5th HEFS workshop, 02/25/2014

Seminar B: hindcasting concepts and requirements

James Brown

james.brown@hydrosolved.com

Contents

1. Motivation for hindcasting
2. Data requirements
3. Experimental design
4. Hindcasting mechanics/configuration
5. Practical tips and considerations
6. Final thoughts and suggestions

1. Motivation for hindcasting

Motivation

Test and improve HEFS (offline)

- To provide a large sample for verification (inc. extremes)
- To provide a consistent sample for verification
- To explore scenarios (compare models, calibrations etc.)
- To allow HEFS users to test their systems (e.g. NYCDEP)

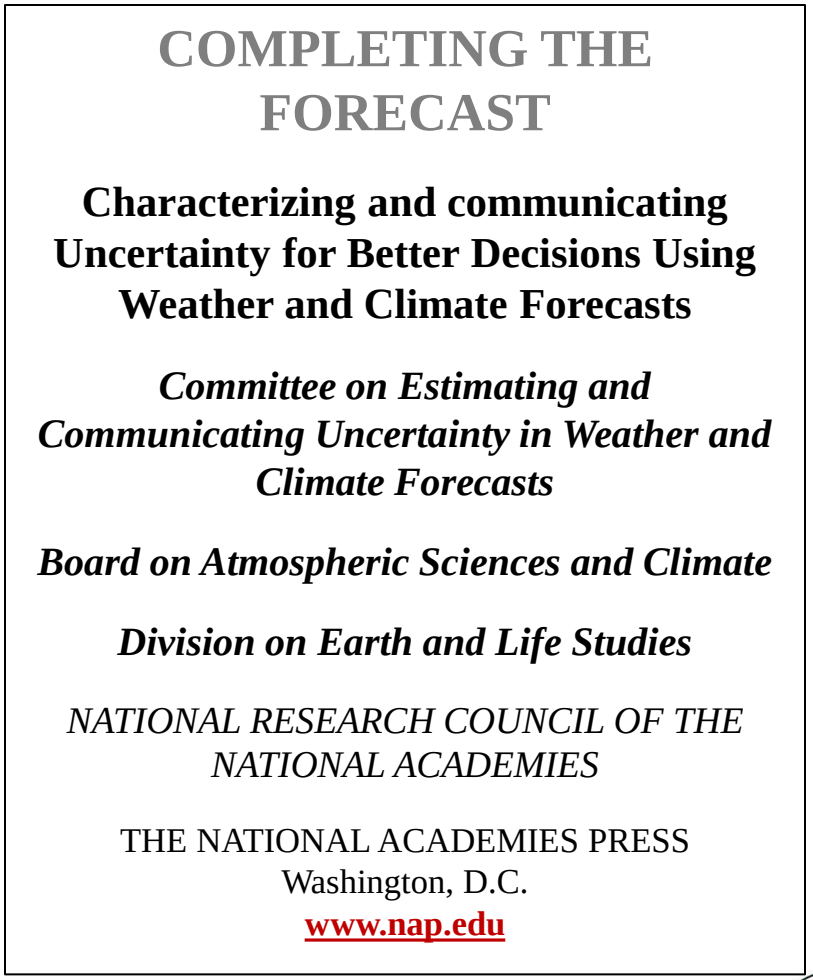
Support forecast operations (online)

- By providing improved guidance, based on experience
- By providing analogs to current forecasting situations
- By building richer products (with quality information)

Motivation: recognized externally

National Research Council, 2006

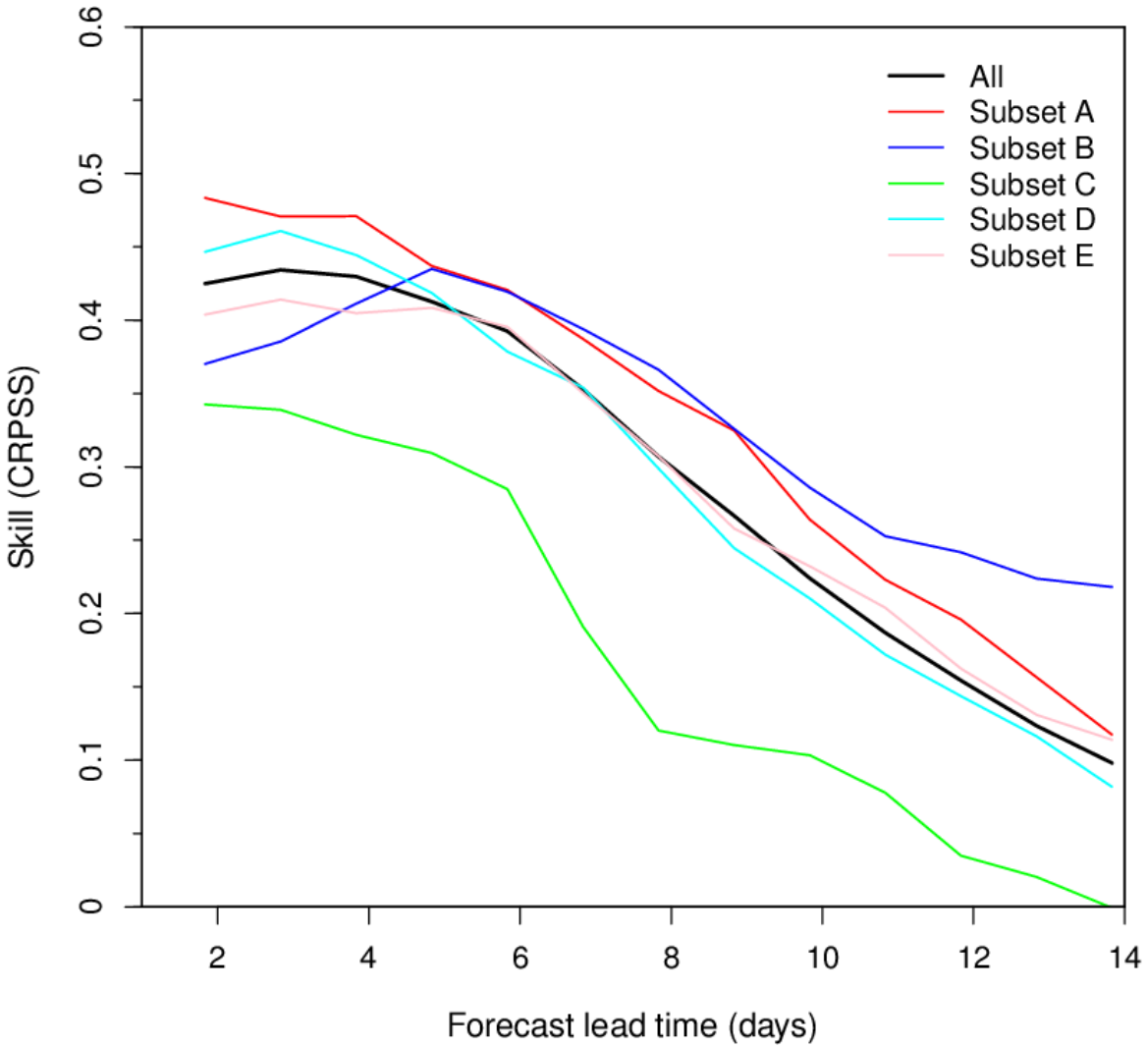
"Recommendation 6: NWS should expand verification of its uncertainty products and make this information easily available to all users in near real time. A variety of verification measures and approaches (measuring multiple aspects of forecast quality that are relevant for users) should be used to appropriately represent the complexity and dimensionality of the verification problem. Verification statistics should be computed for meaningful subsets of the forecasts (e.g. by season, region) and should be presented in formats that are understandable by forecast users. Archival verification information on probabilistic forecasts, including model-generated and objectively generated forecasts and verifying observations, should be accessible so users can produce their own evaluation of the forecasts."



Motivation: support verification

Long/consistent archive

- Skill of HEFS hindcasts with MEFP-GEFS forcing vs. resampled climatology
- FTSC1 in CNRFC
- Hindcasts for 15 years (1985-1999), also split into five, 3-year sub-periods
- RFC QPF/QTF archives may be less than 5 years
- Using a short archive could give noisy and misleading results
- Sampling uncertainties are much higher for 3-year datasets (not shown)



2. Data requirements

Data requirements

HEFS calibration data

- Hindcasting starts w/ operational setup/calibration
- Forcing: MAP/T/PE & raw forecasts (per source)
- Flow: QME/QINE and historical simulations

Other operational forecasting datasets

- Aim of hindcasting is to reproduce operations
- Archived diversions, extractions, releases?
- Other manual modifications archived?
- Probably not, but sensible to minimize differences

Data requirements

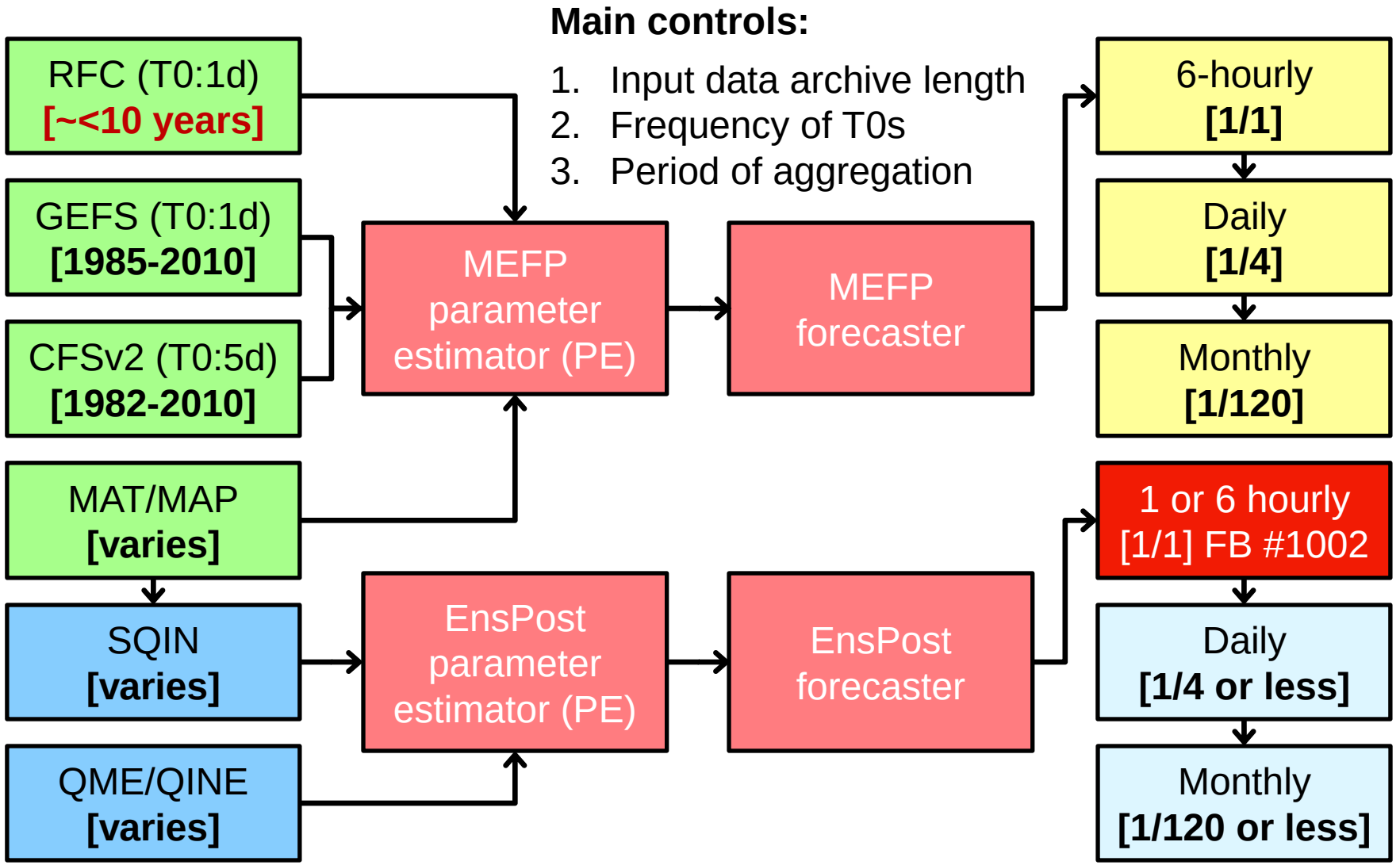
Sample requirements for verification

- Large and consistent historical sample needed
- Hindcasting improves consistency. No guarantee.
- Ideally, no climate trends, shifts in ratings etc.
- Ideally, 10+ years of data, but tough to generalize

Difficult to generalize about sample size

- Not just archive length (aggregation period etc.)
- Effective sample size declines w/ persistence
- Effective sample size declines w/ noise:signal

Sample size for inputs/outputs



Mitigating sample size issues

Less than perfect archives are a reality

- For example, RFC forcing data may be <5 years
- Observed data may be missing or inadequate etc.

What steps can be taken?

- Does the short archive add value (RFC, CFSv2)?
- Explore HEFS diagnostics to identify issues
- See calibration options (MEFP window size etc.)
- For verification, focus on lumped scores, avoid extreme thresholds, assess sampling uncertainty

Mitigating consistency issues

Consistency issues are varied/complex

- Hindcasting removes many issues (v. archiving forecasts)
- But: do hindcasts cover consistent climate & ratings?
- But: do hindcasts represent current operations?

What steps can be taken?

- Update configurations & minimize/archive mods etc.
- Repeat EnsPost calibration & hindcasting if necessary
- Check for potential inconsistencies (time consuming)
  - Check for changes in ratings, climatology etc.
  - Archive operational forecasts & compare to hindcasts

3. Hindcast experimental design

Dependent/independent validation

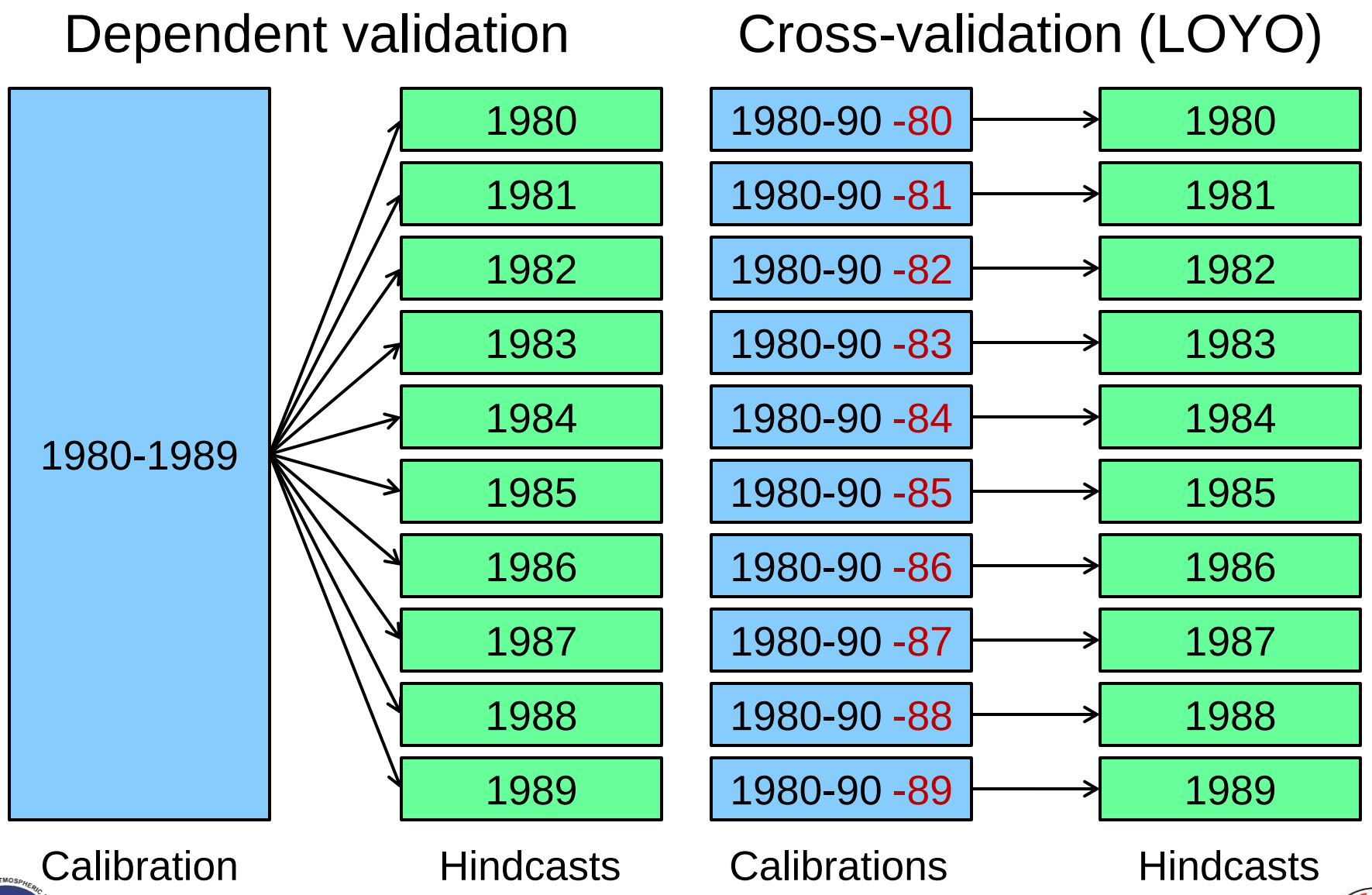
Dependent validation (practical)

- Calibration and validation periods are the same
- Advantage: simple, requires only one hindcast run
- Disadvantage: exaggerates skill, particularly for extremes
- The approach used in the phased evaluation of the HEFS

Independent validation (preferred)

- Validation period does not overlap with calibration period
- Advantage: completely independent test of system
- Disadvantage: requires multiple calibrations/hindcast runs
- Several different flavors of independent validation…

Example: cross-validation (LOYO)

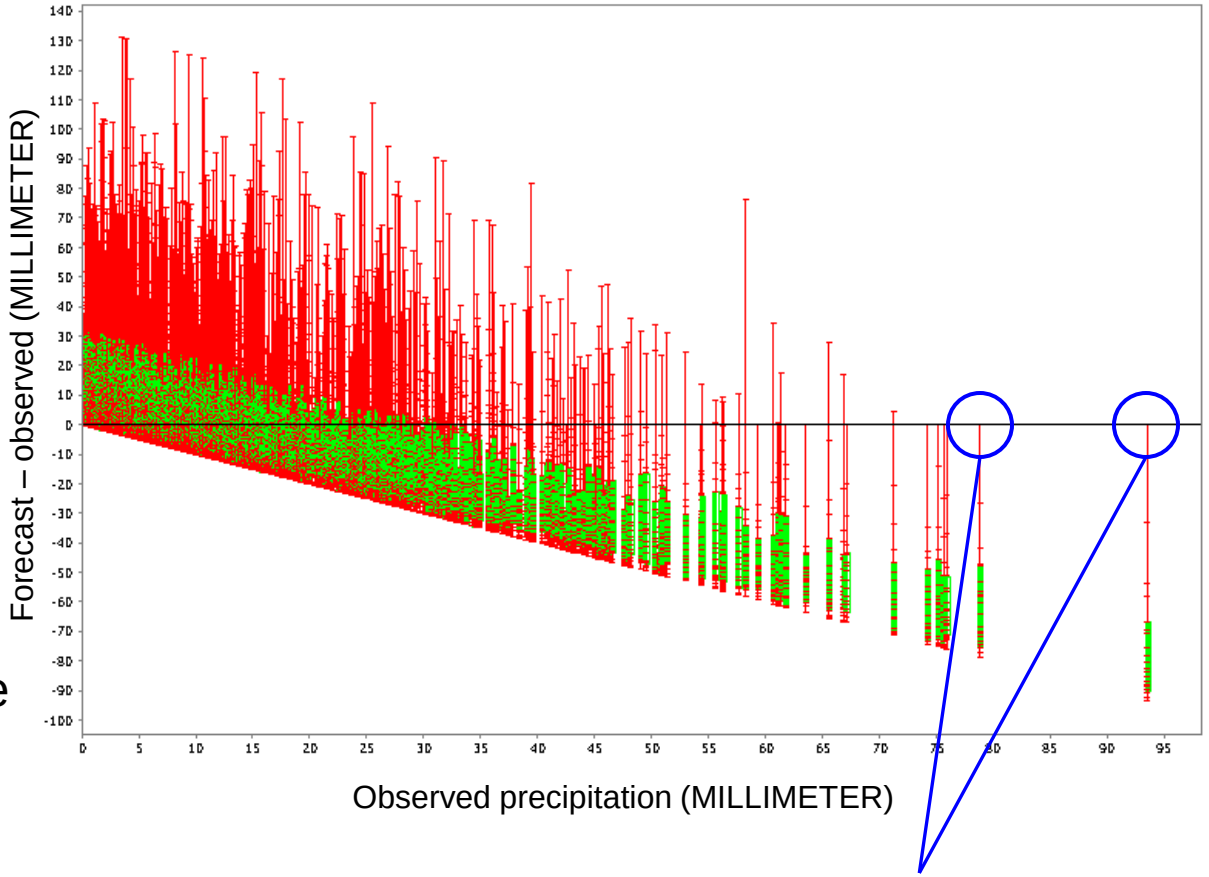


Warning: MEFP raw climate option

Dependent validation

- Example of problems with dependent validation
- MEFP "raw climatology" samples historical observations without fitting/smoothing
- In hindcast mode, one ensemble member is always equal to the verifying observation!
- Not realistic, so do not use MEFP raw climatology in hindcast mode
- Otherwise, dependent validation still "best" option

MEFP raw climatology, precipitation at WALN6

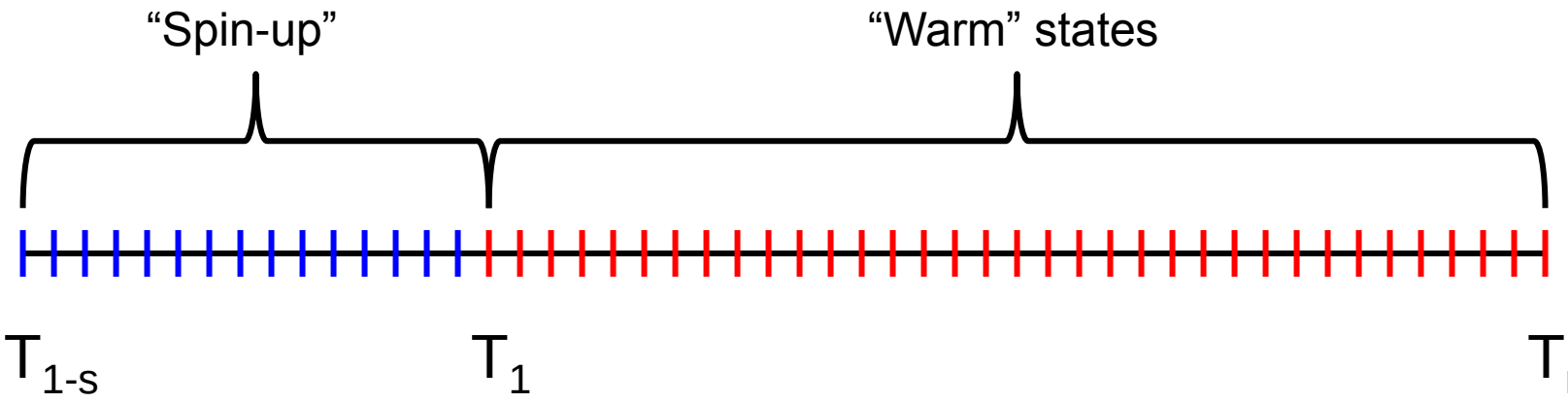


In dependent validation, one member in raw climatology always equals verifying observation!

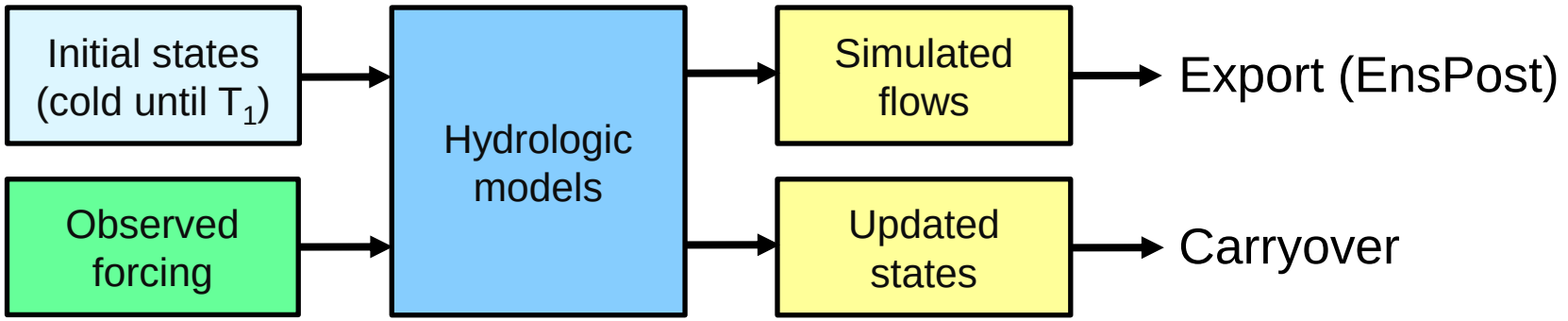
4. Hindcasting mechanics and configuration

HEFS hindcasting mechanics

STEP 1: warm states and simulations for hindcast period

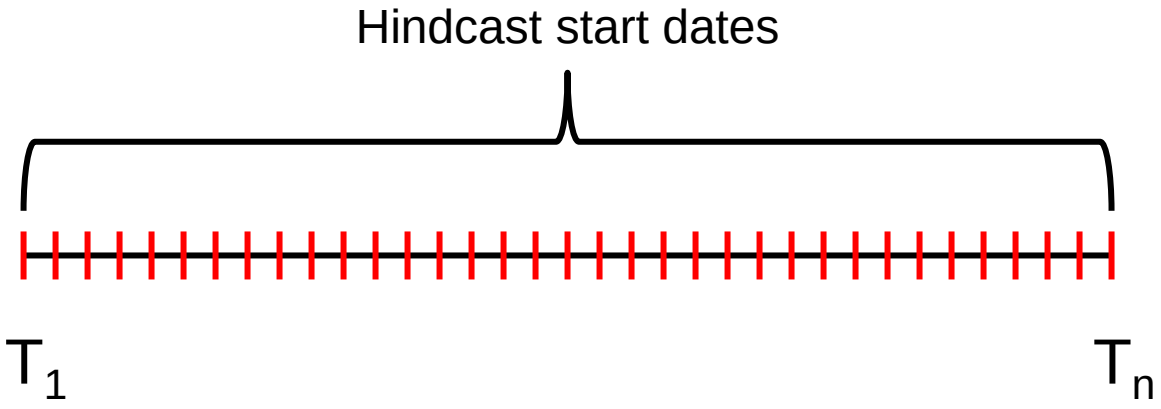


For each of T_{1-s} ,…,T_1,…,T_n:

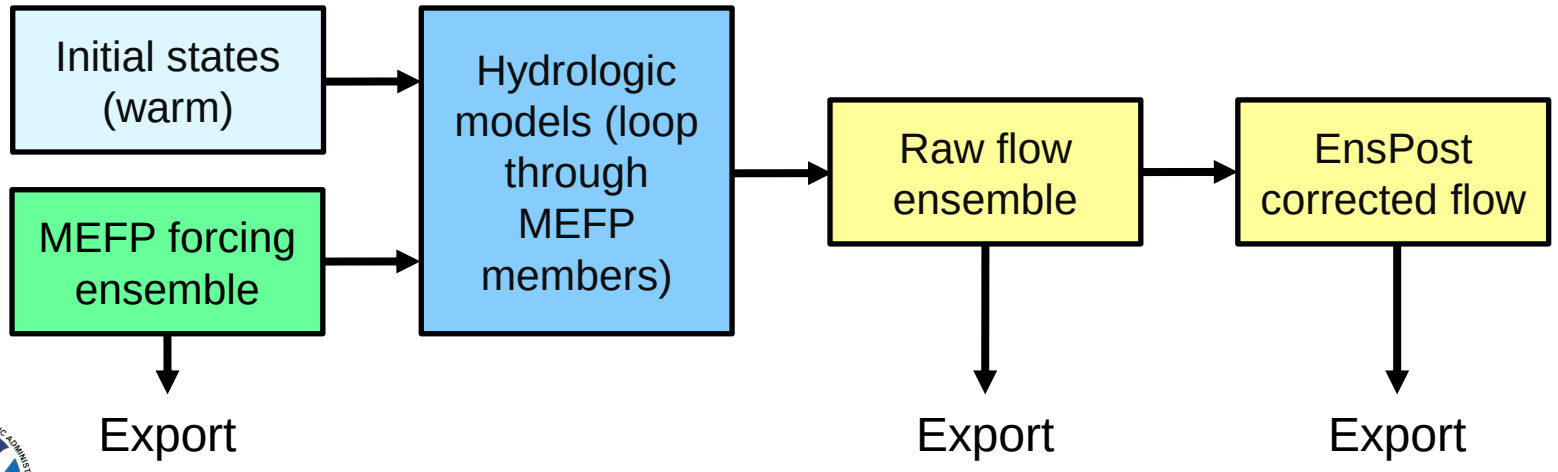


HEFS hindcasting mechanics

STEP 2: generate ensemble hindcasts



For each of $T_1,\ldots,T_n$:



Steps before hindcasting

- Entry point is a working operational HEFS
  - MEFP calibrated and configured in CHPS
  - EnsPost calibrated and configured in CHPS
  - May need to adjust operational configs (tips later)
- Required data available for hindcast period
- Warm states generated (run “UpdateStates” first)
- Hindcasting configuration developed
  - Controls order of activities, such as running models & exporting
  - Configures the exports of the hindcasting datasets
  - Assume dependent validation (single set of parameters)

Main steps (see Demo 1/Ex. 1)

Activity hierarchy (HEFS_Hindcast.xml)

1. Run MEFP I (EXISTS) →
2. Run MEFP II (EXISTS) →
3. Export MEFP **(NEW)** →
4. Run raw flow (EXISTS) →
5. Export raw flow **(NEW)** →
6. Run EnsPost (EXISTS) →
7. Export EnsPost **(NEW)** →

```
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>MEFP_Temp_Forecast</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>MEFP_Precip_Forecast</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>MEFP_Export</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>HEFS_Forecast</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>HEFS_Export</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>EnsPost_Forecast</workflowId>
</activity>
<activity>
    <runIndependent>true</runIndependent>
    <workflowId>EnsPost_Export</workflowId>
</activity>
```

5. Practical considerations and lessons learned

Run times and disk space required

- Many factors will impact resources required
  - Number of forecast points
  - Forecast scenarios (e.g. climatology, flow w/ and w/o EnsPost,…)
  - Forecast time horizon and frequency of T0s
  - Hardware (run times)
  - Output formats (ASCII or compressed)
  - EVS outputs generated (pairs, plots etc.)
  - Many others…
- Various hindcasting tests conducted at OHD…

Resources required

Example runs at OHD (hindcasting only)

Run property	Scenario 1	Scenario 2	Scenario 3
Forecast horizon (days)	15	15	365
RFC	MARFC	CNRFC	MARFC
Number of MEFP basins	14	28	14
Number of flow basins	14	15	14
Years (# years)	1988-1998 (10)	1985-1995 (10)	1988-1998 (10)
HEFS components	All (no G. Gen.)	All (no G. Gen.)	All (no G. Gen.)
Frequency of T0s (days)	1	1	5
Model timestep (hours)	6	1	6
Forcing sources	GEFS	GEFS	GEFS-CFSv2-CLIM
Runtime per T0 (mins)	1.0	1.8	4
Runtime per year (mins)	300	615	200
Total run time (mins)	3250	7680	2000
MEFP as % of run time	32	60	-
localDataStore (GB)	16	38	41
PI-XML total export (GB)	43	71	14.5

Practical tips

General tips (i.e. not RFC specific)

- Data QC
  - Use test runs (e.g. 2 yr) to screen for obvious issues
  - Check exports created for each T0 (see OHD robot script too)
  - Search (e.g. grep) for missing data in export files
  - Verify hindcasts before using them (e.g. warm states issues,…)!
- Manage disk-space requirements
  - Plan for disk-space requirements before run
  - Set time series as “temporary” (targeted but cumbersome)
  - Set workflow as “expiring” (less cumbersome, but care w/ states)
  - Export compressed .fi/.bin files (in EVS after 1.1.1 release)

General tips (i.e. not RFC specific)

- Manage runtimes
  - Configure for parallel processing (FogBugz #1150)
  - Split and parallelize runs (FEWS macros or OHD robot script)
- Manage/avoid runtime failures
  - Break CFSv2 runs around Feb 29th (skip) to retain 5-day cycle
  - Avoid timeout by increasing runtime limit (e.g. 600 secs per T0)
  - Turn off screen saver, as this can cause freezing
  - In DB viewer, use F12 + M to terminate run
- Export considerations (for EVS)
  - If possible, export files per basin/variable (to check EVS pairs)
  - Export numeric (e.g. -999) for missing values, not NaN

Practical tips

Tips that depend on individual RFCs

- Warm states search window (end at 0): e.g. MARFC
- Data import/merge considerations
  - Import 6-hourly MAP/MAT etc. as CST (inc. non-CST RFCs)
  - MERGETS with MAPX priority but no MAPX = zeroes
  - Ensemble MAPE (set to “read all forecast”): e.g. MARFC
- Hindcast configurations out of sync with operations
  - Outdated config may conflict w/ latest binaries (e.g. LagK)
  - Syncing avoids this, but raises other issues
  - For example, EnsPost may require re-calibration

6. Final thoughts and suggestions

Final thoughts

Create checklist (Hindcasting Guide)

- Think about length and consistency of historical record
- Think about consistency with operational practice
- Identify scenarios needed (inc. baselines, avoid raw clim.)
- Choose an experimental design (dependent validation)
- Adapt operational configs. for hindcasting if needed
- Generate warm states (w/ appropriate spin-up time)
- Optimize run settings (parallel process, split runs etc.)
- QC input data/parameters and conduct test runs
- QC and verify final runs (even if verification is limited)

Questions?

Extra slides

Thoughts on runtime modifications

Hindcasts vs. operational forecasts

- Many adjustments made in real-time are not archived
- Thus, hindcasts will differ from operational forecasts
- If possible, compare hindcasts & (archived) forecasts
- How to minimize runtime mods and archive others?

Calibration vs. operational forecasts

- EnsPost: need consistent historical & operational sim.
- If operational simulations differ, EnsPost not optimal
- Again, how to minimize differences or archive mods?