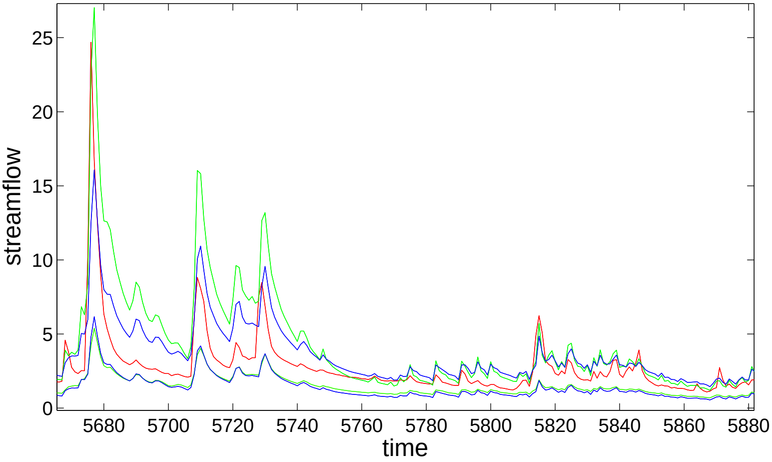
First results (interval forecasts)

- Data set used: 03451500
- 8 ensemble members (9 including constant) non-calibrated
- daily streamflow data, period 1961 to 1997 (13514 days)
- Calibration sample size 6757
- Evaluation sample size 6757

Evaluation criteria: coverage probability (CP) + average predictive interval (PI) length

	nominal CP	CP	CP(lower)	CP(upper)	PI length
GRA	0.950	0.954	0.985	0.969	0.957
QR	0.950	0.902	0.948	0.955	0.890
QR	0.975	0.957	0.958	0.979	1.050

Streamflow, observed + interval forecasts



Conclusions (preliminary)

- Linear regression model provides good point forecasts
- Interval forecasts from linear regression model have correct coverage (after log-transforming streamflow data)
- Interval forecasts from quantile regression are slightly more narrow, but have worse coverage
- Adjusting the QR intervals such that the coverage is correct leads to intervals that are wider than the linear regression intervals
- Overall, the linear regression model performed best

Future research

- Extend quantile regression results
- Local quantile estimation (kernel methods)
- Construct density forecasts
- Density forecast evaluation
- Focus on specific regions of interest (e.g. upper part of ditribution for flood warnings)