Verification of BC Hydro's short- and long-range inflow forecasts

CSHS Operational Forecasting Workshop
October 6-7, 2011

Adam Gobena & Frank Weber
Hydrology & Technical Services Group
Generation Resource Management

BC hydro
FOR GENERATIONS

hydrology & technical services

Outline

- Motivation & Objectives
- Data, methods, assumptions and limitations
- Highlights of results
- Summary and recommendations

Objective & Motivation

✓Verification of BC Hydro's operational reservoir inflow forecasts

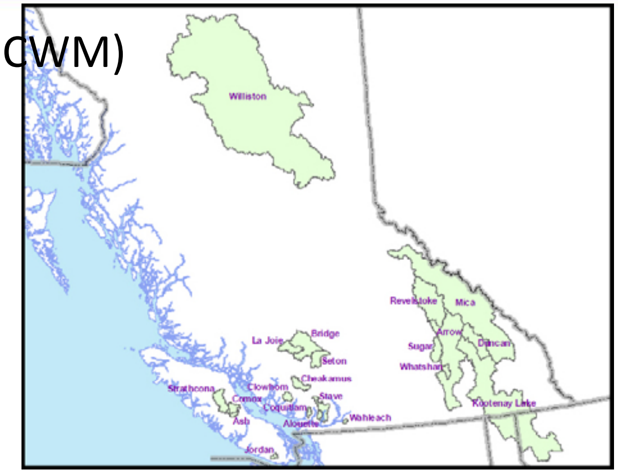
✓ As part of the project, an inflow forecast verification toolbox was designed.

- *Hydrologic Forecast Verification System (HFVS)*: Short- & long-range; deterministic & probabilistic; statistical & process-oriented model forecast verification capabilities

✓***Why do we verify forecasts?***

- User awareness
  - How good (or bad) are our forecasts?
  - How good (or bad) are competing models?
- Understand sources of skill and uncertainty in forecasts
- Identify targets when forecasts are skilful or not
- Track evolution of forecast skill over time (*not performed here*)
- Provide direction for future R&D

Operational Inflow Forecast Archive



- ✓ Operational short-range forecasts for 20 projects (UBCWM)
  - ➢ Jan 1, 2003 – Dec 31, 2010
- ✓ Long-range forecasts for Mica and Williston projects
  - ➢ ESP (UBCWM-based volume forecasts)
    - ✓ Mica: 1980 – 2010 (operational)
    - ✓ Williston: 1986 – 2010 (operational)
  - ➢ VoDCa (Statistical volume forecasts)
    - ✓ Mica: 1966 to 2002 (hindcasts) & 2007 to 2010 (operational)
    - ✓ Williston: 1973 to 2003 (hindcasts) & 2007 to 2010 (operational)
- ✓Verifying “observations” are calculated inflows; QC’d to WY2007 and raw to end of 2010

BC hydro FOR GENERATIONS

hydrology & technical services

Assumptions & Limitations

✓ Observations and forecasts were assumed to be homogeneous in time.

➢ *Short-range forecast verification*

☒Impact of any changes in weather forecasting process on inflow forecasts was not taken into account.

➢ *Long-range forecast verification*

☒ *ESP*: Impact of changes in forecast model versions, input & output processing techniques and forecasters was not accounted for.

☒ *VoDCa*: impact of hindcasts (generated with high quality data but with no forecaster updating) vs. operational forecasts (produced with non-QC'd data but with possible forecaster updating) was not accounted for.

☒ Uncertainty in verification statistics due to sample size not analyzed

✓ Process-based model forecasts are based on old to very old model versions

Analysis details

Short-range verification statistics were computed for each lead time &:

- Entire forecast

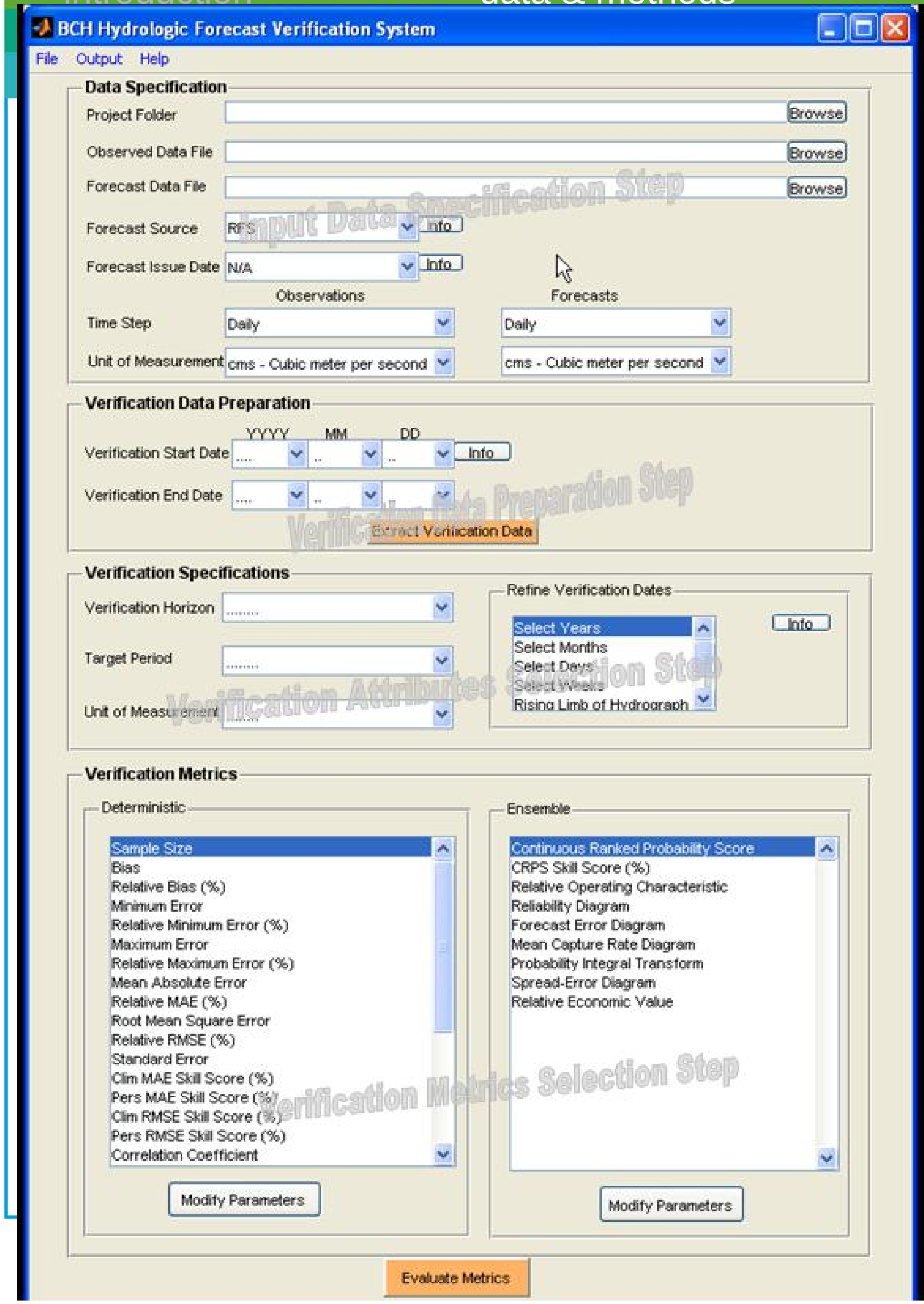
- Hydrologic seasons (Winter: Jan-Mar; Spring: Apr-Jun; Summer: Jul-Sep; Fall: Oct-Dec)

- High flows (observed flows > 90^{th} percentile of 1971-2000 climatology)

- Rising / receding flows (for observed flows > 50^{th} percentile of 1971-2000 climatology)

Long-range verification statistics were computed for:

- Entire forecast for each forecast issue date (Jan – Aug), for monthly & residual Feb-Sep targets



rics used

Metrics selected for this work emphasize forecast bias, accuracy, skill, discrimination, and reliability.

Deterministic forecast verification

***Overall performance*: Bias, Mean absolute error, MAE Skill score**

***High flows*: Hit rate, False alarm ratio**

Probabilistic forecast verification

***Overall performance*: CRPS skill score, ensemble spread-error relationships,**

Diagnostics*: PIT *(not discussed here)

hydrology & technical services

Some of the verification metrics used

Metrics selected for this work emphasize forecast bias, accuracy, skill, discrimination, and reliability.

Bias (or Mean error) – a measure of correspondence between the average observation and the average forecast; is zero for unbiased forecasts.

$$Bias = \overline{y} - \overline{o}$$

Mean absolute error (MAE) – a measure of correspondence between individual forecasts and observations; measure of accuracy; is zero for a perfect forecast.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - o_i\right|$$

Hit rate - proportion of high flow events for which a forewarning was correctly issued.

False alarm ratio - proportion of predicted high flow warnings that turned out to be wrong.

Ranked probability score – analogous to MAE but is computed in probability space; measure of accuracy for probabilistic forecasts; is zero for a perfect forecast.

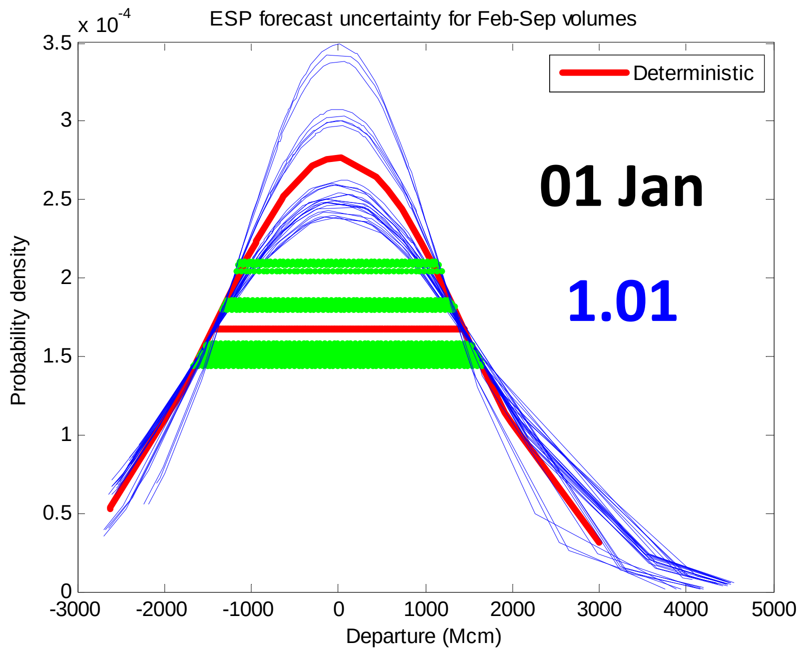
Skill score – relative accuracy measure evaluated against a control forecast; is zero or negative for a naïve forecast

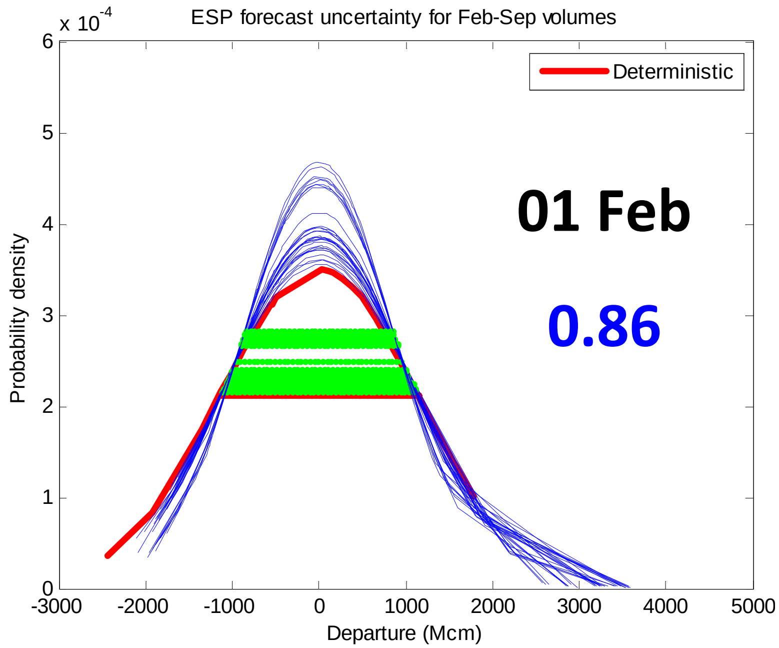
$$MAE\ SS = 1 - MAE_{fcst} / MAE_{reference}$$

Forecast Spread to Error comparison

SD/RMSE ratio as a measure of the reliability of predictive intervals

Small ensemble spread (e.g., small SD) <--> more confidence in ensemble mean (e.g., small RMSE) & the forecast in general.





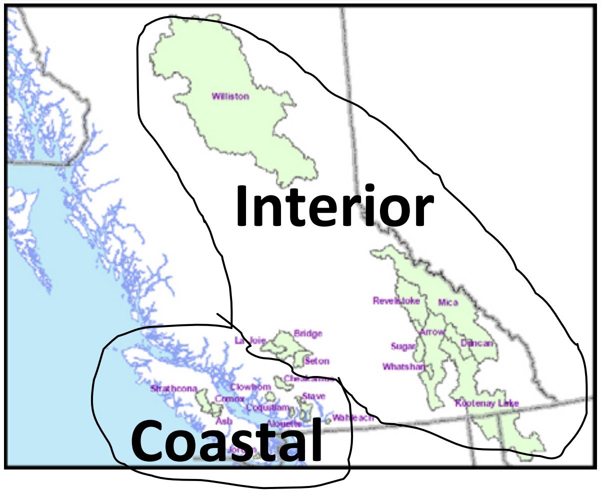
Short-range forecast verification highlights

For the purpose of this presentation, verification results were aggregated into coastal (Vancouver Island, Lower Mainland & South Coast) and interior (Bridge, Columbia/Kootenay and Williston) separately.

Something to note when interpreting regional statistics: Hydrologic characteristics for the two regions are vastly different!

Coastal basins: typically small; are characterized by primarily rainfall-driven, flashy response

Interior basins: typically large; are characterized by primarily snowmelt-driven response



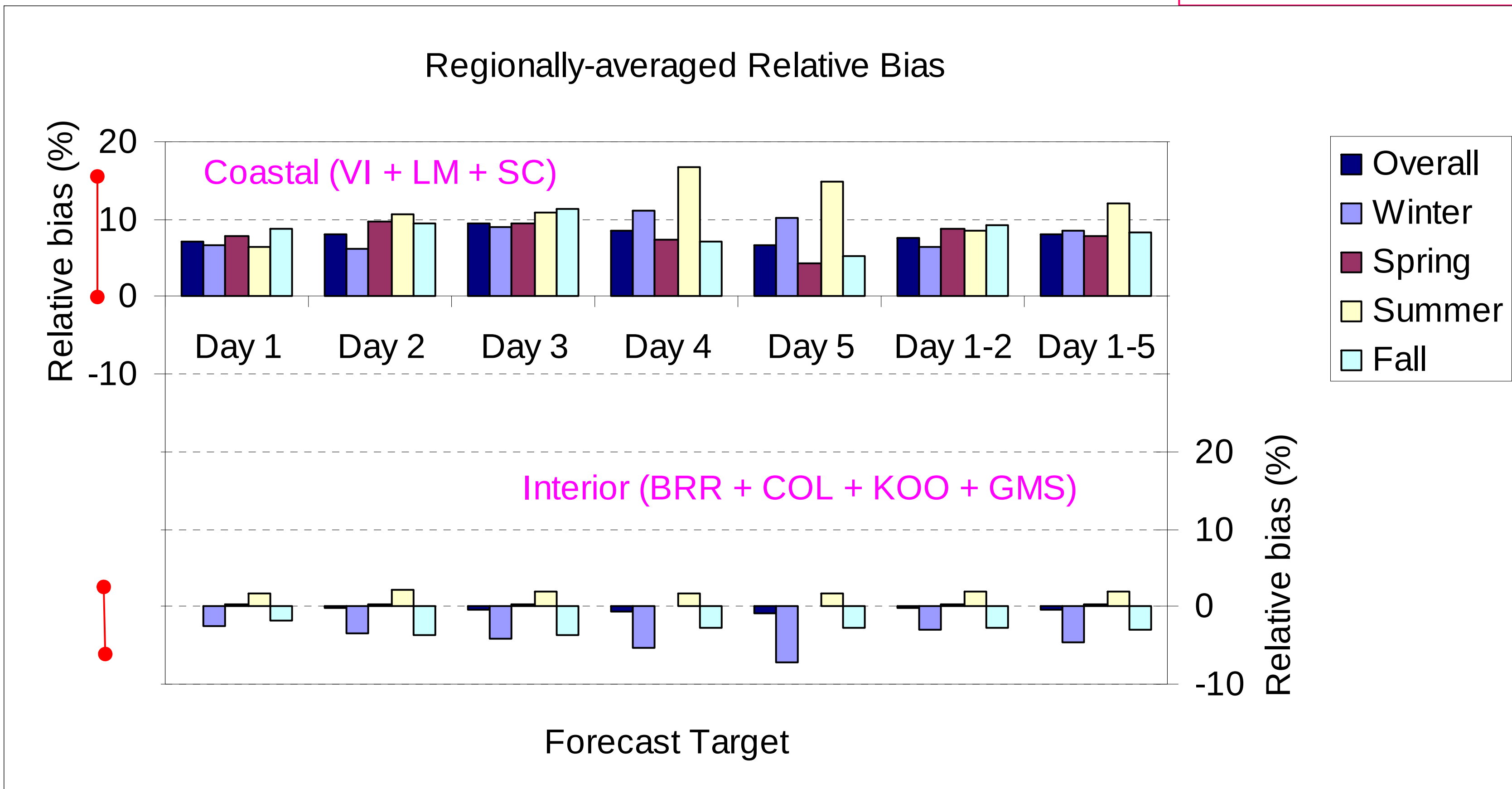
Short-range: Forecast bias relative to observed mean

Unbiased forecasts have bias of zero.

(1) Coastal basins are typically over-forecasted all year round

(2) Basins that stand out: Jordan, Alouette, Stave, Coquitlam

$$rBias(\%) = 100\frac{Bias(cms)}{\overline{Q_o}(cms)}$$

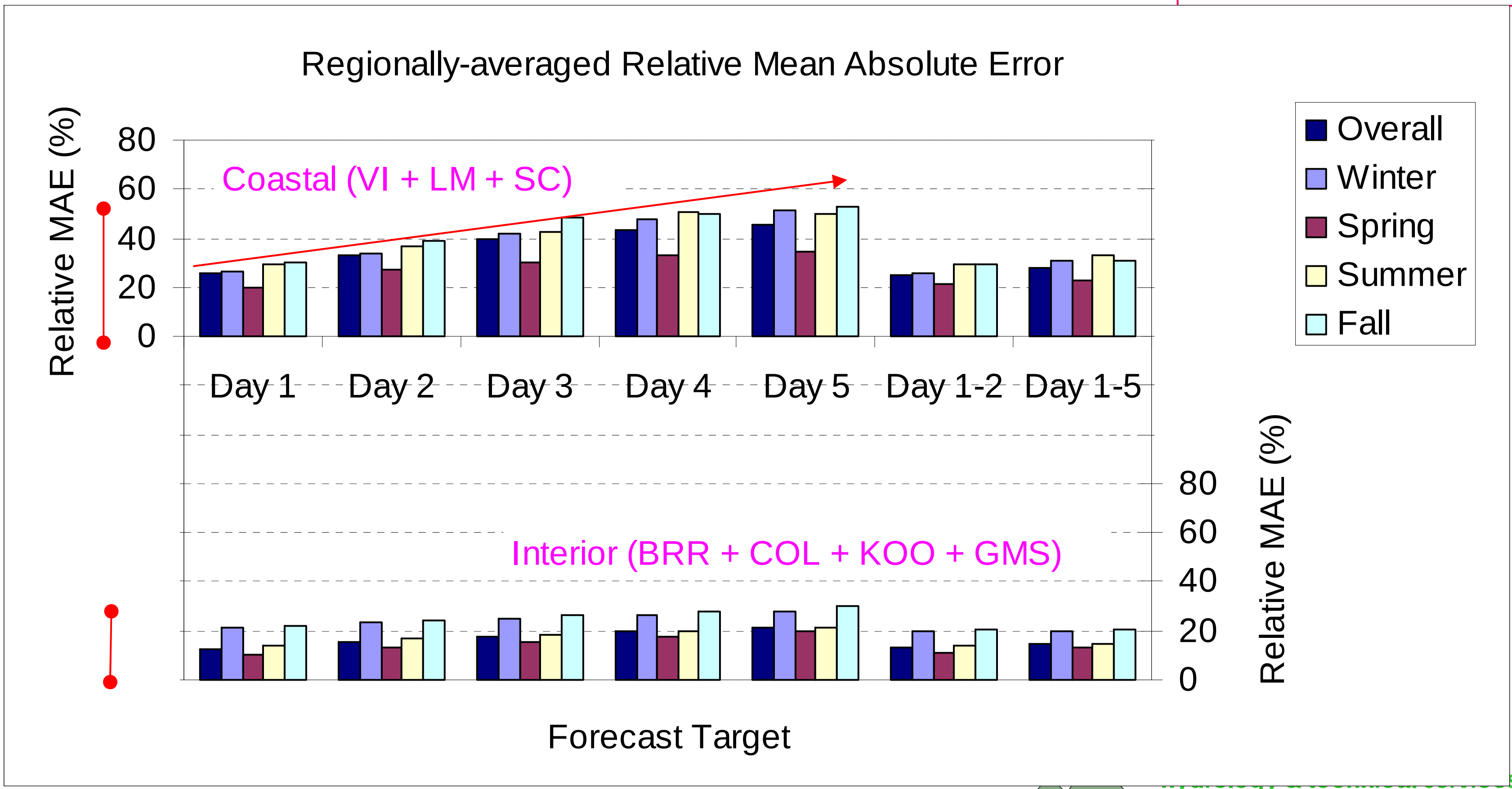


Short-range: Forecast accuracy relative to observed mean

Accurate forecasts have MAE of zero

- Interior projects forecasts are more accurate for all seasons
- Forecast accuracy deteriorates with lead time

$$rMAE\,(\%) = 100\,\frac{MAE\,(cms)}{\overline{Q}_o\,(cms)}$$

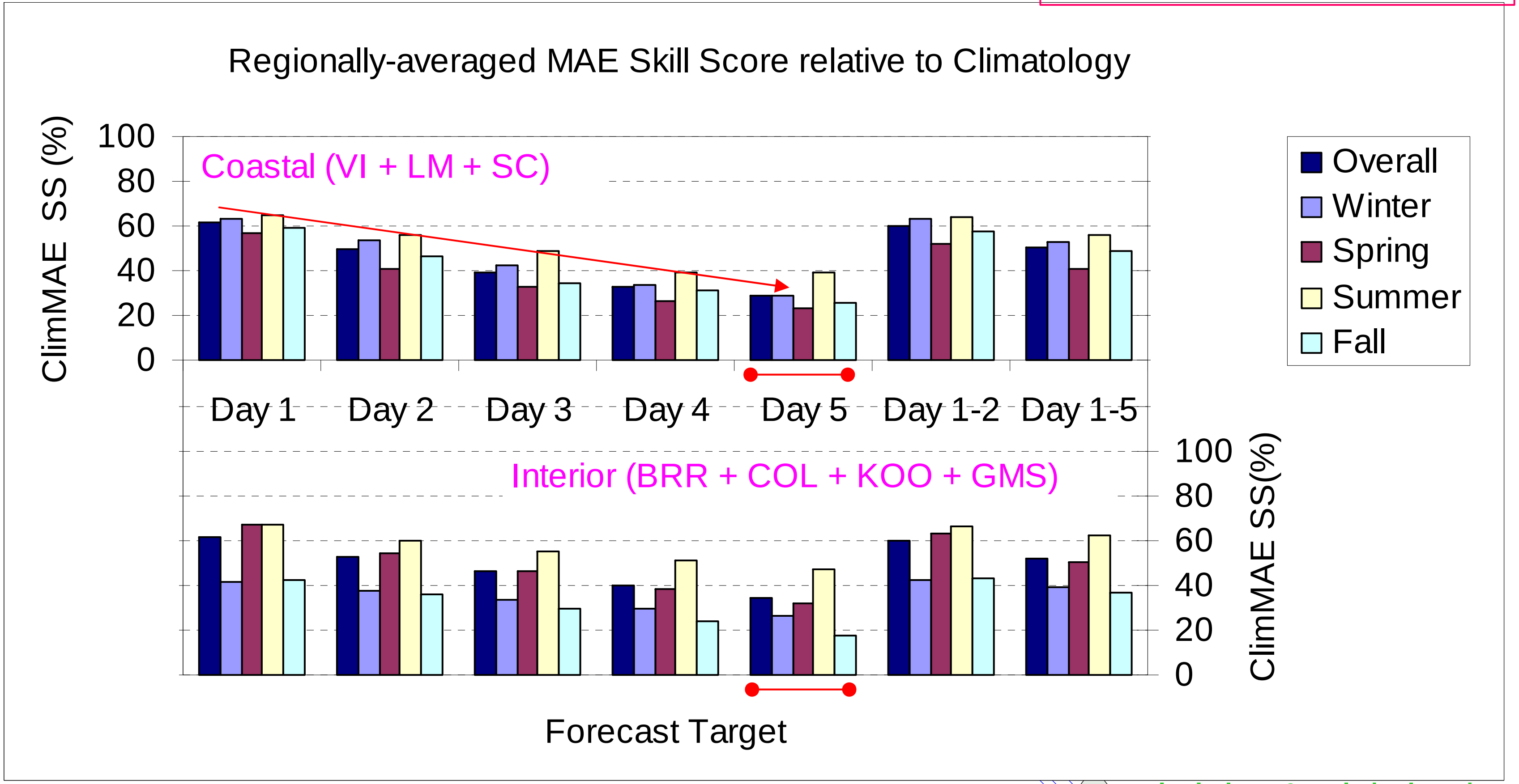


Short-range: Forecast MAE skill score relative to climatology

Perfect forecasts have skill of 100%; –ve skill if climatology is better

- Forecasts for both regions are better than climatology for all lead times & seasons

$$MAESS(\%) = 100\left(1 - \frac{MAE_{Forecast}}{MAE_{Climatology}}\right)$$

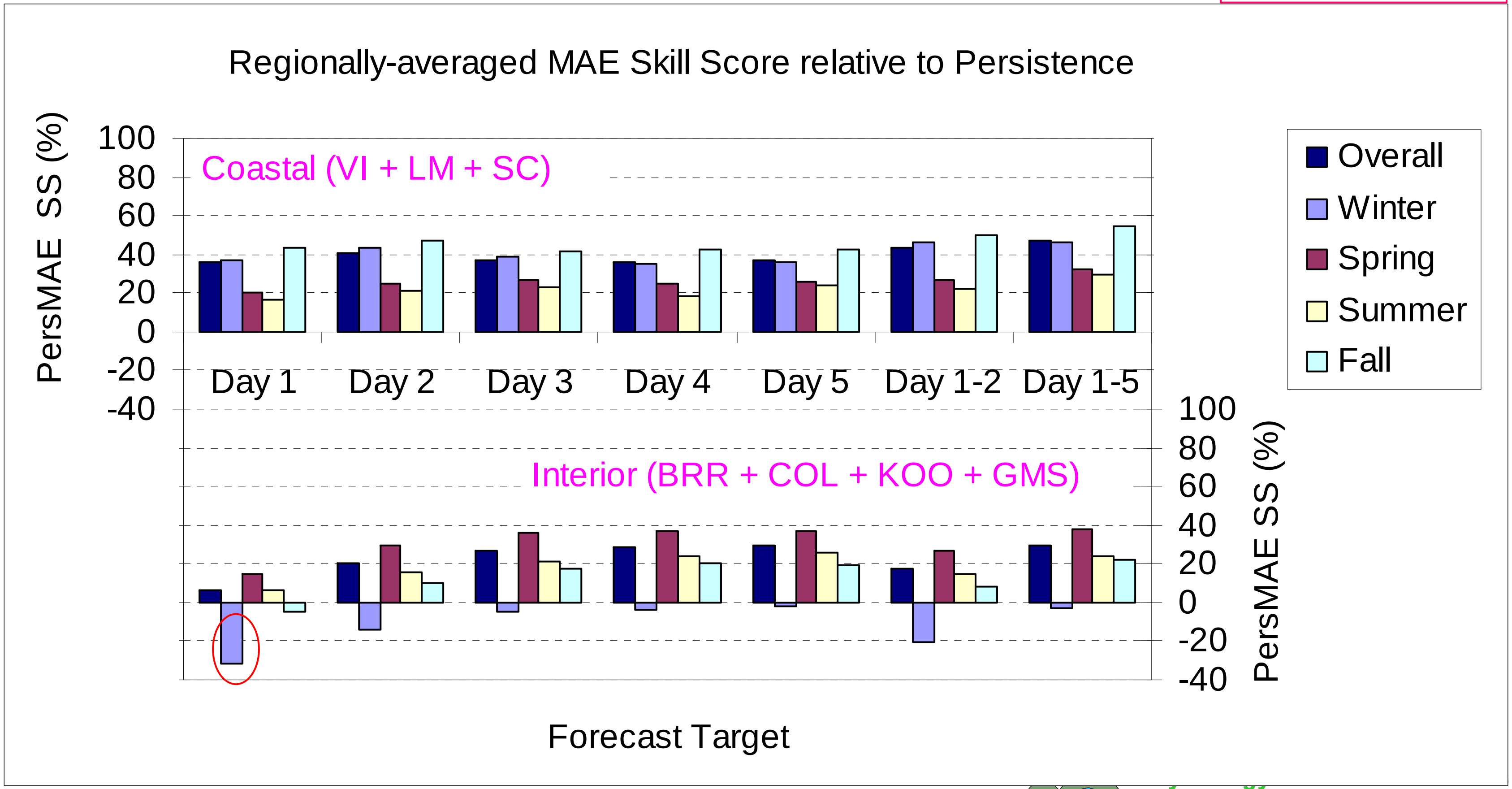


Short-range: Forecast MAE skill score relative to persistence

Perfect forecasts have skill of 100%; –ve skill if persistence is better

- Forecasts are skillful for most lead times & seasons
- Forecast skill for interior basins is affected by uncertainty in observed inflows

$$MAESS(\%) = 100(1 - \frac{MAE_{Forecast}}{MAE_{Persistence}})$$

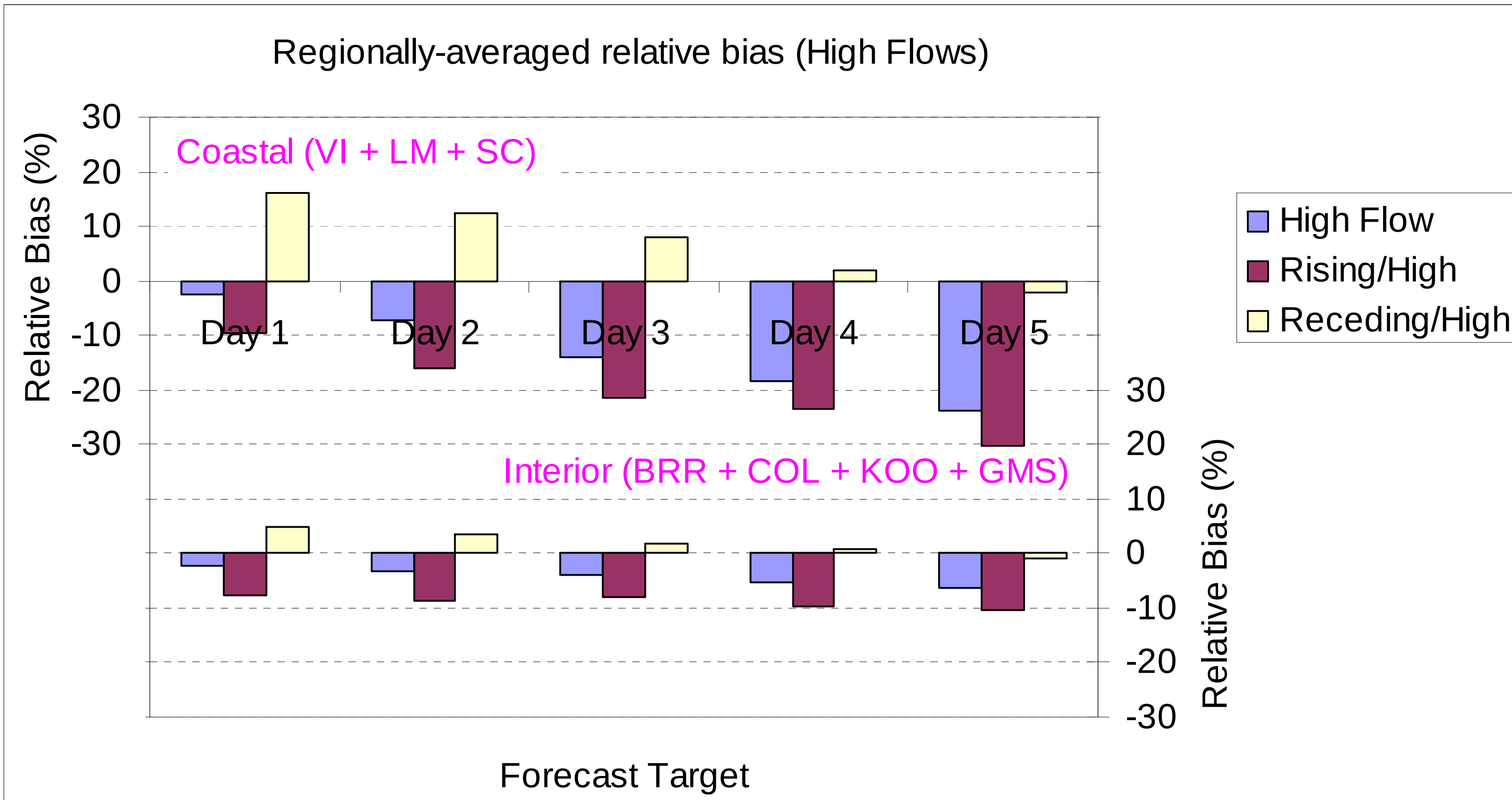


Short-range: High Flow forecast bias relative to observed mean

Unbiased forecasts have bias of zero

- High flows are generally under-forecasted
- Rising high flows are the worst offenders

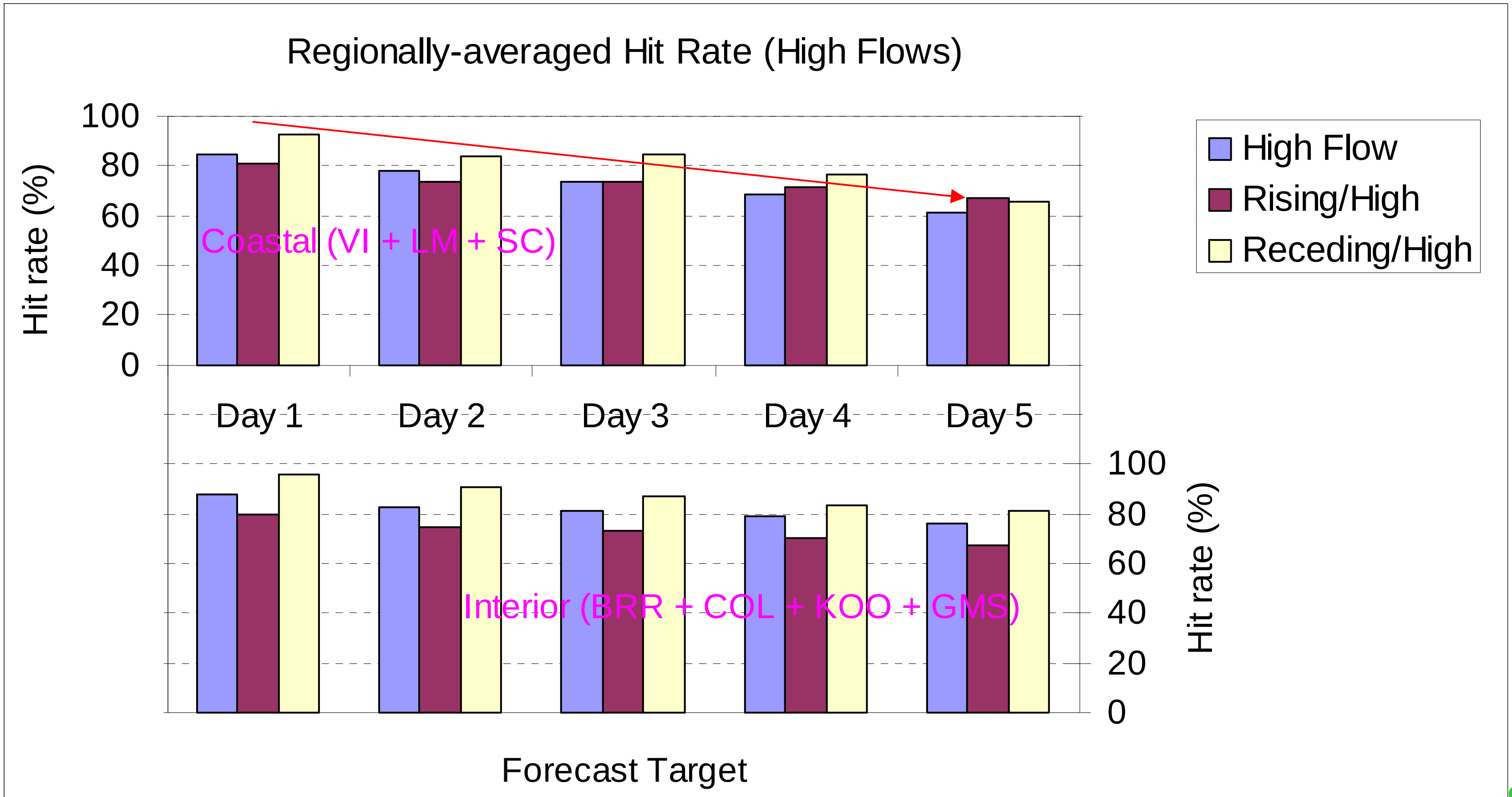
$$rBias\,(\%) = 100\,\frac{Bias\,(cms)}{\overline{Q_o}\,(cms)}$$



Short-range: categorical High Flow forecasts – hit rate

Forecasts have relatively high ability to forewarn high flow event occurrences out to day 5.

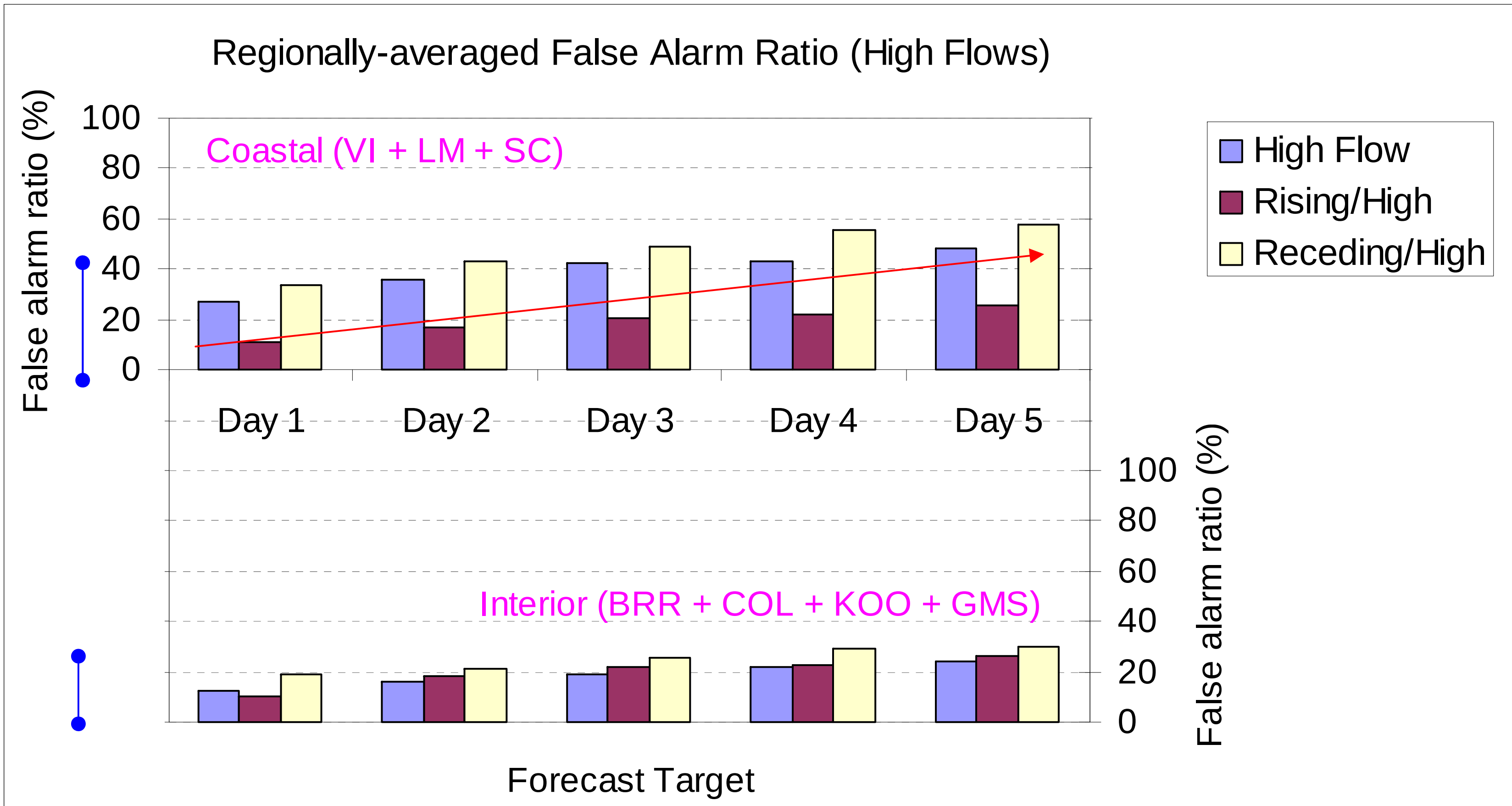
$$HitRate(\%) = 100 \frac{\#hits}{(\#hits + \#misses)}$$



Short-range: categorical High Flow forecasts – false alarm ratio

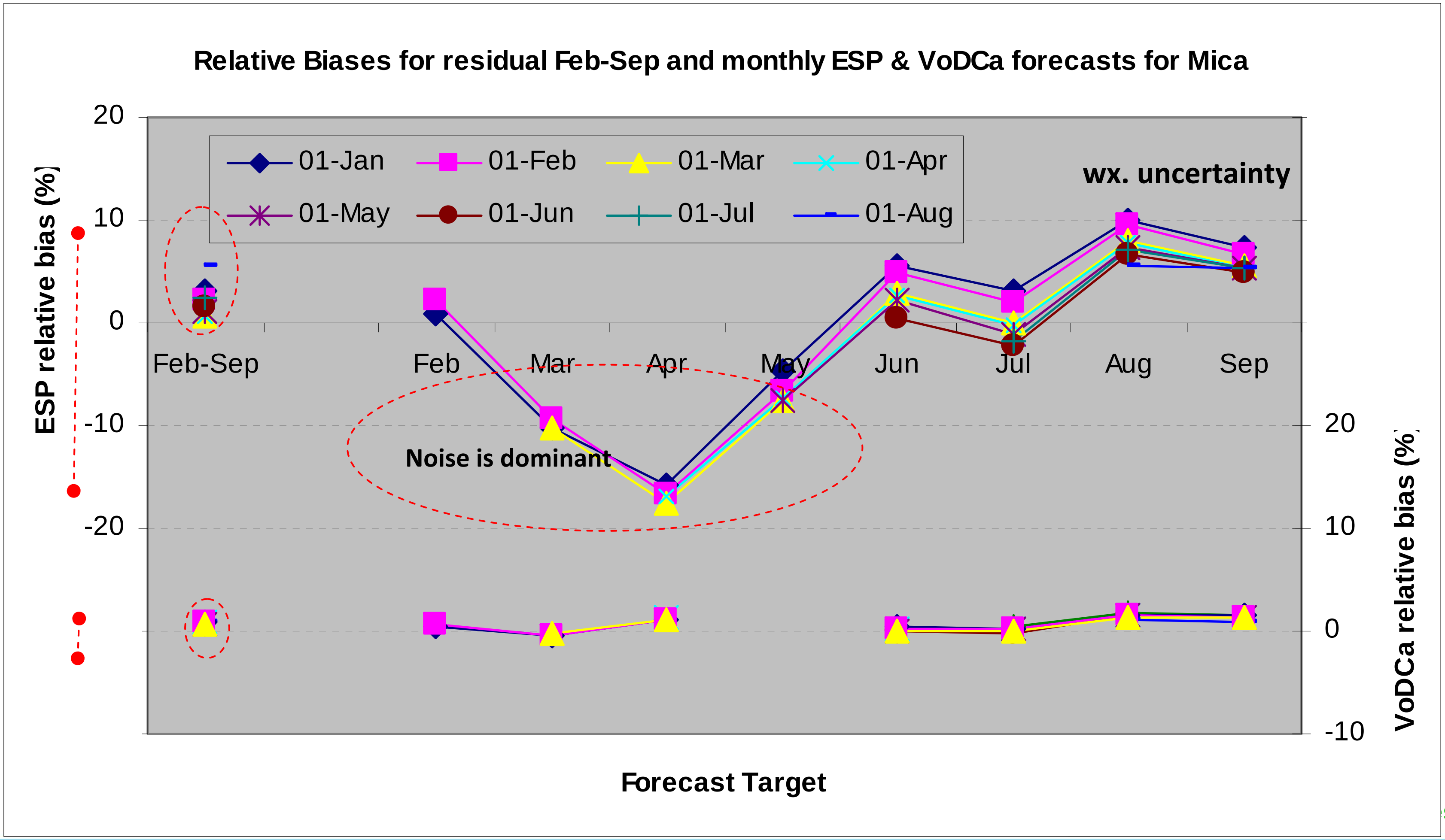
Perfectly reliable: 100% hit rate; 0% false alarm ratio
For coastal basins, high hit rates come at the expense of relatively high false warnings

$$FAR\,(\%) = 100\,\frac{\#\,false\,alarms}{(\#\,hits + \#\,false\,alarms)}$$



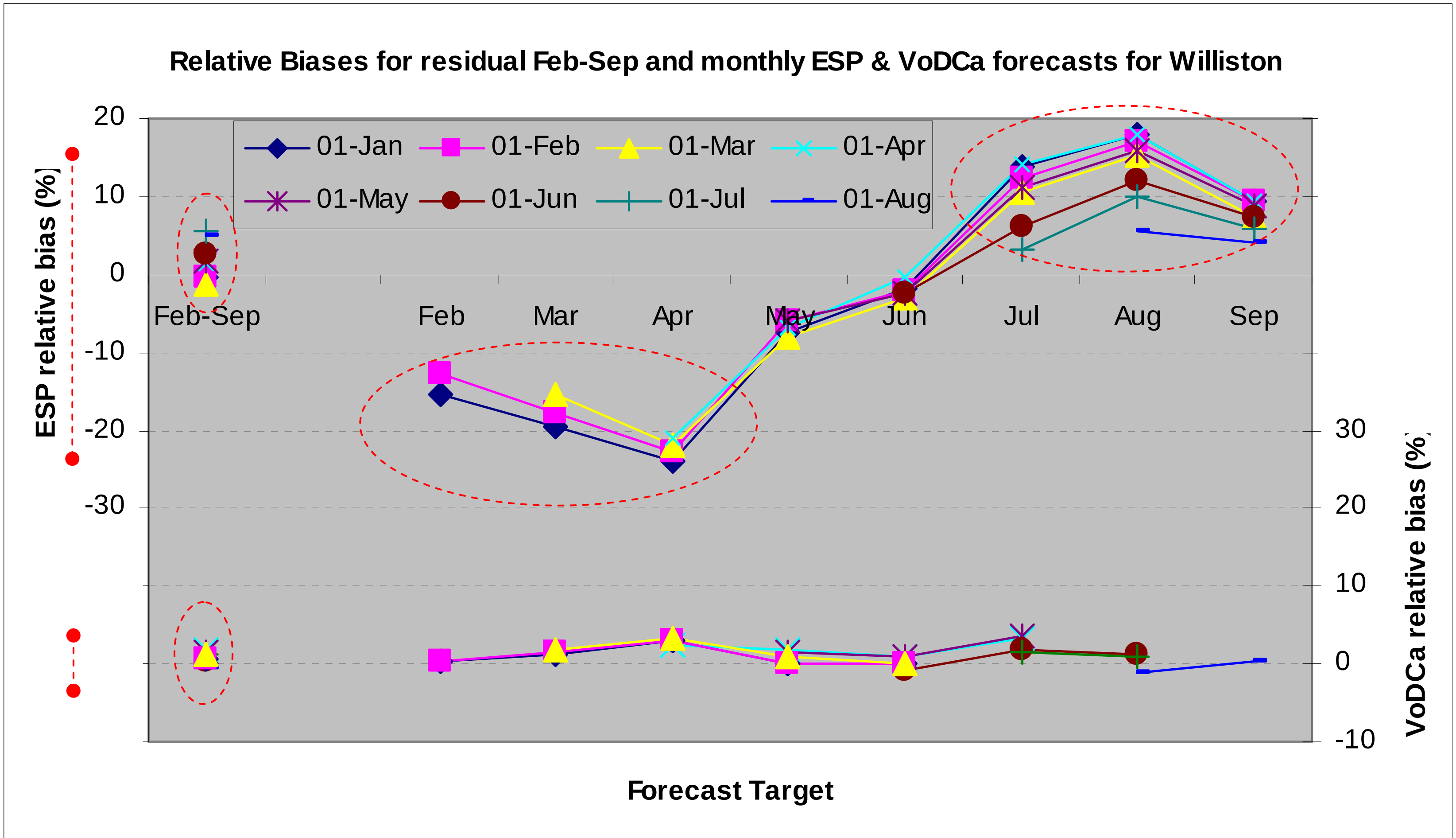
Long-range: Forecast biases for Mica

For ESP forecasts, Feb-Sep biases are small but biases in monthly forecast can be substantial; VoDCa forecasts are essentially unbiased.



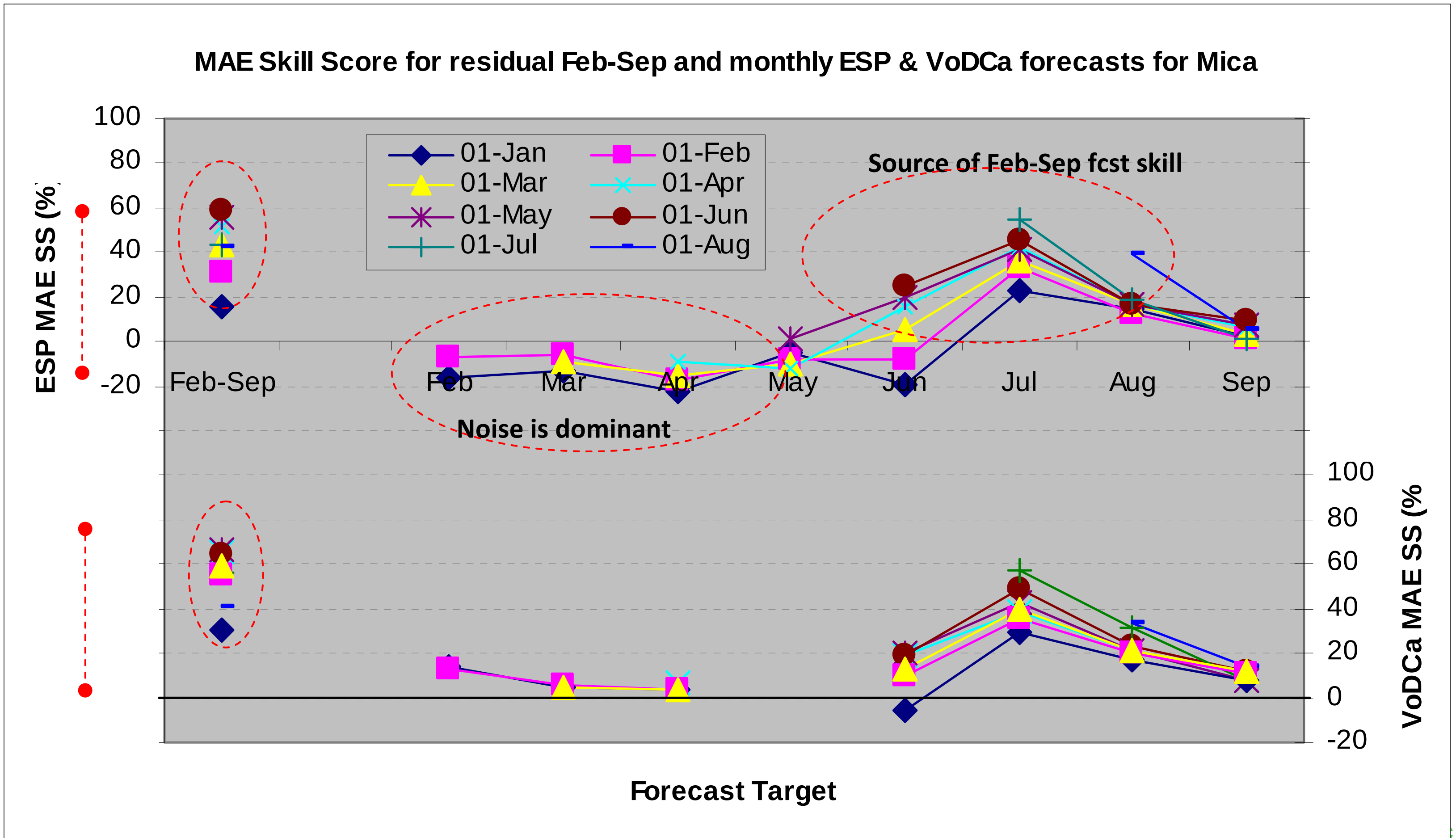
Long-range: Forecast biases for Williston

For ESP forecasts, Feb-Sep biases are small but biases in monthly forecast can be substantial; VoDCa forecasts are essentially unbiased.



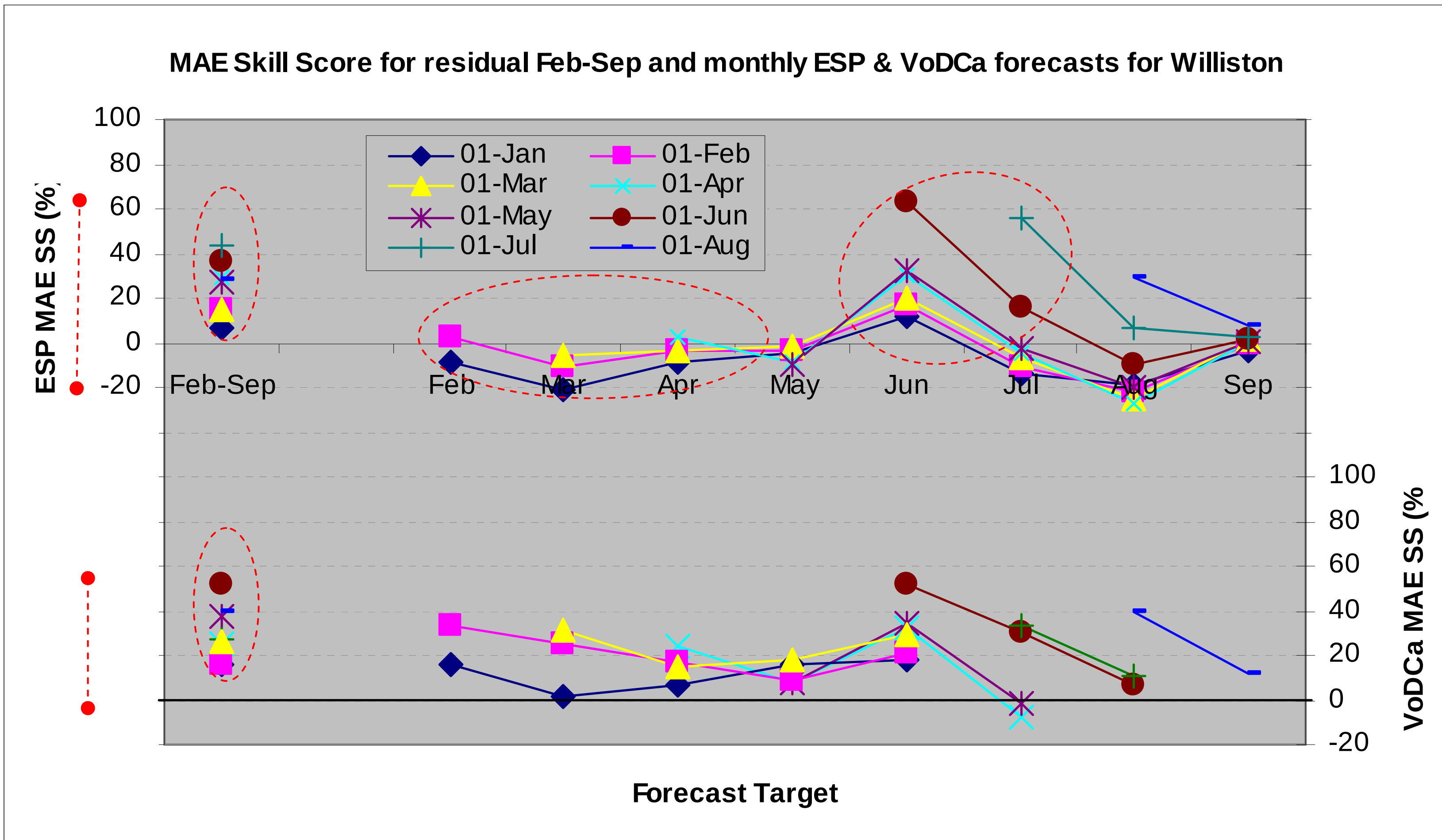
Long-range: Forecast skill relative to climatology for Mica

VoDCa forecast skill is marginally better than ESP. Feb-Sep forecasts from both models draw skill from forecasts for the freshet period.



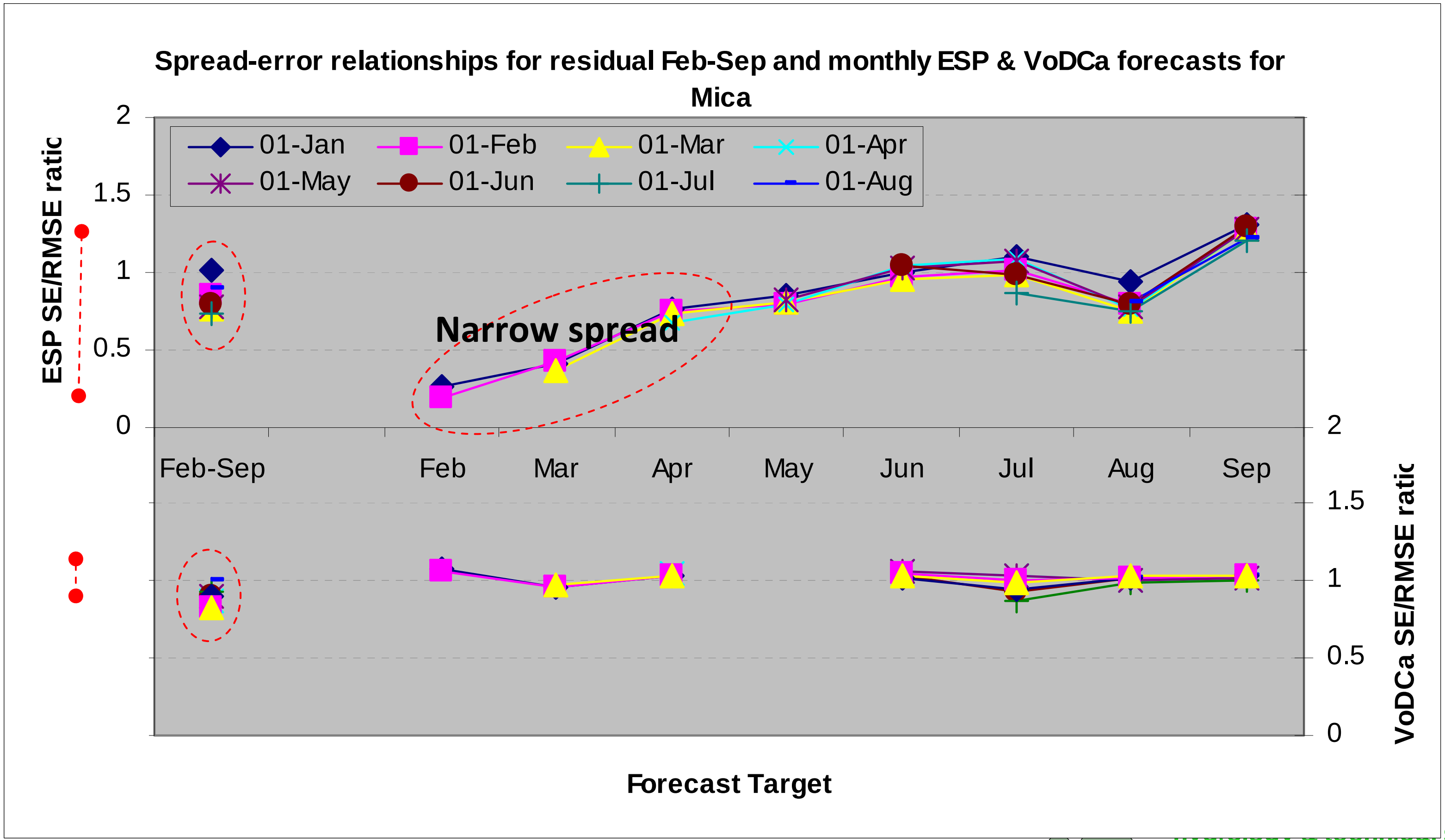
Long-range: Forecast skill relative to climatology for Williston

VoDCa forecast skill is marginally better than ESP. Feb-Sep forecasts from both models draw skill from forecasts for the freshet period.



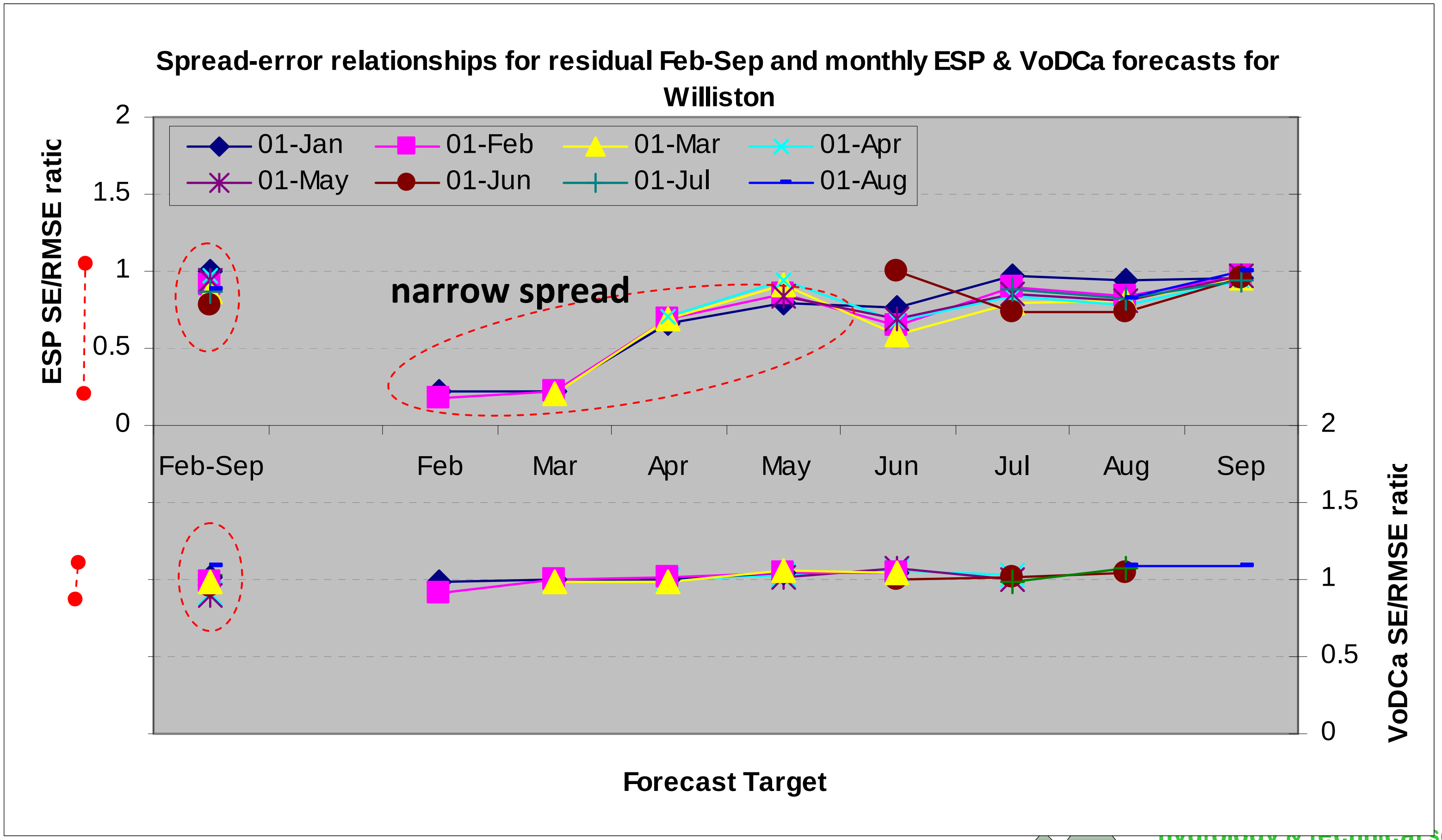
Long-range: Forecast spread-error relationships for Mica

SE/RMSE ratio ~ 1 → predictive uncertainty is reliable.
SE/RMSE > 1 → over-dispersed (under-confident) forecast
SE/RMSE < 1 → under-dispersed (over-confident) forecast



Long-range: Forecast spread-error relationships for Williston

SE/RMSE ratio ~ 1 → predictive uncertainty is reliable.
SE/RMSE > 1: over-dispersed (under-confident) forecast
SE/RMSE < 1: under-dispersed (over-confident) forecast

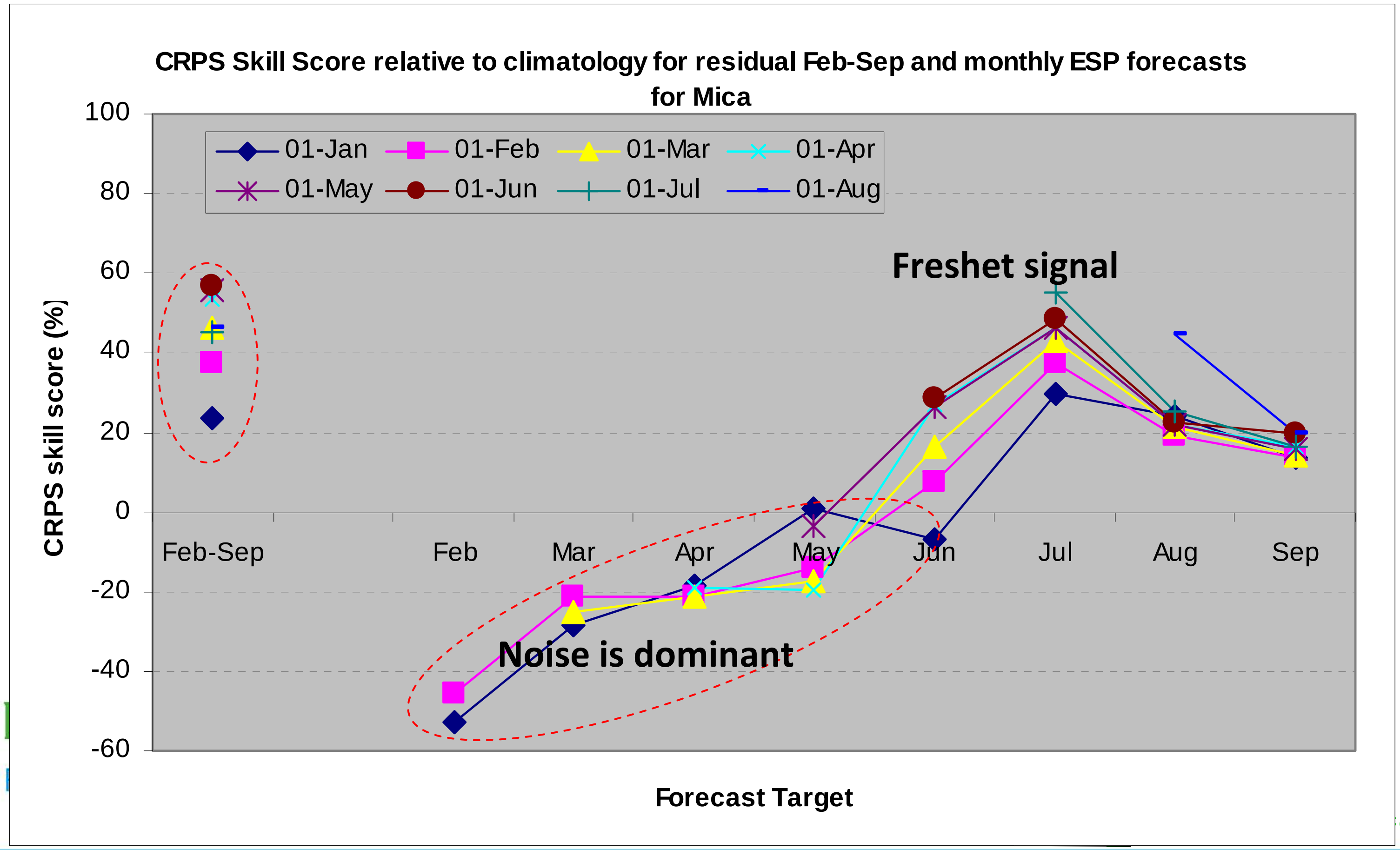


Long-range: CRPS skill score for Mica

CRPSS > 0 ➔ improvement over climatological forecast

In spite of reliability issues, ESP forecasts for Feb-Sep volume forecasts are better than climatology; monthly forecasts for the freshet period are also skillful.

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dx$$

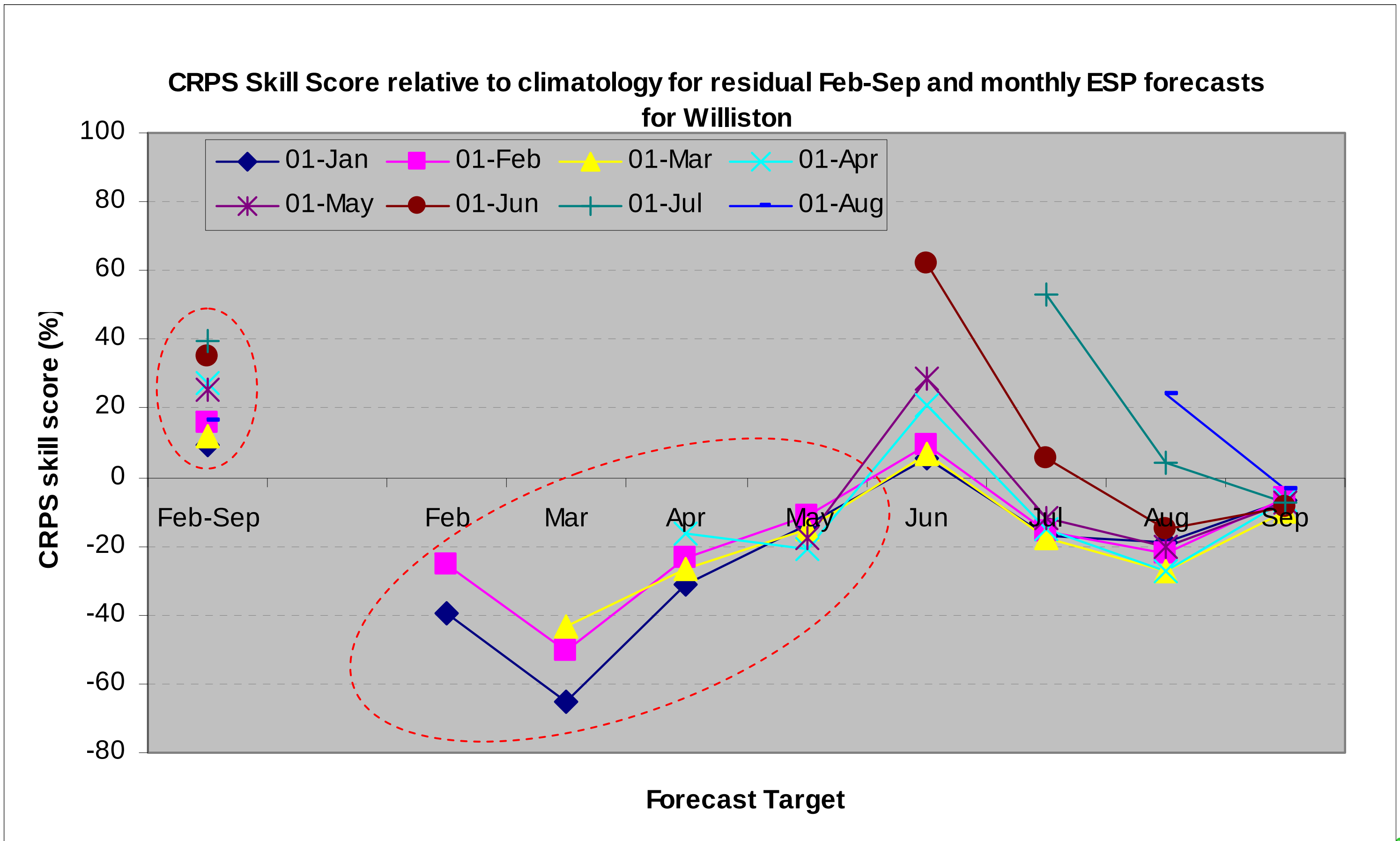


al services

Long-range: CRPS skill score for Williston

CRPS SS > 0 ➔ improvement over climatological forecast

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dx$$



Summary

The Positives	The Negatives
Short-range	
- overall forecast quality is good	- forecast biases for some coastal basins are of concern
- forecasts are skilful out to day 5	- low flow forecasts are less skilful
- good discrimination of high flow occurrences	- large bias in high flow forecasts, esp. for coastal basins
	- high false alarm ratio of high flow forecasts for coastal basins
Long-range	
- VoDCa forecasts are unbiased	
- VoDCa predictive uncertainty is good	
- ESP residual Feb-Sep forecasts have small biases	- ESP monthly forecast biases
	- ESP predictive uncertainty
- skill is good for peak flow months	- low skill for winter and late summer

BC hydro
FOR GENERATIONS

hydrology & technical services

Work to date/ wish list for further work

- **Just completed:**
    - UBCWM recalibration aimed at reducing forecast biases
    - Super-ensemble framework in place (can incorporate uncertainty due to wx fcst and/or observational sample size)

- **Wish List**
    - Analyze performance of new calibrations for forecast bias
    - Analyze the ESP super-ensemble framework for predictive uncertainty
    - Analyze viability of ESP forecasting for hybrid watersheds
    - Track operational ESP forecast performance with new model calibrations & new processing tools (bias correction, SWE adjustment)
    - Attribute forecast errors to weather & hydrologic modeling
    - Implement near real-time inflow quality procedures
    - Evaluate the $$$$$$ value of forecasts**

Thank You!

BC hydro
FOR GENERATIONS

hydrology & technical services