



HAL
open science

L'outil d'analyse des performances evalhyd

Thibault Hallouin, François Bourgin, Charles Perrin, Maria-Helena Ramos,
Vazken Andréassian

► **To cite this version:**

Thibault Hallouin, François Bourgin, Charles Perrin, Maria-Helena Ramos, Vazken Andréassian.
L'outil d'analyse des performances evalhyd. 5es Rencontres HydroGR, Jun 2023, Antony, France.
hal-04144015

HAL Id: hal-04144015

<https://hal.inrae.fr/hal-04144015>

Submitted on 29 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

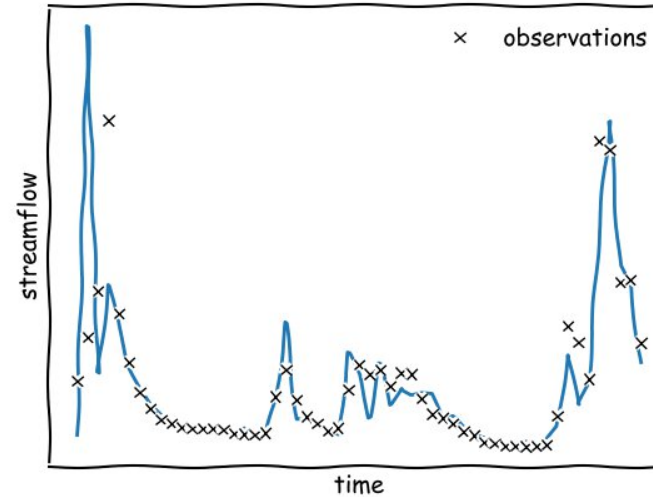
➤ L'outil d'analyse des performances *evalhyd*

Thibault Hallouin¹, François Bourgin¹, Charles Perrin¹,
Maria-Helena Ramos¹, Vazken Andréassian¹

¹Université Paris-Saclay, INRAE, HYCAR, Antony
thibault.hallouin@inrae.fr

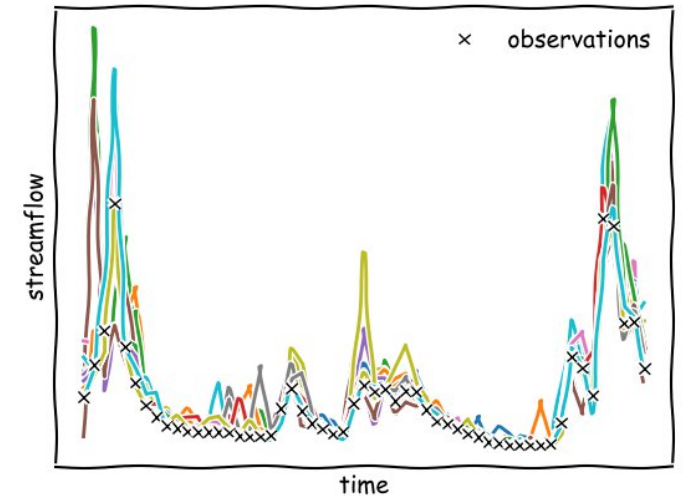
➤ Contexte

- L'étape de validation / vérification / évaluation (prédictions vs. observations) est très courante en hydrologie
- Les critères utilisés sont souvent les mêmes (*par ex. MAE, NSE, KGE, Brier, CRPS*)
- Pourtant, les outils pour les calculer sont rarement les mêmes (*par ex. EVS, xskillscore, scoringRules, scores*) et spécifiques à un langage de programmation
- De plus, pas tout le temps open source ou bien documenté, menant à des problèmes de reproductibilité des études
- D'où le développement d'un nouvel outil ouvert et polyglotte pour l'évaluation des prédictions de débit



évaluation
déterministe

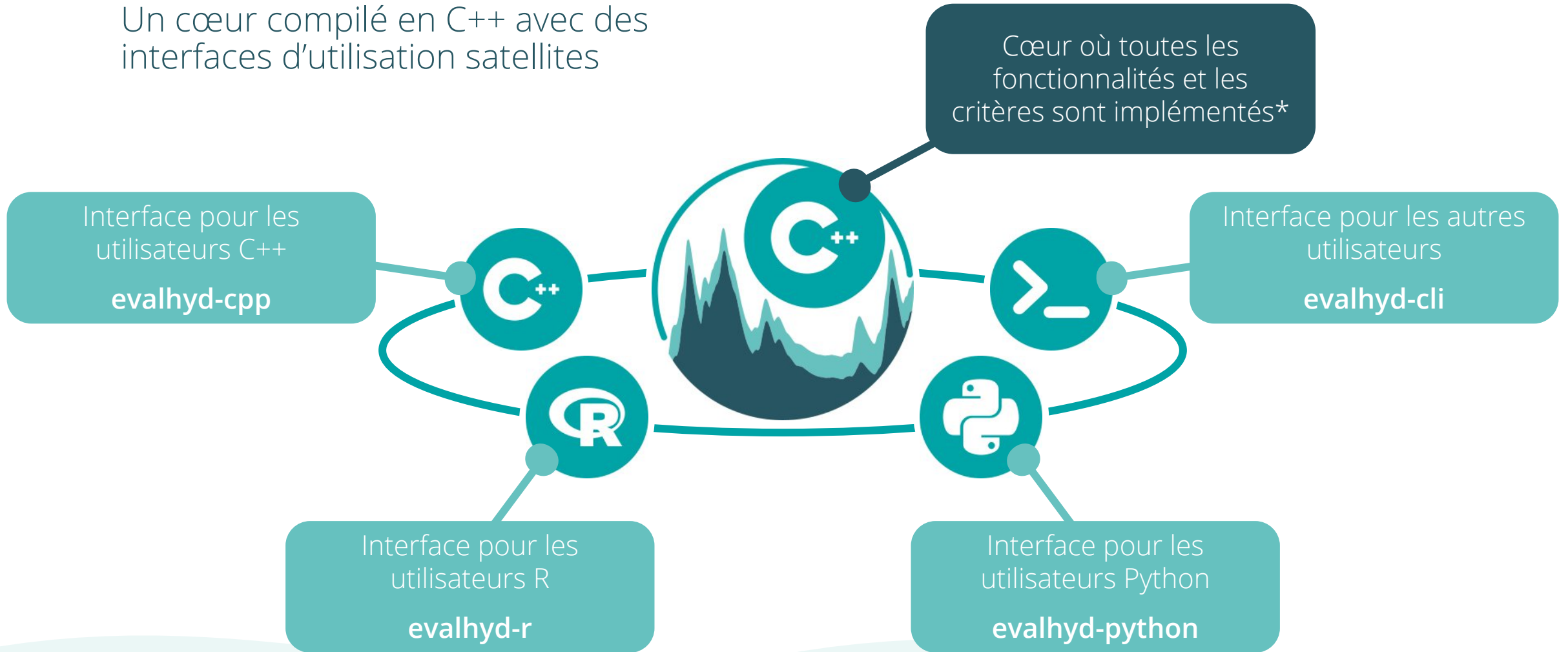
évaluation
probabiliste
(ensembliste)



evalhyd

➤ Une pile logicielle polyglotte nommée evalhyd

Un cœur compilé en C++ avec des interfaces d'utilisation satellites



INRAE

L'outil d'analyse des performances **evalhyd**
Thibault Hallouin et al. | Journées HydroGR | 22 juin 2023, Antony

*Merci à Antoine Prouvost et Johan Mabilie de QuantStack pour leur aide avec **xtensor**

➤ Interfaces harmonisées à travers la pile logicielle

2 points d'entrée **evald** et **evalp**

Déterministe



```
res = evalhyd.evald(obs, prd, ["NSE"])
```



```
res ← evalhyd::evald(obs, prd, c("NSE"))
```



```
auto res = evalhyd::evald(obs, prd, {"NSE"});
```



```
./evalhyd evald "./obs.csv" "./prd.csv" "NSE"
```

Probabiliste



```
res = evalhyd.evalp(  
  obs, prd, ["BS"], q_thr=thr, events="high"  
)
```



```
res ← evalhyd::evalp(  
  obs, prd, c("BS"), q_thr=thr, events="high"  
)
```



```
auto res = evalhyd::evalp(  
  obs, prd, {"BS"}, q_thr=thr, events="high"  
);
```



```
./evalhyd evalp \  
  "./obs/" "./prd/" "BS" \  
  --thr "./thr/" --events "high"
```

➤ Données multidimensionnelles

Déterministe

- Philosophie bidimensionnelle :
 - Observations **{1, temps}**
 - Prédictions **{séries, temps}**
- Pour permettre l'évaluation de plusieurs prédictions, par ex. expériences de Monte Carlo ou approches multi-modèles

Probabiliste

- Philosophie quadridimensionnelle :
 - Observations **{bassins, temps}**
 - Prédictions **{bassins, échéances, membres, temps}**
- Pour permettre l'évaluation de plusieurs bassins et plusieurs échéances de prévision
- Pour permettre le calcul de scores multivariés (par ex. "energy score")



➤ Critères d'évaluation disponibles

Déterministe

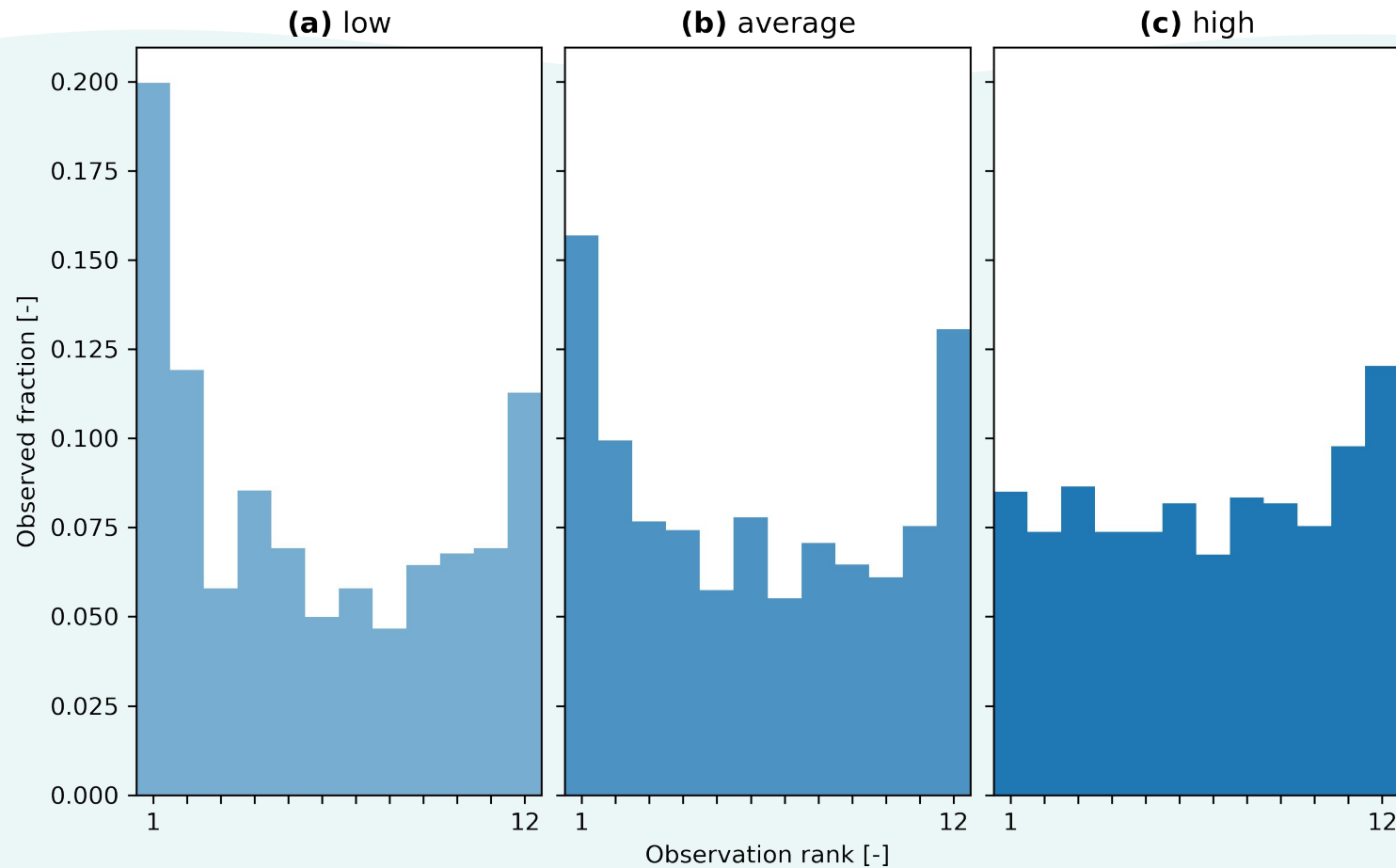
Mean absolute error	MAE
Mean absolute relative error	MARE
Mean square error	MSE
Root mean square error	RMSE
Nash-Sutcliffe efficiency	NSE
Kling-Gupta efficiency	KGE, KGE_D, KGE_PRIME, KGE_PRIME_D
Contingency table	CONT_TBL

Probabiliste

Brier scores and their decompositions	BS, BSS, BS_CRD, BS_LBD
Reliability diagram	REL_DIAG
Continuous rank probability score	CRPS_FROM_BS, CRPS_FROM_QS, CRPS_FROM_ECDF
Contingency table	POD, POFD, CSI, FAR, ROCSS
Quantile scores	QS
Rank scores	RANK_HIST, DS, AS
Interval scores	CR, AW, AWN, WS

➤ Quelques fonctionnalités d'evalhyd: *masquage*

Stratifier par gammes de débits (Bellier et al., 2017)



```
rank_hist, = evalhyd.evalp(  
    q_obs=obs,  
    q_prd=prd,  
    metrics=['RANK_HIST'],  
    m_cdt=np.repeat(  
        np.array(["q_prd_median{<qt10.3}",  
                 "q_prd_median{≥ qt10.3, ≤ qt10.7}",  
                 "q_prd_median{>qt10.7}"]),  
        repeats=obs.shape(0), axis=0  
    )  
)
```

```
rank_hist ← evalhyd::evalp(  
    q_obs=obs,  
    q_prd=prd,  
    metrics=c('RANK_HIST'),  
    m_cdt=cbind(  
        rep("q_prd_median{<qt10.3}",  
           dim(obs)[1]),  
        rep("q_prd_median{≥ qt10.3, ≤ qt10.7}",  
           dim(obs)[1]),  
        rep("q_prd_median{>qt10.7}",  
           dim(obs)[1])  
    )  
)[[1]]
```



INRAE

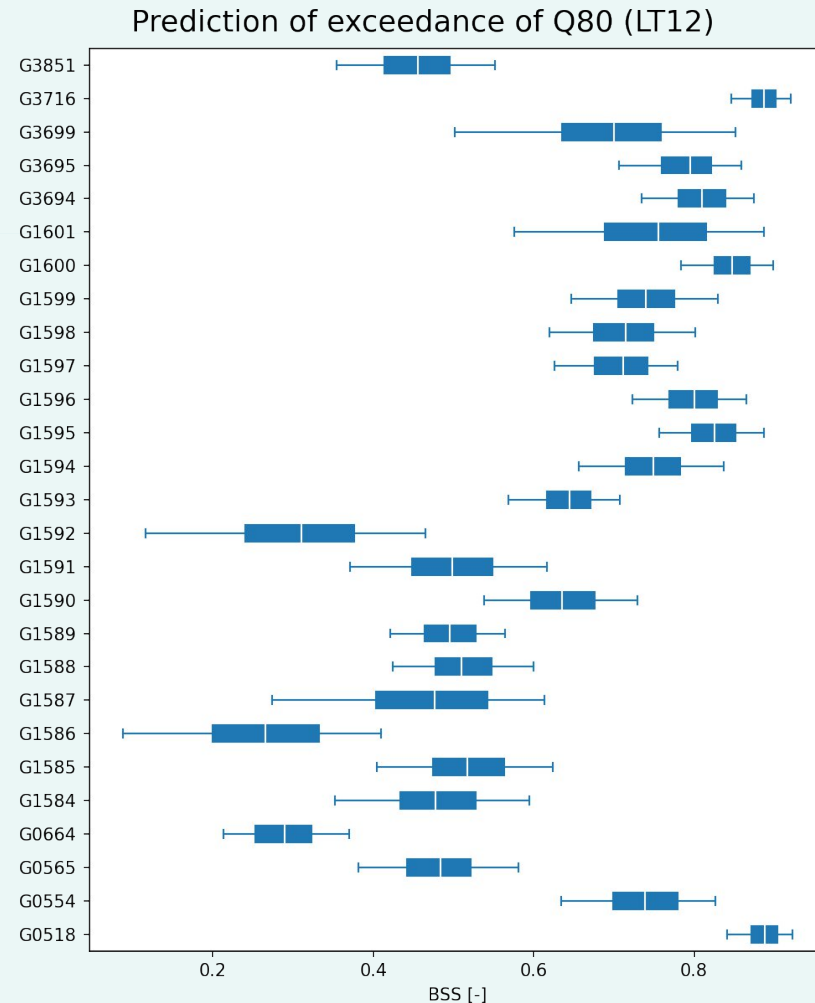
L'outil d'analyse des performances **evalhyd**

Thibault Hallouin et al. | Journées HydroGR | 22 juin 2023, Antony

Bellier, J., et al. (2017). Sample Stratification in Verification of Ensemble Forecasts of Continuous Scalar Variables (...). Mon. Wea. Rev., 145, 9, 3529-3544. DOI:10.1175/MWR-D-16-0487.1

➤ Quelques fonctionnalités d'evalhyd: *bootstrapping*

Estimer l'incertitude sur les valeurs de critères (Clark et al., 2021)



```
bss, = evalhyd.evalp(  
    q_obs=obs,  
    q_prd=prd,  
    metrics=['BSS'],  
    q_thr=thr,  
    events='Low',  
    bootstrap={'n_samples': 1000,  
              'len_sample': 10,  
              'summary': 2},  
    dts=dts  
)
```

```
bss ← evalhyd::evalp(  
    q_obs=obs,  
    q_prd=prd,  
    metrics=c('BSS'),  
    q_thr=thr,  
    events='Low',  
    bootstrap=list(n_samples=1000,  
                  len_sample=10,  
                  summary=2),  
    dts=dts  
)[[1]]
```



INRAE

L'outil d'analyse des performances **evalhyd**

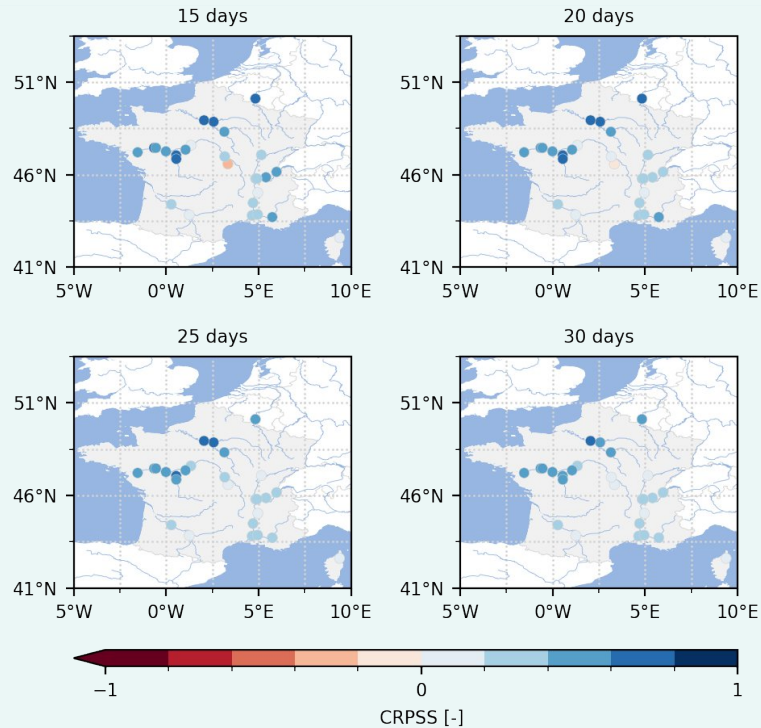
Thibault Hallouin et al. | Journées HydroGR | 22 juin 2023, Antony

Clark, M. P., et al. (2021). The abuse of popular performance metrics in hydrologic modeling. *Water Resources Research*, 57, e2020WR029001. DOI:10.1029/2020WR029001

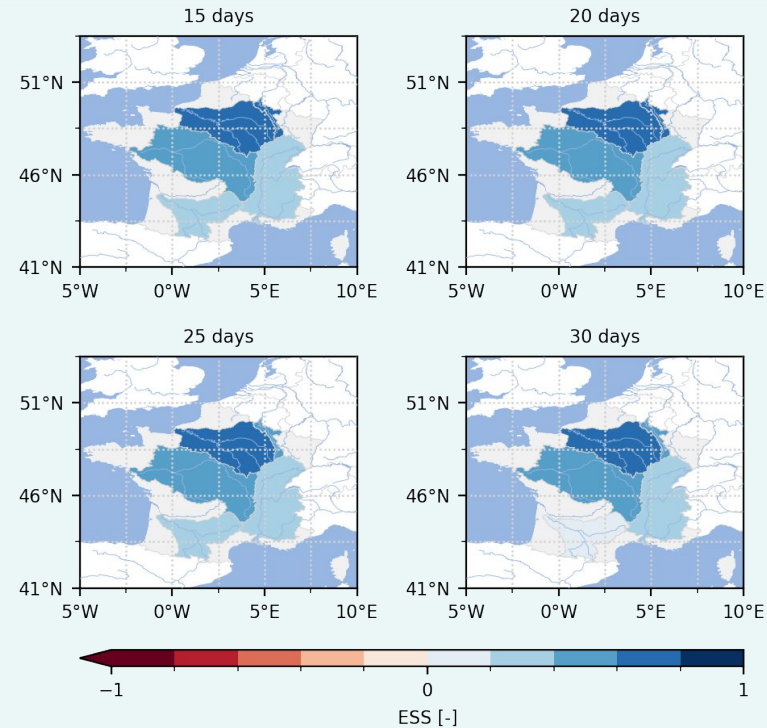
➤ Quelques fonctionnalités d'evalhyd: *multivarié*

Synthétiser les performances par zones géographiques (Gneiting et al., 2008)

GloFAS reforecast vs. climatology benchmark



GloFAS reforecast vs. climatology benchmark



```
es_prd, = evalhyd.evalp(  
    q_obs=obs, q_prd=prd, metrics=["ES"]  
)  
  
es_clm, = evalhyd.evalp(  
    q_obs=obs, q_prd=clm, metrics=["ES"]  
)  
  
ess = 1 - (es_prd / es_clm)
```

```
es_prd ← evalhyd::evalp(  
    q_obs=obs, q_prd=prd, metrics=c("ES")  
)[[1]]  
  
es_clm ← evalhyd::evalp(  
    q_obs=obs, q_prd=clm, metrics=c("ES")  
)[[1]]  
  
ess ← 1 - (es_prd / es_clm)
```



INRAE

L'outil d'analyse des performances **evalhyd**

Thibault Hallouin et al. | Journées HydroGR | 22 juin 2023, Antony

Gneiting, T., et al. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. TEST, 17, 211–235. DOI:10.1007/s11749-008-0114-x

> Disponibilité

<https://hydrogr.github.io/evalhyd>

Source à compiler



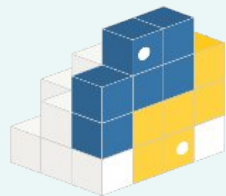
GitLab



Binaire précompilé



conda-forge



PyPI



INRAE

➤ **Merci**



<https://hydrogr.github.io/evalhyd>

thibault.hallouin@inrae.fr