

The Hubverse: Streamlining Collaborative Infectious Disease Modeling

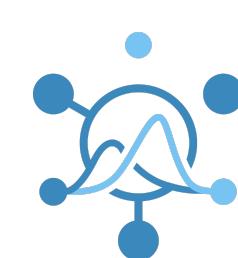
US-RSE Conference 2025

Anna Krystalli 

info@r-rse.eu

R-RSE SMPC

Consortium of Infectious Disease Modeling Hubs



7 October 2025

Background

✖ The problem

Infectious disease modeling has scaled rapidly...

- But the landscape is fragmented:
 - Inconsistent formats
 - Redundant or conflicting forecasts
 - Lack of coordination between modelers and stakeholders

“Comparing the accuracy of forecasting applications is difficult because forecasting methods, forecast outcomes, and reported validation metrics varied widely.”

- Chretien et al., PLOS ONE, 2014

✨ The promise of modeling hubs

Modeling hubs coordinate collaborative forecasting:

- Provide **centralised location** for effort coordination
- Define **data standards** and **modeling targets**
- Improve **transparency** and **comparability**
- Aggregate forecasts **enabling ensembles**
- Facilitate timely **public health decision-making**

***"Collaborative Hubs: Making the Most of Predictive Epidemic Modeling"**, American Journal of Public Health Reich, et al. 2022*

⌚ Project origins

- **Pre-COVID:** Forecasting code base existed for CDC **influenza** hubs
- **During COVID:** That code was reused for new **COVID-19** hubs + demand internationally (e.g. Europe) for similar setups
- ! Problem: Each hub required **manual editing** of source code

➡ Need for generalisation, modularity, and configurability

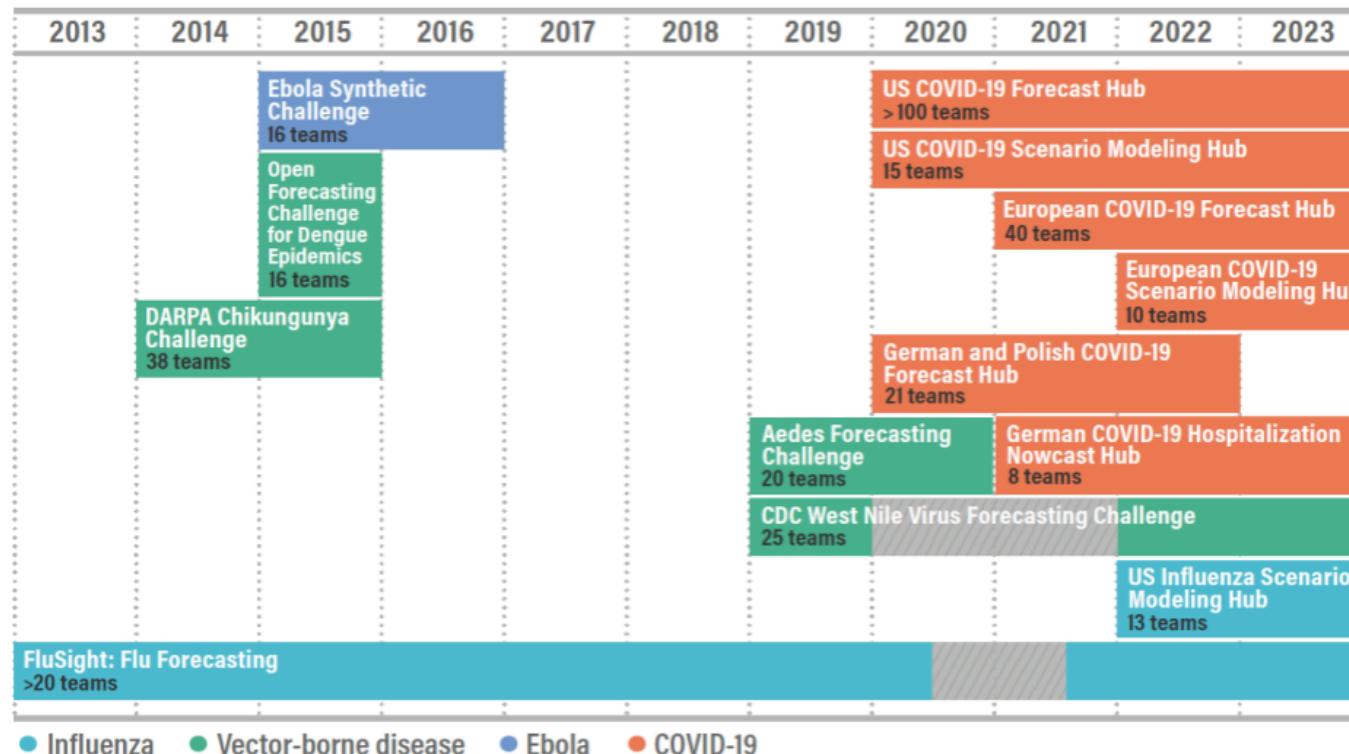


Figure credits: Alex Vespignani and Nicole Samay

Enter the hubverse

An open-source **software ecosystem** to power modeling hubs:

- GitHub repositories for **centralising hub activity**
- **Data standards** for infectious disease modeling data
- **Schema-driven configuration** for modeling tasks + hub setup
- Modular tools for validation, access, evaluation, ensembling, communication and hub administration
- **Supports full lifecycle:** from hub set up, data submission to decision-making

Hubverse overview

Standardised Data

Modeling hubs are **built around a shared data standard**:

- **Modeling task definition:** targets (response variables), standard predictors, output types (e.g. [mean](#), [quantiles](#))
- **Structured hub layout:** consistent file system for organizing submissions
- **Standard model output format:** for file content and naming

 Enables comparability, validation, and streamlined data access



Config-driven hub setup

Hub administrators configure hubs using structured JSON config files:

- `admin.json`: hub-level metadata.
- `task.json`: modeling task specification:
 - **Task IDs**: Targets (response), horizons, locations (predictors) etc.
 - **Output types**: accepted model outputs e.g. `mean`, `median`, `quantiles`, `cdf`, `pmf`, `samples`.
- Configs are validated against a shared JSON `schema`

The R (and friends) package stack

The hubverse package ecosystem is organized by **role**. Each tool is designed to support a particular group of users in the hub workflow.

Hub roles

-  Hub administrators
-  Modelers
-  Analysts
-  Policy makers

Tools & packages

- [hubAdmin](#) : config creation + validation 
- [hubValidations](#) : submission checks (structure, schema, content)  
- [hubData](#) (R)  / [hubdata](#) (Python) : access multi-file model output via Arrow    
- [hubEvals](#) : compute evaluation metrics   
- [hubEnsembles](#) : build weighted/unweighted ensembles   
- [hubVis](#) : visualise model outputs   

Dashboards & communication

- Built with [Quarto](#) so easily customisable via Quarto configuration
- Deployed as a **fully static site**, no backend required
- Powered by JSON data prepared via GitHub workflows
- Interactive UI built with client-side JavaScript (fast!)
- New instances can be set up by copying/configuring the [hub-dashboard-template](#)



Cloud hub storage and access

- Hubs **mirrored to public AWS S3 buckets**
- Multi-file data can be **opened as Arrow datasets**
- Enables **query-able data access** via R  `hubData` and Python  `hubdata`.

GitHub workflows

We automate everything we can:

-  PR-level model output validation
-  Hub configuration validation
-  AWS Cloud hub data synching
-  dashboard data regeneration and model evaluation with each update

All hubverse actions stored in the [hubverse-actions](#) repo and can be installed with `hubCI :: use_hub_github_action()`



List of adopting hubs

🔗 <https://hubverse.io/community/hubs.html>

Search

List of Hubs

Current Hubs

The hubs listed below are all active. Many of these hubs have **open source data** available and ready to analyze in hubverse format.

Do you have a hub, but it's not listed? [Tell us about your hub!](#)

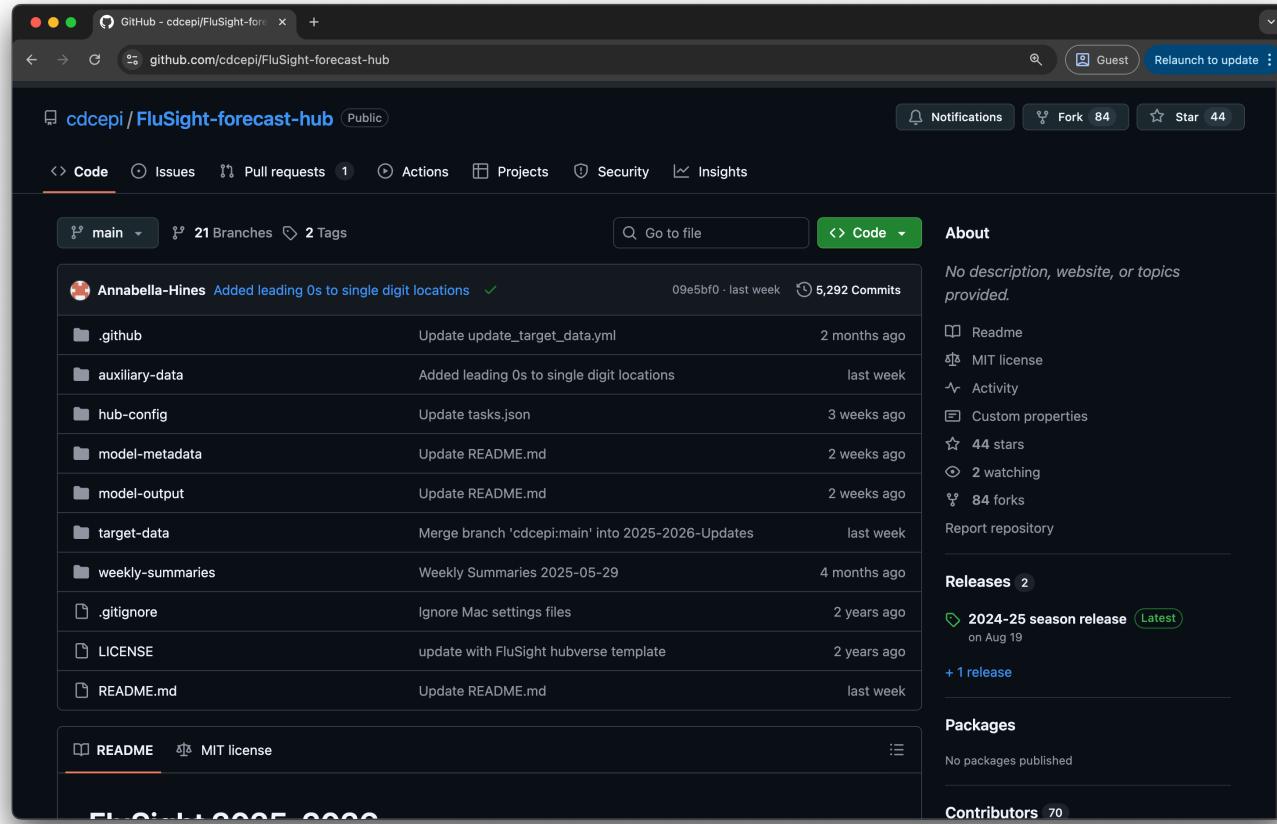

US Centers for Disease Control



Real-world example: CDC FluSight Hub

Real-world example: CDC FluSight Hub

 <https://github.com/cdcepi/FluSight-forecast-hub>



The screenshot shows the GitHub repository page for `cdcepi/FluSight-forecast-hub`. The repository is public and has 5,292 commits. The commit history includes updates to `update_target_data.yml`, `tasks.json`, `README.md`, and weekly summaries. The repository also contains files like `.github`, `auxiliary-data`, `hub-config`, `model-metadata`, `model-output`, `target-data`, `weekly-summaries`, `.gitignore`, `LICENSE`, and `README.md`. The repository has 44 stars, 84 forks, and 2 releases. The latest release is the "2024-25 season release" from August 19, 2024.

- Used by US CDC to monitor influenza severity
- Weekly forecasts from 40 teams across 70 different models.
- Hosted on GitHub + S3 cloud mirror.
- Managed using full hubverse stack since 2023/2024 season.





File structure: model output (CDC FluSight)

Model outputs committed by teams to versioned directories > one directory per model > one file per modeling round.

FluSight-forecast-hub/model-... X +

github.com/cdcepi/FluSight-forecast-hub/blob/main/model-output/CADPH-FluCAT_Ensemble/2023-10-14-CADPH-FluCAT_Ensemble.csv

Notifications Fork 84 Star 44 Relaunch to update :

cdcepi / FluSight-forecast-hub Public

Code Issues Pull requests Actions Projects Security Insights Copy path

Files

main Go to file

One file for each round

CADPH-FluCAT_Ensemble

2023-10-14-CADPH-FluCAT_...
2023-10-21-CADPH-FluCAT_...
2023-10-28-CADPH-FluCAT_...
2023-11-04-CADPH-FluCAT_...
2023-11-11-CADPH-FluCAT_...
2023-11-18-CADPH-FluCAT_...
2023-11-25-CADPH-FluCAT_...
2023-12-02-CADPH-FluCAT_...
2023-12-09-CADPH-FluCAT_...
2023-12-16-CADPH-FluCAT_...
2023-12-23-CADPH-FluCAT_...
2023-12-30-CADPH-FluCAT_...
2024-01-06-CADPH-FluCAT_...
2024-01-13-CADPH-FluCAT_...
2024-01-20-CADPH-FluCAT_...

FluSight-forecast-hub model-output CADPH-FluCAT_Ensemble / **2023-10-14-CADPH-FluCAT_Ensemble.csv** Model ID Round ID Model ID White and White 2023-10-14 CADPH-FluCAT forecast c8b9ecb · 2 years ago History

Preview Code Blame 116 lines (116 loc) · 9.06 KB Raw

Task IDs	Output Type	Predicted value						
reference_date	target	horizon	target_end_date	location	output_type	output_type_id	value	
1	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.01	39.9515155457952
2	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.025	40.8861508526788
3	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.05	41.6981179874296
4	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.1	42.6341620242193
5	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.15	43.2577513366505
6	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.2	43.7493906806784
7	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.25	44.1734137688546
8	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.3	44.5573279620668
9	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.35	44.9120344544207
10	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.4	45.2470881870173
11	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile	0.45	45.5762623516782
12	2023-10-14	wk inc flu hosp	-1	2023-10-07	06	quantile		

✓ Model output validation with hubValidations

Model outputs submitted through PRs and validated through GitHub Actions

Pull requests · cdcepj/FluSight-forecast-hub · GitHub

<https://github.com/cdcepj/FluSight-forecast-hub/pulls?q=is%3Aopen>

91 Next >

- Update target-hospital-admissions.csv
 - #2267 by Annabella-Hines was merged on Jun 18
 - #2266 by jcblemai was closed on Jun 5
- Weekly Summaries 2025-05-29
 - #2265 by Annabella-Hines was merged on May 29
- Add new ensemble forecasts
 - #2264 by github-actions bot was merged on May 29
- 2025-05-31 Forecast
 - #2263 by cvanbommel was merged on May 29
- MIGHTE 2025-05-31
 - #2262 by fl16180 was merged on May 29
- 2025-05-31
 - #2261 by ShitalAdhikari was merged on May 29
- Update Forecasting Submission 0528
 - Gatech-ensemble
 - #2260 by LJC-FVNR was merged on May 29
- NIH TS Model 2025-05-31
 - #2259 by aperofsky was merged on May 29
- results for 2025-05-31
 - #2258 by hieuvt29 was merged on May 29
- UVAFluX models
 - #2257 by aniruddhadiga was merged on May 29

Code Issues Pull requests Actions Projects Security Insights

← Hub Submission Validation (R) ✓ SGroup-RandomForest submission for Jan 13th, 2024 #620

Summary

Jobs validate-submission

validate-submission succeeded yesterday in 1m 9s

Search logs

Set up job 9s

Run actions/checkout@v3 9s

Run r-lib/actions/setup-r@v2 1s

Update R 5s

Run r-lib/actions/setup-r-dependencies@v2 27s

Run validations 15s

```

1  ➤ Run library("hubValidations")
20
21 — Individual check results —
22
23 Notice: ✓ FluSight-forecast-hub: All hub config files are valid.
24 ✓ 2024-01-13-SGroup-RandomForest.csv: File exists at path
25   'model-output/SGroup-RandomForest/2024-01-13-SGroup-RandomForest.csv'.
26 ✓ 2024-01-13-SGroup-RandomForest.csv: File name
27   "2024-01-13-SGroup-RandomForest.csv" is valid.
28 ✓ 2024-01-13-SGroup-RandomForest.csv: File directory name matches `model_id` metadata in file name.
29 ✓ 2024-01-13-SGroup-RandomForest.csv: `round_id` is valid.
30 ✓ 2024-01-13-SGroup-RandomForest.csv: File is accepted hub format.
31 ✓ 2024-01-13-SGroup-RandomForest.csv: Metadata file exists at path
32   'model-metadata/SGroup-RandomForest.yml'.
33 ✓ 2024-01-13-SGroup-RandomForest.csv: File could be read successfully.
34 ✓ 2024-01-13-SGroup-RandomForest.csv: `round_id_col` name is valid.
35 ✓ 2024-01-13-SGroup-RandomForest.csv: `round_id` column "reference_date"
36 contains a single, unique round ID value.
37 ✓ 2024-01-13-SGroup-RandomForest.csv: All `round_id_col` "reference_date" values match submission `round_id` from file name.
38 ✓ 2024-01-13-SGroup-RandomForest.csv: Column names are consistent with expected
39 round task IDs and std column names.
40 ✓ 2024-01-13-SGroup-RandomForest.csv: Column data types match hub schema.
41 ✓ 2024-01-13-SGroup-RandomForest.csv: `tbl` contains valid values/value
42 combinations.
43 ✓ 2024-01-13-SGroup-RandomForest.csv: All combinations of task ID
44 column `/output_type` / `output_type_id` values are unique.
45 ✓ 2024-01-13-SGroup-RandomForest.csv: Required task ID/output type/output type
46 ID combinations all present.
47 ✓ 2024-01-13-SGroup-RandomForest.csv: Values in column `value` all valid with
48 respect to modeling task config.
51 ✓ 2024-01-13-SGroup-RandomForest.csv: Values in `value` column are
52 non-decreasing as output_type_ids increase for all unique task ID
53 value/output type combinations of quantile or cdf output types.
54 ✓ 2024-01-13-SGroup-RandomForest.csv: Values in `value` column do sum to 1 for
55 all unique task ID value combination of pmf output types.
56 ✓ 2024-01-13-SGroup-RandomForest.csv: Time differences between t0 var
57   'reference_date' and t1 var 'target_end_date' all match expected period of 7d
58   0H 0M 0S * `horizon`.
59 ✓ 2024-01-13-SGroup-RandomForest.csv: Target counts are less than location
60   population sizes.
61 ✓ 2024-01-13-SGroup-RandomForest.csv: Submission time is within accepted
62   submission window for round.
63
64 — Overall validation result —
65 ✓ All validation checks have been successful.

    ➤ Post Run r-lib/actions/setup-r-dependencies@v2 0s
    ➤ Post Run actions/checkout@v3 0s
    ➤ Complete job 0s
  
```

hubverse

Accessing model output via hubData

Connect to Arrow dataset of forecast submissions

```

1 library(hubData)
2
3 hub_path <- s3_bucket(
4   "cdcepi-flusight-forecast-hub"
5 )
6 hub_con <- connect_hub(
7   hub_path,
8   skip_checks = TRUE
9 )
10 hub_con

```

```

hub_connection
9 columns
reference_date: date32[day]
target: string
horizon: int32
target_end_date: date32[day]
location: string
output_type: string
output_type_id: string
value: double
model_id: string

```

Query and collect data

```

1 # Filter for one model and forecast date using dplyr
2 library(dplyr)
3 hub_con |>
4   filter(
5     model_id == "CADPH-FluCAT_Ensemble",
6     target_end_date == "2023-10-28"
7   ) |>
8   collect_hub()

```

```

# A tibble: 92 × 9
  model_id    reference_date target horizon target_end_date location output_type
* <chr>        <date>      <chr>   <int>   <date>          <chr>    <chr>
1 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
2 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
3 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
4 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
5 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
6 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
7 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
8 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
9 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
10 CADPH-Flu... 2023-10-14   wk in...     2 2023-10-28      06 quantile
# i 82 more rows
# i 2 more variables: output_type_id <chr>, value <dbl>

```

See more in [Accessing data vignette](#).

Python analogue [hub-data](#) also available.

Ensembling with hubEnsembles

Combine models using simple or weighted rules

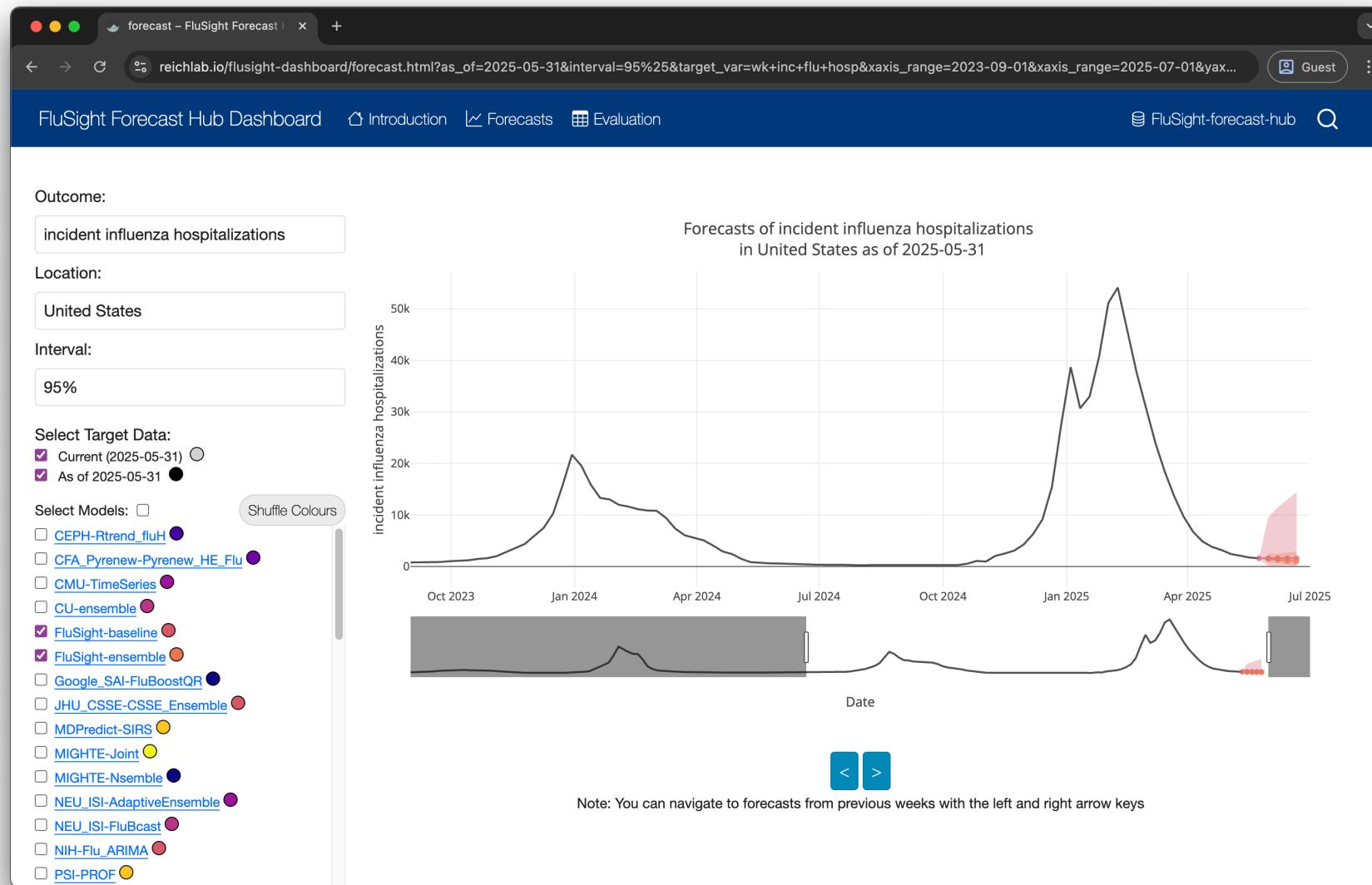
```

1 forecast_df <- hub_con |>
2   filter(
3     model_id %in%
4     c(
5       "CADPH-FluCAT_Ensemble",
6       "CEPH-Rtrend_fluH",
7       "CFA_Pyrenew-Pyrenew_HE_Flu"
8     ),
9     output_type == "quantile"
10   ) |>
11   collect_hub()
12
13
14 hubEnsembles::simple_ensemble(
15   forecast_df,
16   agg_fun = median,
17   model_id = "simple-ensemble-median"
18 )

```

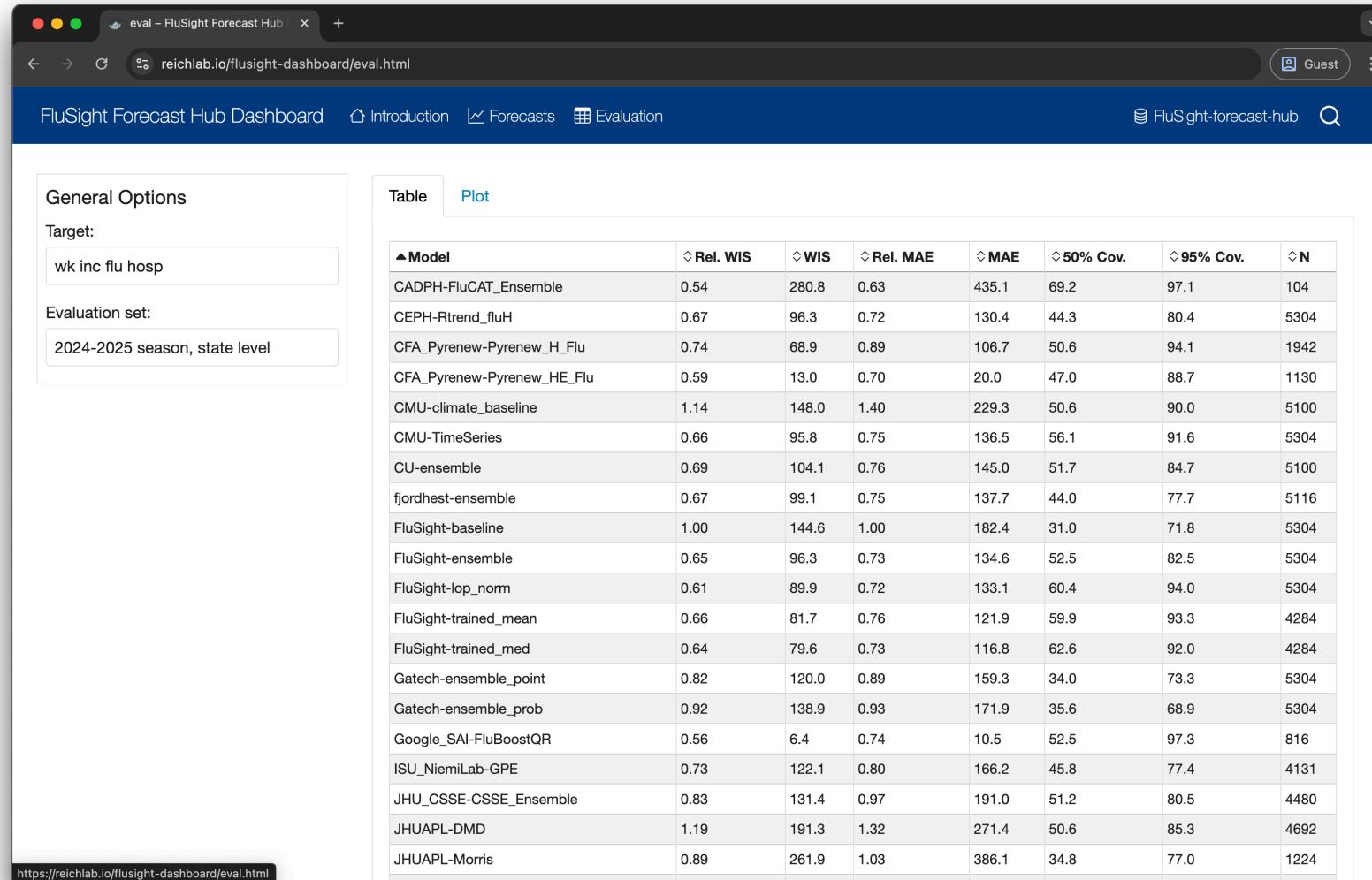
```
# A tibble: 282,716 × 9
  model_id reference_date target_horizon target_end_date location output_type
* <chr>      <date>        <chr>      <int> <date>        <chr>      <chr>
1 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
2 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
3 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
4 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
5 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
6 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
7 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
8 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
9 simple-en... 2023-10-14 wk in...      -1 2023-10-07    01 quantile
10 simple-en... 2023-10-14 wk in...      -1 2023-10-07   01 quantile
# i 282,706 more rows
# i 2 more variables: output_type_id <chr>, value <dbl>
```

Dashboard - forecasts



Dashboard - model evaluations

Evaluates forecasts against target (observed) data.



The screenshot shows a web browser window for the FluSight Forecast Hub Dashboard at reichlab.io/flusight-dashboard/eval.html. The dashboard has a dark blue header with tabs for 'Introduction', 'Forecasts', 'Evaluation', and a search bar. On the left, there's a sidebar with 'General Options' for 'Target' (set to 'wk inc flu hosp') and 'Evaluation set' (set to '2024-2025 season, state level'). The main content area is titled 'Evaluation' and contains two tabs: 'Table' (selected) and 'Plot'. The 'Table' tab displays a table of evaluation metrics for various models. The columns are labeled: Model, Rel. WIS, WIS, Rel. MAE, MAE, 50% Cov., 95% Cov., and N. The table includes rows for models like CADPH-FluCAT_Ensemble, CEPH-Rtrend_fluH, CFA_Pyrenew-Pyrenew_H_Flu, etc., with values ranging from 0.54 to 1.19 and sample sizes from 104 to 5304.

Model	Rel. WIS	WIS	Rel. MAE	MAE	50% Cov.	95% Cov.	N
CADPH-FluCAT_Ensemble	0.54	280.8	0.63	435.1	69.2	97.1	104
CEPH-Rtrend_fluH	0.67	96.3	0.72	130.4	44.3	80.4	5304
CFA_Pyrenew-Pyrenew_H_Flu	0.74	68.9	0.89	106.7	50.6	94.1	1942
CFA_Pyrenew-Pyrenew_HE_Flu	0.59	13.0	0.70	20.0	47.0	88.7	1130
CMU-climate_baseline	1.14	148.0	1.40	229.3	50.6	90.0	5100
CMU-TimeSeries	0.66	95.8	0.75	136.5	56.1	91.6	5304
CU-ensemble	0.69	104.1	0.76	145.0	51.7	84.7	5100
fjordhest-ensemble	0.67	99.1	0.75	137.7	44.0	77.7	5116
FluSight-baseline	1.00	144.6	1.00	182.4	31.0	71.8	5304
FluSight-ensemble	0.65	96.3	0.73	134.6	52.5	82.5	5304
FluSight-lop_norm	0.61	89.9	0.72	133.1	60.4	94.0	5304
FluSight-trained_mean	0.66	81.7	0.76	121.9	59.9	93.3	4284
FluSight-trained_med	0.64	79.6	0.73	116.8	62.6	92.0	4284
Gatech-ensemble_point	0.82	120.0	0.89	159.3	34.0	73.3	5304
Gatech-ensemble_prob	0.92	138.9	0.93	171.9	35.6	68.9	5304
Google_SAI-FluBoostQR	0.56	6.4	0.74	10.5	52.5	97.3	816
ISU_NiemiLab-GPE	0.73	122.1	0.80	166.2	45.8	77.4	4131
JHU_CSSE-CSSE_Engsemble	0.83	131.4	0.97	191.0	51.2	80.5	4480
JHUAPL-DMD	1.19	191.3	1.32	271.4	50.6	85.3	4692
JHUAPL-Morris	0.89	261.9	1.03	386.1	34.8	77.0	1224

Lessons & wider relevance

-  Standards + automation reduce friction
-  Open source keeps it free & accessible
-  Collaborative infrastructure empowers public health
-  Standardised, open data fuels downstream use cases like training, education, and reproducible research

NFIDD SI... Getting started Sessions ▾ Reference ▾ Authors  

Evaluating real-world outbreak forecasts

Introduction

So far in this course we have focused on building, visualising and combining “toy” forecast models in somewhat synthetic settings. In this session you will work with real forecasts from an existing modeling framework and learn how to evaluate them. We will also discuss many of the challenges involved with real-time forecasting, including how to handle uncertainty and how to incorporate new information as it becomes available.

🔗 <https://nfidd.github.io/sismid/sessions/real-world-forecasts.html>

🙏 Thank you!

- 🌐 [@ https://hubverse.io](https://hubverse.io)
- 📄 [📄 https://docs.hubverse.io](https://docs.hubverse.io)
- 💬 [💬 hubverse-org](https://hubverse-org)
- 📩 [✉️ info@r-rse.eu](mailto:info@r-rse.eu)



Tip

Interested in getting involved in the community? Check out our [Getting Involved](#) page!