Vienna, Austria & Online | 3–8 May 2026

Evaluating loss functions for extreme streamflow predictions

Georgia Papacharalampous, Francesco Marra, Eleonora Dallan, and Marco Borga

georgia.papacharalampous@unipd.it, papacharalampous.georgia@gmail.com

Acknowledgements: The present study was developed in the framework of the ATLAS project, with financial support provided by the Autonomous Province of Bolzano/Bozen, Italy.

Designing "fit-for-purpose" hydrological models (1)

- **General focus:** Calibration and validation of hydrological models

- **Core concept:**

$$z = \mathcal{HM}\left(\boldsymbol{x}_p; \boldsymbol{\theta}\right)$$

(z = statistical functional of the conditional streamflow distribution $\underline{y}|\underline{\boldsymbol{x}}_p$,

$\mathcal{HM}$ = hydrological model, $\boldsymbol{x}_p$ = predictor variables, $\boldsymbol{\theta}$ = parameters)

- **Core choice:** Loss function that is strictly consistent for the target functional

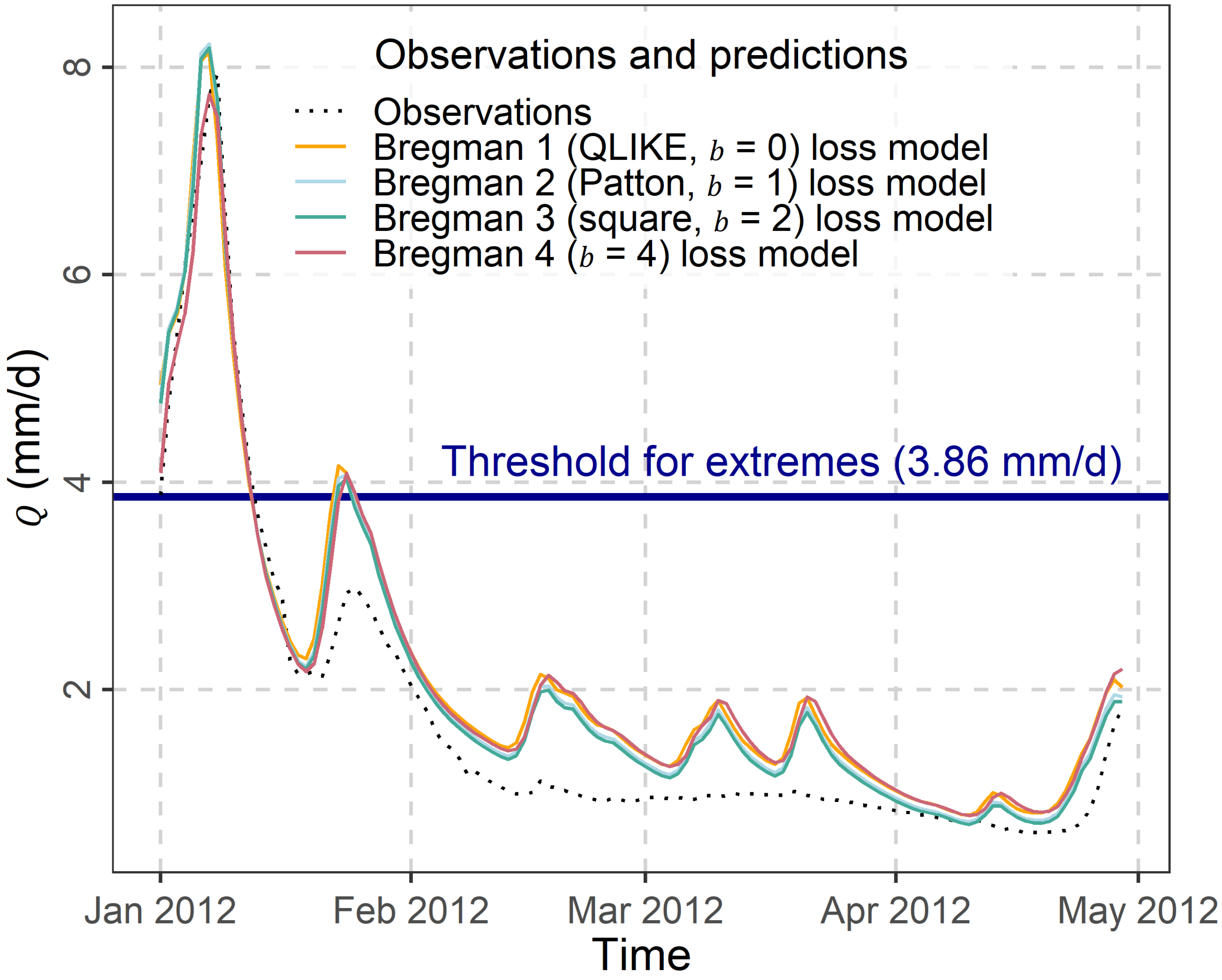
Designing "fit-for-purpose" hydrological models (2)

- **Specific focus:** Variable transformations (e.g., log(y) instead of y) in strictly consistent loss functions for:
  - ✓ Mean
  - ✓ Median
  - ✓ Other quantiles

  Functionals of interest
- **Key target:** Extreme streamflow predictions

Progress so far and next steps

- **Now:** GR4J + 3 functionals × 5 loss functions + 1 case study
- **Next:** Large-sample dataset → Generalized findings across hydroclimatic regimes



Foundations and domain

- z: prediction
- y: observation
- $\boldsymbol{z}_n = (z_1, \dots, z_n)^{\mathrm{T}}$: vector of predictions in train or test set
- $\boldsymbol{y}_n = (y_1, \dots, y_n)^{\mathrm{T}}$: vector of observations in train or test set
- $L(z, y)$: loss function when we predict z and y realizes
- $\bar{L}(\boldsymbol{z}_n, \boldsymbol{y}_n) = (1/n) \sum_{i=1}^{n} L(z_i, y_i)$: average loss in train or test set when we predict $\boldsymbol{z}_n$ and $\boldsymbol{y}_n$ realizes
- $V(z, y)$: identification function when we predict z and y realizes
- $\bar{V}(\boldsymbol{z}_n, \boldsymbol{y}_n) = (1/n) \sum_{i=1}^{n} V(z_i, y_i)$: empirical identification function in train or test set when we predict $\boldsymbol{z}_n$ and $\boldsymbol{y}_n$ realizes
- With monotonicity and convexity hereinafter referring to the domain $z, y > 0$

Loss and identification functions for the mean (1)

- Bregman loss function:

$$L_{\mathrm{Br}}(z, y; \varphi) = \varphi(y) - \varphi(z) - \varphi'(z)(y - z), \text{ with:}$$

φ convex implying L consistent and φ strictly convex implying L strictly consistent

- For $\varphi(t) = |t|^b, t > 0$, we have the following family of loss functions:

$$L_{\mathrm{Br}}(z, y; b) = (1/b(b-1))\left(y^b - z^b\right) - (1/(b-1))z^{b-1}(y - z), b \in \mathbb{R}\backslash\{0,1\}$$

- Special cases arise for specific values of b:

$$L_{\mathrm{Br}}(z, y) = \begin{cases} L_{\text{Bregman 1 (QLIKE)}}(z, y) = (y/z) - \log(y/z) - 1, b = 0 \\ L_{\text{Bregman 2 (Patton)}}(z, y) = y \log(y/z) - y + z, b = 1 \\ L_{\text{Bregman 3 (square)}}(z, y) = (1/2)(z - y)^2, b = 2 \\ L_{\text{Bregman 4}}(z, y) = (1/12)(y^4 - z^4) - (1/3)z^3(y - z), b = 4 \end{cases}$$

Loss and identification functions for the mean (2)

- A special case of the Bregman loss function for evaluating predictions of extremes:

$$L_{\text{Bregman 5 (Taggart) }[\alpha \in \mathbb{R}]}(z, y) = (y - a)^2 \mathbb{1}\{y \geq a\} + [(y - z)^2 - (y - a)^2] \mathbb{1}\{z \geq a\}$$

- An intuitive interpretation of this loss function follows:
  - Both observation and prediction are in the extreme region ($z \geq a$ and $y \geq a$): The $(y - a)^2$ terms cancel out, leaving exactly the standard squared error:

$$L_{\text{Bregman 5 (Taggart) }[\alpha \in \mathbb{R}]}(z, y) = (y - z)^2$$

  - False alarm ($z \geq a$ and $y < a$): The first term is zero, leaving:

$$L_{\text{Bregman 5 (Taggart) }[\alpha \in \mathbb{R}]}(z, y) = (y - z)^2 - (y - a)^2$$

  - Missed extreme ($z < a$ and $y \geq a$): The second term is zero, leaving:

$$L_{\text{Bregman 5 (Taggart) }[\alpha \in \mathbb{R}]}(z, y) = (y - a)^2$$

Loss and identification functions for the mean (3)

- Neither is in the extreme region ($z < a$ and $y < a$): Both indicator functions are zero, meaning no penalty is applied:

$$L_{\text{Bregman 5 (Taggart) } [\alpha \in \mathbb{R}]}(z, y) = 0$$

- Mean identification function:

$$V_{\text{mean}}(z, y) = z - y$$

Loss and identification functions for the median (1)

- GPL loss function for the median:

$$L_{\text{GPL }[\tau=0.5]}(z, y; g) = |g(z) - g(y)|, \text{ with:}$$

g non-decreasing implying L consistent and g strictly increasing implying L strictly consistent

- Special cases arise for specific functions g:

$$L_{\text{GPL 1 }[\tau = 0.5]}(z, y; g = \log(t)) = |\log(z) - \log(y)|$$

$$L_{\text{GPL 2 }[\tau = 0.5] \text{ (absolute)}}(z, y; g = t) = |z - y|$$

$$L_{\text{GPL 3 }[\tau = 0.5]}(z, y;\ g = t^2) = |z^2 - y^2|$$

$$L_{\text{GPL 4 }[\tau = 0.5]}(z, y;\ g = t^3) = |z^3 - y^3|$$

Loss and identification functions for the median (2)

- A special case of the GPL loss function for the median for evaluating predictions of extremes:

$$L_{\text{GPL 5 (Taggart) } [\tau = 0.5,\, a \in \mathbb{R}]}(z, y) = |(z - a)\mathbb{1}\{z \geq a\} - (y - a)\{y \geq a\}|$$

- An intuitive interpretation of this loss function follows:
  - Both in the extreme region ($z \geq a$ and $y \geq a$): The score is exactly the standard absolute error:

$$L_{\text{GPL 5 (Taggart) } [\tau = 0.5,\, a \in \mathbb{R}]}(z, y) = |z - y|$$

  - False alarm ($z \geq a$ but $y < a$): The observation term becomes zero. The penalty is strictly how far the forecast pushed into the extreme zone:

$$L_{\text{GPL 5 (Taggart) } [\tau = 0.5,\, a \in \mathbb{R}]}(z, y) = (z - a)$$

Loss and identification functions for the median (3)

- Missed extreme ($z < a$ but $y \geq a$): The forecast term becomes zero. The penalty is how far the observation pushed into the extreme zone:

$$L_{\text{GPL 5 (Taggart) } [\tau = 0.5,\, a \in \mathbb{R}]}(z, y) = (y - a)$$

- Neither is extreme ($z < a$ and $y < a$): Both terms are zero, yielding:

$$L_{\text{GPL 5 (Taggart) } [\tau = 0.5,\, a \in \mathbb{R}]}(z, y) = 0$$

No penalty is applied because the event occurred entirely outside the region of interest.

- Quantile identification function for the median:

$$V_{\text{GPL } [\tau=0.5]}(z, y) = \mathbb{1}\{z \geq y\} - 0.5$$

Loss and identification functions for the τ-quantile (1)

- GPL loss function:

$$L_{\text{GPL}}(z, y; g, \tau) = (\mathbb{1}\{z \geq y\} - \tau)(g(z) - g(y)), \text{ with:}$$

g non-decreasing implying L consistent and g strictly increasing implying L strictly consistent

- Special cases arise for specific functions g:

$$L_{\text{GPL 1 }[\tau]}\big(z, y; g = \log(t)\big) = (\mathbb{1}\{z \geq y\} - \tau)(\log(z) - \log(y))$$

$$L_{\text{GPL 2 (pinball) }[\tau]}(z, y; g = t) = (\mathbb{1}\{z \geq y\} - \tau)(z - y)$$

$$L_{\text{GPL 3 }[\tau]}(z, y;\ g = t^2) = (\mathbb{1}\{z \geq y\} - \tau)(z^2 - y^2)$$

$$L_{\text{GPL 4 }[\tau]}(z, y;\ g = t^3) = (\mathbb{1}\{z \geq y\} - \tau)(z^3 - y^3)$$

Loss and identification functions for the τ-quantile (2)

- A special case of the GPL loss function for evaluating predictions of extremes:

$$L_{\text{GPL 5 (Taggart) }[\tau,\,\alpha\,\in\,\mathbb{R}]}(z,y) = (\mathbb{1}\{z \geq y\} - \tau)\big((z-a)\mathbb{1}\{z \geq a\} - (y-a)\{y \geq a\}\big)$$

- An intuitive interpretation of this loss function follows:
  - Both in the extreme region ($z \geq a$ and $y \geq a$): The score is exactly the standard pinball loss:

$$L_{\text{GPL 5 (Taggart) }[\tau,\,\alpha\,\in\,\mathbb{R}]}(z,y) = (\mathbb{1}\{z \geq y\} - \tau)(z-y)$$

  - False alarm ($z \geq a$ but $y < a$): The observation term becomes zero. The penalty is strictly how far the forecast pushed into the extreme zone:

$$L_{\text{GPL 5 (Taggart) }[\tau,\,\alpha\,\in\,\mathbb{R}]}(z,y) = (\mathbb{1}\{z \geq y\} - \tau)(z-a)$$

Loss and identification functions for the τ-quantile (3)

- Missed extreme ($z < a$ but $y \geq a$): The forecast term becomes zero. The penalty is how far the observation pushed into the extreme zone:

$$L_{\text{GPL 5 (Taggart) } [\tau,\, \alpha \in \mathbb{R}]}(z, y) = (\mathbb{1}\{z \geq y\} - \tau)(y - a)$$

- Neither is extreme ($z < a$ and $y < a$): Both terms are zero, yielding:

$$L_{\text{GPL 5 (Taggart) } [\tau,\, \alpha \in \mathbb{R}]}(z, y) = 0$$

No penalty is applied because the event occurred entirely outside the region of interest.

- Quantile identification function:

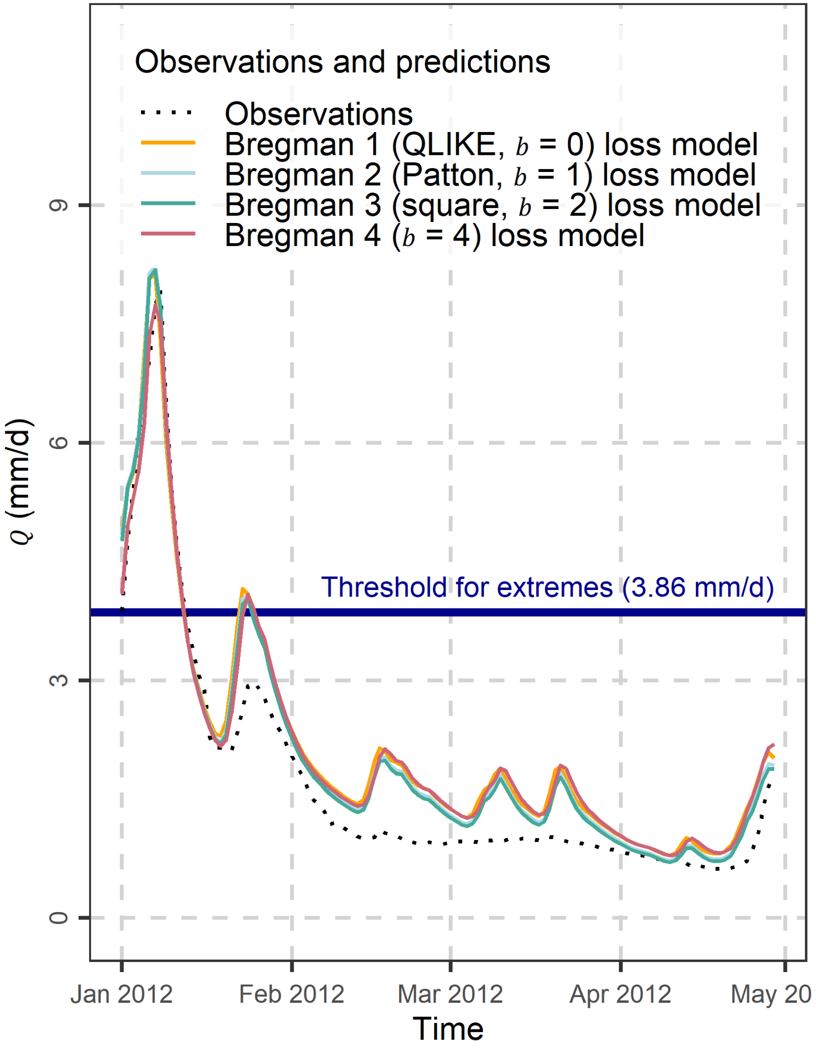
$$V_{\text{quantile}}(z, y; \tau) = \mathbb{1}\{z \geq y\} - \tau$$

Case study design

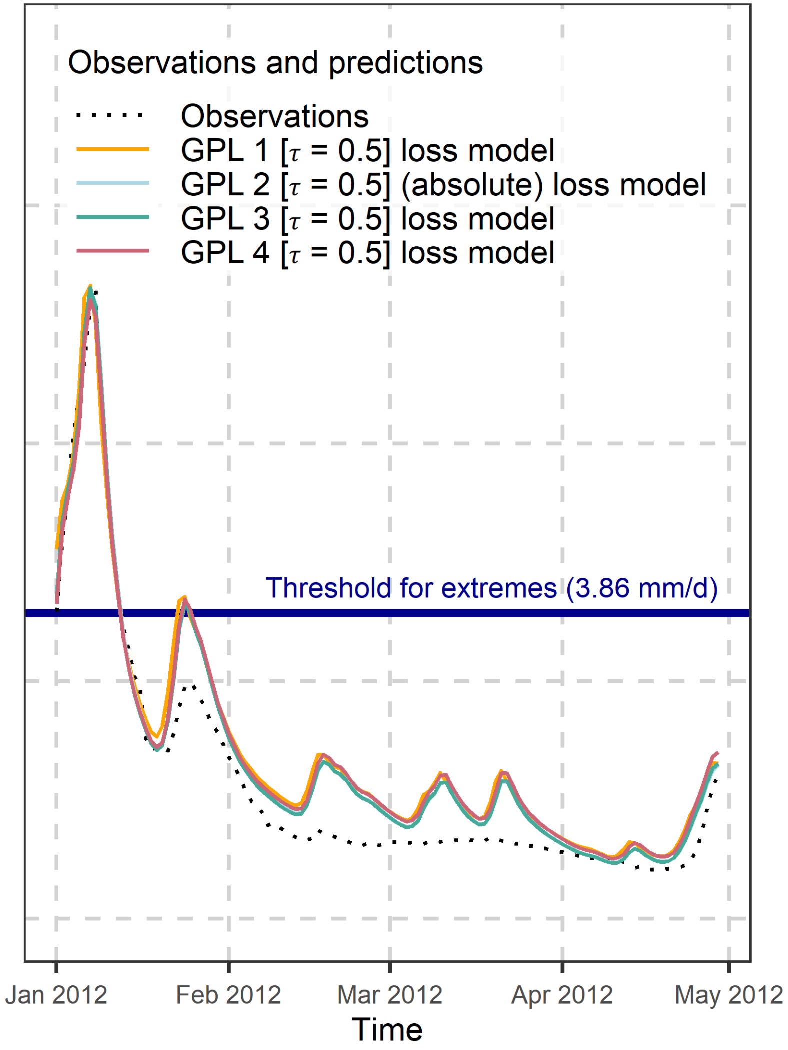
- **Model:** GR4J via `airGR`
- **Dataset:** 20-year daily rainfall-runoff dataset from `airGRdatasets`
- **Target functionals:** Mean, median, 0.90-quantile
- **Loss functions:** Subcases of the Bregman and Generalized Piecewise Linear (GPL) loss functions; five per functional
- **Calibration:** Using the non-extreme loss functions per functional
- **Evaluation:** Using all loss and identification functions per functional; extremes defined via the Taggart loss threshold (here set to the 95th percentile in the train set)

Examples of predictions and threshold for extremes

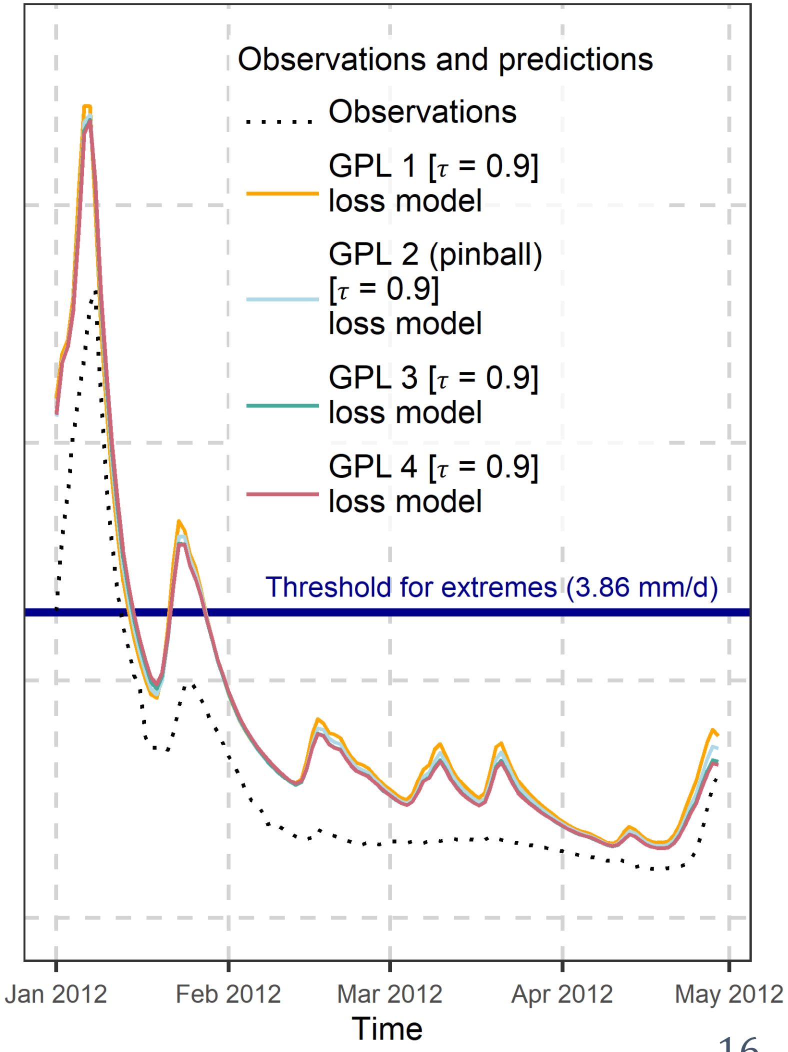
Predictive mean



Predictive median



Predictive 0.90-quantile



Predictive performance (1)

Predictive mean

Loss or identification function	Train: Bregman 1 (QLIKE, $b = 0$) loss model	Train: Bregman 2 (Patton, $b = 1$) loss model	Train: Bregman 3 (square, $b = 2$) loss model	Train: Bregman 4 ($b = 4$) loss model	Test: Bregman 1 (QLIKE, $b = 0$) loss model	Test: Bregman 2 (Patton, $b = 1$) loss model	Test: Bregman 3 (square, $b = 2$) loss model	Test: Bregman 4 ($b = 4$) loss model
Bregman 1 (QLIKE, $b = 0$) loss	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Bregman 2 (Patton, $b = 1$) loss	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Bregman 3 (square, $b = 2$) loss	0.16	0.16	0.16	0.16	0.18	0.17	0.17	0.18
Bregman 4 ($b = 4$) loss	0.99	0.99	0.99	0.93	2.07	1.83	1.81	1.97
Bregman 5 (Taggart) loss	0.04	0.04	0.04	0.04	0.06	0.05	0.05	0.06
Mean identification	0.06	0.01	-0.01	0.04	0.02	-0.03	-0.06	-0.01
Quantile identification [$\tau = 0.5$]	0.08	-0.01	-0.06	0.06	0.05	-0.03	-0.07	0.04

Model

Rank
(1 = best)
1
2
3
4
NA

Predictive performance (2)

Predictive median

Train

Loss or identification function	GPL 1 [τ = 0.5] loss model	GPL 2 [τ = 0.5] (absolute) loss model	GPL 3 [τ = 0.5] loss model	GPL 4 [τ = 0.5] loss model
GPL 1 [τ = 0.5] loss	0.20	0.21	0.21	0.21
GPL 2 [τ = 0.5] (absolute) loss	0.25	0.25	0.25	0.25
GPL 3 [τ = 0.5] loss	1.07	1.04	1.04	1.05
GPL 4 [τ = 0.5] loss	5.10	5.03	5.02	4.97
GPL 5 (Taggart) [τ = 0.5] loss	0.03	0.03	0.03	0.03
Mean identification	0.03	-0.03	-0.03	0.01
Quantile identification [τ = 0.5]	0.04	-0.06	-0.07	0.02

Test

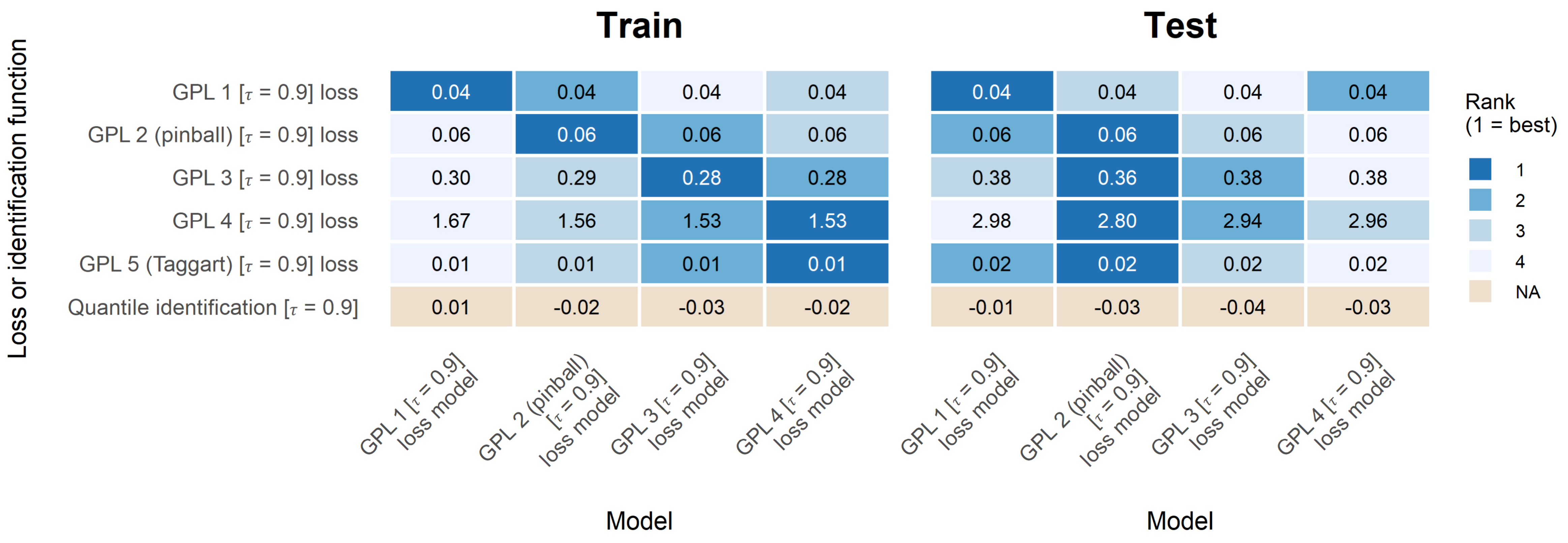
Loss or identification function	GPL 1 [τ = 0.5] loss model	GPL 2 [τ = 0.5] (absolute) loss model	GPL 3 [τ = 0.5] loss model	GPL 4 [τ = 0.5] loss model
GPL 1 [τ = 0.5] loss	0.22	0.23	0.23	0.23
GPL 2 [τ = 0.5] (absolute) loss	0.27	0.26	0.26	0.26
GPL 3 [τ = 0.5] loss	1.28	1.24	1.24	1.28
GPL 4 [τ = 0.5] loss	7.88	7.80	7.78	7.93
GPL 5 (Taggart) [τ = 0.5] loss	0.04	0.04	0.04	0.04
Mean identification	-0.01	-0.07	-0.07	-0.03
Quantile identification [τ = 0.5]	0.02	-0.07	-0.08	-0.00

Model

Rank (1 = best): 1, 2, 3, 4, NA

Predictive performance (3)

Predictive 0.90-quantile



Key takeaways

- When the data-generating process is unknown, both the target functional and the specific loss function of interest must be specified to reach optimal performance.
- To better predict extremes, consider transformations in loss functions.
- The transformation should be introduced into the Bregman and GPL loss functions.
- To assess extreme predictions, consider specialized loss functions.
- Uncertainties might notably affect case study results, especially in the test set.
- Therefore, moving toward large-sample datasets is necessary.

Related toy experiments

Communicating hydrological model calibration with toy examples

EGU26-5507 poster (Wed, 06 May, 14:00–15:45 at Hall X5 | X5.224)