EGU General Assembly 2021

NVE

DESIGN FLOOD ESTIMATION AT LOCATIONS WITH NO DATA OR SHORT RECORDS IN A BAYESIAN FRAMEWORK

Kolbjørn Engeland, Trond Reitan, Seija Maria Stenius, and Per Glad
The Norwegian Water Resources and Energy Directorate, Oslo, Norway

Photo: Grethe Holm Midttømme, NVE

Background

- New guidelines for flood frequency analysis in Norway are needed.
  - More data available since the previous model was published in 1997
  - Newer and more flexible methods could be used (i.e. Bayesian approaches)
- Operational guidelines need to be
  - simple and robust models
  - make the best possible use of available data

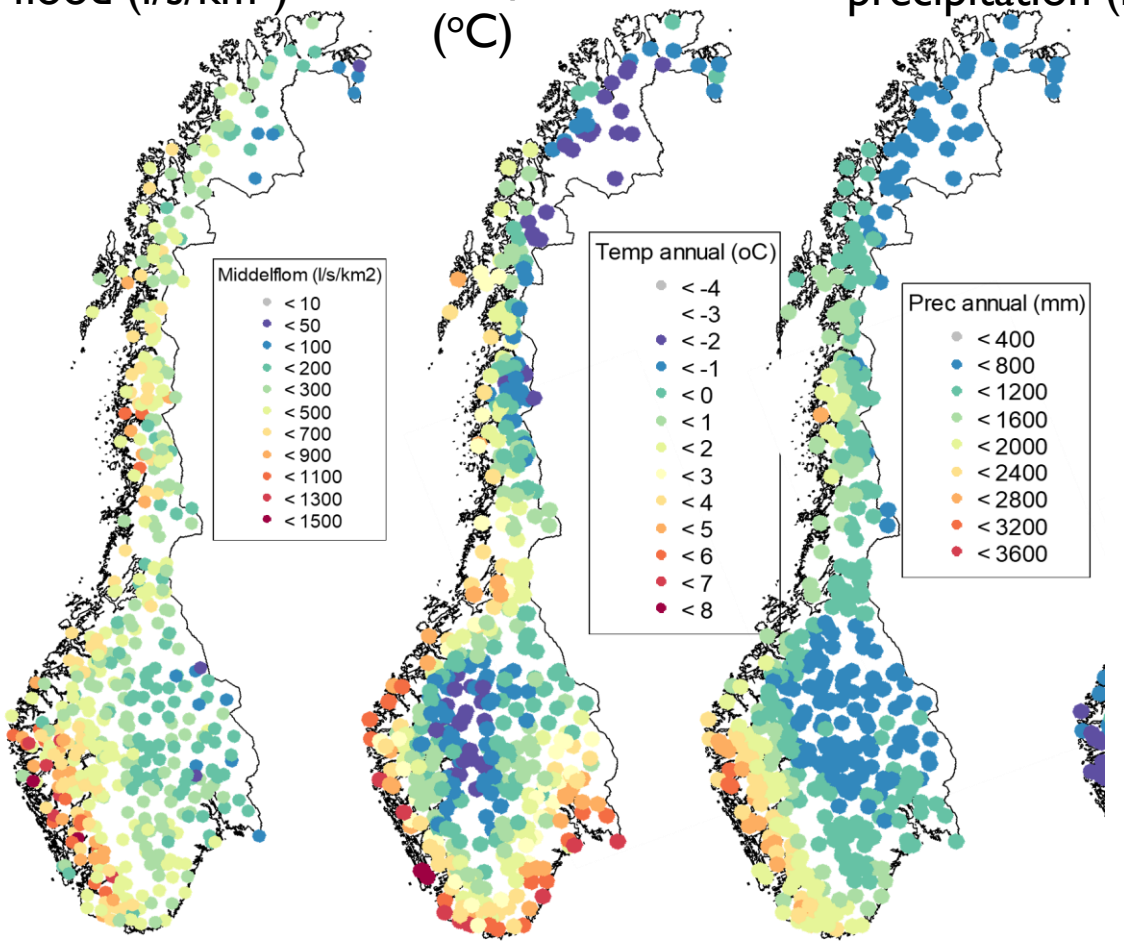
Photo: Arne T. Hamarsland, NVE

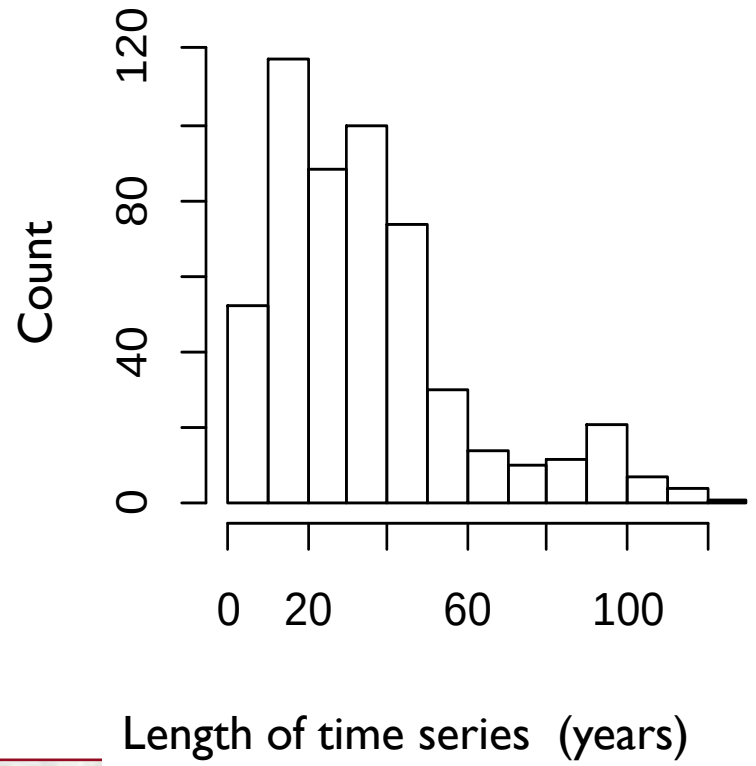
Aims

- Develop a regional flood frequency model that consists of a regression model for the index flood and a growth curve model.
- Assess and attribute the uncertainty to the components of the regional flood frequency model.
- Develop flexible approaches for combining a regional flood frequency model with local data and provide recommendations for how to combine local and regional data.

Photo: Øystein Nøtsund, NVE

Data – 528 streamflow gauging station





Methods – re-parameterization of GEV distribution

- Original quantile function: $q(T) = \mu - \frac{\sigma}{\xi}\left[1 - \left(ln\left(\frac{T}{T-1}\right)\right)^{-\xi}\right]$
- Use median flood *M* as a parameter
- New scale parameter: $\alpha = ln\left(\frac{\sigma}{M}\right)$
- Shape parameter: $\xi(k) = -\frac{1}{2} + \frac{1}{1+e^{-k}}$
  - Will always be between -0.5 and 0.5
- New quantile function: $q(T) = M\left[1 + e^{\alpha}\frac{\log\left(\frac{T}{T-1}\right)^{-\xi(k)} - \log(2)^{-\xi(k)}}{\xi(k)}\right]$
- We can now model index flood *M* and the growth curve parameters α and k independently

Regression model for index flood

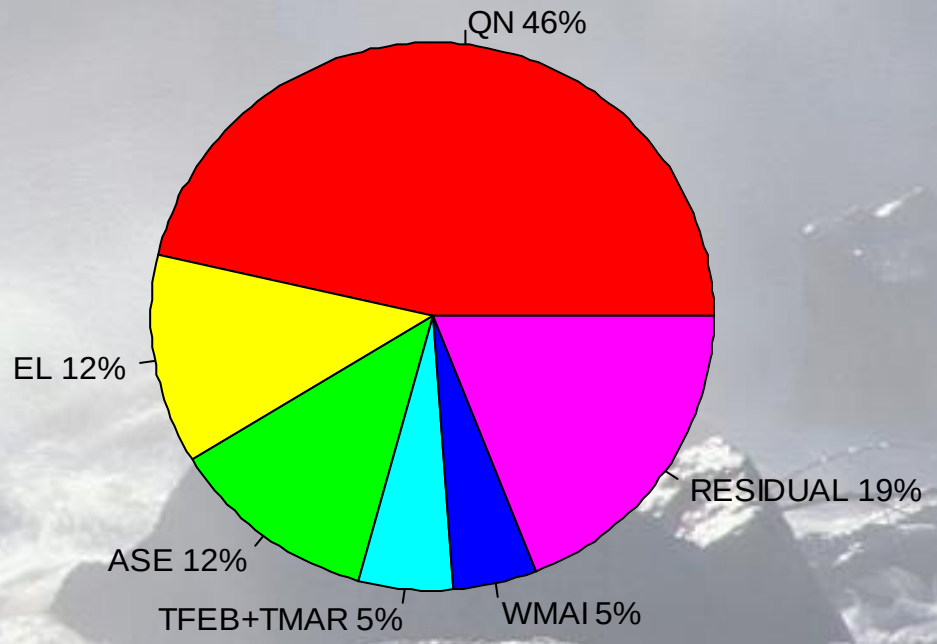
- Dependent variable: log-transformed index flood
- Independent variables: Transformed or untransformed catchment properties
- Bayesian regression
  - Data split into training set, validation set and test set
  - Stepvise procedure
  - Evaluated different penalties for model complexity by setting different priors
- Selected the simplest model where the RMSE was within 10% of the best model for the validation set
- Final evaluation for the test set

Results – index flood model

$$q_{ind} = \exp\left[\begin{array}{c} 4.196 + 0.473 * \sqrt[3]{Q_N} - 0.0632 * \sqrt[2]{E_L} - 0.0520 * A_{SE} \\ -0.00751 * T_{Feb}{}^{2} - 0.000942 * T_{Mar}{}^{3} + 0.0376 * \sqrt{W_{Mai}} \end{array}\right]$$

Uncertainty: */ 1.72

Q_N (l s^{-1} km^{-2}) is the mean annual discharge,
E_L (km) is the length of the main river,
A_{SE} (%) is the effective percentage of the catchment covered by lakes,
T_{Feb}(°C) is the mean February temperature,
T_{Mar}(°C) is the mean March temperature,
W_{May} (mm) is the mean May rain and snow melt



Growth curve parameters

- Regression-like equations for the scale and shape parameters.
- Median flood estimated from the local flood data.
- Bayesian model selection of the explanatory variables applied to the re-parameterize GEV model.
- We selected models using a threshold of 50%, 75% and 90% usage ratios in MCMC sampling, and evaluated them using the Bayesian model likelihood (marginal likelihood).

Growth curve parameters

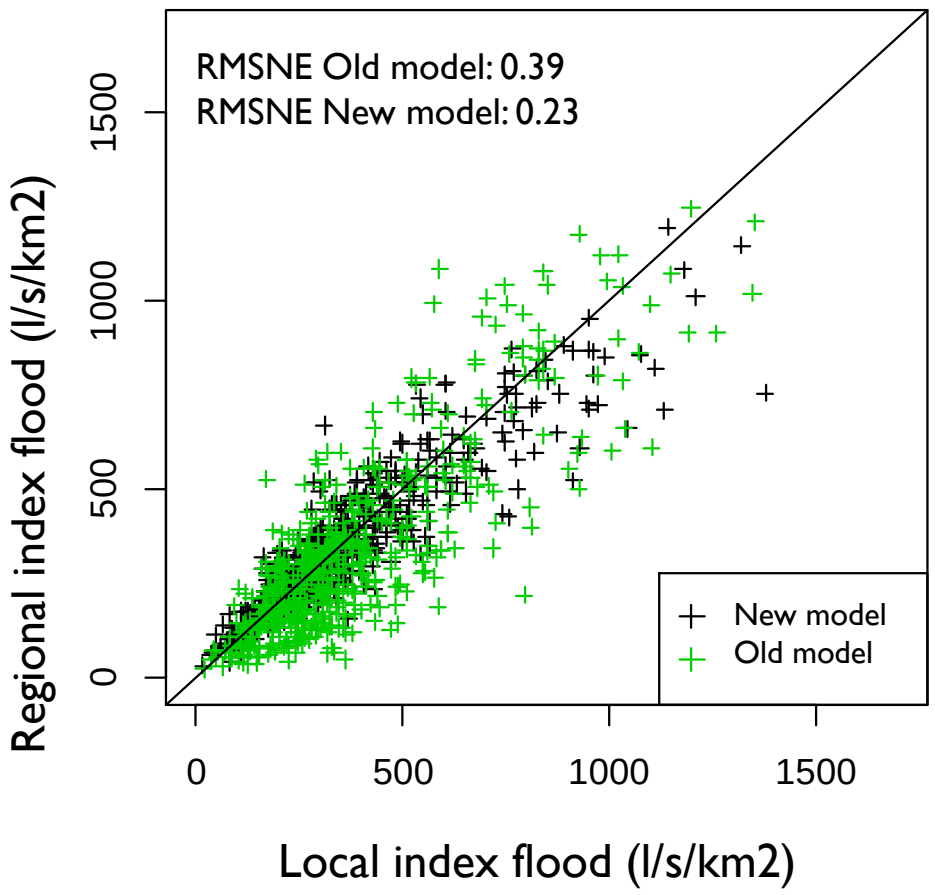
$$\alpha = -1.562 - 0.00361 * A_{Glac} + 0.00227 * A_{For} + 0.000238 * H_{10} + 0.00157 * P_{Jul} - 0.000580 * W_{Jun} + 0.164 * \delta$$

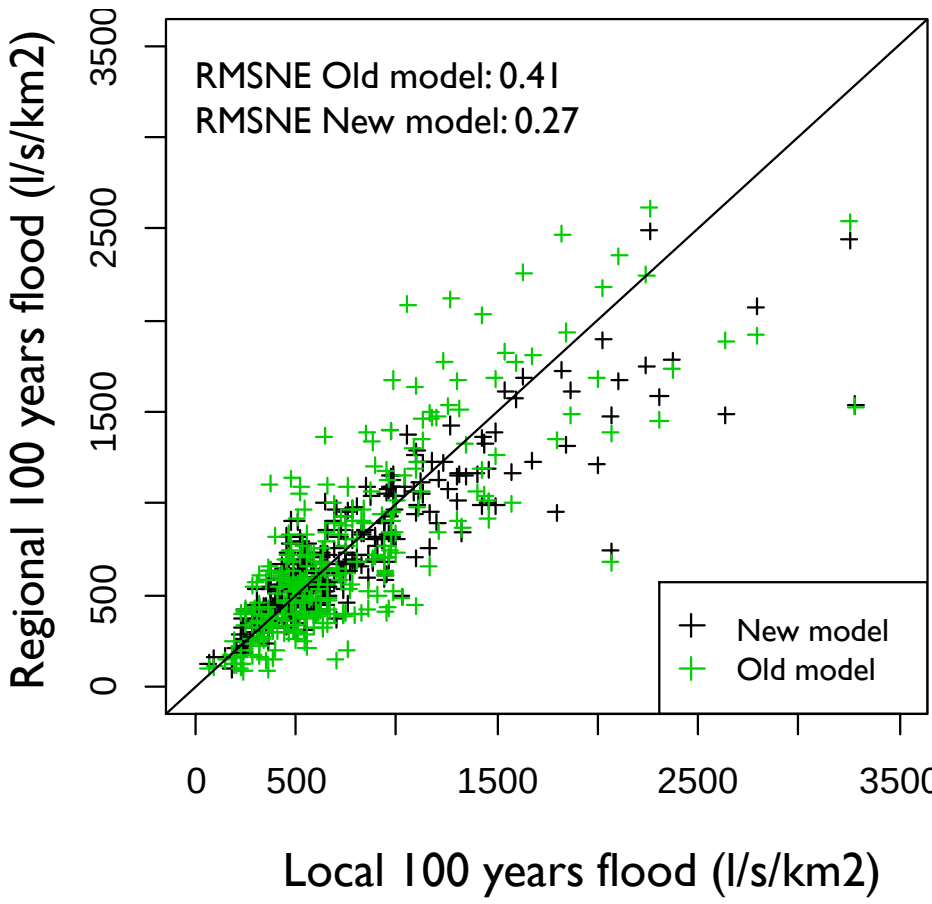
$$\mathrm{k} = 0.111 - 0.0173 * A_{SE} - 5.79 * 10^{-5} * E_{TL,net} + 0.165 * \epsilon$$

A_{Glac} is the percentage of the catchment covered by glaciers,
A_{For} is the percentage of the catchment covered by forest,
H_{10} is the 10% quantile of the elevation distribution within the catchment,
P_{Jul} is the mean July precipitation,
W_{Jun} is the mean June rain and snow melt,
$E_{TL,net}$ is the length of all rivers in the catchment excluding lakes
δ and ϵ are independent standard normally distributed noise terms.

Results – estimation of index flood and 100 year flood

- Compare to old RFFA model in Norway
- RMSE is good compared to other RFFA models summarized in Salinas et al (2013)

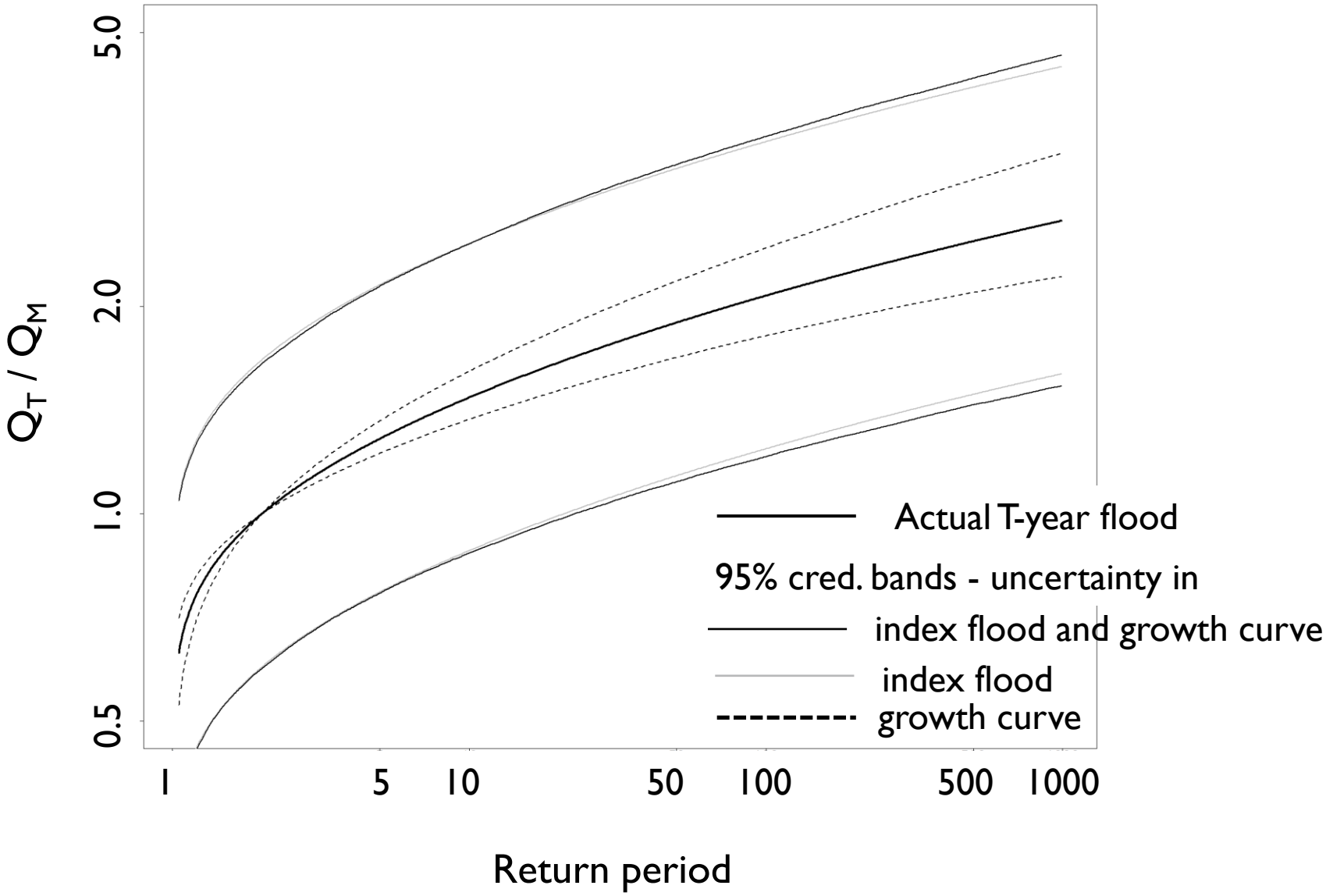




Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., & Blöschl, G. (2013) Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci., 17, 2637–2652, doi:10.5194/hess-17-2637-2013

Attribution of uncertainty

- Use the mean of catchment characteristics
- The index flood model has the largest uncertainty contribution



Methods – regional-compliant local analysis

- We used the regional model to establish general priors for Norway.
- Used the sample of catchment properties to establish a sample of flood frequency models using the regional models.
- Used this sample to establish a prior.
- This prior can be used at any ungauged location in Norway.
- This prior use no local knowledge about the catchment.

Methods – simple combination

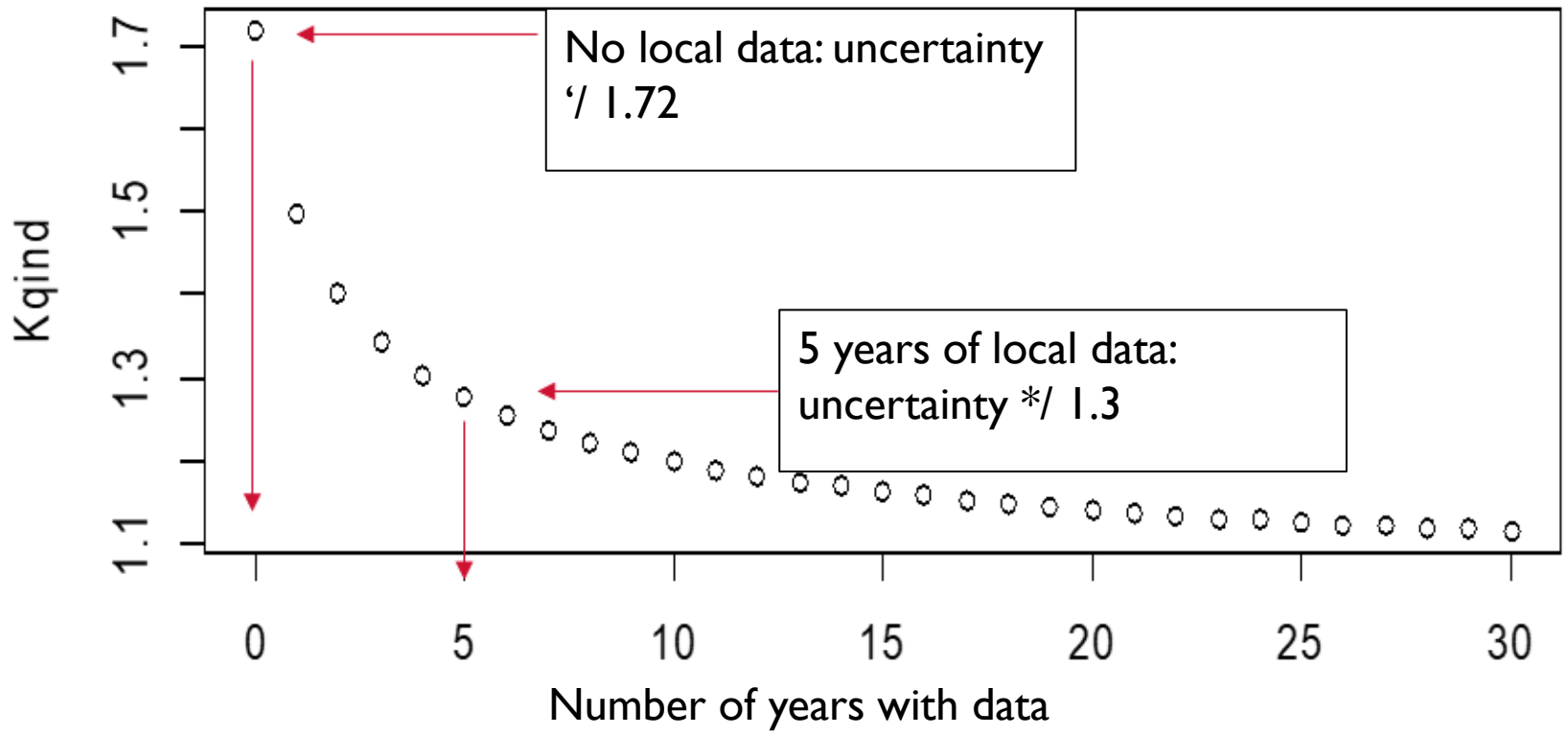
- Combine regression model for median flood with index flood estimated from local data on the log-scale. The posterior median flood is then given as:

$$q|D \sim N\left(\frac{m\left(x_{m_1}, x_{m_2}, \cdots, x_{m_{N_m}}\right)\sigma_q^2/n + \bar{q}s_m^2}{\sigma_q^2/n + s_m^2}, \frac{s_m^2\sigma_q^2/n}{\sigma_q^2/n + s_m^2}\right)$$

- Used the regional growth curve to get the index flood $m\left(x_{m_1}, x_{m_2}, \cdots, x_{m_{N_m}}\right)$ and the estimation variance s_m^2.

- Used $\sigma_{\hat{q}}^2 = 0.31^2$ - typical value for the dataset

Simple combination – posterior uncertainty of index flood



Methods – full combination

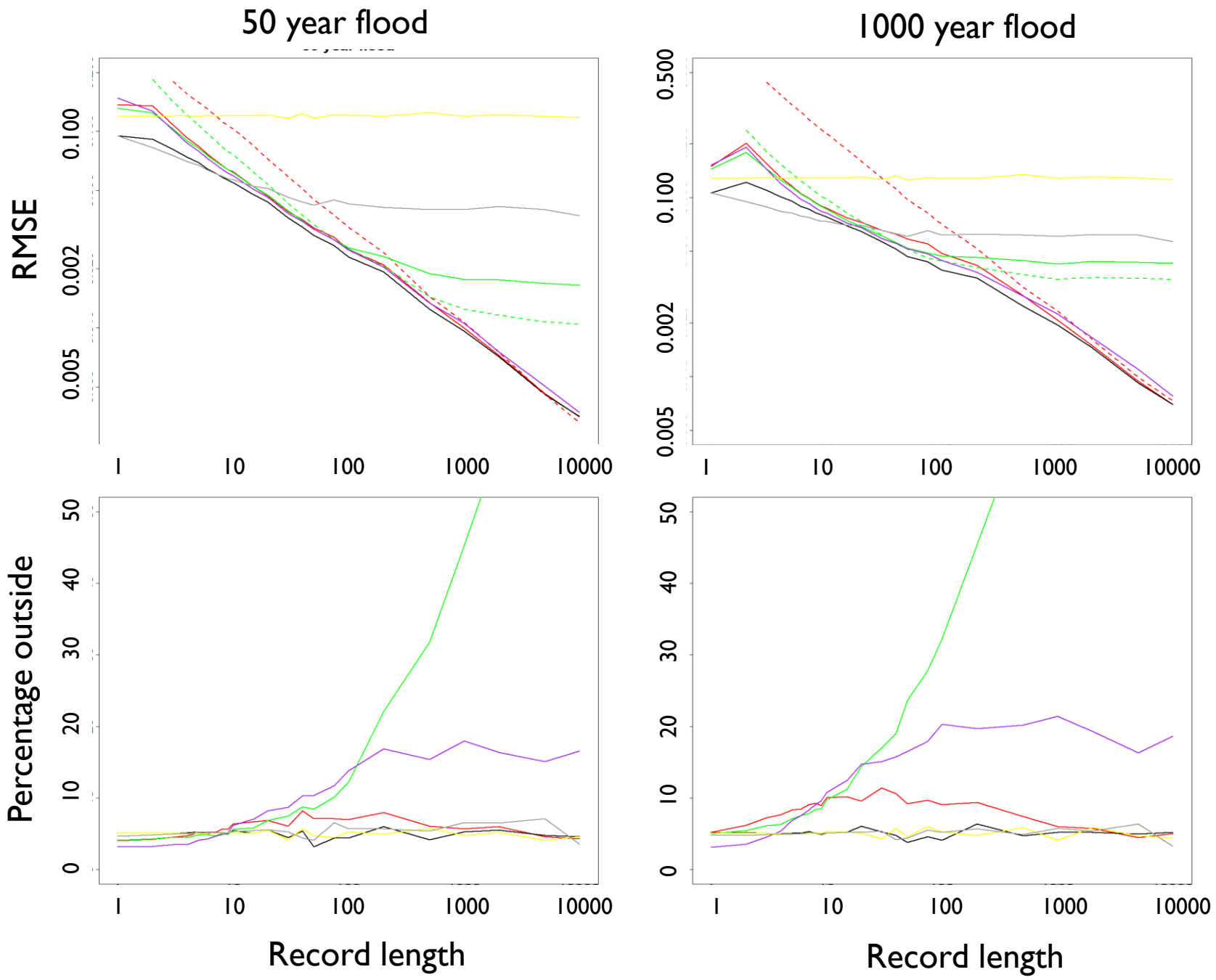
- Use the full regional models estimate of distribution parameters as the prior
- Estimate the posterior distribution of the GEV distribution using MCMC

Simulation study

- Assuming the regional model to be valid, while taking its uncertainty into account
- 10000 datasets were generated for each data size, where the data size varied from 1 to 10000 years.

Results from simulations

- Regional model
- Full local+regional analysis
- Simple local+regional analysis
- Local Bayesian Gumbel analysis
- Local l-moment Gumbel analysis
- Local Bayesian GEV analysis
- Local l-moment GEV analysis
- Composite Bayes



Key findings

- Local methods quickly outperforms the regional model already for just 3 data points for the Bayesian GEV and composite Bayes methods
- Local+regional methods outperform the purely local methods for small datasets (n<20)
- The simple local+regional methods is the best choice for n< 10
- The simple local+regional method starts to perform worse than the local methods for n>30
- The full local+regional method outperforms or performs just as well as the local methods for all sample sizes

Summary

- Re-parameterization allowed us to estimate the index flood model and the growth curve models separately
- Robust fit compared to previous studies
- Largest uncertainty contribution from index flood model
- Recommendations:
  - No local data: use the regional model
  - Less than 10 years of data: use the simple local+regional model
  - More than 10 years of data: use the full local+regional model