How well can we model river flows across Great Britain?

Learning from the UK Hydro-MIP model intercomparison project

Rosie Lane, Helen Baron, Emma Robinson
(and many others!)

UKCEH

Our planet.
Decoded.

We need to be able to model river flows

- Water resource assessment
- Flood forecasting
- Short and long-term flow forecasts (e.g. UK Hydrological Outlook)
- Providing river flow estimates for ungauged locations / time periods
- Evaluating the impact of climate and land-use change on flow
- Also improved understanding of catchment functioning by treating models as “working hypotheses”

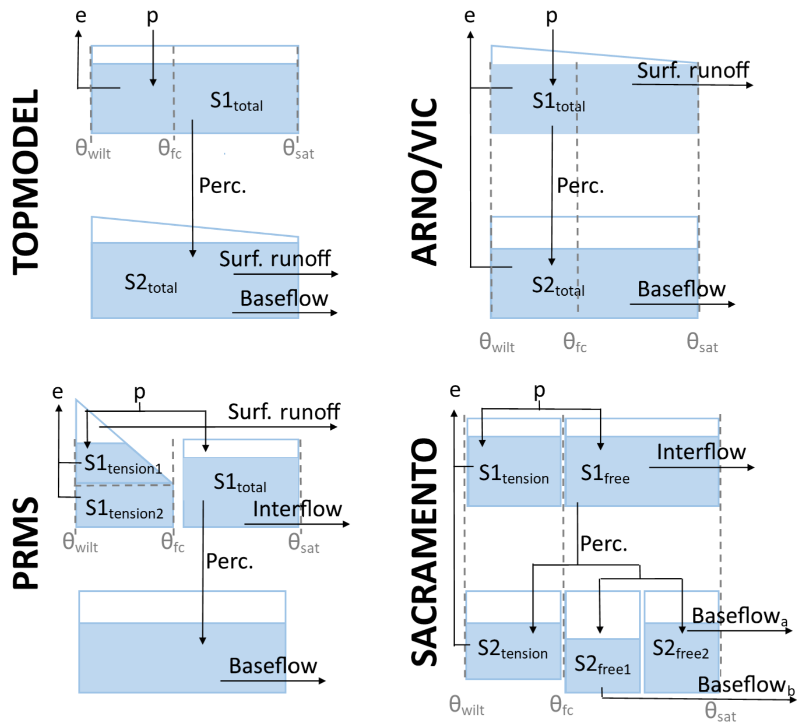
There are many different models available

- Different types of models for different purposes

Lumped hydrological models	Semi-distributed and fully distributed hydrological models
Land-surface models	Machine learning models

There are many different models available

- Different types of models for different purposes



Source: Lane et al. (2019) HESS

Lumped hydrological models

Semi-distributed and fully distributed hydrological models

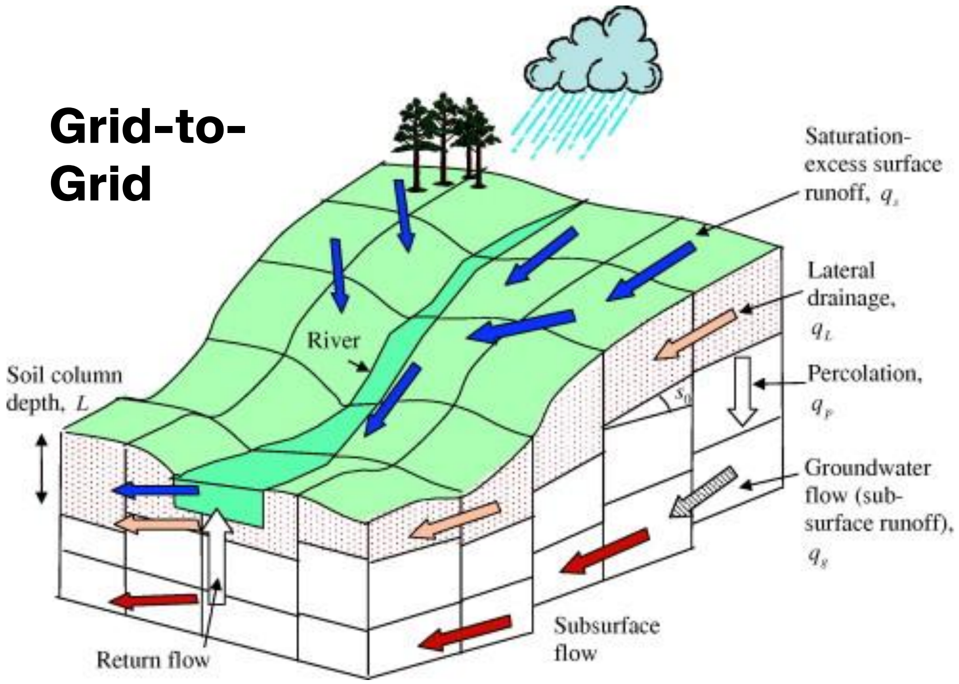
Land-surface models

Machine learning models

There are many different models available

- Different types of models for different purposes

Grid-to-Grid



Source: *Bell et al. (2009) Journal of Hydrology.*

Lumped hydrological models

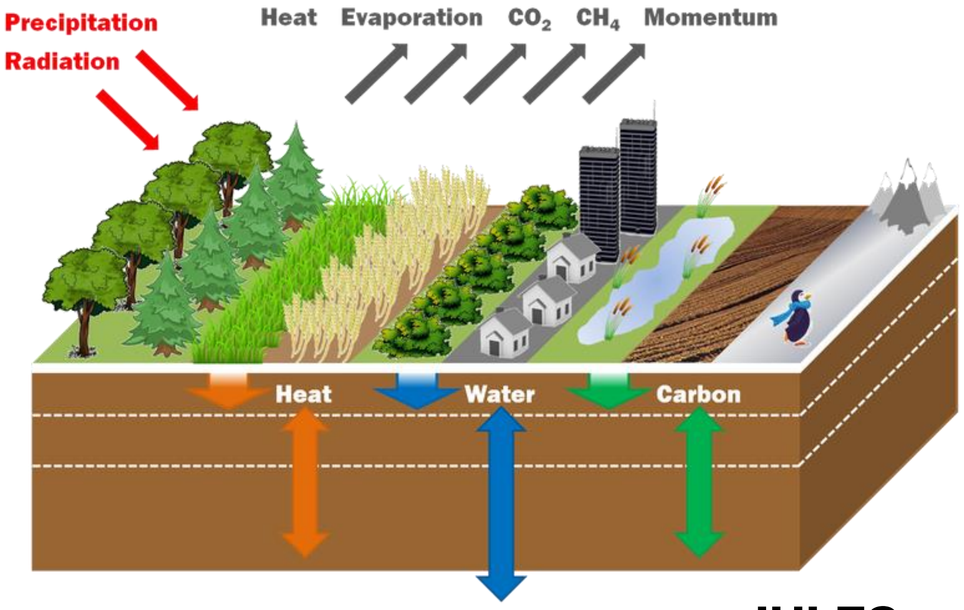
Semi-distributed and fully distributed hydrological models

Land-surface models

Machine learning models

There are many different models available

- Different types of models for different purposes



JULES

Source: https://jules.jchmr.org/about

Lumped hydrological models

Semi-distributed and fully distributed hydrological models

Land-surface models

Machine learning models

There are many different models available

- Different types of models for different purposes

Lumped hydrological models

Semi-distributed and fully distributed hydrological models

Land-surface models

Machine learning models

With all these available models, how to choose?

- Model choice often the result of “legacy rather than adequacy”[1]

Lumped hydrological models

Semi-distributed and fully distributed hydrological models

Land-surface models

Machine learning models

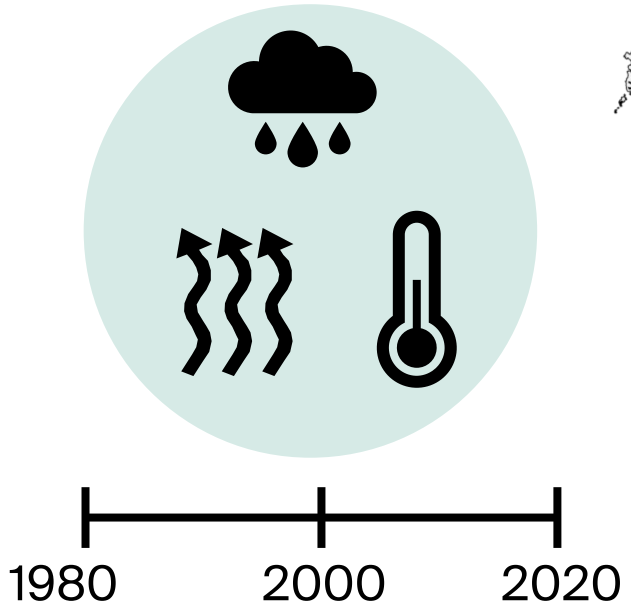
1 Addor and Melsen (2019) WRR

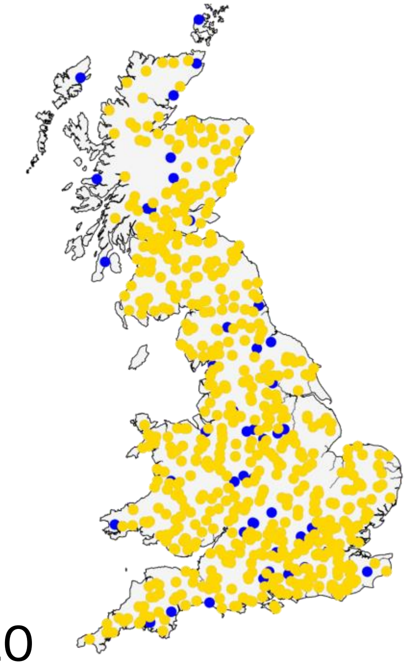
With all these available models, how to choose?

- Model choice often the result of “legacy rather than adequacy”[1]
- Need for rigorous model evaluation and intercomparison to inform model selection and understand model differences

Lumped hydrological models	Semi-distributed and fully distributed hydrological models
Land-surface models	Machine learning models

We've carried out a community model intercomparison project (the UK Hydro-MIP)





Modellers were invited to submit flows following our community-agreed protocol:

- Consistent meteorology
- 1980-2020
- 673 catchments (628 core)
- Daily mean flows (m^3/s)

Other decisions left to expert judgement of modellers.

Submissions included a diverse range of model structures

Machine Learning Models 3	Lumped Hydrological Models 6
Distributed hydrological models 5	Land surface models 2

14+ model structures

27+ separate flow submissions

Collaboration across 14+ different research groups

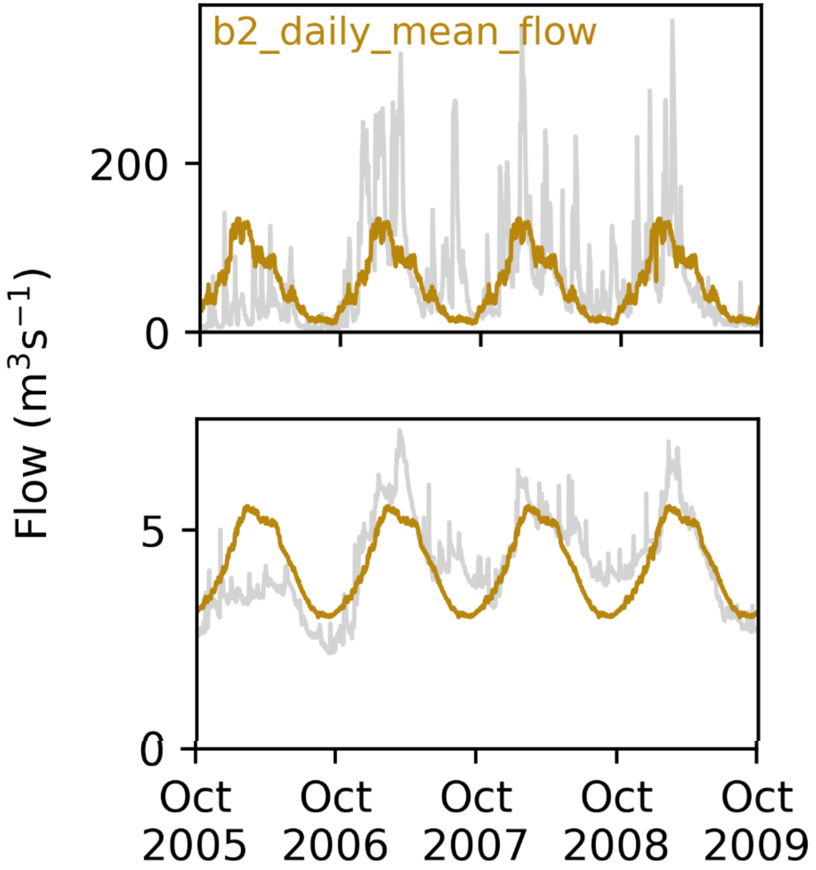
Represents a diverse set of models which are widely used for UK hydrology

We evaluated modelled flows against lower benchmarks

- A range of performance metrics were first calculated for each model/catchment
- Lower benchmarks aim to be intuitive reference points which the models should be able to beat, setting a lower bar for each performance metric

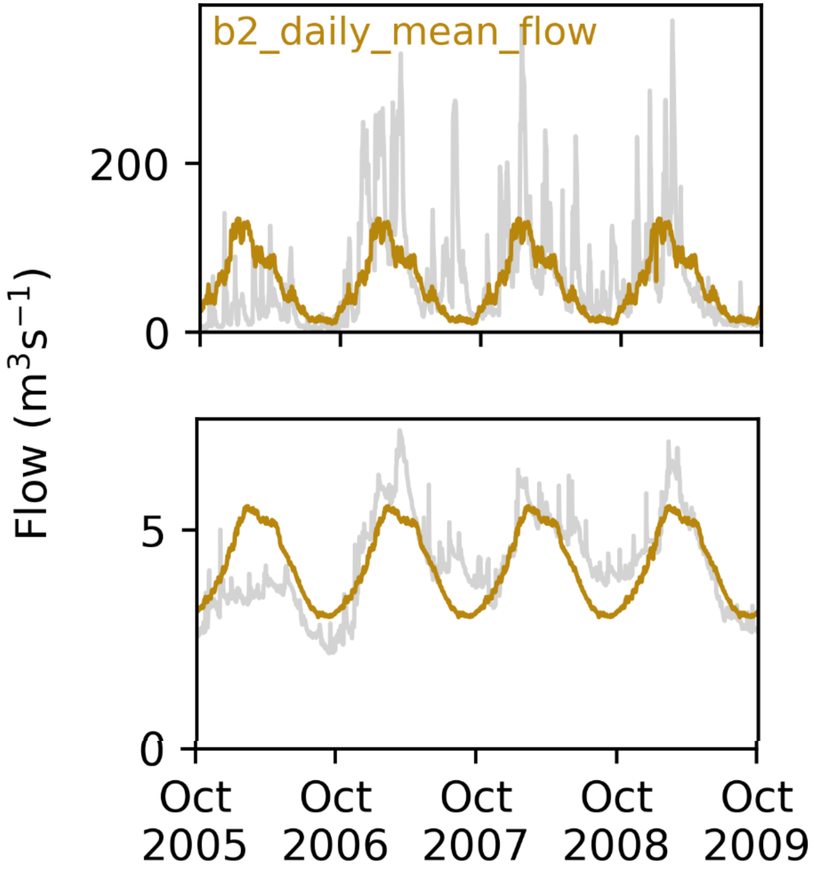
We evaluated modelled flows against lower benchmarks

- A range of performance metrics were first calculated for each model/catchment
- Lower benchmarks aim to be intuitive reference points which the models should be able to beat, setting a lower bar for each performance metric
- Here, we will focus on *mean flow for each day of the year* as a benchmark

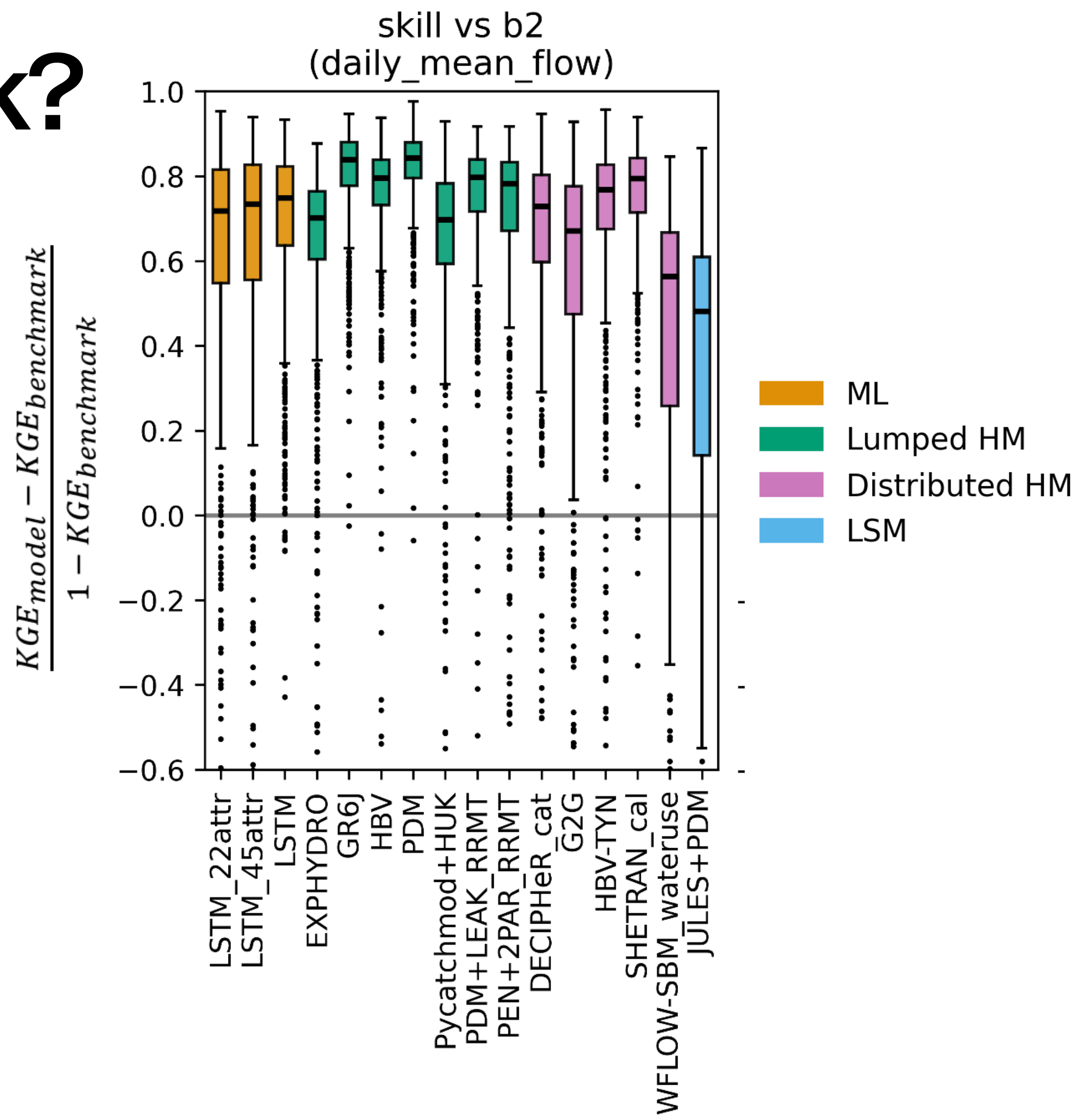


We evaluated modelled flows against lower benchmarks

- A range of performance metrics were first calculated for each model/catchment
- Lower benchmarks aim to be intuitive reference points which the models should be able to beat, setting a lower bar for each performance metric
- Here, we will focus on *mean flow for each day of the year* as a benchmark, and the KGE performance metric

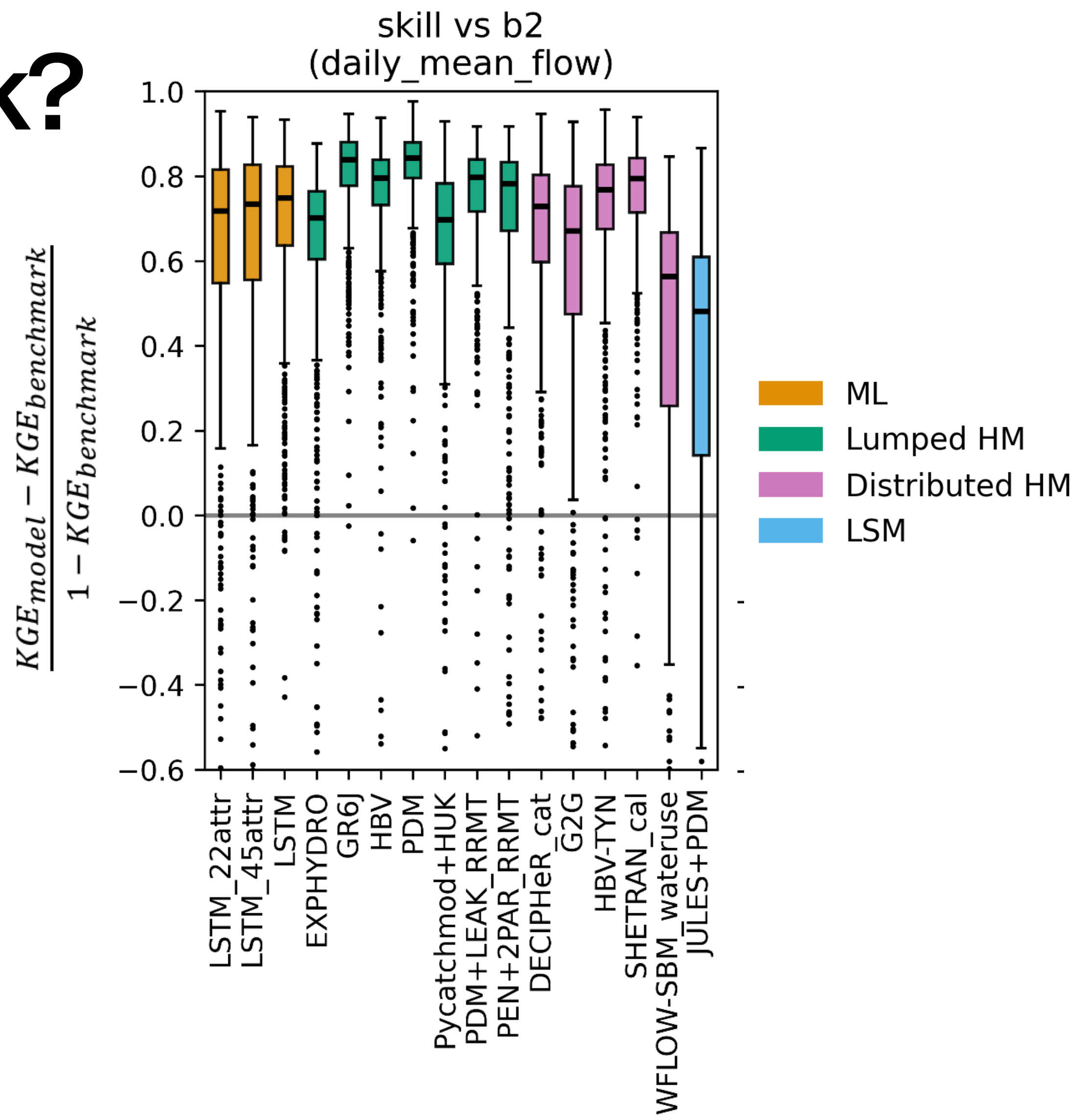


Do we beat the daily mean flow benchmark?



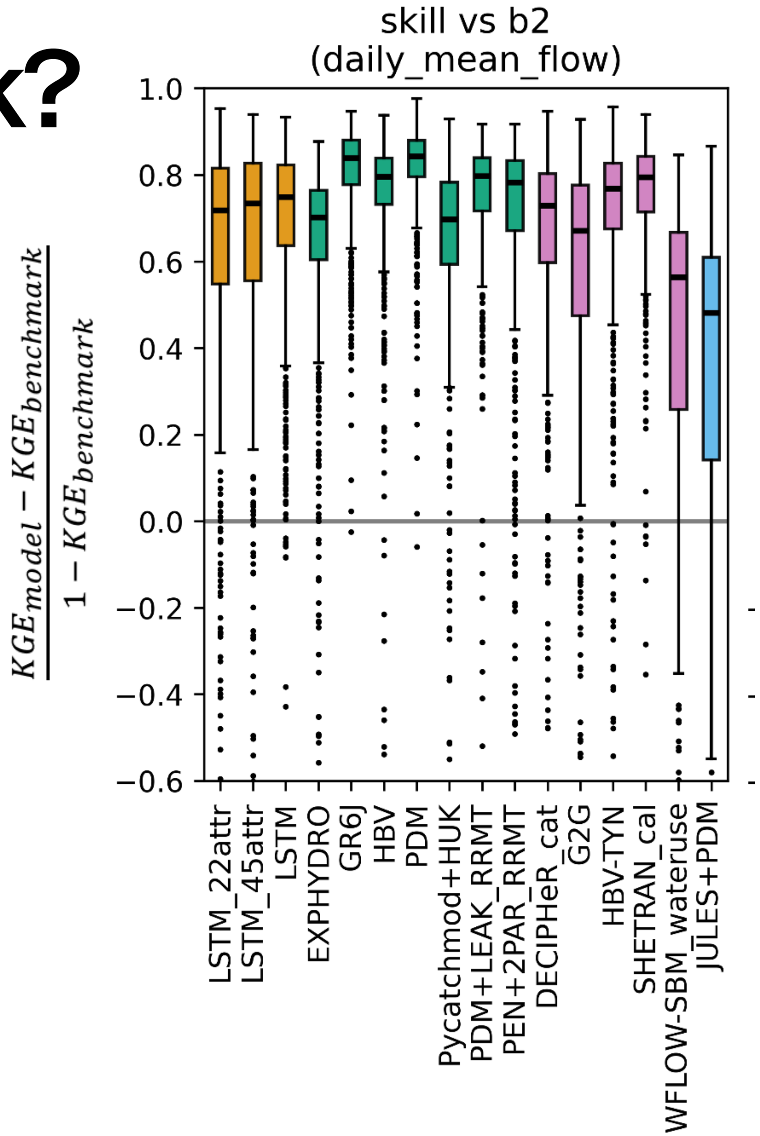
Do we beat the daily mean flow benchmark?

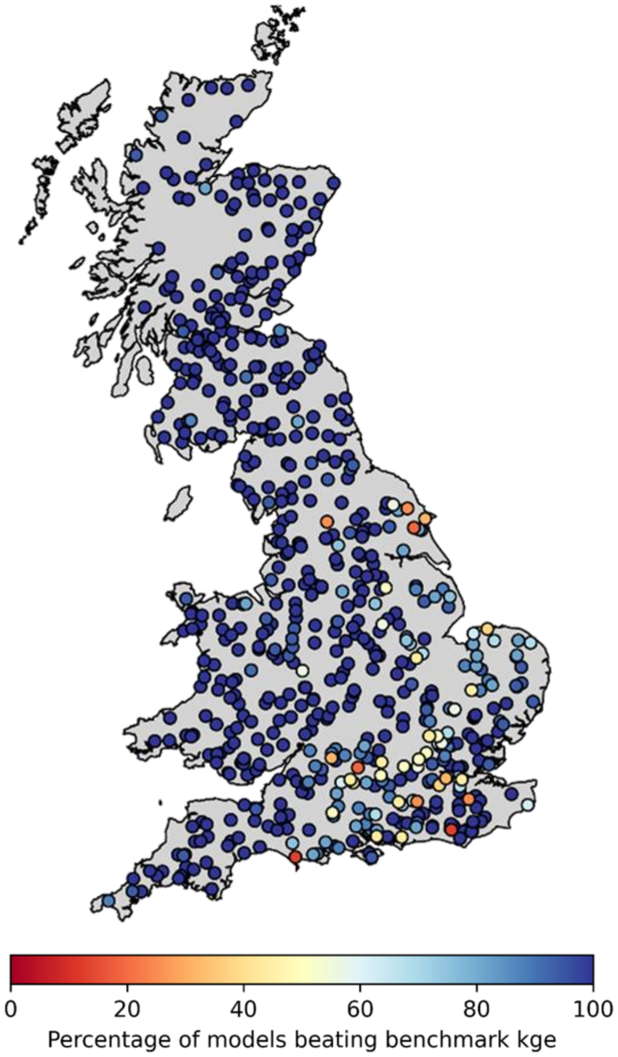
- Models do not always add value compared to our simple daily mean flow benchmark…



Do we beat the daily mean flow benchmark?

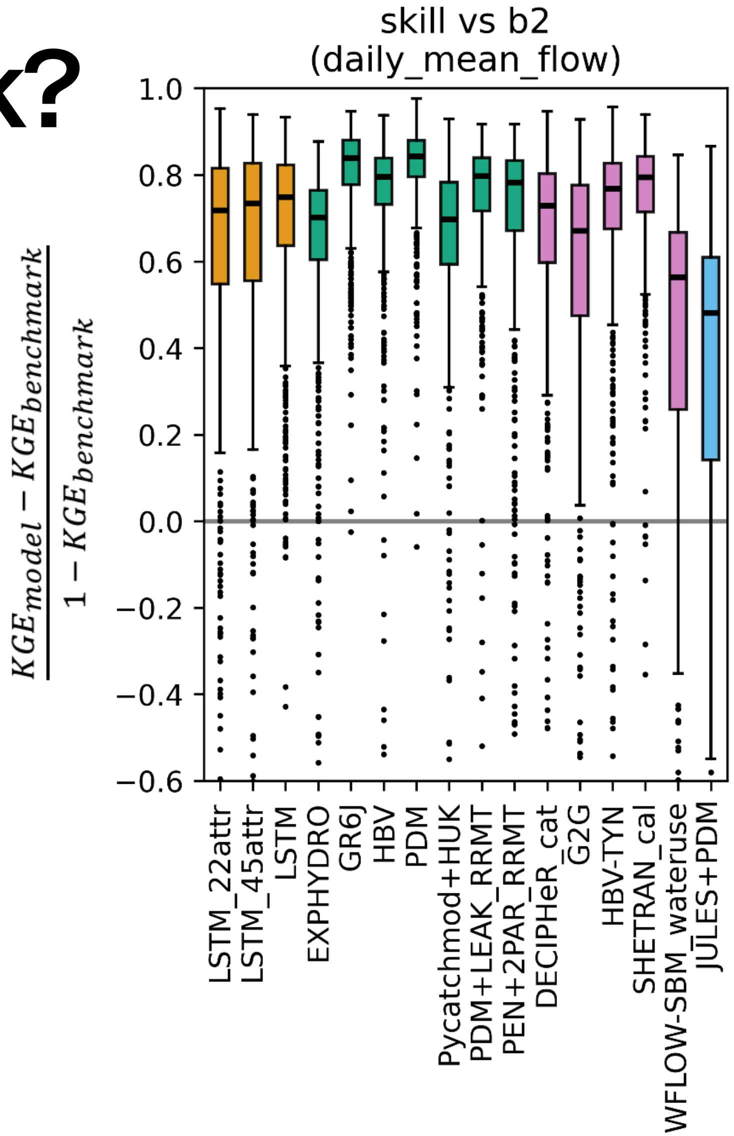
- Models do not always add value compared to our simple daily mean flow benchmark…
- Especially for a band of catchments across the southeast.

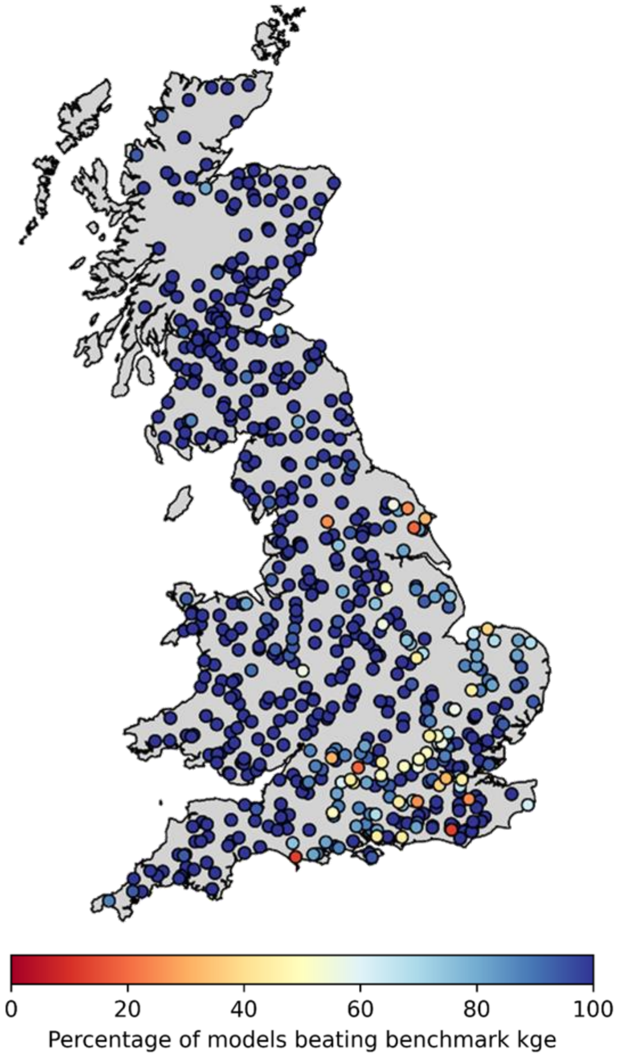




Do we beat the daily mean flow benchmark?

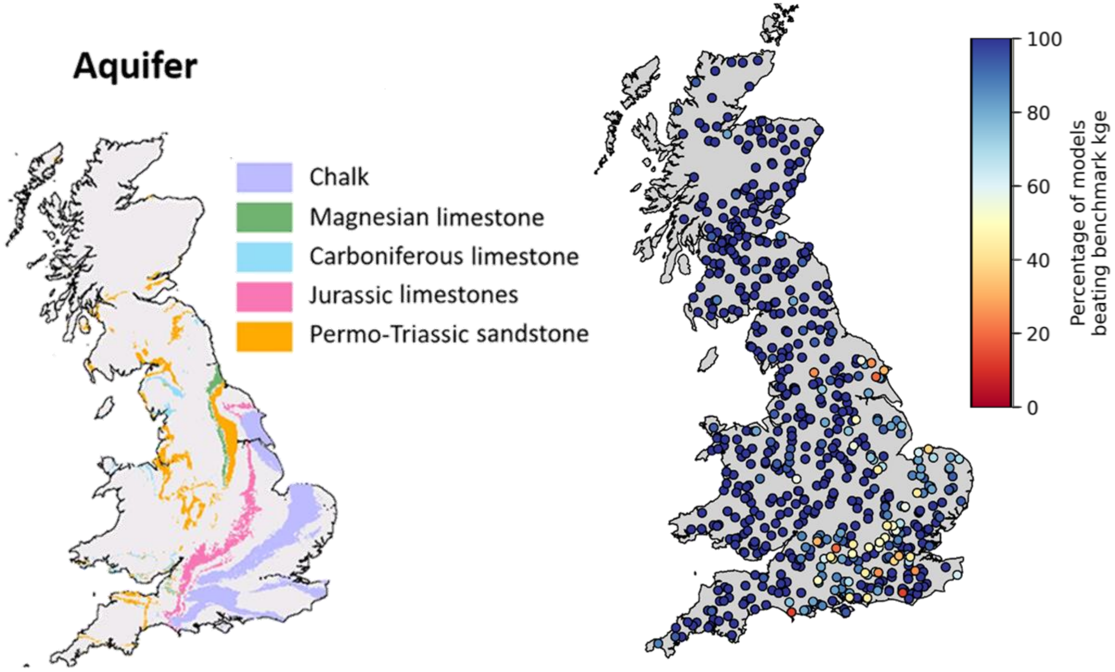
- Models do not always add value compared to our simple daily mean flow benchmark
- Especially for a band of catchments across the southeast
- What properties of these catchments make them challenging to model?





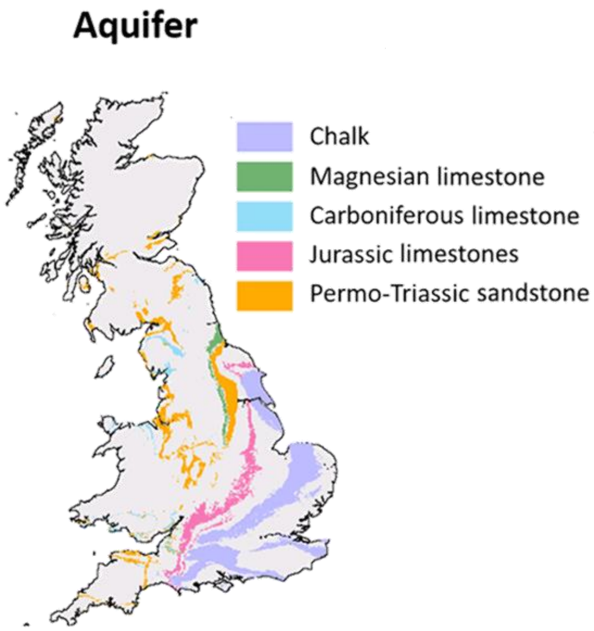
Common characteristics of catchments where models fail to beat the benchmark:

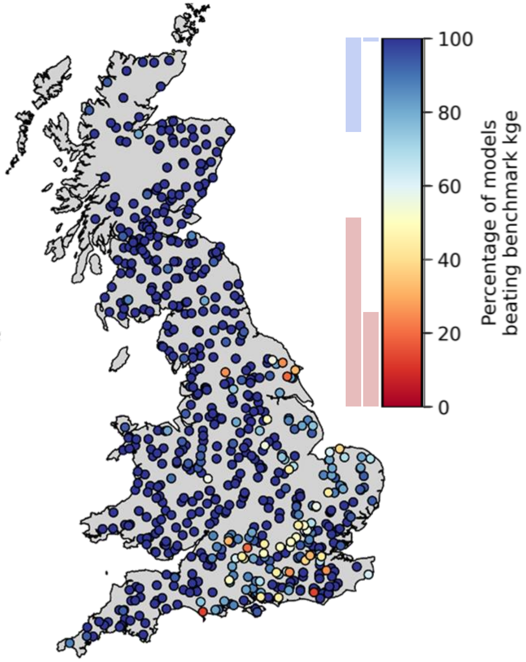
- 1. High baseflow / presence of highly productive aquifers.

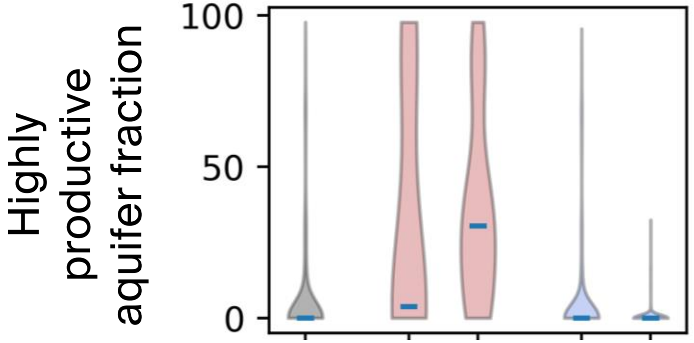


Common characteristics of catchments where models fail to beat the benchmark:

- 1. High baseflow / presence of highly productive aquifers.







Violin plots show distributions across catchments:

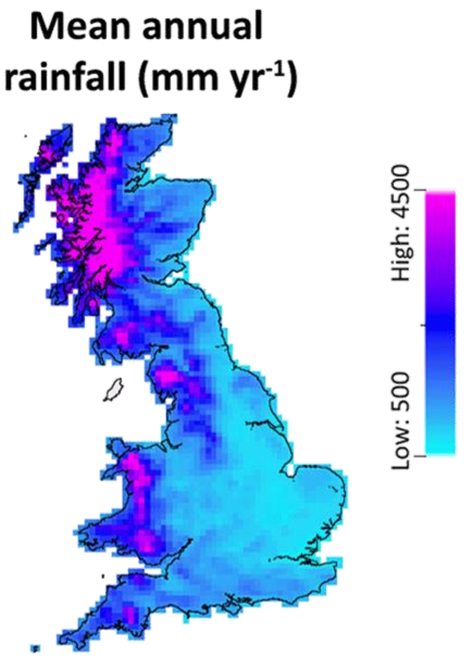
All

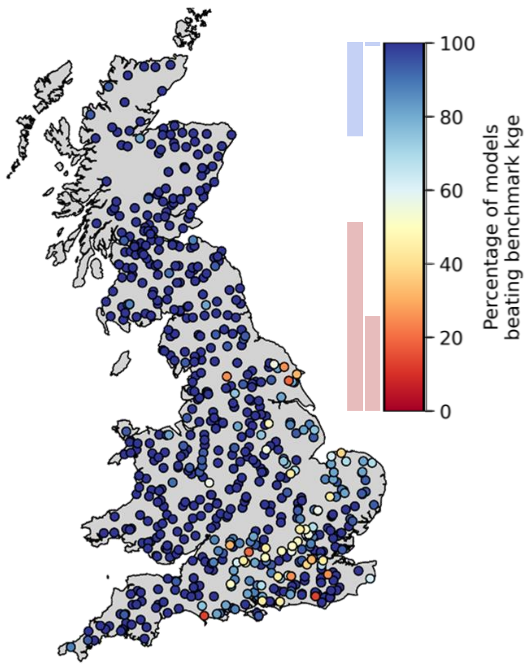
Where models struggle (<25% or <50% beat benchmark)

Where models do well (>75% or 100% beat benchmark)

Common characteristics of catchments where models fail to beat the benchmark:

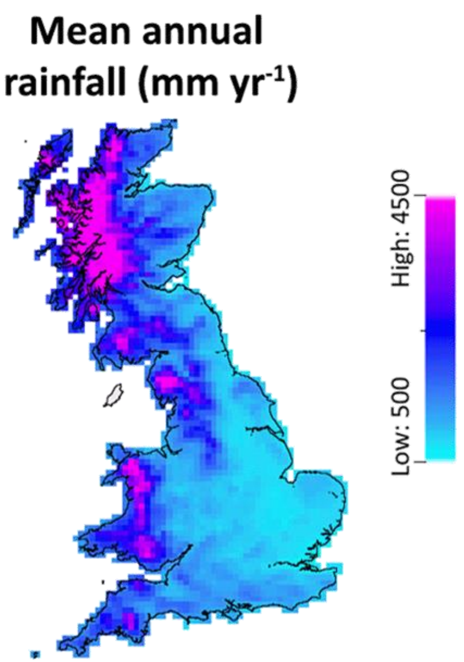
- 1. High baseflow / presence of highly productive aquifers.
- 2. Drier catchments

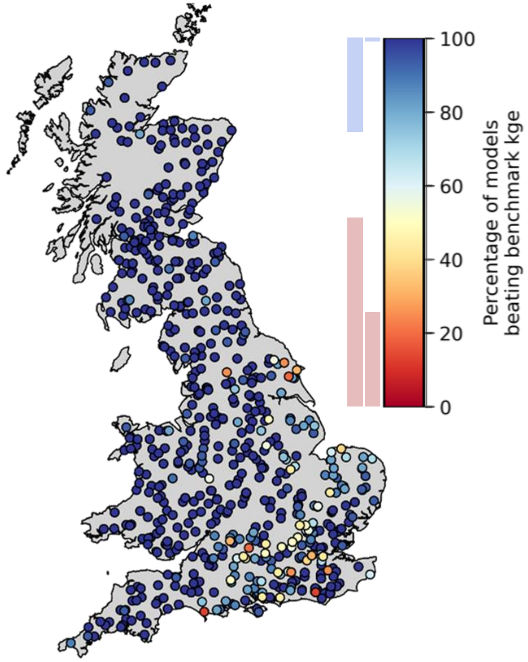


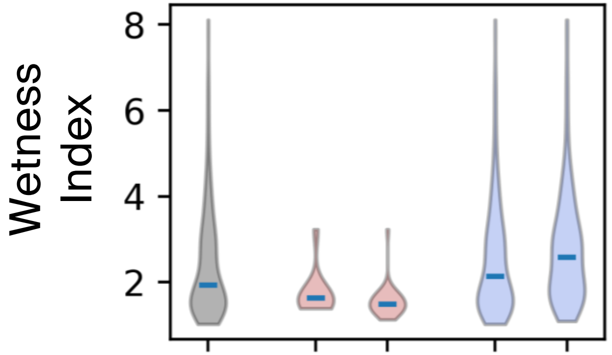


Common characteristics of catchments where models fail to beat the benchmark:

- 1. High baseflow / presence of highly productive aquifers.
- 2. Drier catchments





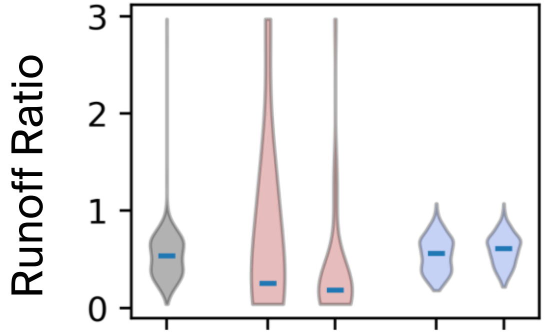


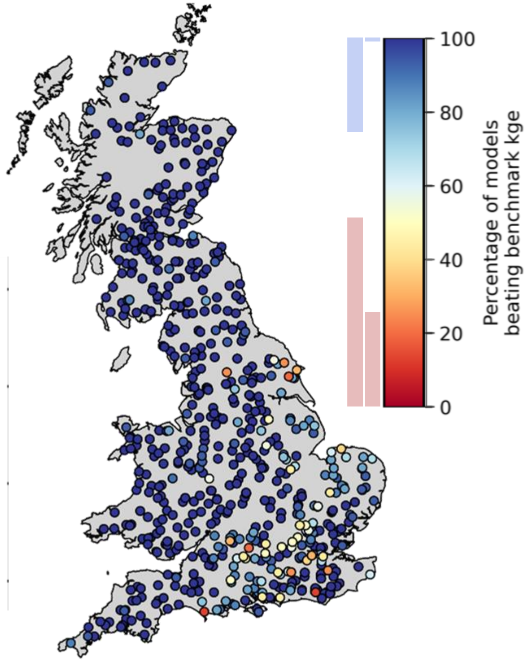
Violin plots show distributions across catchments:

- All
- Where models struggle (<25% or <50% beat benchmark)
- Where models do well (>75% or 100% beat benchmark)

Common characteristics of catchments where models fail to beat the benchmark:

- 1. High baseflow / presence of highly productive aquifers.
- 2. Drier catchments
- 3. Low (or very high) runoff ratios.



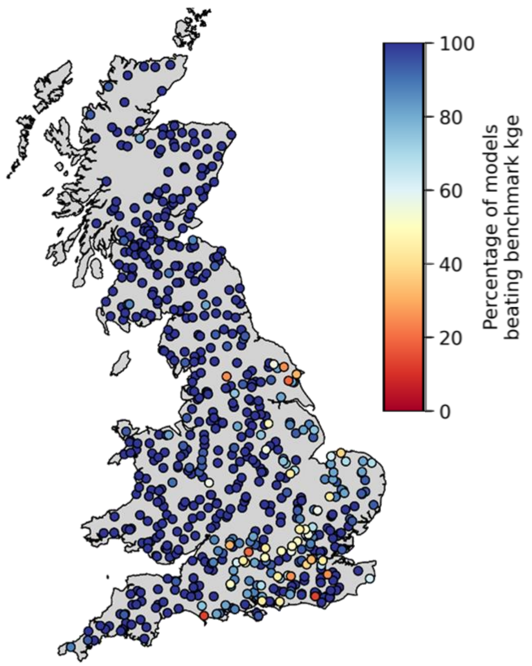


Violin plots show distributions across catchments:

- All
- Where models struggle (<25% or <50% beat benchmark)
- Where models do well (>75% or 100% beat benchmark)

Other issues that make catchments challenging

We’ve looked at catchment characteristics that are challenging to models on a large scale, but there are also catchment-specific issues

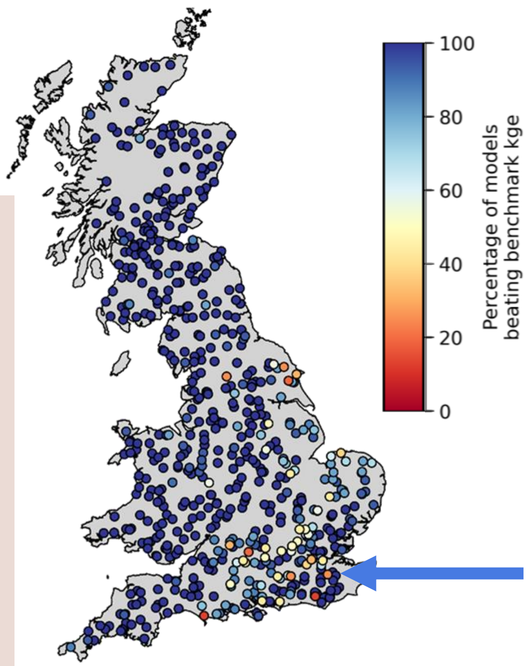


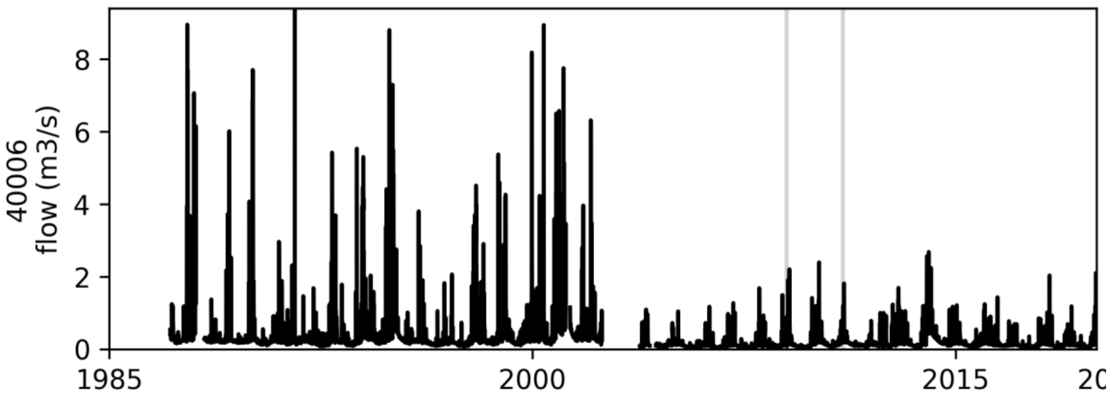
Other issues that make catchments challenging

We've looked at catchment characteristics that are challenging to models on a large scale, but there are also catchment-specific issues, e.g.

- Unreliable/ poor quality observations

Electromagnetic (EM) gauge installed 2002 – step change in the record. Flow estimates for both the original flume and EM are poor.



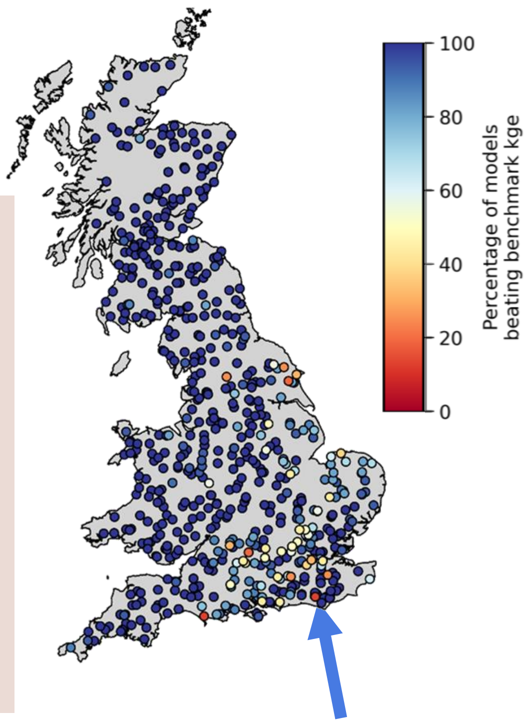


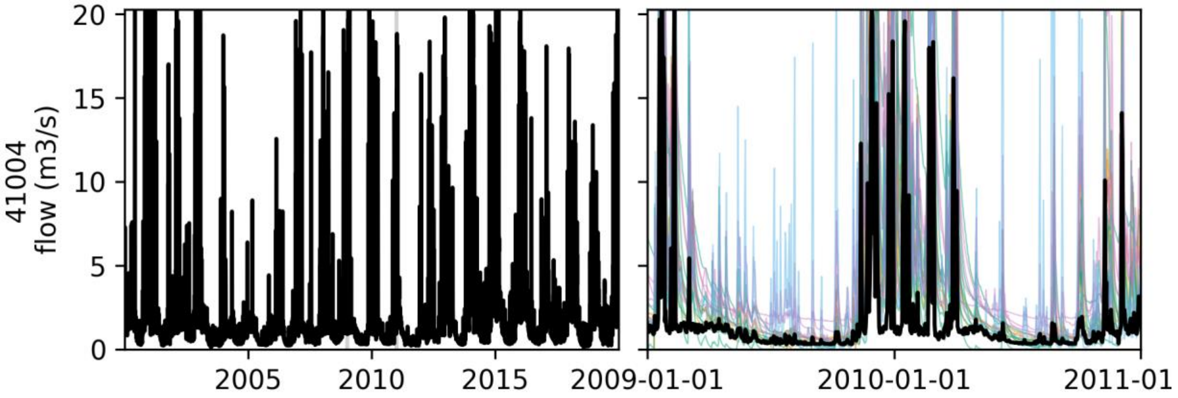
Other issues that make catchments challenging

looked at catchment characteristics that are challenging to models on a large scale, but there are also catchment-specific issues, e.g.

- Unreliable/ poor quality observations
- Processes missing from models (e.g. reservoirs, abstractions/discharges, urban drainage)

High flow measurement problems, major abstraction immediately upstream, and reservoir regulates flows

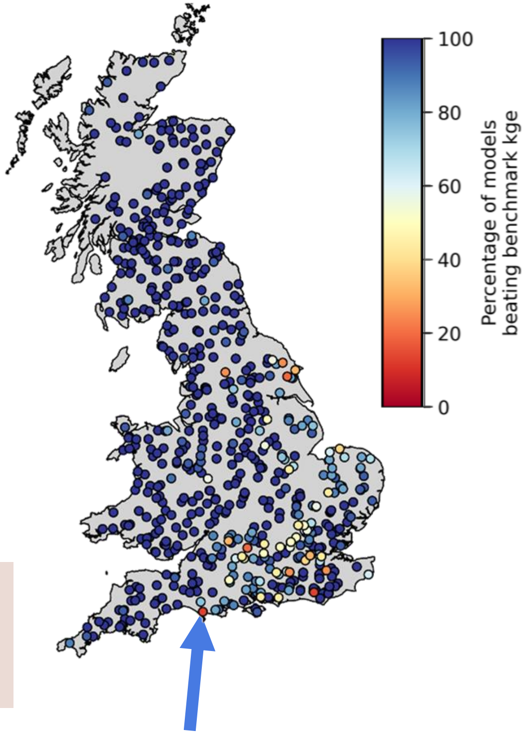




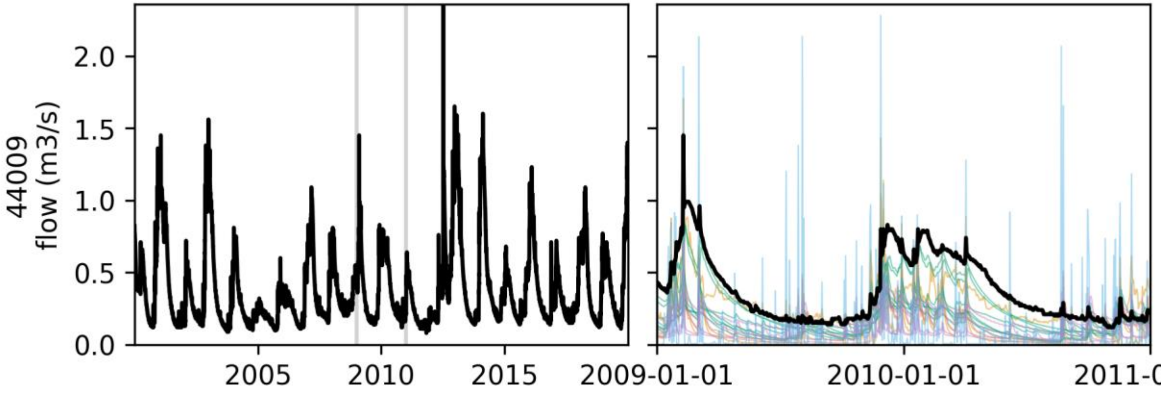
Other issues that make catchments challenging

looked at catchment characteristics that are challenging to models on a large scale, but there are also catchment-specific issues, e.g.

- Unreliable/ poor quality observations
- Processes missing from models (e.g. reservoirs, abstractions/discharges, urban drainage)
- Scale mismatch



Catchment area only 7 km^2



Summary

- The UK Hydro-MIP compared simulated flows from machine-learning, hydrological and land surface models, all following the same protocol
- The highest model performance is seen for calibrated lumped models (KGE) – but models have other strengths
- Models add value compared to the daily mean flow benchmark for the majority of catchments
- Some common challenges: modelling catchments that were drier, overlaying chalk aquifers, very high/low runoff ratios, or data quality issues.

Thank you to all who have contributed to and engaged with the UK Hydro-MIP!

Gabriele Arduini, Olivia Atkins, Ioana Balabasciuc, Helen Baron, Giovanni Bernardi, Segolene Berthou, Stephen Birkinshaw, Eleanor Blyth, Jamie Brown, Burak Bulut, Wilson Chan, Amulya Chevuturi, Douglas Clark, Steven Cole, Elizabeth Cooper, Gemma Coxon, Simon Dadson, Brendan Dalmijn, Gurpreet Das, Richard Davis, Laura Devitt, Adi Dhillon, Rich Ellis, Katie Facer-Childs, Sonja Folwell, Emma Ford, Nicole Forstenhaeusler, Matt Fry, Neven Fučkar, Helen Harfoot, Verity Kington, Melike Kiraz-Safari, Hadush Meresa, Rosanna Lane, Elizabeth Lewis, Eugene Magee, Ritesh Moon, Robert Moore, Simon Moulds, Sopan Patil, Christel Prudhomme, Ponnambalam Rameshwaran, Clarissa Rizzo, Emma Robinson, Heather Rumbold, Lorna Shires, Ben Smith, Paula Stella Teixeira, Mike Vaughan, Gianni Vesuviano, John Wallbank, Albrecht Weerts, Tom Willis, Yanchen Zheng, Qiuyu Zhu

UKCEH

Our planet.
Decoded.