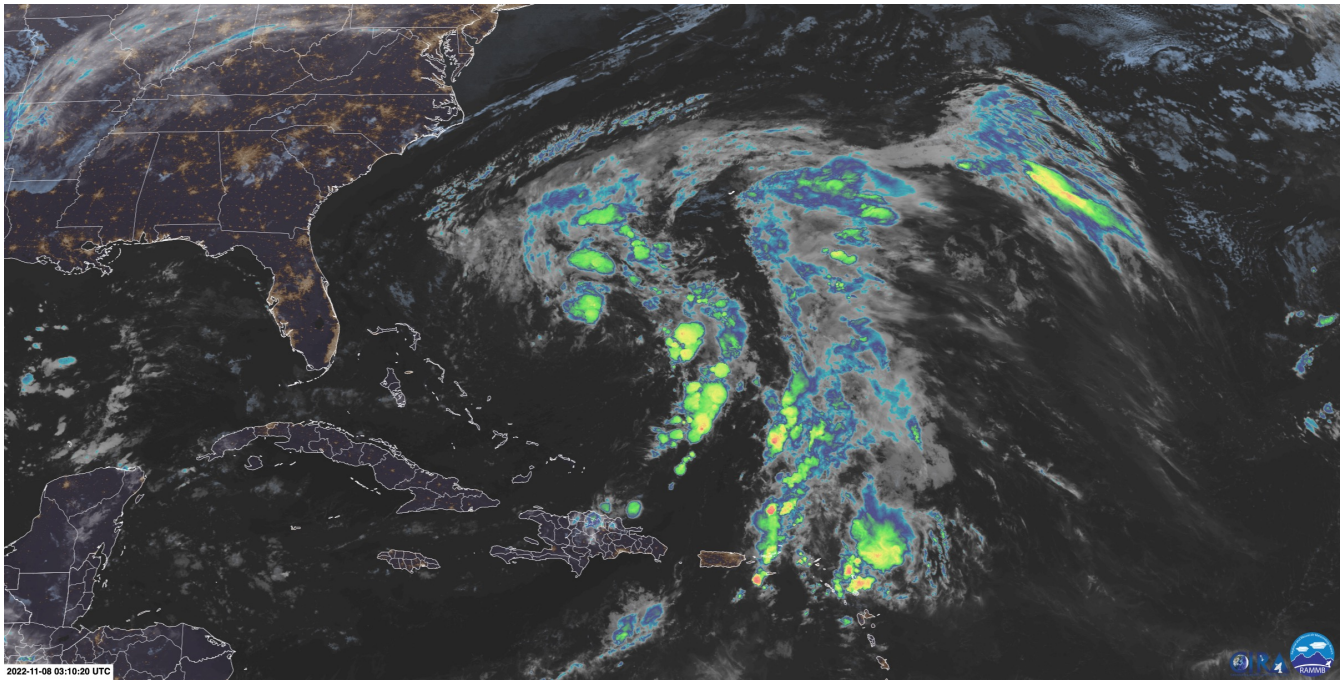
Preliminary Evaluation of the Hurricane Analysis and Forecast System (HAFS) Models for 2020–2022

Jonathan Martinez[1,2]
Benjamin Trabing[1]

[1]Cooperative Institute for Research in the Atmosphere
[2]National Hurricane Center



STS Nicole @ 0320 UTC 08 Nov 2022

A 3-Part Series

Part I – Track Verification

Part II – Intensity Verification

Part III – Wind Radii Verification (R34 analyses; R50 and R64 highlights)

Downstream Impacts of HAFS

Model	Description	Model(s) to Replace
TVCN	Track Variable Consensus	HWRF (+HMON in EPAC)
IVCN	Intensity Variable Consensus	HWRF+HMON
RVCN	Wind Radii Variable Consensus	HWRF
HCCA	HFIP Corrected Consensus Approach Model	HWRF+HMON
DTOPS	Deterministic to Probabilistic Statistical Model for Rapid Intensification	HWRF
NNIC	Neural Network Intensity Combination Model	HWRF

Verification Design

- Homogeneous Verification – a prediction from each model must be present at a given lead time

- Three-year evaluation period from 2020 through 2022
  - North Atlantic – 66 TCs
  - Eastern Pacific – 59 TCs
  - Last TC in North Atlantic – AL142022
  - Last TC in Eastern Pacific – EP192022

- Operational interpolator offsets implemented for all models
  - Optimal interpolator offsets determined for HAFS at 06/30 h

Part I – Track Verification

Track verification for consensus models

TVCN – Variable consensus with minimum 2 members

North Atlantic basin (NATL) – 6 models with only HWRF to be replaced

- TVCN – AVNI, EGRI, HWFI, EMHI, CTCI, EMNI
- TVCA – Replace HWFI with HFAI in TVCA (HAFS-A test)
- TVCS – Replace HWFI with HFSI in TVCA (HAFS-S test)

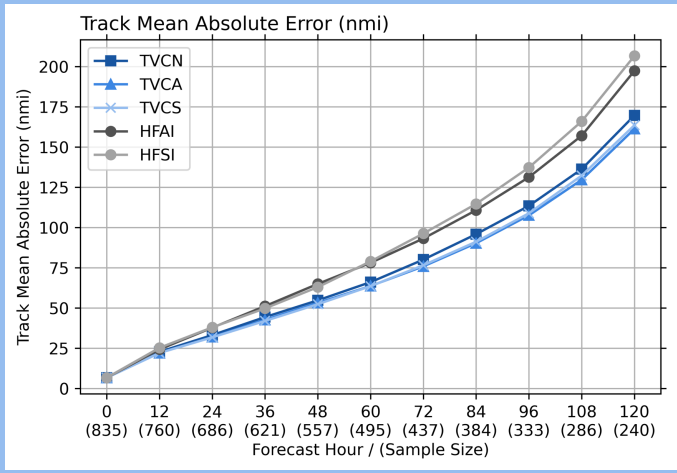
East Pacific basin (EPAC) – 7 models with HWRF + HMON to be replaced

- TVCN – AVNI, EGRI, HWFI, EMHI, CTCI, EMNI, HMNI
- TVCB – Replace HWFI and HMNI with HFAI and HFSI (HAFS test)

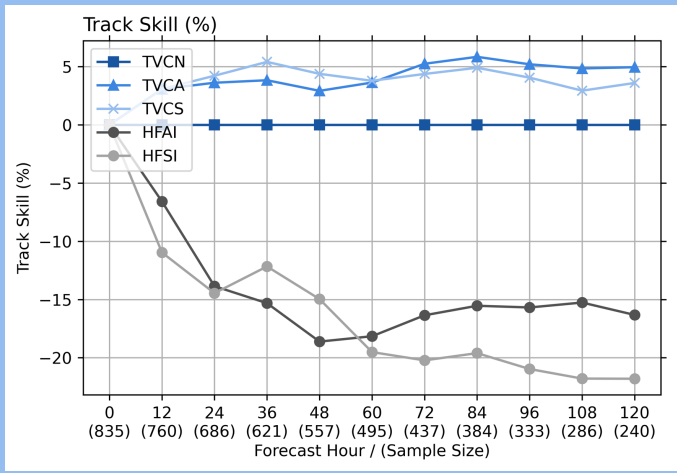
Track verification for consensus models

North Atlantic

Mean Absolute Error

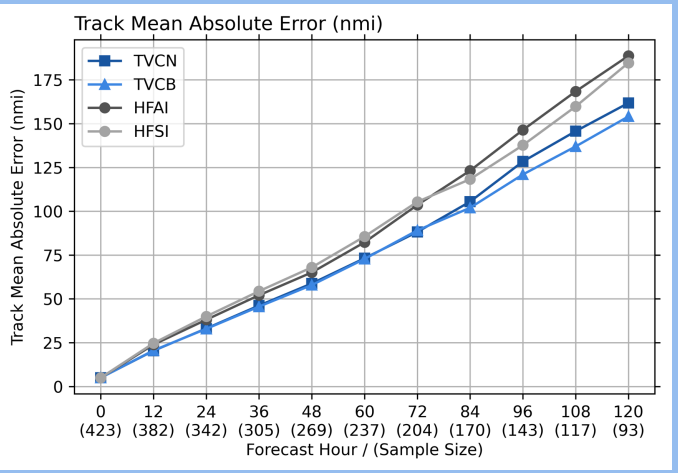


Relative Skill

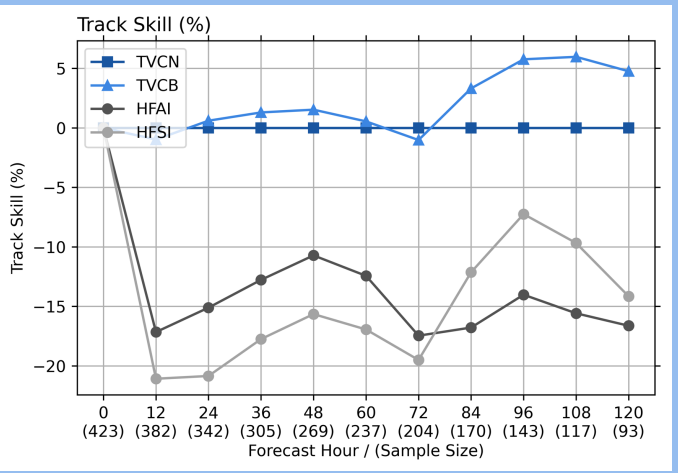


East Pacific

Mean Absolute Error



Relative Skill

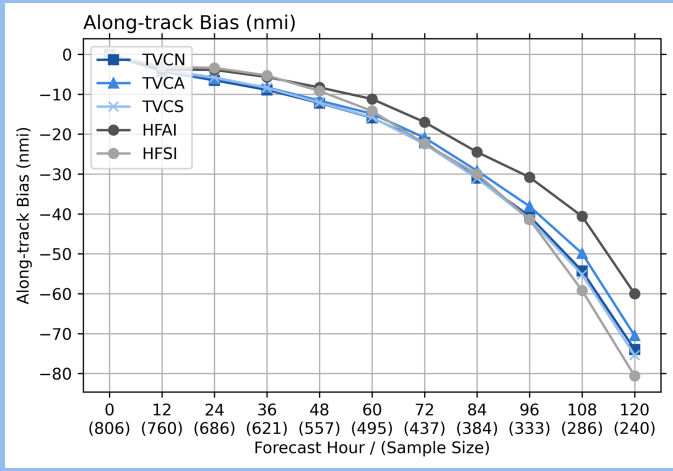


- HAFS models improve track skill of TVCN in NATL by ~5% at all lead times
- HAFS models improve track skill of TVCN in EPAC most at days 4–5 (~5%)

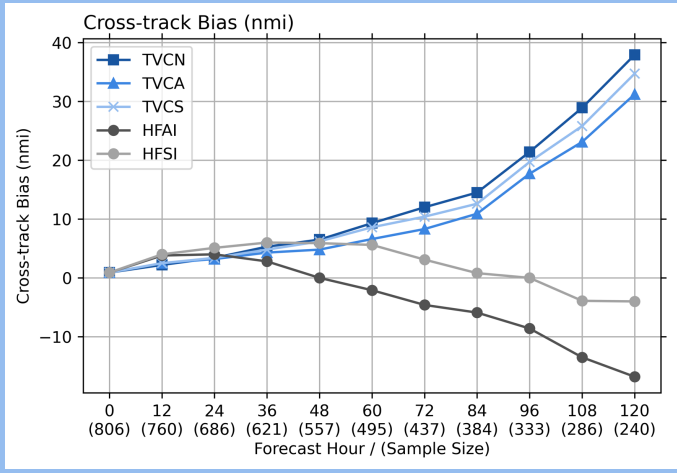
Track verification for consensus models

North Atlantic

Along-track Bias

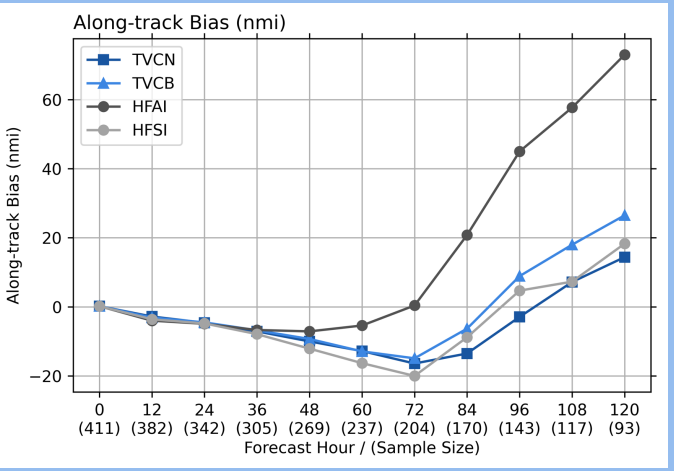


Cross-track Bias

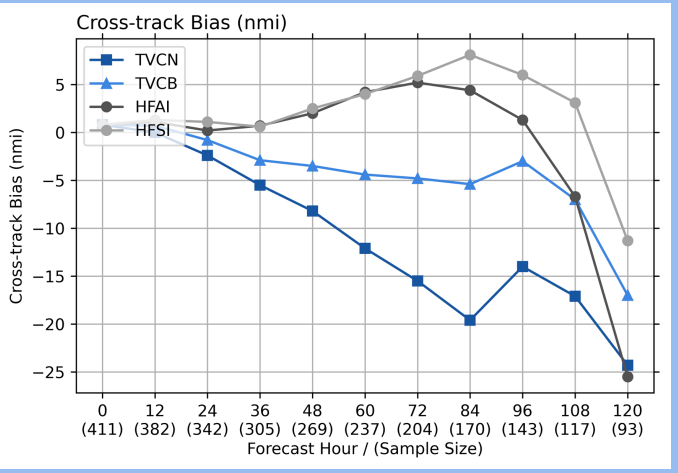


East Pacific

Along-track Bias



Cross-track Bias



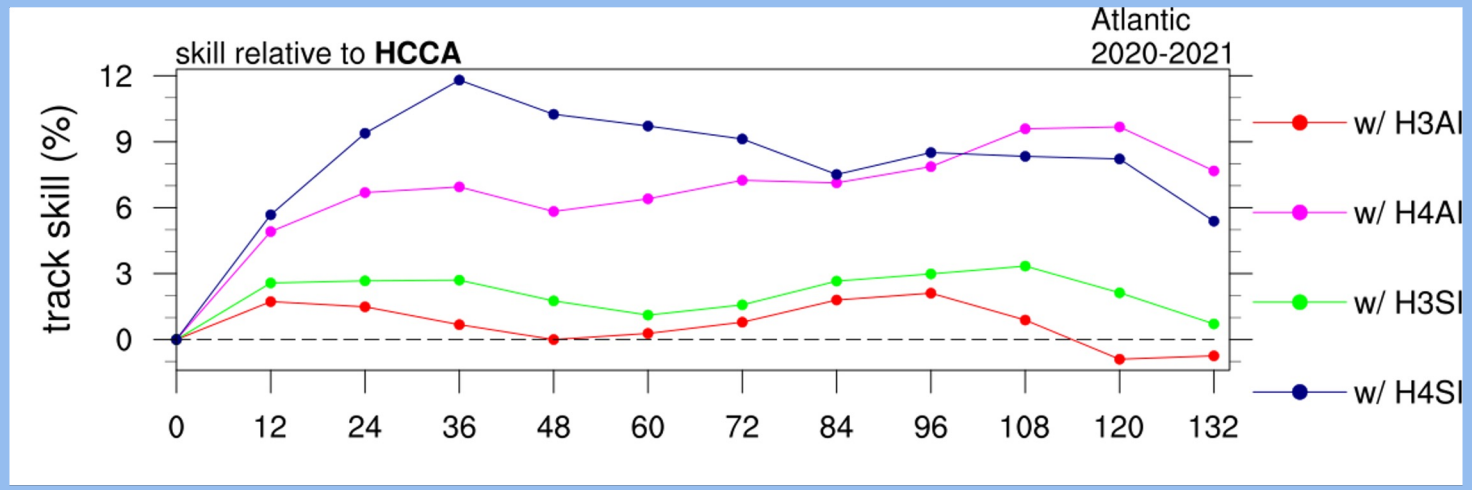
- HAFS models slightly reduce along/cross-track biases in NATL and EPAC at ~all lead times
- Note right-of-track biases at early lead times potentially offsetting predominant left-of-track model biases

HCCA NATL track verification

- Leave-one-out training method – All TCs included in training data except the one that is being forecasted
- H3AI/H3SI – merge HWRF/HMON retrospectives with HAFS-A/HAFS-S
  - Assumes the models have similar error characteristics and behavior
- H4AI/H4SI – remove HWRF/HMON and add HAFS-A/HAFS-S from 2020–2021

HCCA NATL track verification

- Leave-one-out training method – All TCs included in training data except the one that is being forecasted
- H3AI/H3SI – merge HWRF/HMON retrospectives with HAFS-A/HAFS-S
  - Assumes the models have similar error characteristics and behavior
- H4AI/H4SI – remove HWRF/HMON and add HAFS-A/HAFS-S from 2020–2021



- Replacing HWRF/HMON with HAFS-A/HAFS-S in HCCA improves NATL track skill by 6–12%
  - Skill comes at expense of consistency due to small HAFS sample relative to HWRF/HMON

Part I – Track verification summary

NATL Summary

- Replacing HWRF with either HAFS-A or HAFS-S improves track skill of TVCN by ~5% at all lead times
- HAFS slightly reduces the along/cross-track biases of TVCN at ~all lead times
- HAFS improvement in track skill translates to improvements in HCCA track skill of 6–12%

EPAC Summary

- Replacing HWRF/HMON with HAFS-A/HAFS-S improves track skill of TVCN most at days 4–5 (~5%)
- HAFS reduces along/cross-track biases at all lead times
  - Right-of-track bias might help offset predominant left-of-track biases

Part II – Intensity Verification

Intensity verification for consensus models

Variable consensus with minimum 2 members

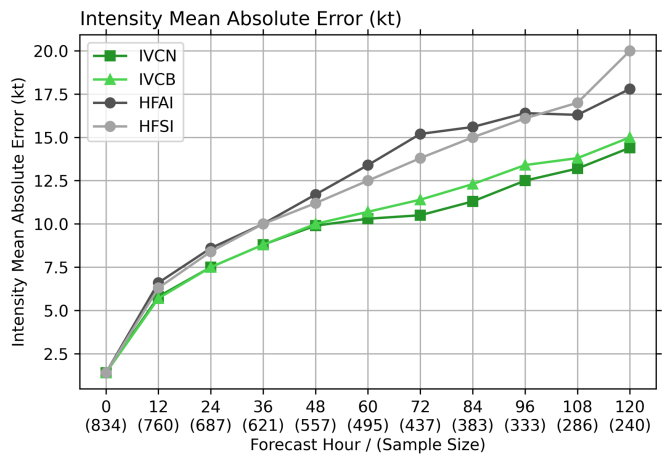
North Atlantic basin (NATL) and Eastern Pacific basin (EPAC)

- IVCN – DSHP, LGEM, HWFI, HMNI, CTCI
- IVCB – Replace HWFI and HMNI with HFAI and HFSI (HAFS test)

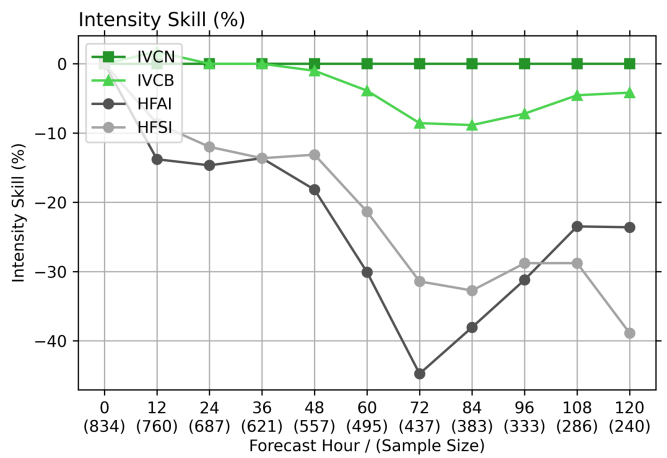
Intensity verification for individual models

North Atlantic

Mean Absolute Error

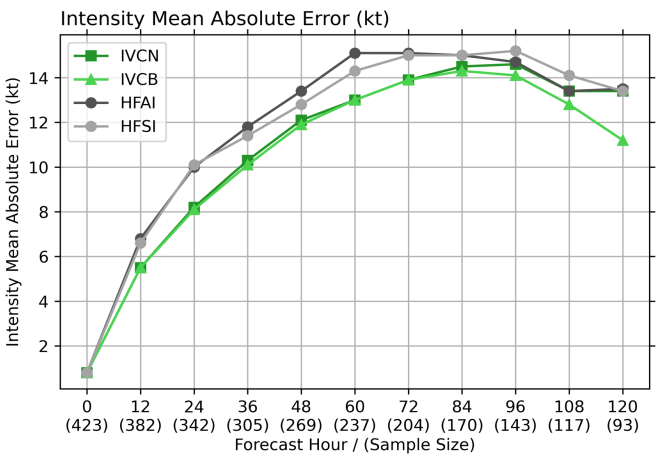


Relative Skill

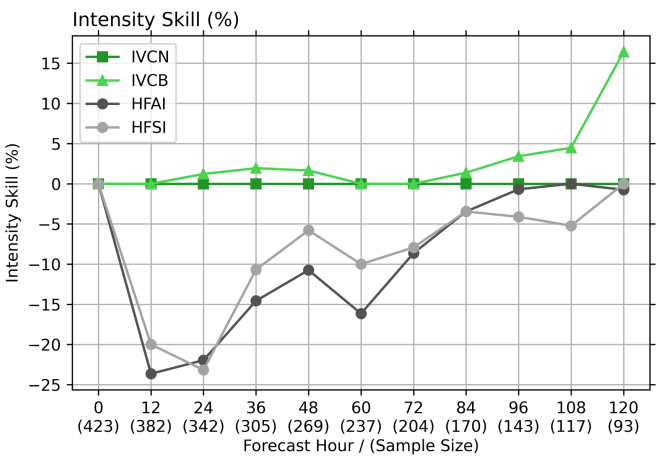


East Pacific

Mean Absolute Error



Relative Skill

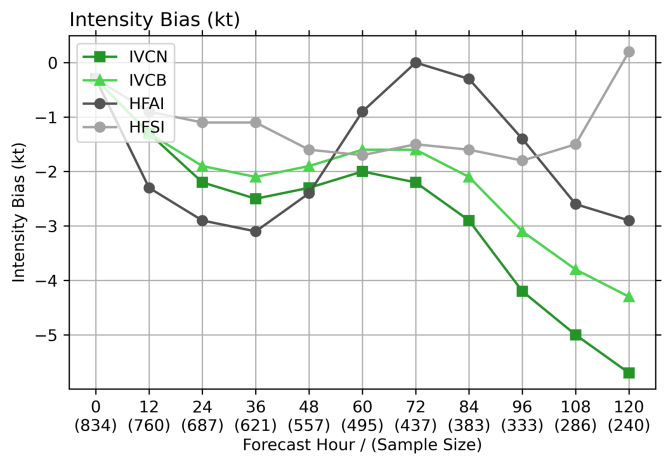


- HAFS models reduce intensity skill of IVCN in NATL at days 2–5 up to 10%
- HAFS models improve intensity skill of IVCN in EPAC by 2% at days 1–4; greater improvements at days 4–5 (2–14%)
- *Note the dip in skill improvements near day 3 was also present in TVCN

Intensity verification for individual models

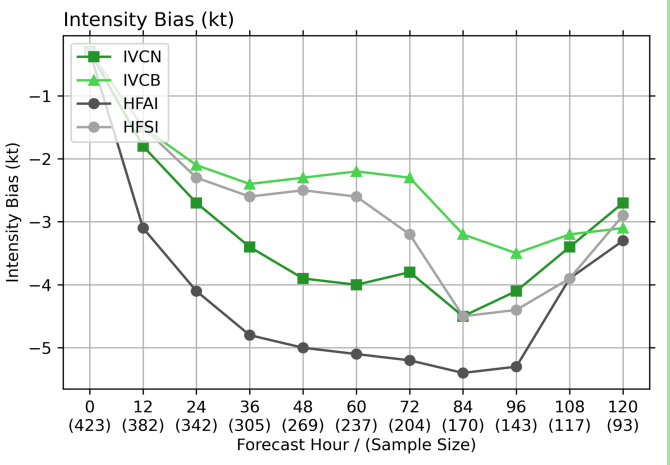
North Atlantic

Intensity Bias



East Pacific

Intensity Bias



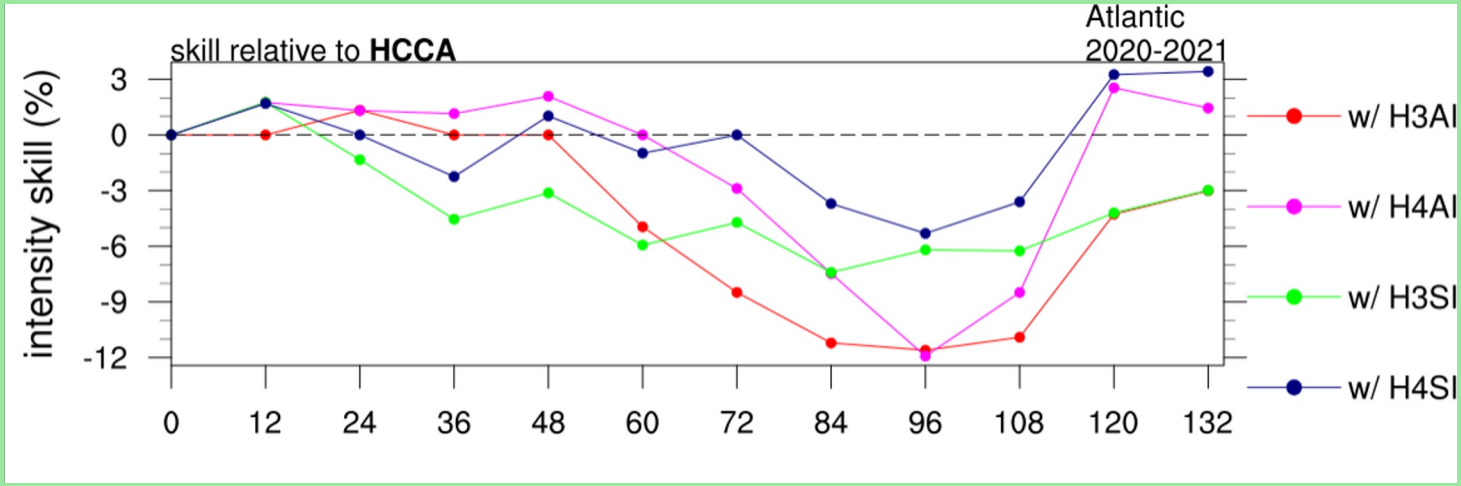
- HAFS models reduce low-intensity bias in both NATL and EPAC at ~all lead times

HCCA intensity verification

- Leave-one-out training method – All TCs included in training data except the one that is being forecasted
- H3AI/H3SI – merge HWRF/HMON retrospectives with HAFS-A/HAFS-S
  - Assumes the models have similar error characteristics and behavior
- H4AI/H4SI – remove HWRF/HMON and add HAFS-A/HAFS-S from 2020–2021

HCCA intensity verification

- Leave-one-out training method – All TCs included in training data except the one that is being forecasted
- H3AI/H3SI – merge HWRF/HMON retrospectives with HAFS-A/HAFS-S
  - Assumes the models have similar error characteristics and behavior
- H4AI/H4SI – remove HWRF/HMON and add HAFS-A/HAFS-S from 2020–2021



- Replacing HWRF/HMON with HAFS-A/HAFS-S in HCCA generally reduces NATL intensity skill; some minor improvements at days 0–2
  - All iterations show relative minimum in skill near day 4

Part II – Intensity verification summary

NATL Summary

- Replacing HWRF/HMON with HAFS models reduces intensity skill of IVCN at days 2–5 up to 10%
- HAFS models reduce low-intensity bias in IVCN at ~all lead times
- Reduced intensity skill from HAFS models generally translates to reduced HCCA intensity skill

EPAC Summary

- Replacing HWRF/HMON with HAFS models improves intensity skill of IVCN by 2% at days 1–4; greater improvements at days 4–5 (2–14%)
- HAFS models reduce low-intensity bias in IVCN at ~all lead times

Part III – Wind Radii Verification

Wind radii verification for consensus models

Variable consensus with minimum 1 member

Different techs given different interpolator offsets (06/18) compared to track/intensity

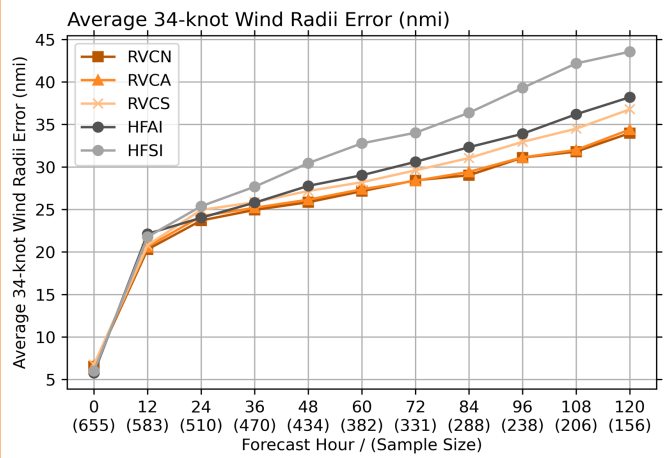
North Atlantic basin (NATL) and Eastern Pacific basin (EPAC)

- RVCN – AHNI, HHFI, EHHI, CHCI
- RVC2 – Replace HHFI with HHAI (HAFS-A test)
- RVC3 – Replace HHFI with HHSI (HAFS-S test)

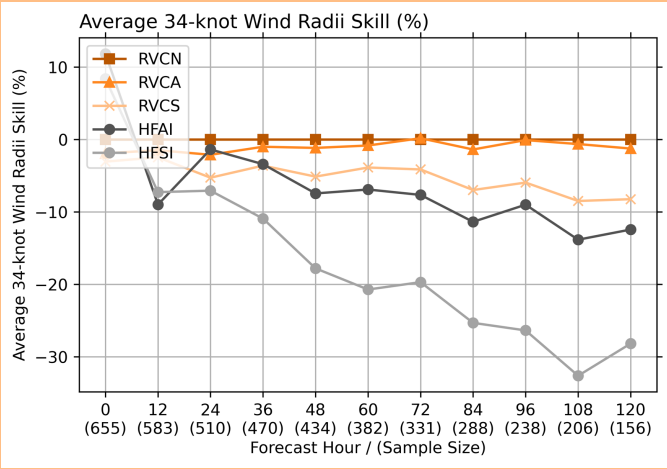
R34 verification for consensus models

North Atlantic

Average R34 Error

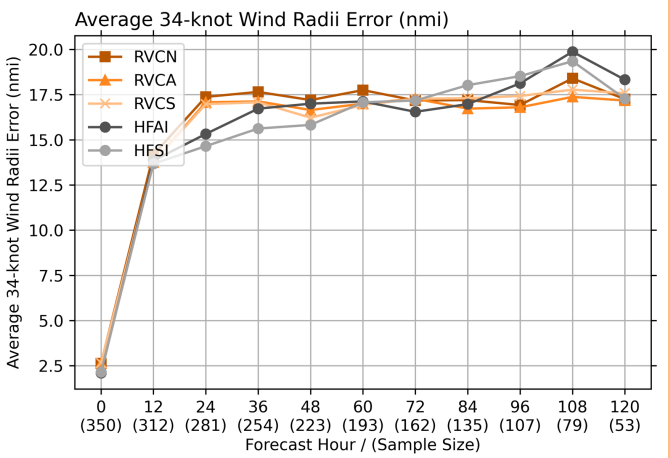


Relative Skill

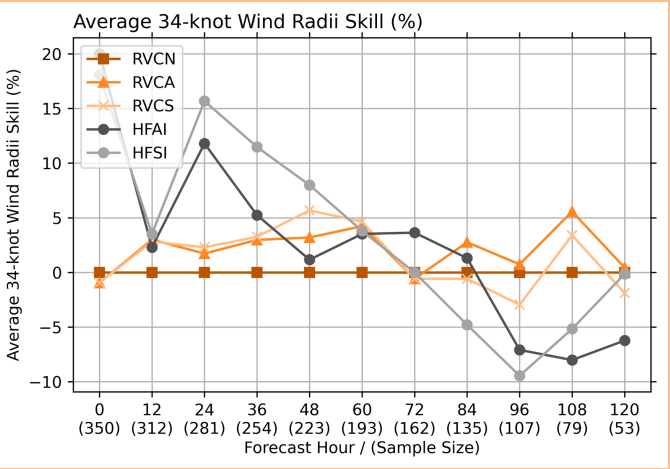


East Pacific

Average R34 Error



Relative Skill

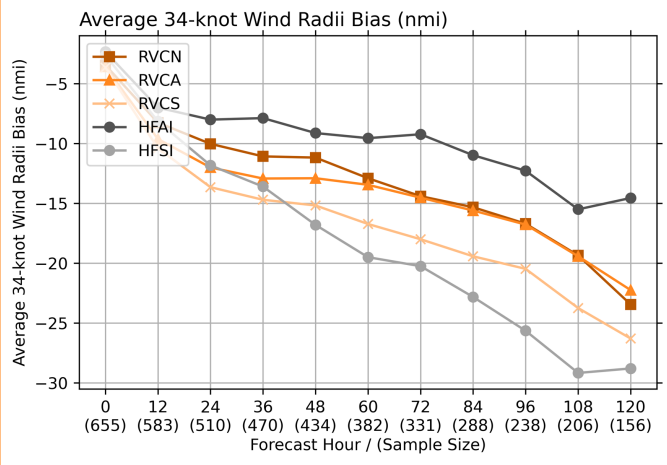


- HAFS-A has relatively neutral impact on RVCN R34 skill in NATL; HAFS-S reduces R34 skill in NATL by 2–8%
- HAFS models improve RVCN R34 skill in EPAC by ~2–5% at days 0–3

R34 verification for consensus models

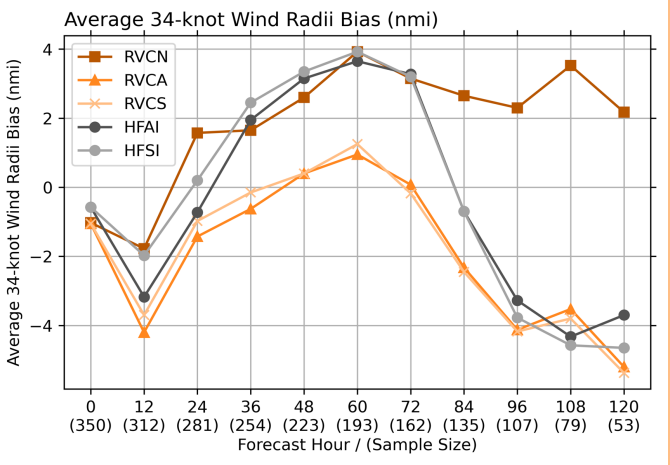
North Atlantic

Average R34 Bias



East Pacific

Average R34 Bias



- HAFS-A has comparable, small R34 bias in RVCN at days 3–5; HAFS-S amplifies small bias of RVCN at all lead times in NATL
- HAFS models produce R34 small bias in RVCN relative to current large bias at all lead times in EPAC

Wind radii verification summary

NATL Summary

- Replacing HWRF with either HAFS-A or HAFS-S in RVCN reduces wind radii skill in range of ~2–10% for all lead times
- HAFS models amplify small wind radii biases at all lead times

EPAC Summary

- Replacing HWRF with either HAFS-A or HAFS-S in RVCN improves wind radii skill in range of ~2–15% at most lead times
- HAFS models produce R34 small bias relative to large bias in RVCN; amplify R50 small biases; and reduce R64 biases

Towards a (more) Comprehensive HAFS Evaluation

- Consider stratifying the forecast verification into various composites
  - Weak vs. strong TCs (set an intensity threshold)
  - Regional verifications (e.g., main development region vs. full NATL basin)
  - Intensity forecasts with/without aircraft observations
  - Rapid intensification (RI) vs. non-RI cases

- Identify potential outlier cases that might be contributing to a significant reduction in skill from HAFS
  - Diagnose source(s) of errors from model fields (contingent upon availability)
  - Assessing median absolute error would relax sensitivity to outliers

- Diagnose downstream implications of replacing HWRF/HMON with HAFS
  - Additional testing required for HCCA
  - Impacts on DTOPS, HCCA, and NNIC need to be assessed

Preliminary 2020–2022 HAFS Evaluation Summary

- Greatest improvement from HAFS noted for track consensus
  - HAFS improves track skill of TVCN by ~5% at all lead times
  - HAFS track skill translates into 6–12% improvements for HCCA in NATL

- HAFS has basin-dependent impacts on intensity consensus
  - HAFS reduces intensity skill of IVCN in NATL at days 2–5 up to ~10%
  - HAFS improves intensity skill of IVCN in EPAC at days 1–4 by ~2%; greater improvements at days 4–5 (2–14%)

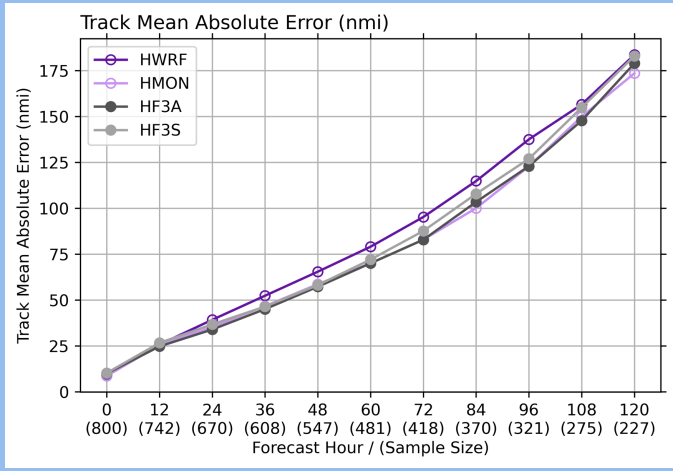
- HAFS has basin-dependent impacts on wind radii skill
  - HAFS reduces wind radii skill of RVCN in NATL in range of ~2–10% for all lead times
  - HAFS improves wind radii skill of RVCN in EPAC in range of ~2–15% at most lead times

Extra Slides

Track verification for individual models

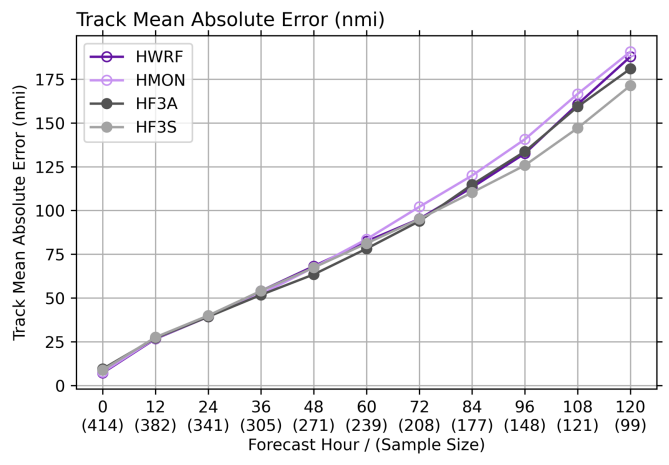
North Atlantic

Mean Absolute Error

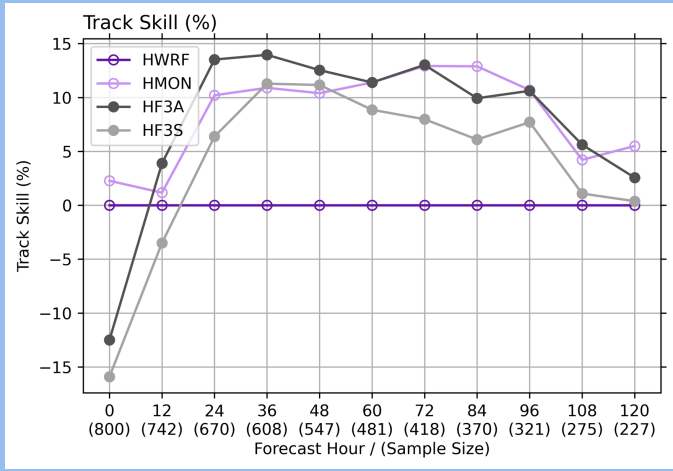


East Pacific

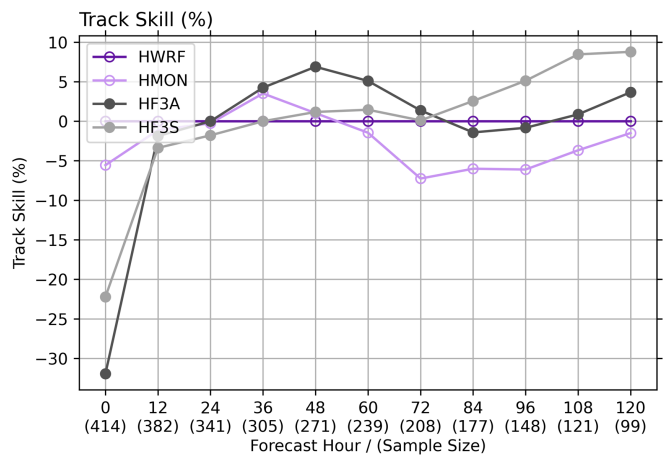
Mean Absolute Error



Relative Skill



Relative Skill

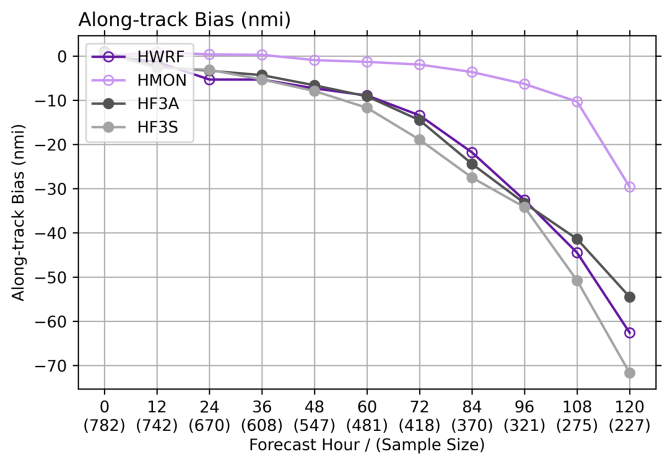


- Consistent improvement of track skill for HAFS at ~all lead times (0–14%)
- Track skill improvement in NATL largest at days 1–3
- Bimodal track skill improvement in EPAC greatest at days 1–2 and 4–5

Track verification for individual models

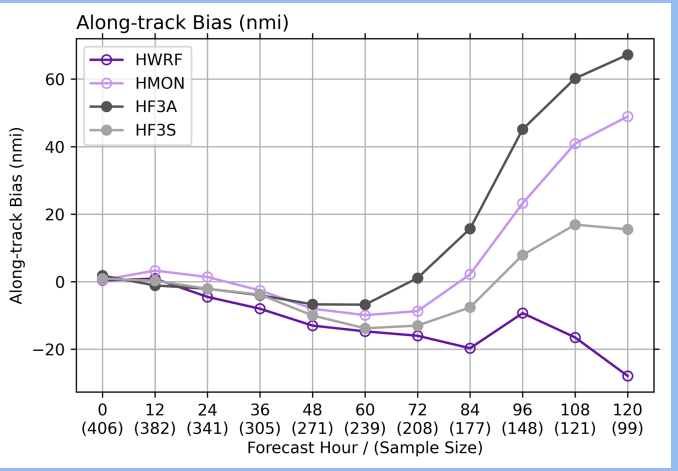
North Atlantic

Along-track Bias

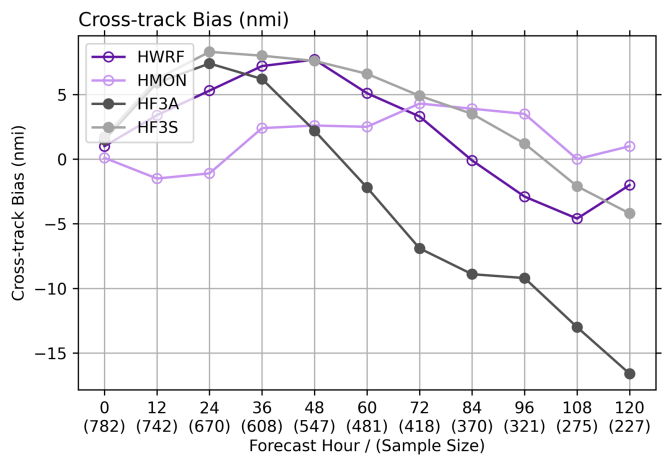


East Pacific

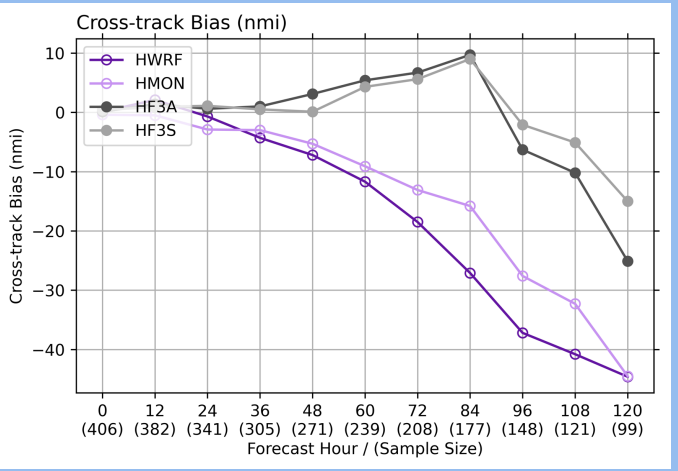
Along-track Bias



Cross-track Bias



Cross-track Bias

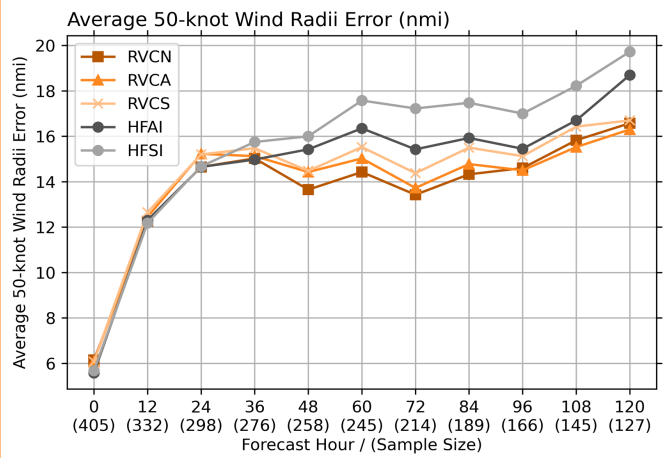


- HAFS slow bias in NATL comparable to HWRF
- HAFS-A growing left-of-track bias in NATL at days 2–5
- HAFS slow bias in EPAC between days 1–3 comparable to HWRF/HMON
- HAFS cross-track bias in EPAC improved at all lead times

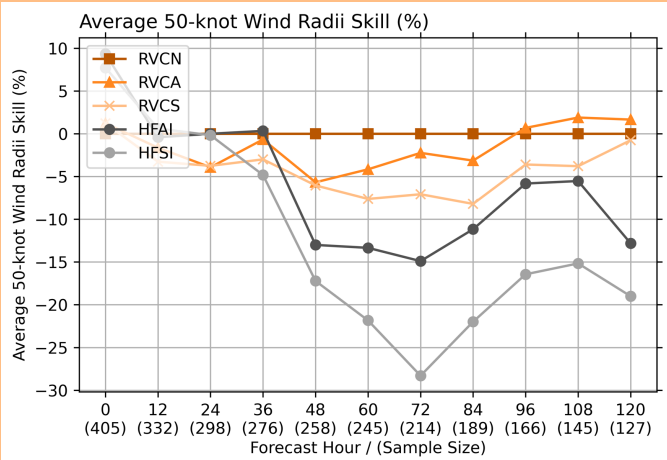
R50 verification for consensus models

North Atlantic

Average R50 Error

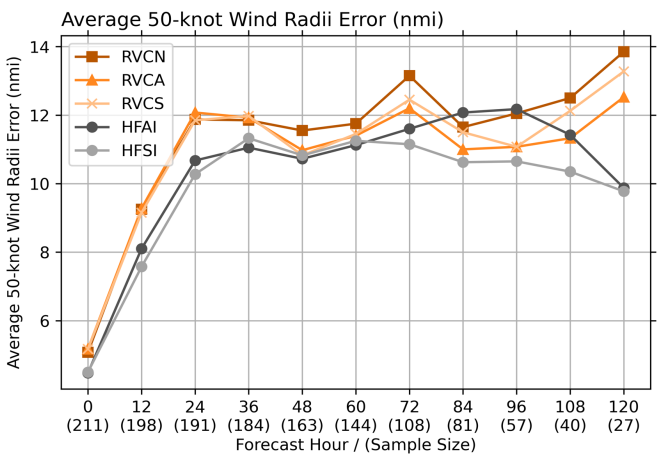


Relative Skill

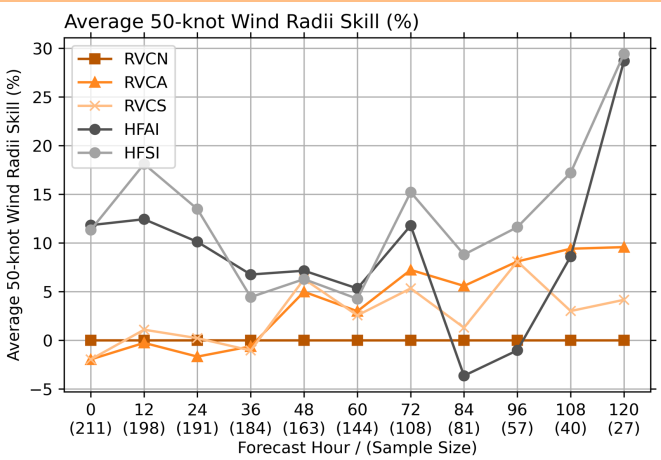


East Pacific

Average R50 Error



Relative Skill

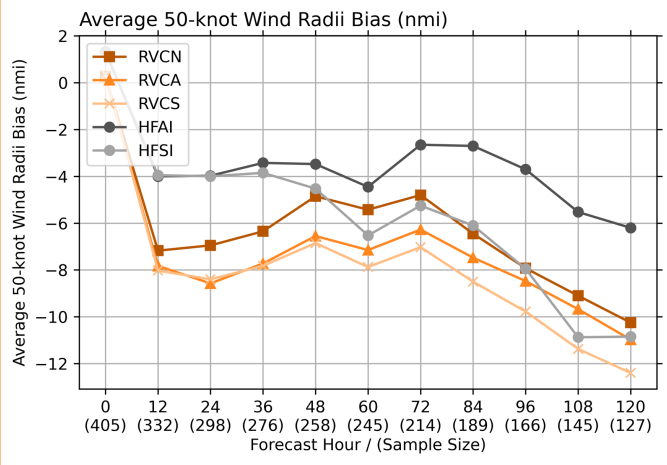


- HAFS models reduce RVCN R50 skill in NATL by ~4–8% at most lead times
- HAFS models improve RVCN R50 skill in EPAC by ~5–10% at days 2–5

R50 verification for consensus models

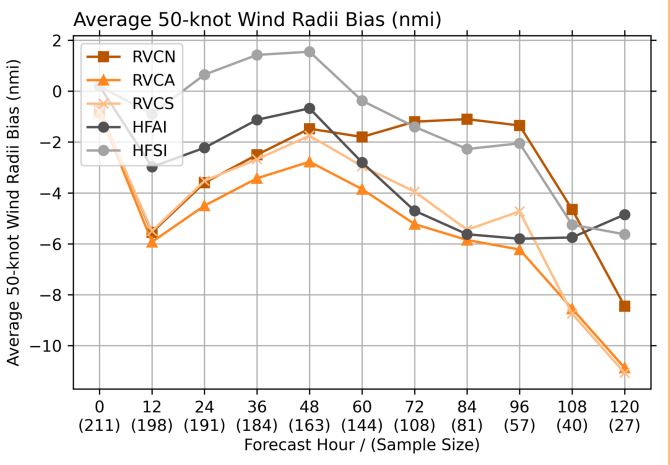
North Atlantic

Average R50 Bias



East Pacific

Average R50 Bias

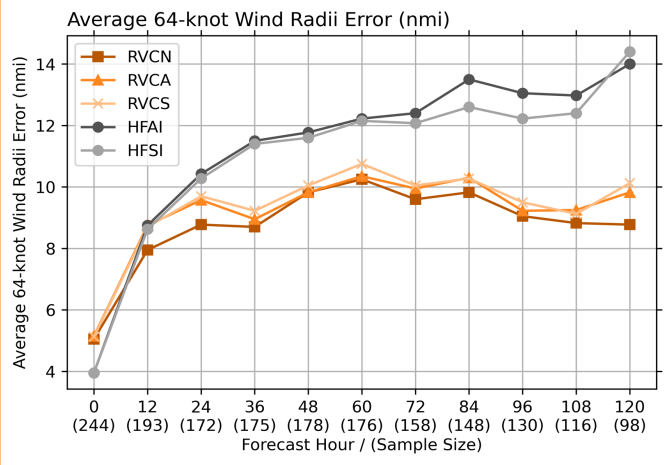


- HAFS models amplify small R50 bias in RVCN at all lead times in NATL and EPAC

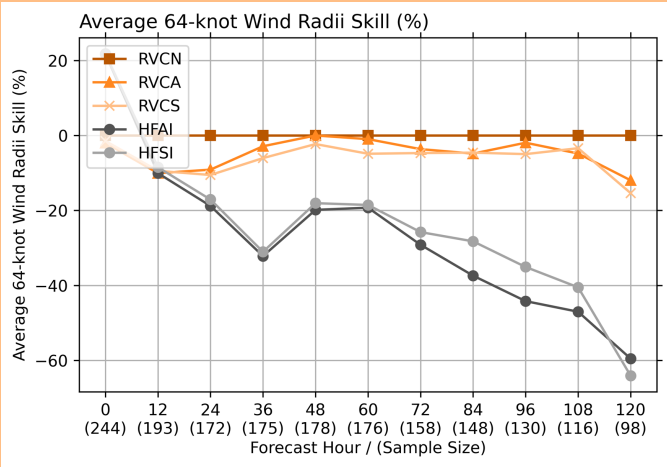
R64 verification for consensus models

North Atlantic

Average R64 Error

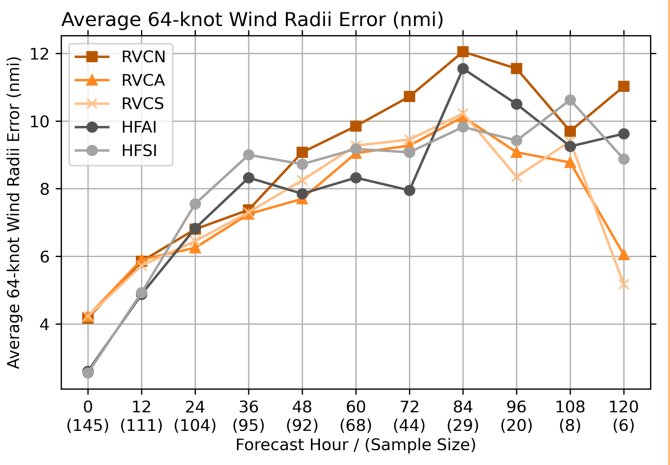


Relative Skill

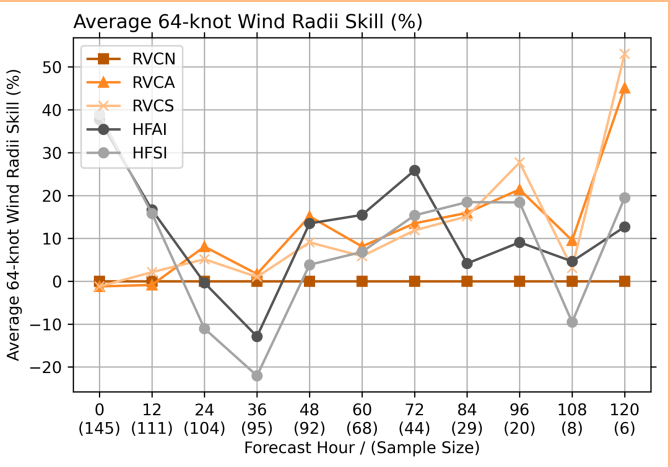


East Pacific

Average R64 Error



Relative Skill

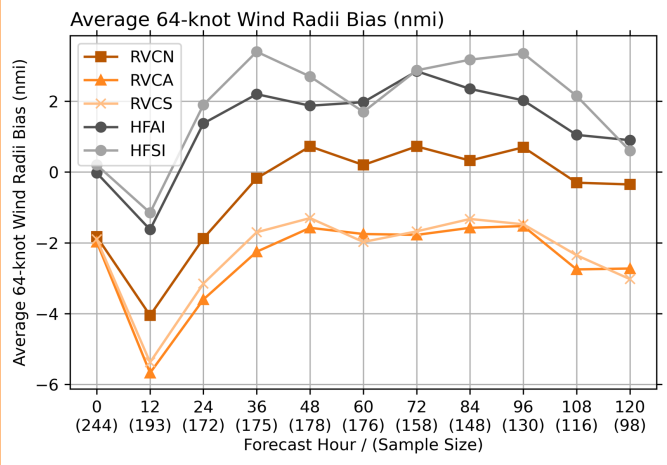


- HAFS models reduce RVCN R64 skill in NATL by ~5–15% at all lead times
- HAFS models improve RVCN R64 skill in EPAC by ~10–20% at most lead times
- *Note the small sample sizes

R64 verification for consensus models

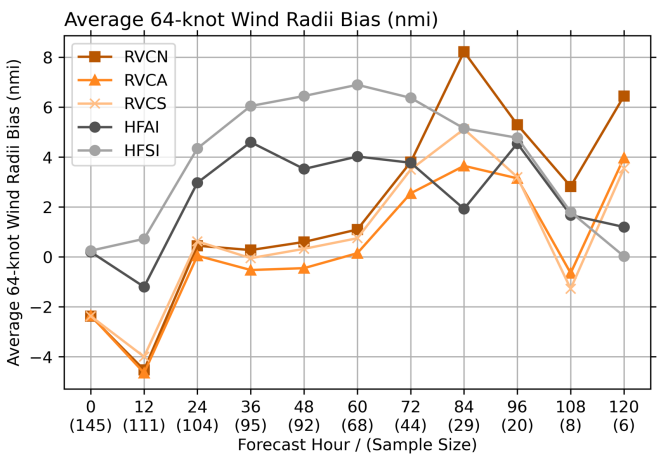
North Atlantic

Average R64 Bias

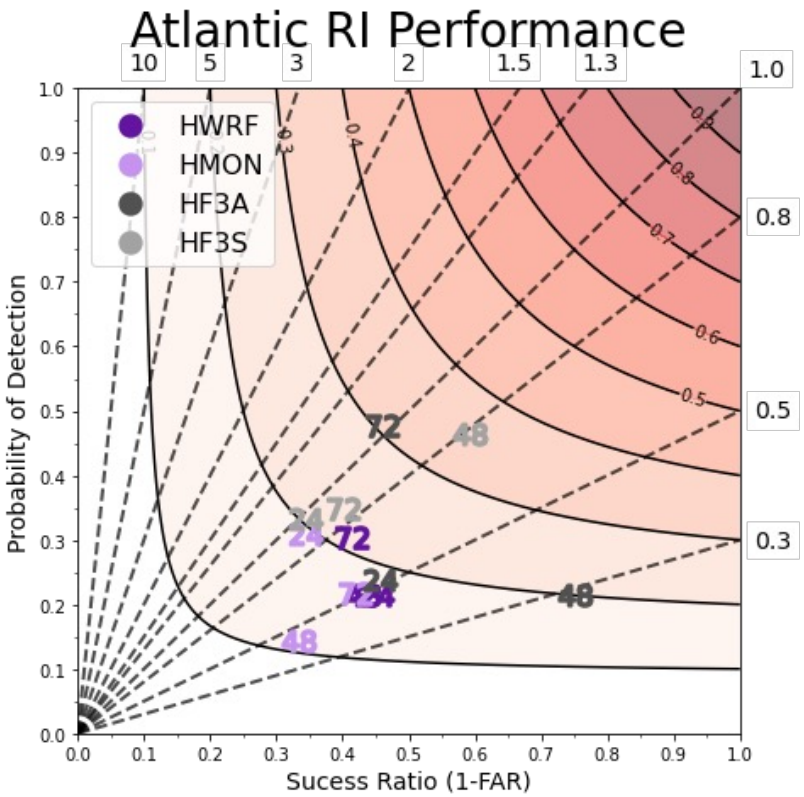


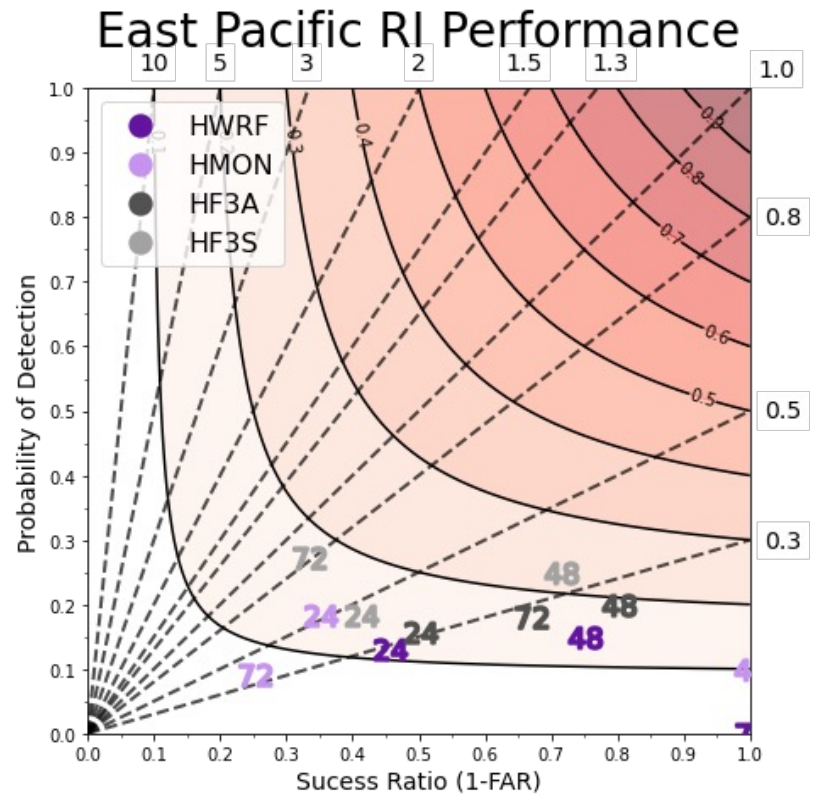
East Pacific

Average R64 Bias



- HAFS models amplify small R64 bias in RVCN at all lead times in NATL
- HAFS models have less R64 biases in RVCN at days 1–5 in EPAC


Atlantic RI Performance
HWRF
HMON
HF3A
HF3S
Probability of Detection
Sucess Ratio (1-FAR)


East Pacific RI Performance
HWRF
HMON
HF3A
HF3S
Probability of Detection
Sucess Ratio (1-FAR)