

Reports
Tutorial

Page Loaded:
04/22/2013 19:25 UTC

Year		
2007	2008	2009
2010	2011	2012
2013	2014	2015

Month					
Jan	Feb	Mar	Apr	May	Jun
Jul	Aug	Sep	Oct	Nov	Dec

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	

Using mPING Data to Generate Random Forests for Precipitation Type Forecasts

Kimberly L. Elmore, OU CIMMS & NOAA/NSSL,
Heather M. Grams, OU CIMMS & NOAA/NSSL,

and

David John Gagne II, OU

Out (Full) Zoom Out (x2) Show History On

Active Window Duration: 10 m 20 m 30 m 1 hr 2 hr 6 hr

03 06 09 12 15 18 21

2/16/2013 [UTC]

de Precipitation Types:

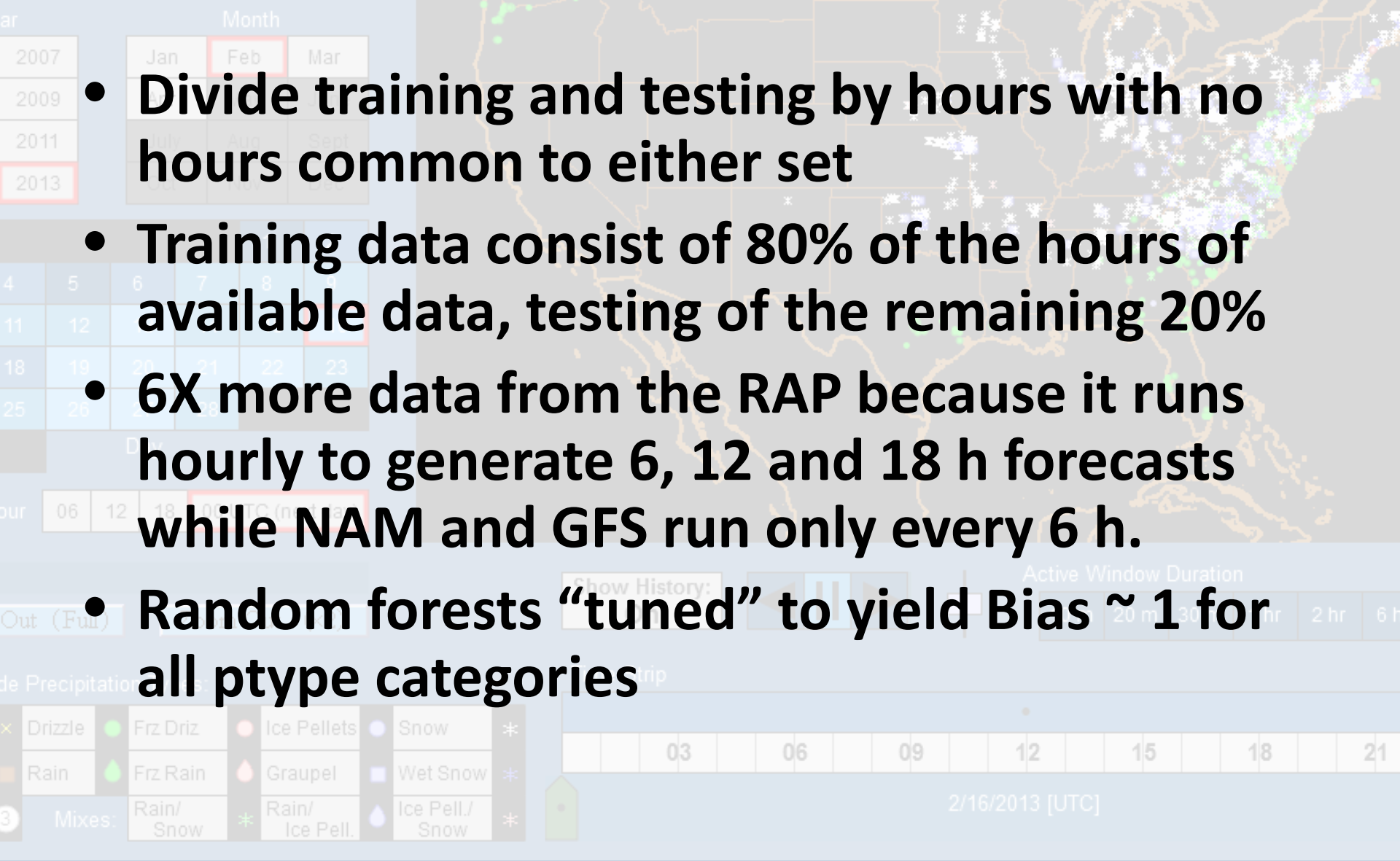
Drizzle	Frz Driz	Ice Pellets	Snow
Rain	Frz Rain	Graupel	Wet Snow
Mixes: Rain/Snow	Rain/Ice Pell.	Ice Pell./Snow	

Can We Significantly Improve Forecast Ptypes Using AI Techniques?

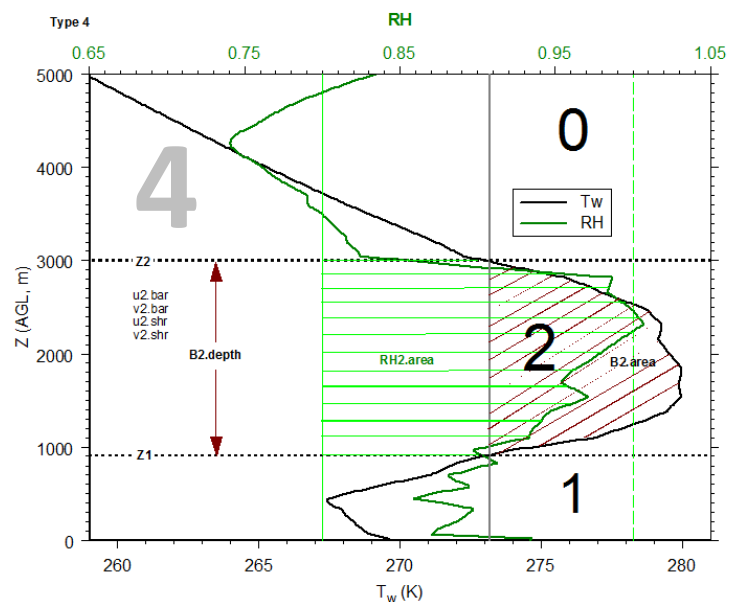
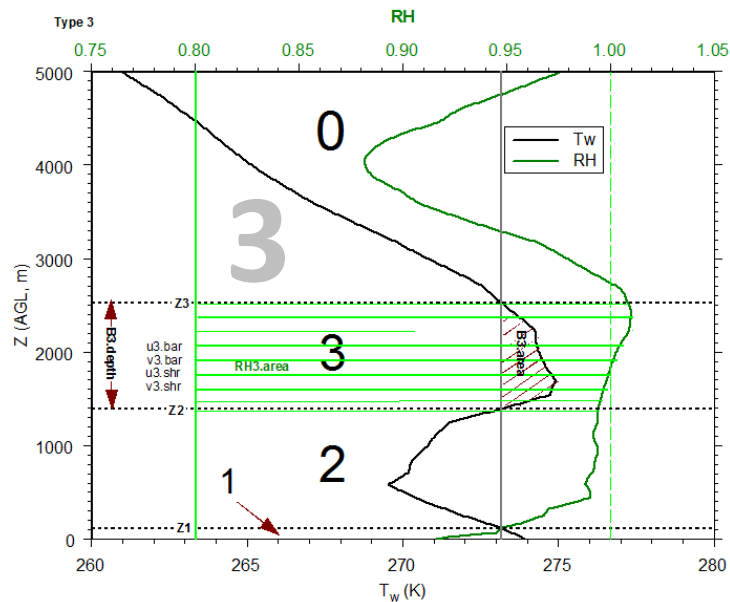
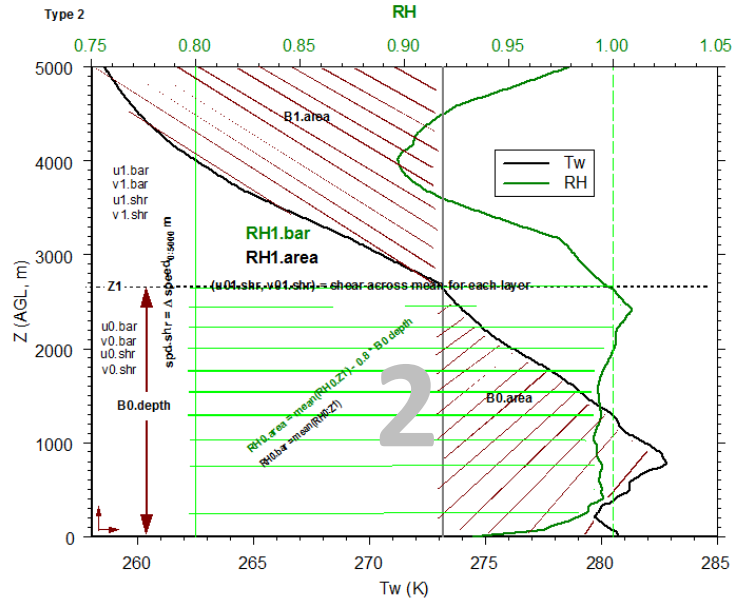
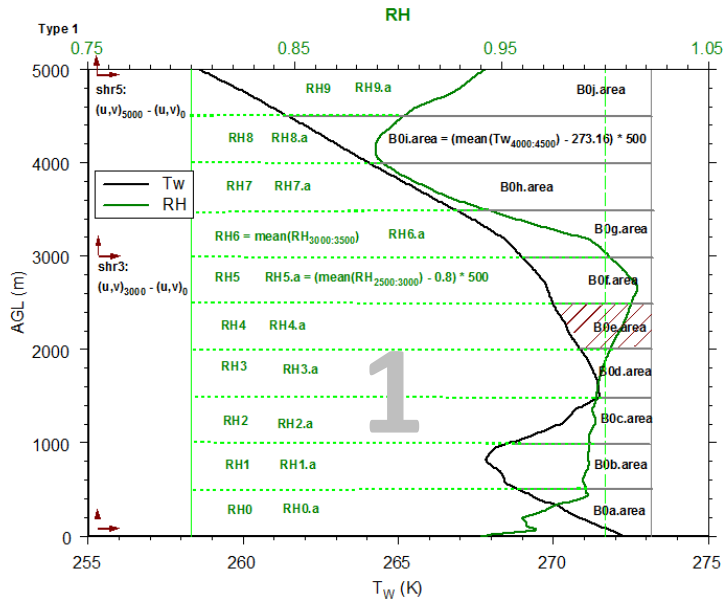
- Generate a Random Forest classifier that uses forecast soundings at mPING observations for each model
- Use mPING obs at closest model grid point ± 30 min from the forecast valid time to evaluate Random forests generated from the RAP, NAM and GFS every 6 h out to 18 h lead time
 - use pressure-level data (native vertical coordinate data unavailable)

Training and Testing Data

- Divide training and testing by hours with no hours common to either set
- Training data consist of 80% of the hours of available data, testing of the remaining 20%
- 6X more data from the RAP because it runs hourly to generate 6, 12 and 18 h forecasts while NAM and GFS run only every 6 h.
- Random forests “tuned” to yield Bias ~ 1 for all ptype categories

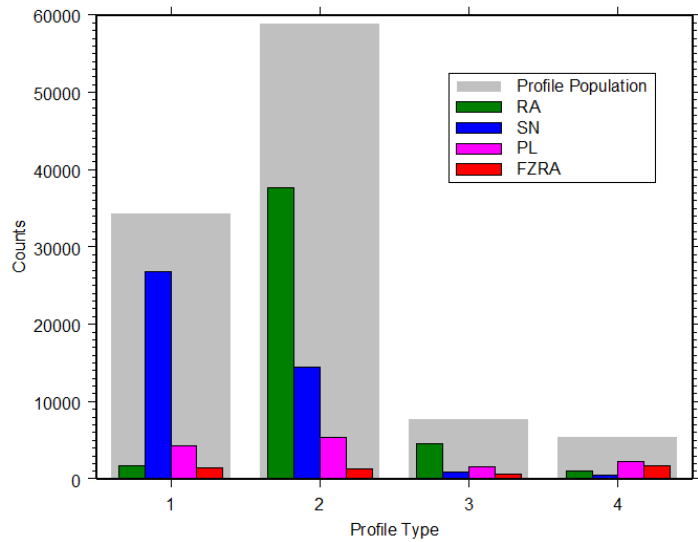


Random Forest Attributes

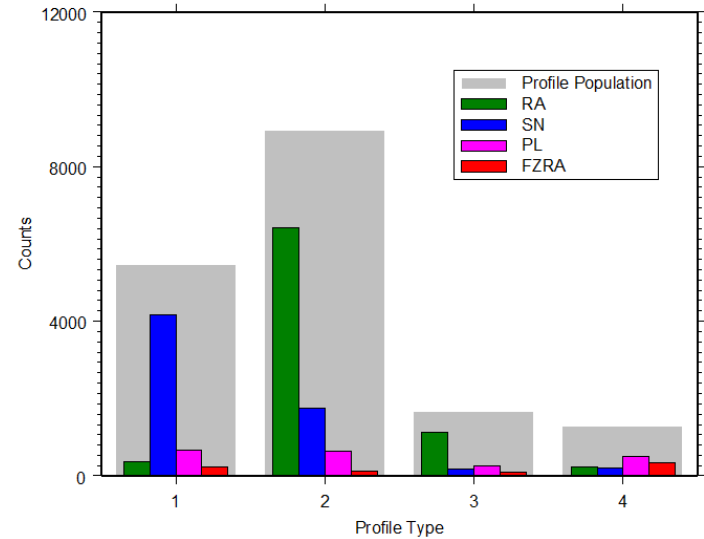


Data Set Size

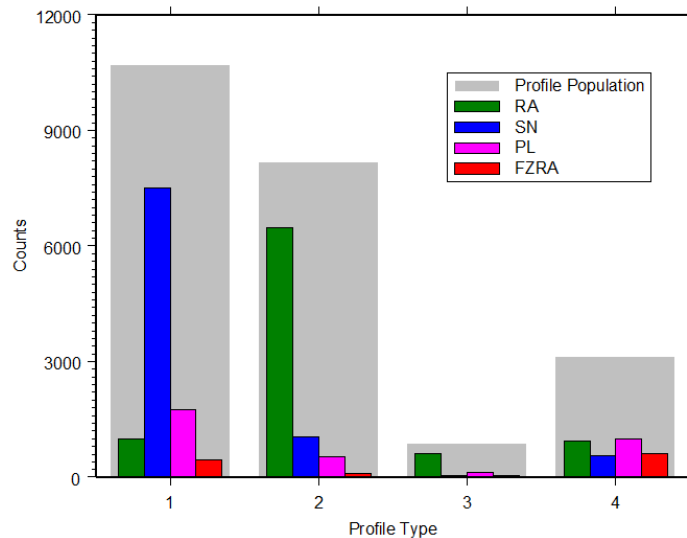
RAP Ptype and Profile Populations



NAM Ptype and Profile Populations



GFS Profile and Ptype Populations



Active Window Duration

10 m 20 m 30 m 1 hr 2 hr 6 hr

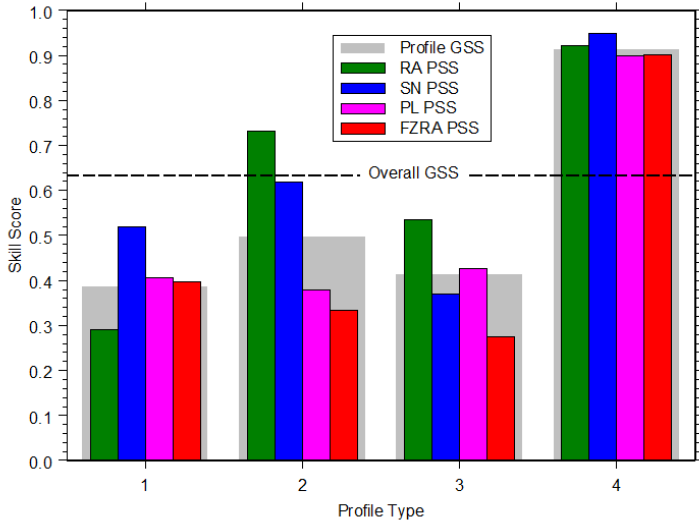
12 15 18 21

013 [UTC]

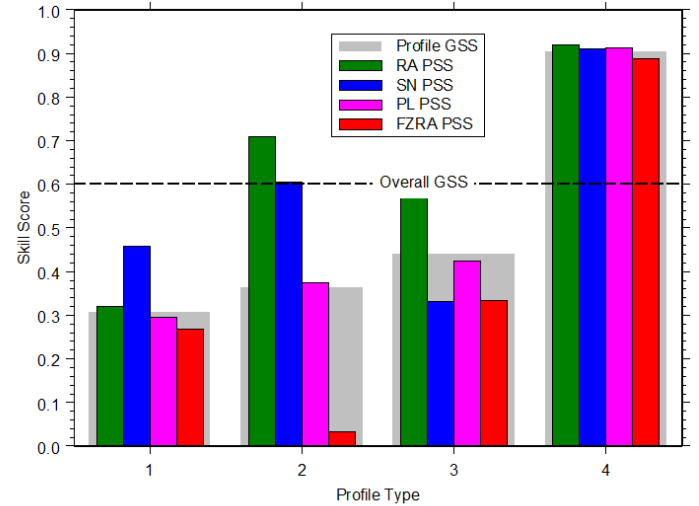
How Well do Random Forests Perform?

Page Loaded:
04/22/2013 19:25 UTC

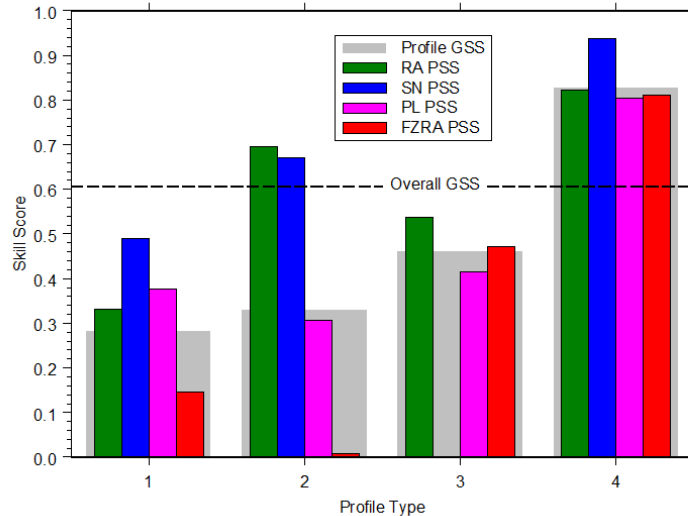
RAP Scores by Profile and Ptype



NAM Scores by Profile and Ptype



GFS Scores by Profile and Ptype



Active Window Duration

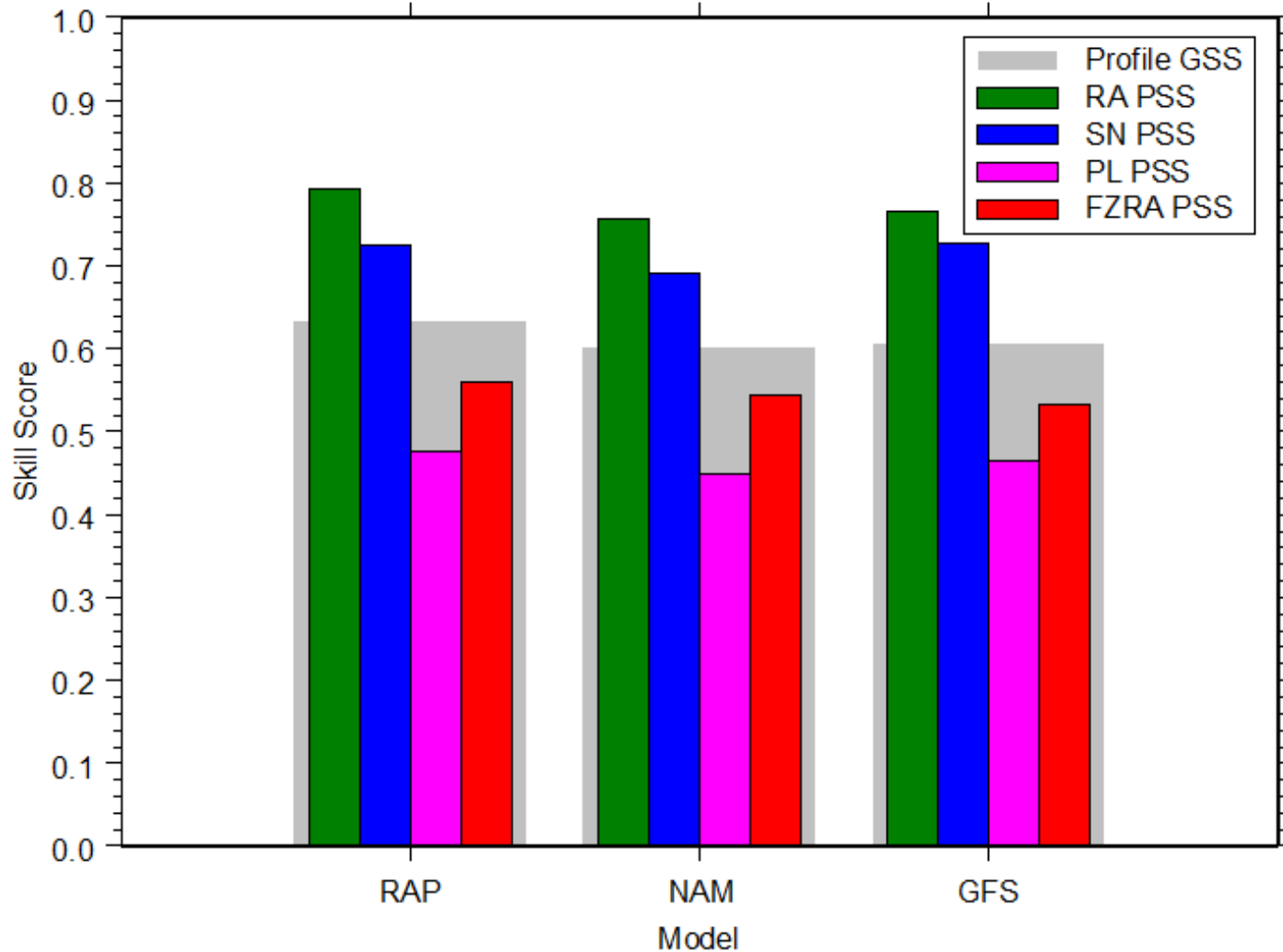
10 m 20 m 30 m 1 hr 2 hr 6 h

12 15 18 21

2013 [UTC]

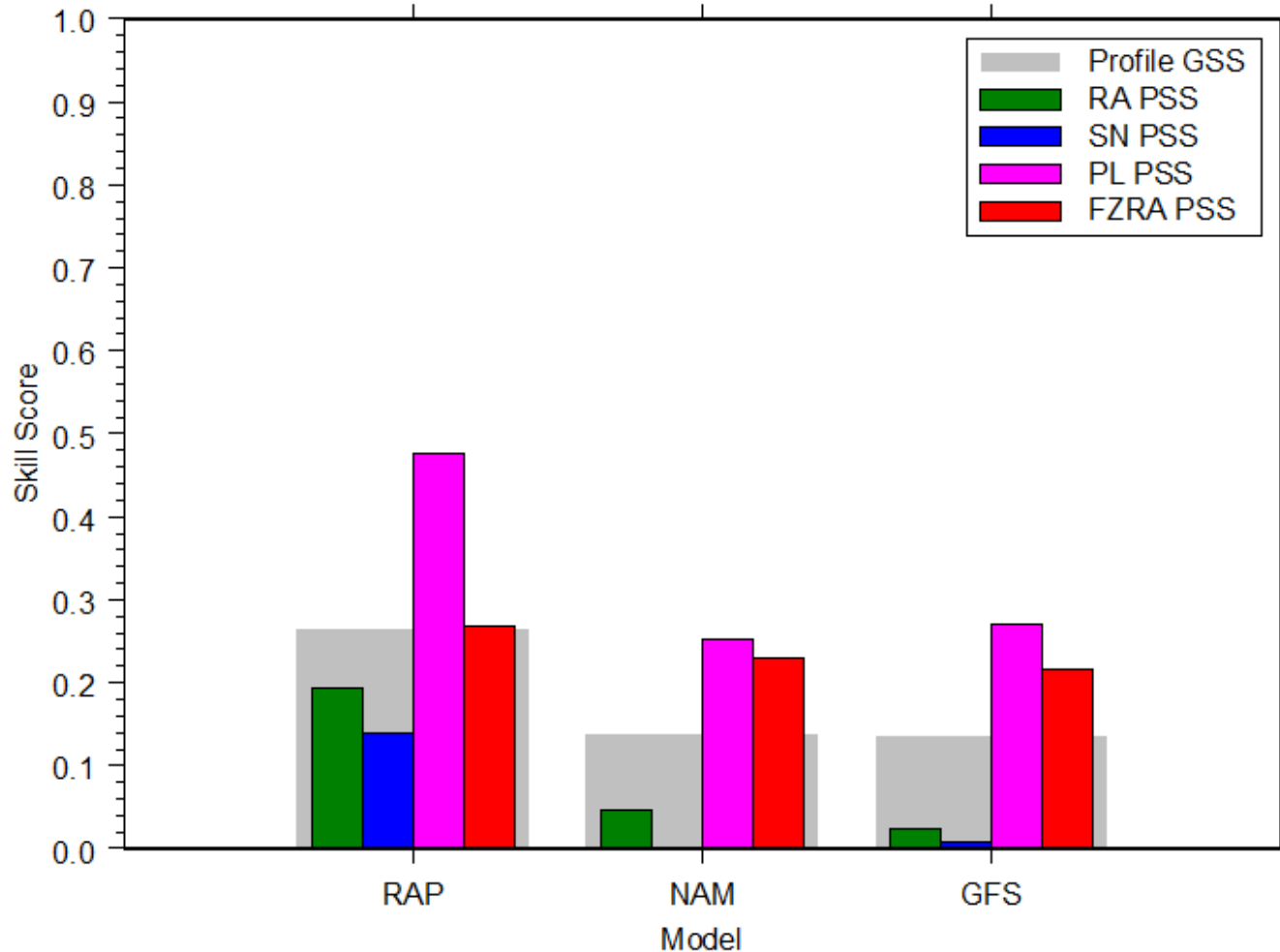
Aggregate Random Forest Performance

Performance by Model



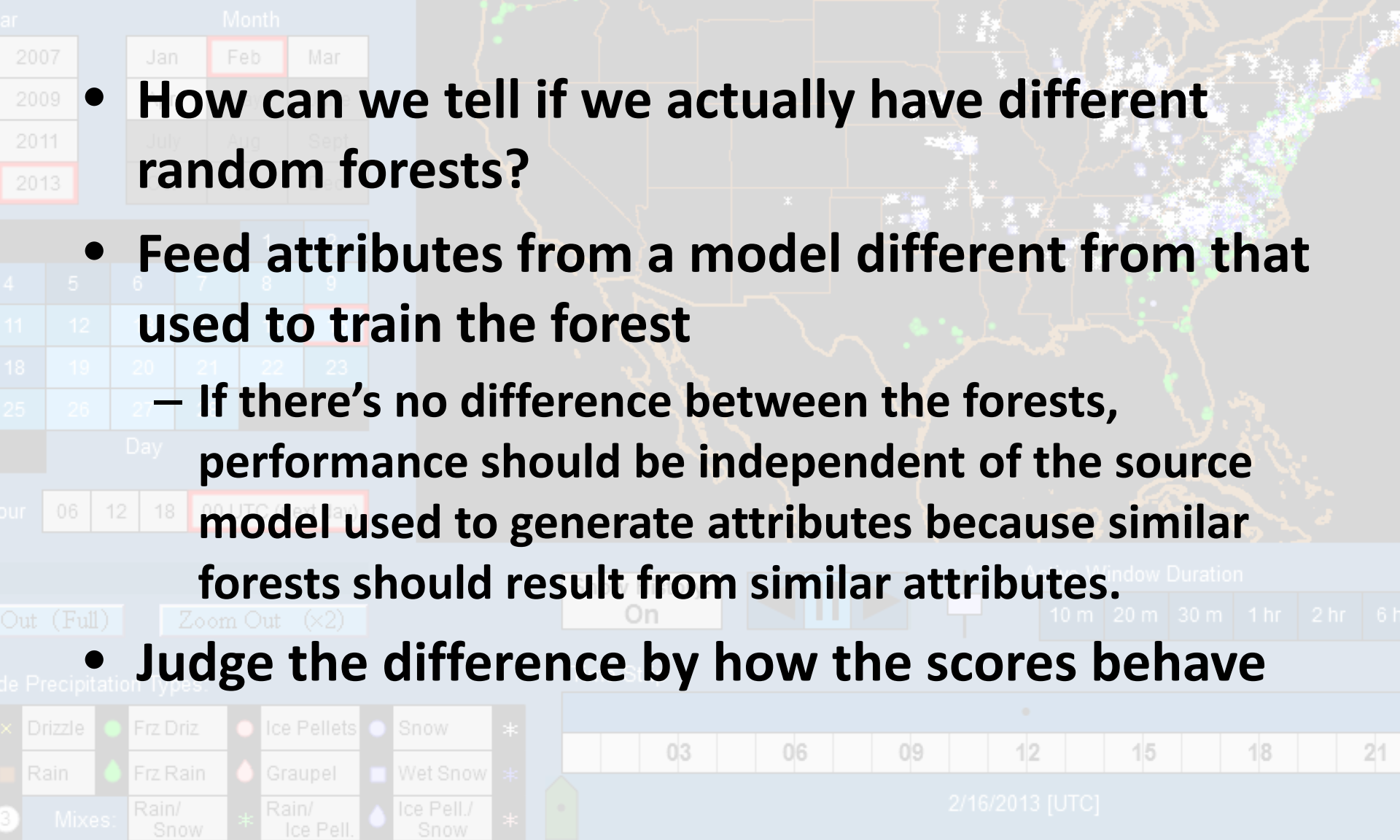
Random Forest Improvement

RF Improvement over Current Algorithms



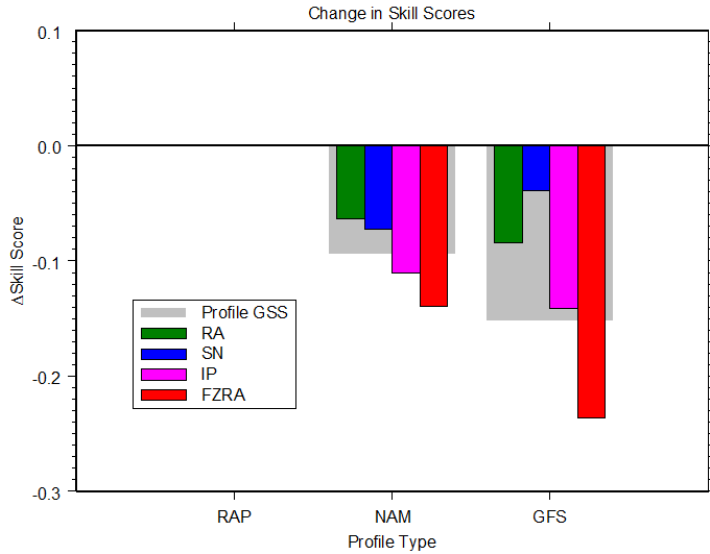
Not Much Difference Between Models!

- How can we tell if we actually have different random forests?
- Feed attributes from a model different from that used to train the forest
 - If there's no difference between the forests, performance should be independent of the source model used to generate attributes because similar forests should result from similar attributes.
- Judge the difference by how the scores behave

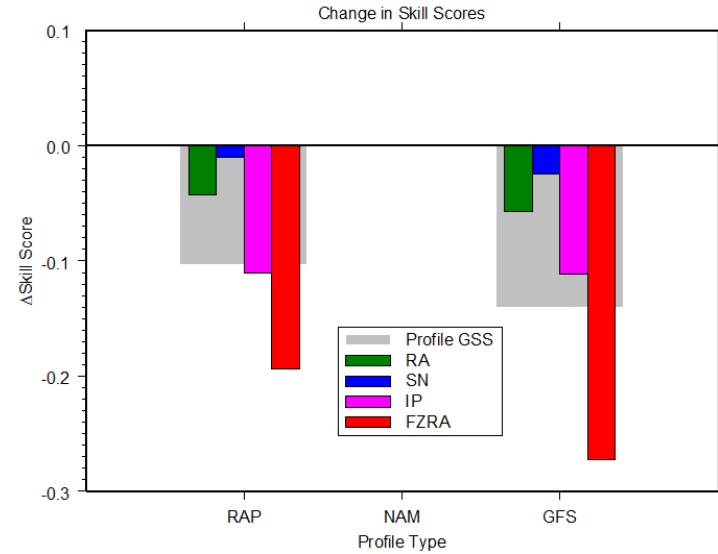


And Results show...

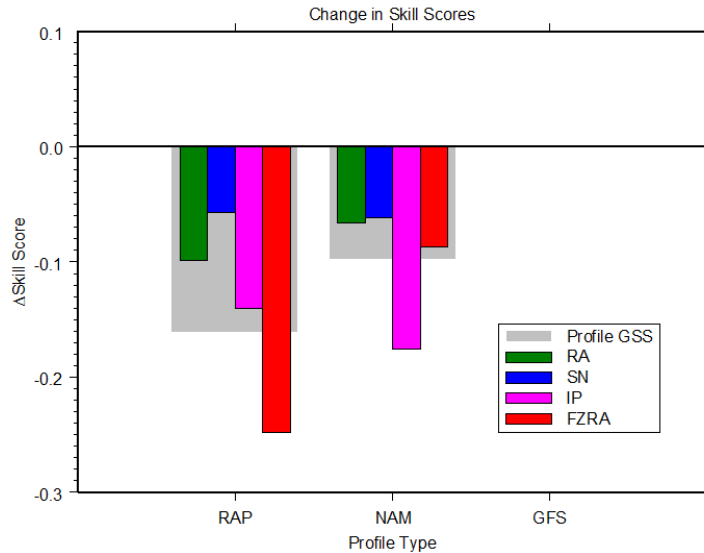
Random Forest Trained on RAP



Random Forest Trained on NAM



Random Forest Trained on GFS



Report

Victoria

ar

200

200

201

201

4

11

18

25 26 27 28

Day

06 12 18 00 UTC (next day)

Out (Full) Zoom Out (x2)

de Precipitation Types:

Drizzle	Frz Driz	Ice Pellets
Rain	Frz Rain	Graupel
Mixes:	Rain/Snow	Rain/Ice Pell.

Active Window Duration

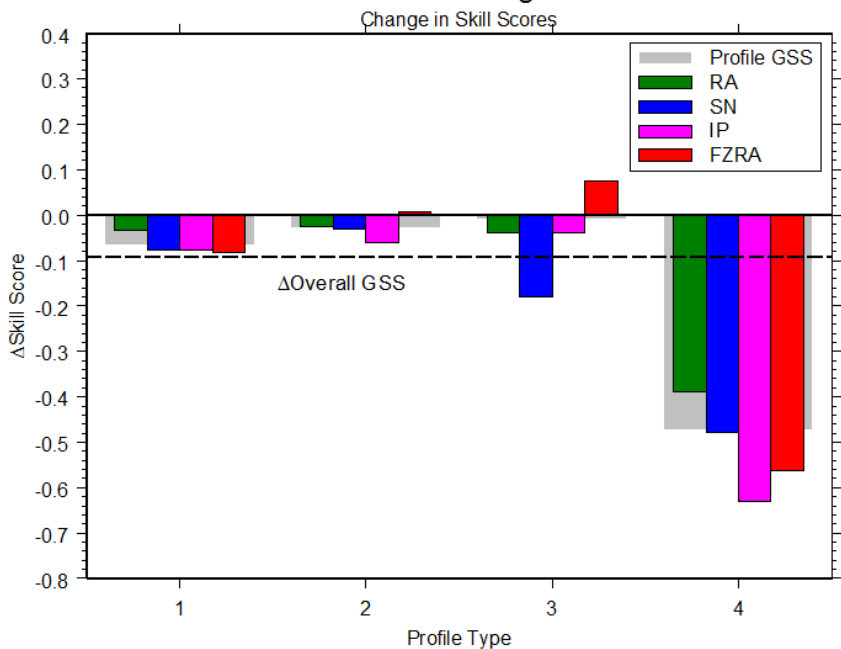
10 m 20 m 30 m 1 hr 2 hr 6 hr

12 15 18 21

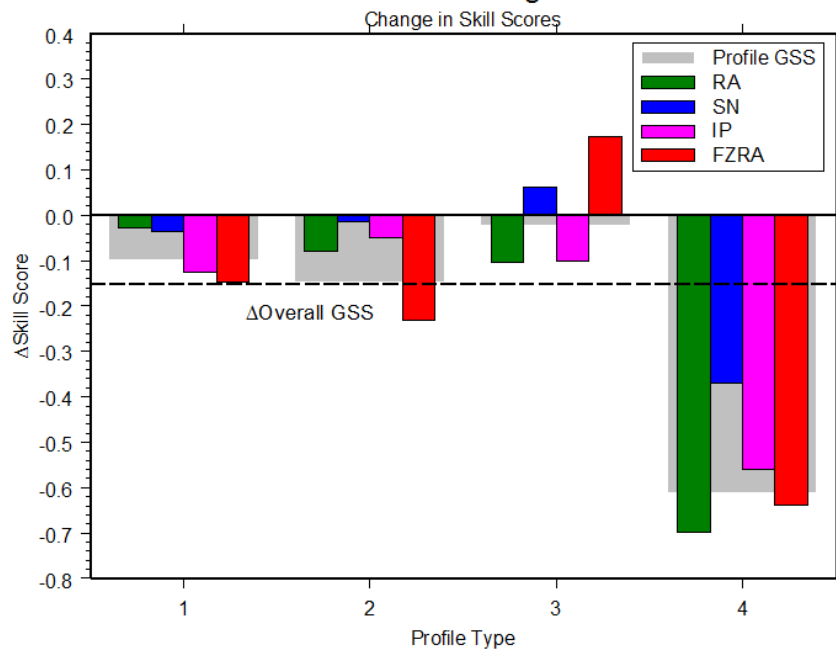
2013 [UTC]

Details of Degradation

Random Forest from RAP Using Attributes from NAM



Random Forest from RAP Using Attributes from GFS



Identify Precipitation Types:

Drizzle	Frz Driz	Ice Pellets	Snow	*
Rain	Frz Rain	Graupel	Wet Snow	*
Mixes:	Rain/Snow	Rain/Ice Pell.	Ice Pell./Snow	*

Time Strip

03 06 09 12 15 18 21

2/16/2013 [UTC]

Conclusions

- Random forests applied to forecast soundings are effective at generating skillful forecasts of surface ptype
- Random forests are able to extract essentially equivalent information from different forecast models
 - The random forest for each model, and each profile type is unique to the particular forecast model
 - Random forests developed using a particular model suffer significant degradation when given attributes derived from a different model
 - Implies that no single algorithm can perform well across all forecast models
- Random forests extract information unavailable to “physically based” methods because the physical information in the models does not appear as we expect
- Ptype results from the classic “warm nose” (Type 4) profile are most sensitive to the forecast model, but this profile is also the one for which random forests are most skillful
- **What’s next? Ptype probabilities!**