CIROH Testbed for Advancing Hydrological Prediction: Templates

Andy Wood, Colorado School of Mines

Updated: Dec 8, 2023

# Overview

The CIROH Hydrologic Prediction Testbed (CHPT) is an initiative to establish rigorous, quantitative intercomparison and benchmarking operational hydrologic forecasts, including the multiple elements (models, methods, datasets) involved in producing them. It is led through a CIROH-funded project at the Colorado School of Mines (Mines), and collaborates with multiple CIROH forecasting projects as well as external research and operational partners. The CHPT promotes overall benchmark-based development paradigm, and collects and tracks the performances of innovations in each area. It defines multiple community protocols for different forecasting objectives and also forecasting sub-component objectives to facilitate consistent intercomparison and evaluation of innovations arising from distinct projects across CIROH, as benchmarked against operational and non-operational baseline capabilities, as appropriate. As it matures, the CHPT will enable the US to quantify and track the current performance of its hydrologic forecasting capabilities, many for the first time, enabling evidence-driven decisions regarding the adoption of new forecast elements into operational practice.

# Testbed Implementation

Each Testbed protocol documents the following elements:

* **Objective or focus**: e.g., predictands, analyses, forecast information
* **Observational datasets** that can be used for validation and verification
* **Reference capabilities** against which to assess strength of innovation (i.e., the baselines)
* **Experimental design**:  the periods, catchments or domain, lead times if applicable, and other relevant details to ensure consistency in evaluating candidate innovations
* **Metrics** of performance (absolute and relative)
* **Other factors** to be taken into consideration in the capability evaluation (e.g., computational expense, portability, complexity, dependencies, CONUS-wide relevance, potential operational latency, and so on)
* **Key references**, studies, and relevant external activities
* **Contact(s): L**ead(s) or active contributors, e.g., helping connect new participants, identify new relevant work, ensuring that testbed results are up-to-date, and so on.
* **Metadata:** Machine-readable information structure to exactly identify each element used in each innovation within a protocol, including details such as model or dataset version, algorithm variation, parameter set variation, generation date, author, among others.

The experimental protocols are used to test and benchmark innovations – e.g., datasets, methods, modeling strategies – arising from different research efforts. Some of the protocols are led and coordinated by the CHPT lead(s), and other protocols are contributed by other CIROH projects and leads. The core design objective of the protocols is to enable the evaluation of a wide range of project outcomes related to the specific protocol objective, hence each protocol is expected to be jointly developed and refined by groups of relevant researchers and partners, with an opportunity for review by the broader CIROH community.

The following pages contain templates for defining the experimental Testbed protocols.

## P-#. Short/Descriptive Title

### Protocol

|  |  |  |
| --- | --- | --- |
| Element | | Description |
| Focus / Objective |  | |
| Observation(s) |  | |
| Reference capabilities (baselines) |  | |
| Experimental design |  | |
| Metrics of performance |  | |
| Other considerations |  | |
| References and related activities |  | |
| Metadata elements |  | |
| Contact(s) |  | |

### Innovation(s) Tested

|  |  |
| --- | --- |
| Element | Description |
| Short title |  |
| Lead(s) and institution(s) |  |
| Description |  |
| Metadata |  |
| References / Links |  |

## Reference Tables

|  |  |
| --- | --- |
| Element | Description |
| Focus / Objective | Predictands, analyses, simulation capability, forecast product, method, dataset |
| Observation(s) | Observations used for validation and verification |
| Reference capabilities (baselines) | Baseline or reference capabilities against which to assess strength of innovation, prioritizing operational baselines as well as current Nextgen capability baselines. |
| Experimental design | The experiment that will be conducted for each innovation, including details such as the time periods, catchments or domain, lead times if applicable, and other relevant details to ensure consistency in evaluating candidate innovations |
| Metrics of performance | Metrics or statistics that will be used to assess the quality of the innovation, including both absolute measures of performance and relative measures – i.e.., skill scores relative to a baseline |
| Other considerations | Other factors to be taken into consideration in the capability evaluation (e.g., computational expense, portability, complexity, dependencies, CONUS-wide relevance, potential operational latency, and so on) |
| References and related or prior activities | Key references, studies, prior activities (context, background) and relevant external activities |
| Contact(s) | Lead(s) or active contributors, e.g., helping connect new participants, identify new relevant work, ensuring that testbed results are up-to-date, and so on. |
| Metadata | Machine-readable information structure to exactly identify each element used in each innovation within a protocol, including details such as model or dataset version, algorithm variation, parameter set variation, generation date, author, among others. This can be designed to configure evaluation routines, for instance. |

|  |  |
| --- | --- |
| Element | Description |
| Short title | Short descriptive title for this capability or dataset alternative |
| Lead(s) and institution(s) | Note that the leads/institutions may differ from the protocol leads |
| Description | What sets this experimental innovation apart from others that will be tested |
| Metadata flags | Machine parse-able flags for the metadata that will be associated with this innovation in the evaluation, e.g., ‘vers=b21, label=no\_veg, method=lstm3.66’ |
| References / Links |  |