Use of reforecasts in the production of CPC's 6-10 day temperature forecasts

Tom Hamill and Jeff Whitaker

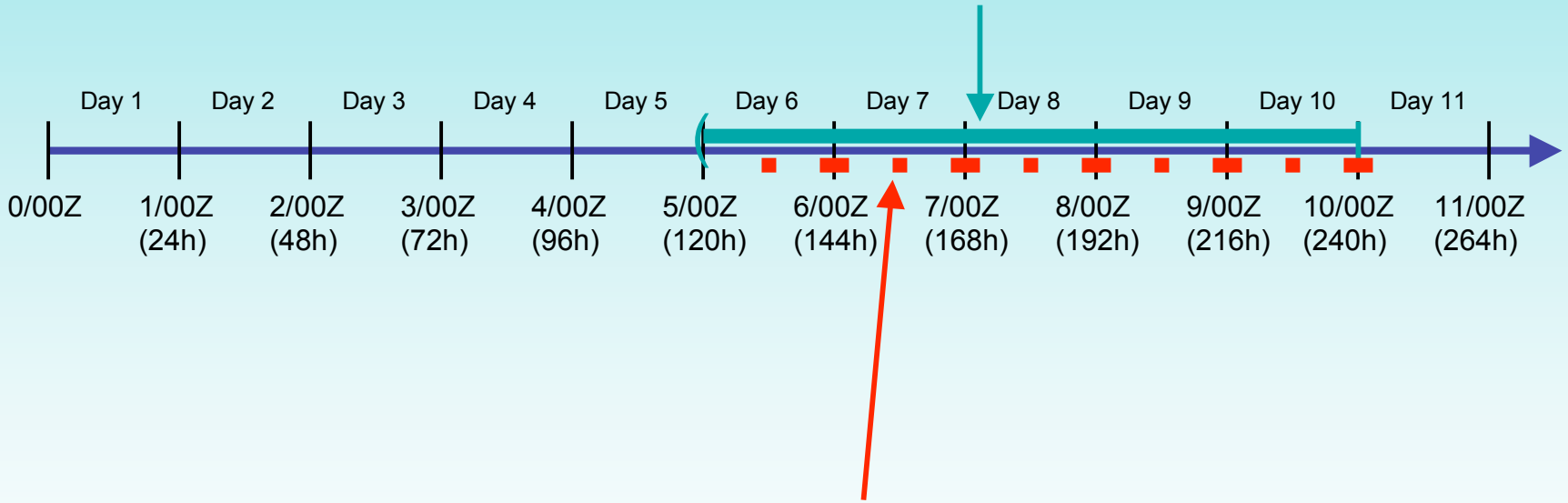
NOAA / ESRL / PSD

Tom.hamill@noaa.gov

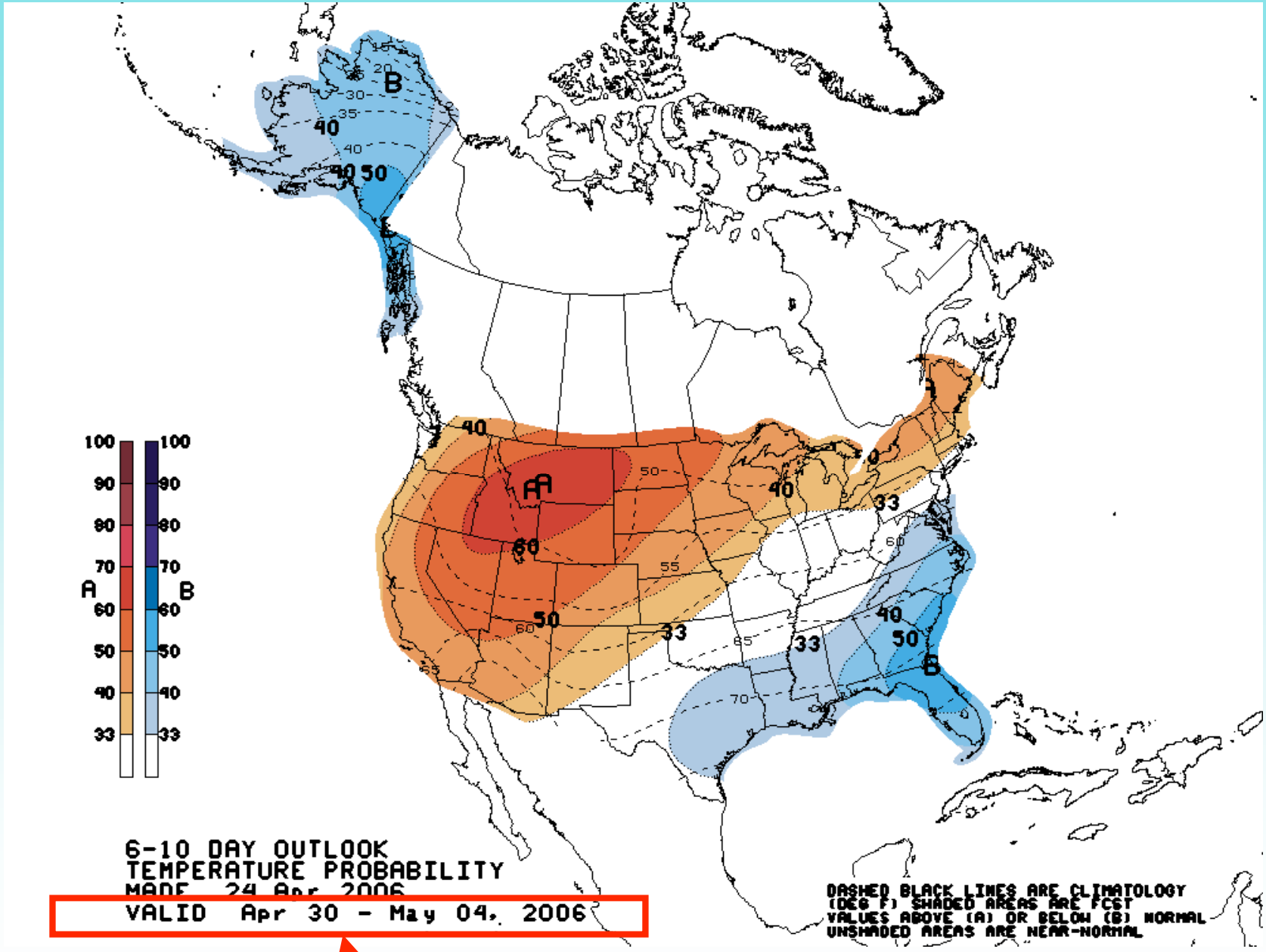
Jeffrey.s.whitaker@noaa.gov

What IS a 6-10 day forecast?

Most straightforward definition:
average value between 120 h & 240 h,
assuming 0000 UTC start.



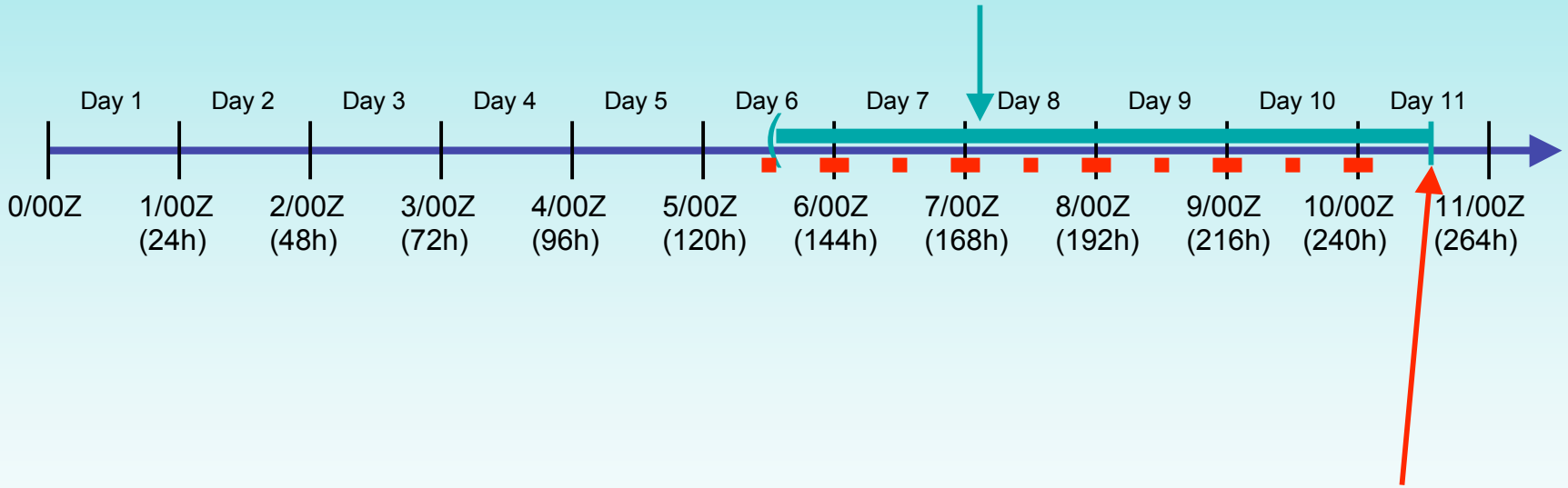
(CPC forecast tools and ESRL reforecast
tools extract forecast data from 132 h to 240 h)



Ambiguous: 00Z - 00Z? 12Z - 12Z? 00 LST - 00 LST?

What IS a 6-10 day forecast?

Alternative definition:
average value between 132 h & 252 h,
1200 UTC to 1200 UTC



(then right now we're not training including last time)

And then there's ambiguity in the observations, also.

1979010200	-3.90
1979010201	-3.80
1979010202	-3.80
1979010203	-3.90
1979010204	-3.80
1979010205	-4.30
1979010206	-4.40
1979010207	-4.90
1979010208	-4.90
1979010209	-5.60
1979010210	-6.60
1979010211	-7.10
1979010212	-7.20
1979010214	-7.70
1979010215	-7.80
1979010216	-8.20
1979010217	-9.30
1979010218	-10.60
1979010219	-12.10
1979010220	- Missing
1979010221	-14.40
1979010222	-16.00
1979010223	-17.10
1979010300	-18.30
1979010301	-19.30
1979010302	-19.90
1979010303	-20.00
1979010304	-20.50
1979010305	-21.00
1979010306	-21.10
1979010307	-21.00
1979010308	-20.50
1979010309	-20.60
1979010310	-19.90
1979010312	-20.00
1979010313	-19.90
1979010314	-19.30
1979010315	-17.80
1979010316	-16.60
1979010317	-16.00
1979010318	-15.00
1979010319	-14.30
1979010320	-14.30
1979010321	-13.90
1979010322	-13.80
1979010323	-14.30

(Toledo, OH)

Coop max for 19790102: -4.44, which is consistent with midnight - midnight next day.

More recently, Coop data consistent with max in subsequent 00Z-00Z period

2004022500	-0.60
2004022501	-2.00
2004022502	-2.00
2004022503	-2.00
2004022504	-3.00
2004022505	-4.00
2004022506	-4.40
2004022507	-4.00
2004022508	-4.00
2004022509	-5.00
2004022510	-5.00
2004022512	-6.00
2004022513	-3.00
2004022514	-1.00
2004022515	1.00
2004022516	2.00
2004022517	3.00
2004022518	2.80
2004022519	4.00
2004022520	4.00
2004022521	4.00
2004022522	2.00
2004022523	1.00

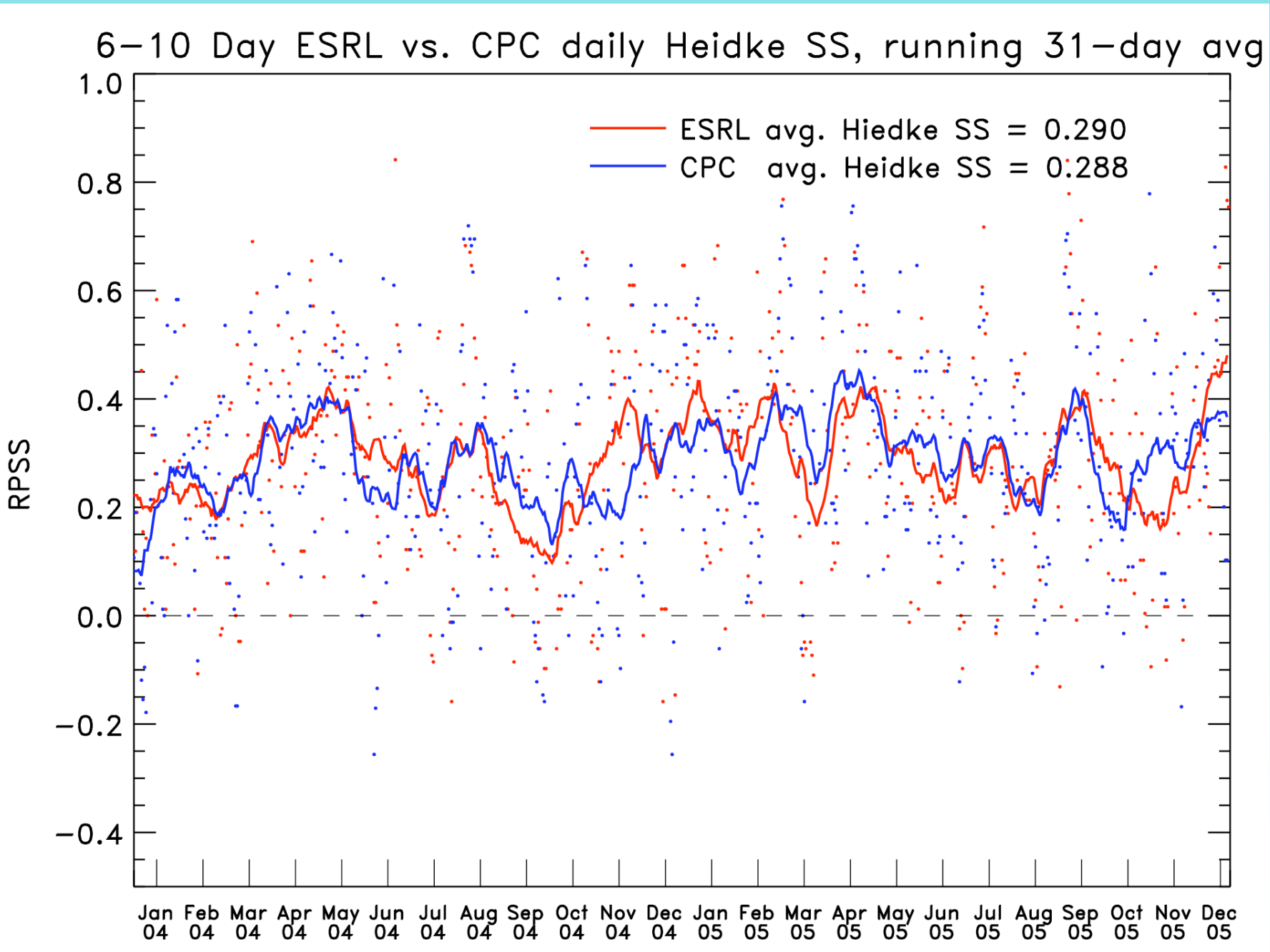
Coop max: 4.44

Implication: Our reforecasts,trained on coop max/min data, do not have a consistent time of max/min.
1 of the 5 days of our training data may be not what we expect. Unclear whether this will mean current skill estimates too large/too small.

Suggestions:
(1) Let's decide on 00Z-00Z or 12Z-12Z
(2) Use hourly data to remove ambiguity

Comparison of skill

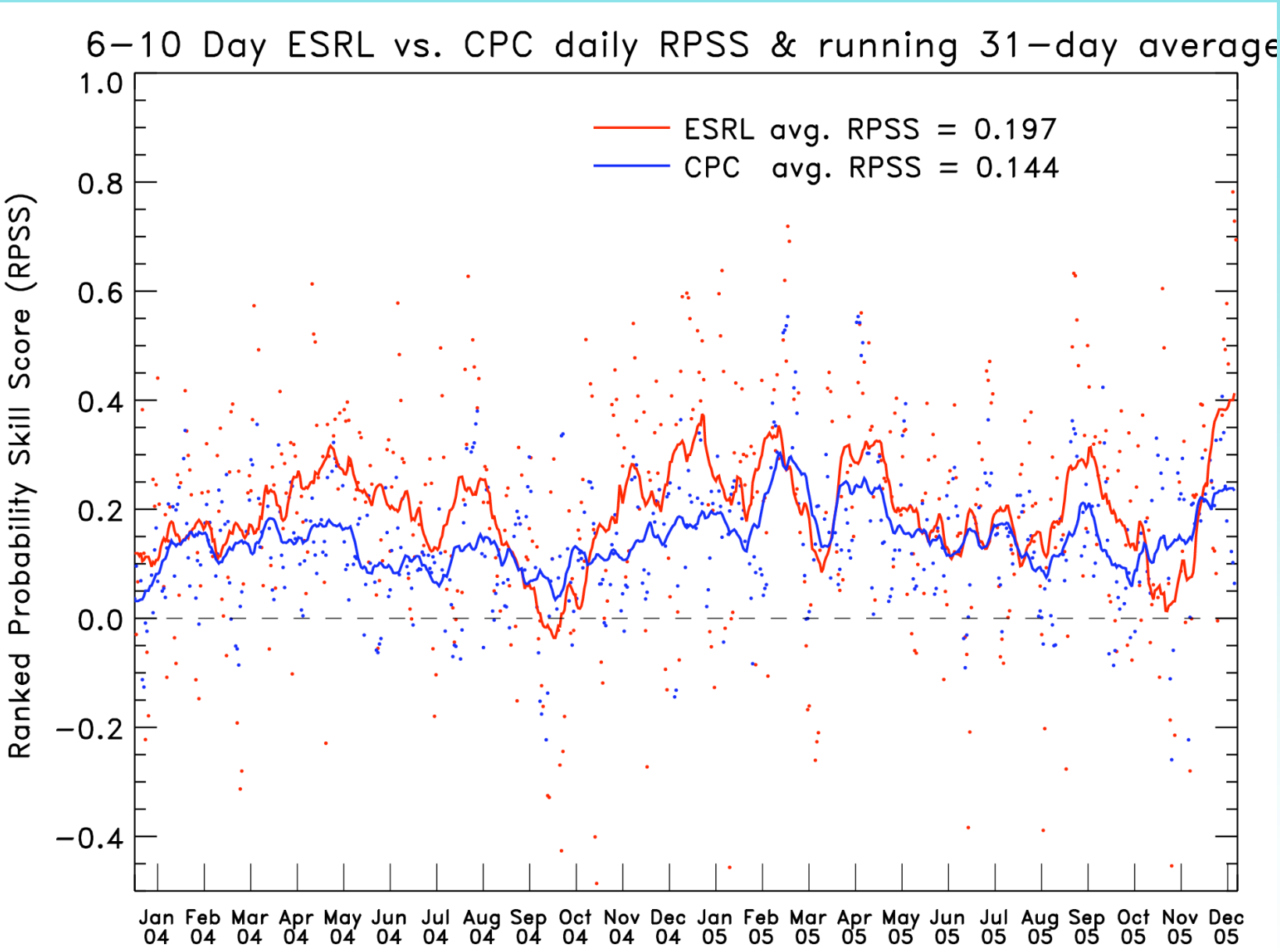
- Heidke SS and Brier SS, 6-10 day temp.
- NOTE: Preliminary result. Given ambiguity in observations, would like us to agree upon observation data set, we'll retrain and re-do comparison.


6–10 Day ESRL vs. CPC daily Heidke SS, running 31–day avg
ESRL avg. Hiedke SS = 0.290
CPC avg. Heidke SS = 0.288
RPSS
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
Jan 04
Feb 04
Mar 04
Apr 04
May 04
Jun 04
Jul 04
Aug 04
Sep 04
Oct 04
Nov 04
Dec 04
Jan 05
Feb 05
Mar 05
Apr 05
May 05
Jun 05
Jul 05
Aug 05
Sep 05
Oct 05
Nov 05
Dec 05

Problems with Heidke SS

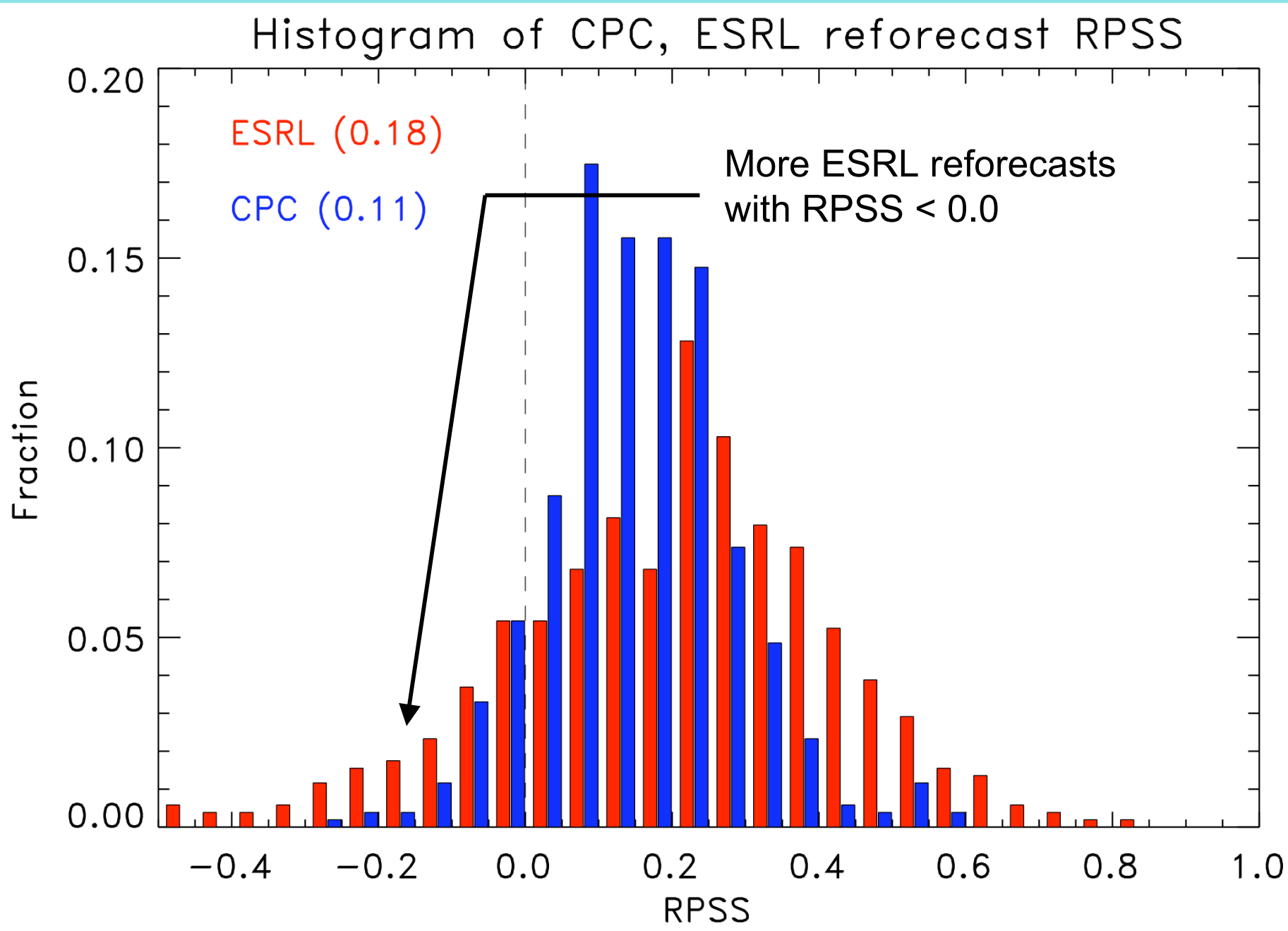
- Not “strictly proper” - i.e., maximum skill should be awarded when forecaster issues forecast of true belief of the probability. HSS encourages hedging. Example: F1, F2 awarded the same score:
  - F1 = [0.32, 0.33, 0.35] --> [0,0,1]
  - F2 = [0.00, 0.10, 0.90] --> [0,0,1]
  - V = [0, 0, 1]
- Won’t penalize greater misses more than smaller misses. Same penalty for F1, F2 when
  - F1 = [0,1,0]
  - F2 = [1,0,0]
  - V = [0,0,1]

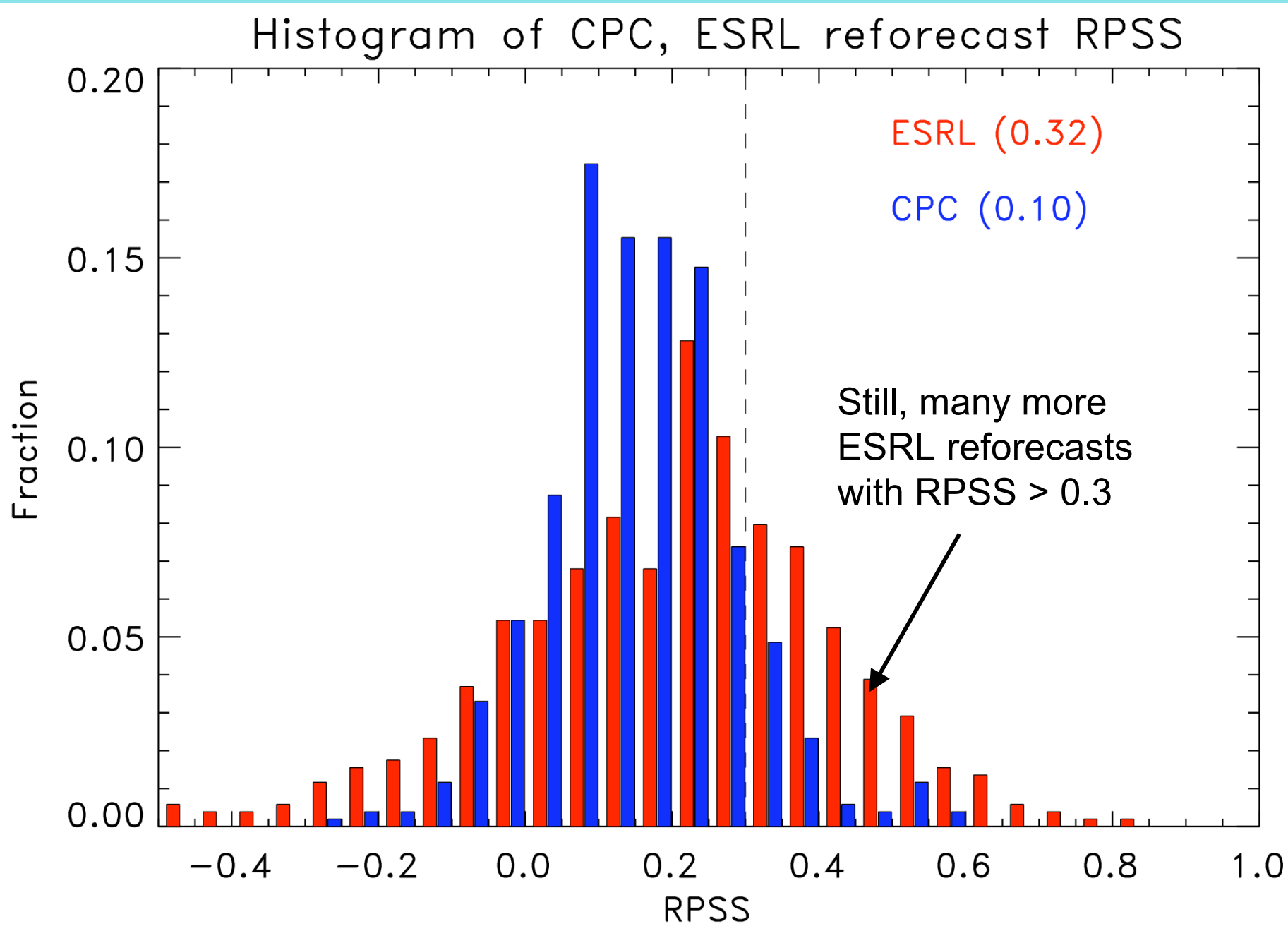
[inspiration from R. Livezey,
Chapter 4 of “*Forecast Verification:*
A Practitioner’s Guide in Atmospheric Science”

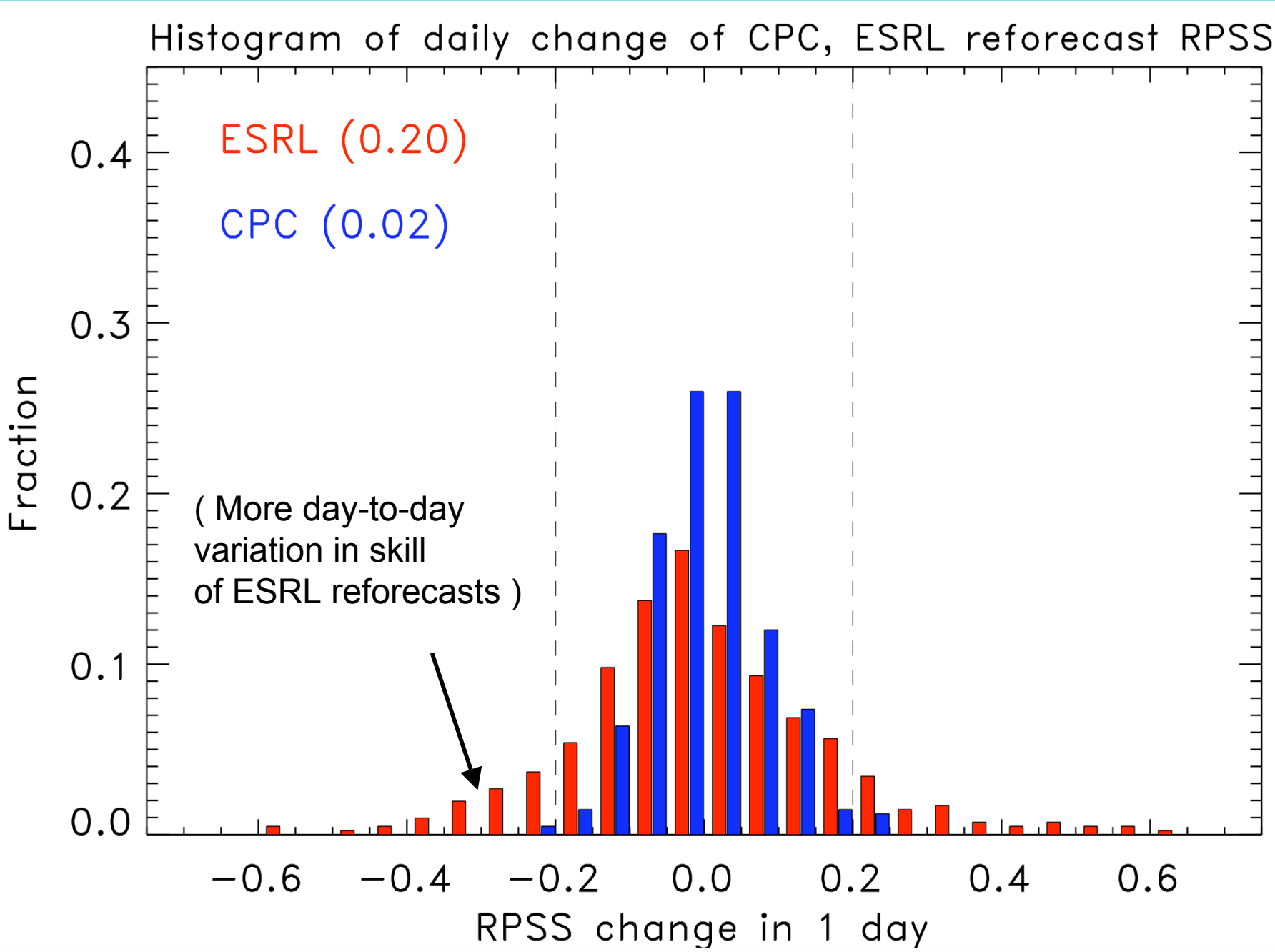

6–10 Day ESRL vs. CPC daily RPSS & running 31–day average
ESRL avg. RPSS = 0.197
CPC avg. RPSS = 0.144
Ranked Probability Skill Score (RPSS)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
Jan 04
Feb 04
Mar 04
Apr 04
May 04
Jun 04
Jul 04
Aug 04
Sep 04
Oct 04
Nov 04
Dec 04
Jan 05
Feb 05
Mar 05
Apr 05
May 05
Jun 05
Jul 05
Aug 05
Sep 05
Oct 05
Nov 05
Dec 05

Ed O'Lenic's comment about comparing ESRL, CPC RPSS

"The graphs also indicate that CPC true 6-10 day probability forecasts are more conservative than ESRL true 6-10 day forecasts, with ESRL forecasts showing extreme swings in skill. Understandably, this latter property continues to make it difficult for forecasters to use the ESRL product."


Histogram of CPC, ESRL reforecast RPSS
ESRL (0.18)
CPC (0.11)
More ESRL reforecasts
with RPSS < 0.0
Fraction
0.20
0.15
0.10
0.05
0.00
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
RPSS


Histogram of CPC, ESRL reforecast RPSS
ESRL (0.32)
CPC (0.10)
Still, many more
ESRL reforecasts
with RPSS > 0.3
Fraction
RPSS
0.20
0.15
0.10
0.05
0.00
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0


Histogram of daily change of CPC, ESRL reforecast RPSS
ESRL (0.20)
CPC (0.02)
(More day-to-day
variation in skill
of ESRL reforecasts)
Fraction
0.0
0.1
0.2
0.3
0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
RPSS change in 1 day

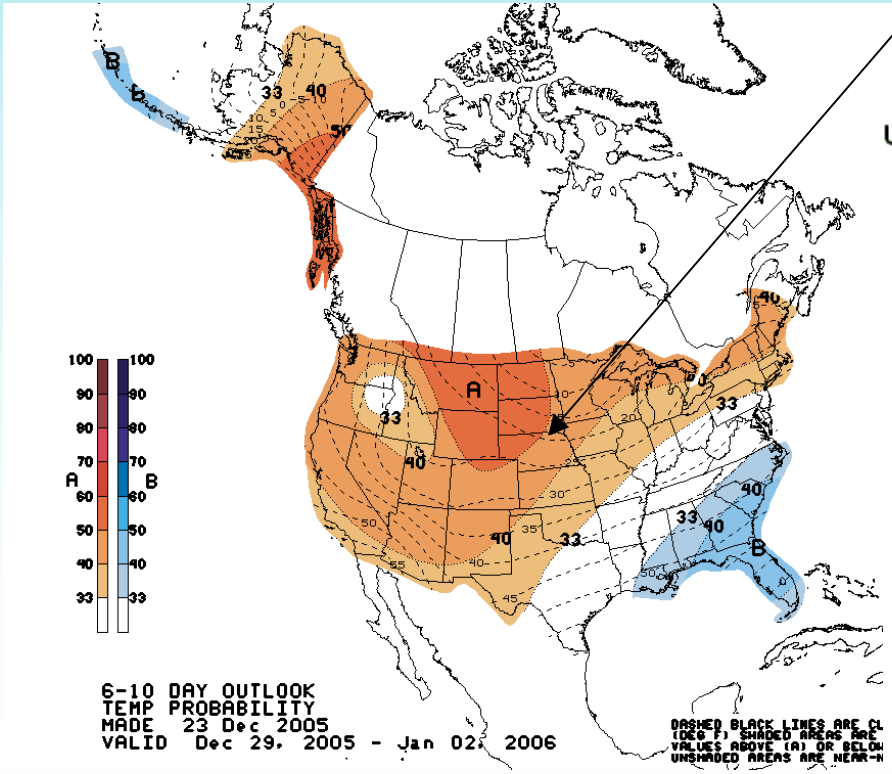
Are daily swings (in forecast probabilities, and thus in skill) good or bad?

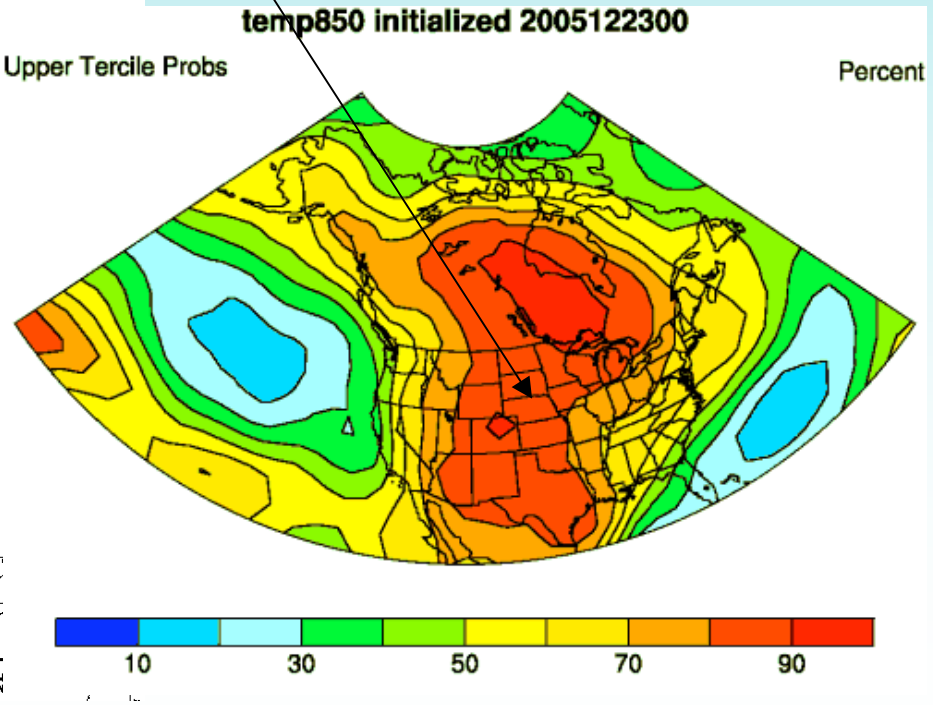
- Users prefer consistency of forecasts from one day to the next…may wonder why they jumped [BAD].
- If you strongly suspect yesterday's forecast was lousy, why should you aim for consistency with it? [GOOD].

How might we combine forecasts to be better on average than either individually?

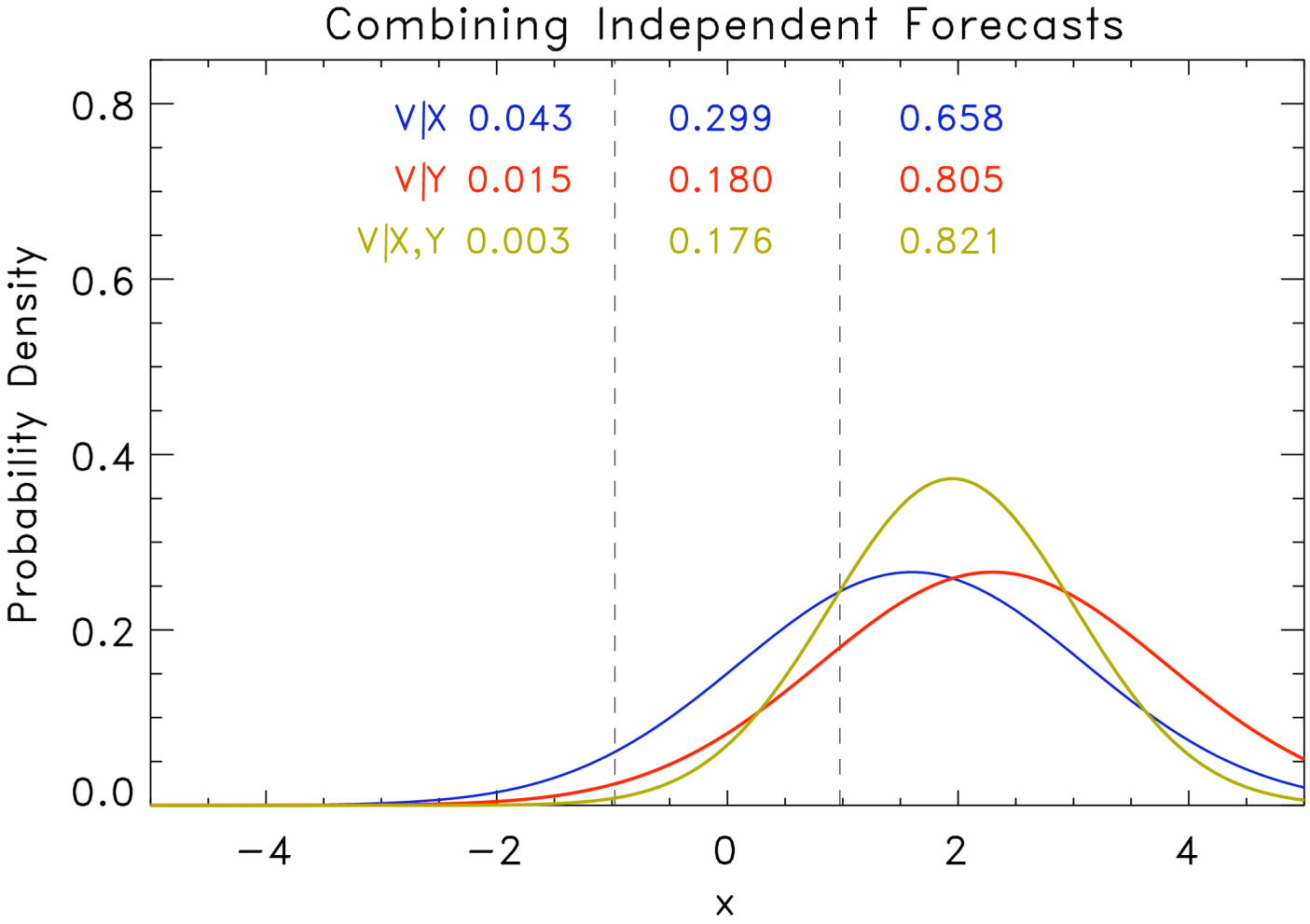
Here, 50% CPC, 80+% ESRL

→ 65% Combined?

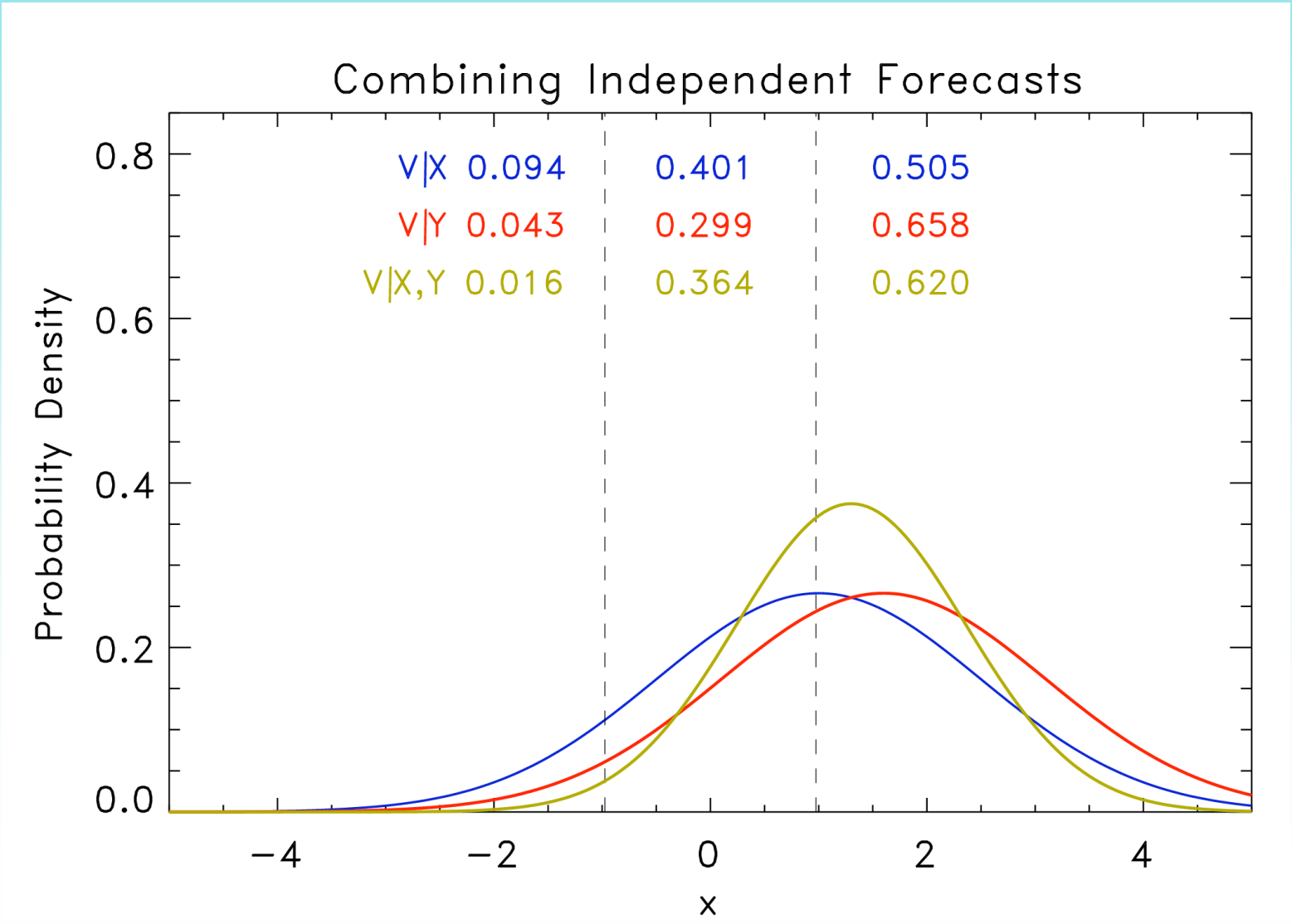




IF P(V|X), P(V|Y) independent,
THEN P(V|X,Y) ∝ P(V|X) P(V|Y)



However…


Combining Independent Forecasts
V|X 0.094 0.401 0.505
V|Y 0.043 0.299 0.658
V|X,Y 0.016 0.364 0.620
Probability Density
0.0
0.2
0.4
0.6
0.8
−4
−2
0
2
4
x

Using Bayes' Rule to sharpen forecasts and improve skill

- Let $T_{2/3}$ = Temperature threshold between middle, upper tercile
  V = Observed 6-10 day temperature
  F_{CPC} = CPC's P($V > T_{2/3}$)
  F_{CDC} = Reforecast-based P($V > T_{2/3}$)

$$P\left(V > T_{2/3} \mid F_{CPC}, F_{CDC}\right) = \frac{P\left(F_{CPC}, F_{CDC} \mid V > T_{2/3}\right) P(V > T_{2/3})}{P\left(F_{CPC}, F_{CDC} \mid V > T_{2/3}\right) P(V > T_{2/3}) \;+\; P\left(F_{CPC}, F_{CDC} \mid V \le T_{2/3}\right) P(V \le T_{2/3})}$$

Combining CDC and ESRL forecasts

Lesson 1: If CPC, ESRL agree on the sign of the anomaly, issue a forecast sharper than either.

Adjusted probabilities using both ESRL, CPC

CPC \ ESRL	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-00
90-**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
80-90	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	1.000
70-80	0.000	1.000	0.857	0.000	0.500	0.889	1.000	0.817	0.898	0.827
60-70	0.000	0.385	0.769	0.806	0.728	0.861	0.917	0.922	0.909	0.961
50-60	0.444	0.412	0.631	0.626	0.677	0.696	0.764	0.825	0.893	0.961
40-50	0.259	0.436	0.465	0.537	0.600	0.634	0.715	0.736	0.831	0.917
30-40	0.127	0.220	0.335	0.423	0.471	0.519	0.599	0.662	0.721	0.938
20-30	0.087	0.145	0.237	0.317	0.383	0.440	0.513	0.586	0.799	0.814
10-20	0.037	0.060	0.099	0.110	0.293	0.256	0.245	0.467	0.250	1.000
0-10	0.017	0.012	0.019	0.043	0.171	0.000	0.000	0.000	0.000	0.000

0.000 or 1.000 = Untrustworthy because of small sample size

Combining CDC and ESRL forecasts

Lesson 2: Often (not always) when best category differs, issue intermediate probability.

Adjusted probabilities using both ESRL, CPC

CPC \ ESRL	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-00
90-**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
80-90	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	1.000
70-80	0.000	1.000	0.857	0.000	0.500	0.889	1.000	0.817	0.898	0.827
60-70	0.000	0.385	0.769	0.806	0.728	0.861	0.917	0.922	0.909	0.961
50-60	0.444	0.412	0.631	0.626	0.677	0.696	0.764	0.825	0.893	0.961
40-50	0.259	0.436	0.465	0.537	0.600	0.634	0.715	0.736	0.831	0.917
30-40	0.127	0.220	0.335	0.423	0.471	0.519	0.599	0.662	0.721	0.938
20-30	0.087	0.145	0.237	0.317	0.383	0.440	0.513	0.586	0.799	0.814
10-20	0.037	0.060	0.099	0.110	0.293	0.256	0.245	0.467	0.250	1.000
0-10	0.017	0.012	0.019	0.043	0.171	0.000	0.000	0.000	0.000	0.000

0.000 or 1.000 = Untrustworthy because of small sample size

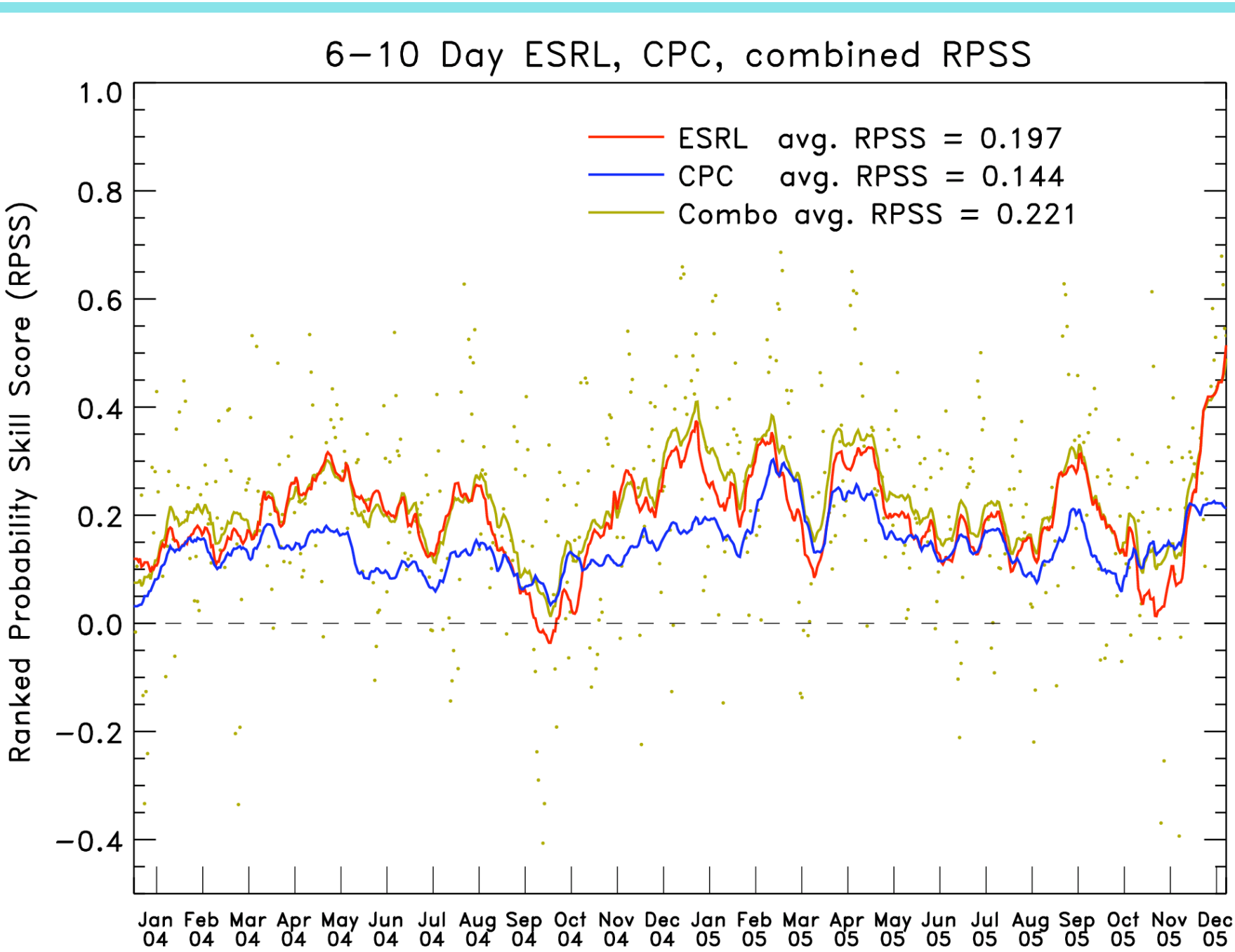
Combining CDC and ESRL forecasts

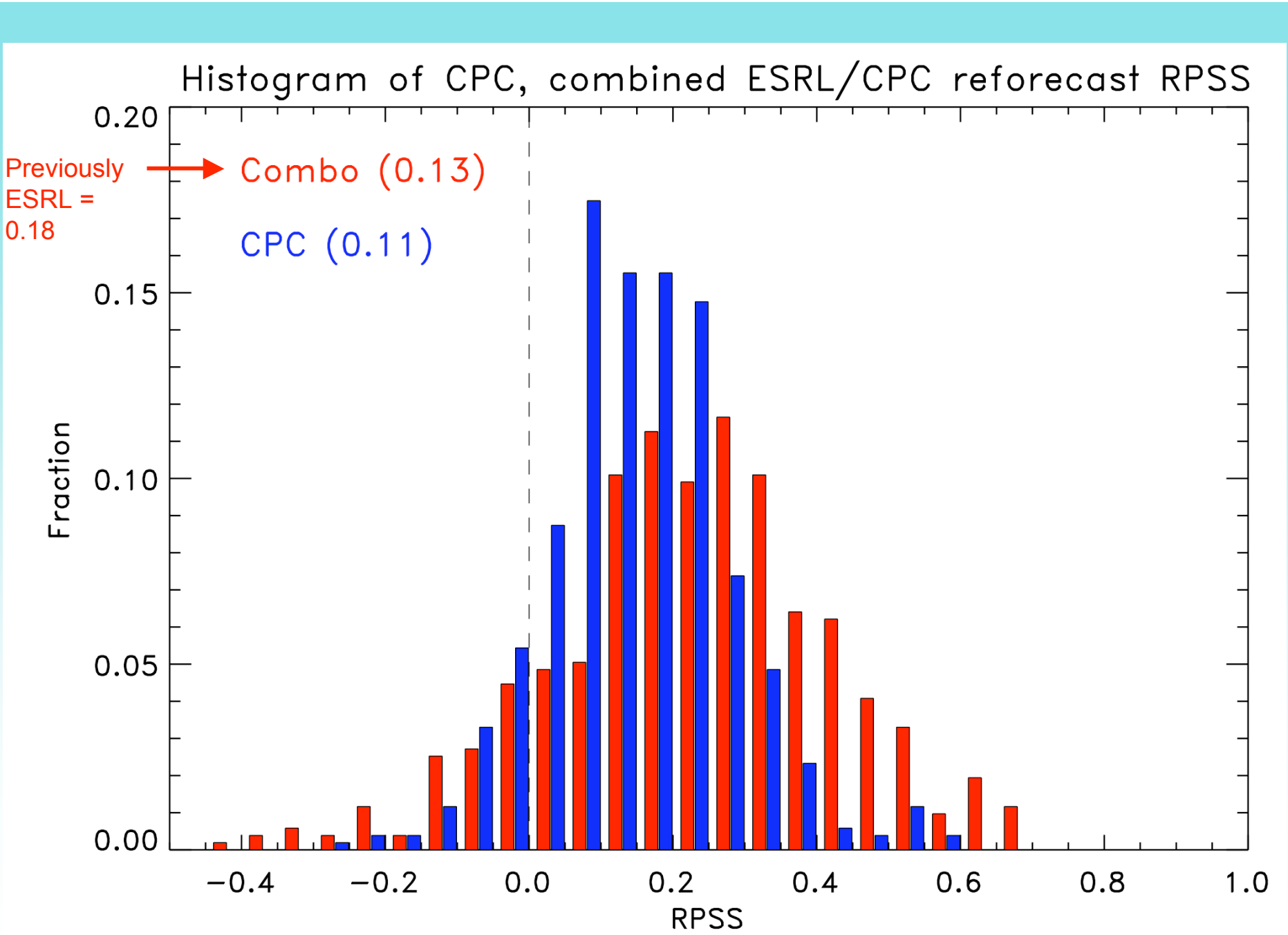
Lesson 3: Apart from ESRL, it looks like in some instances CPC forecasts could be sharpened.

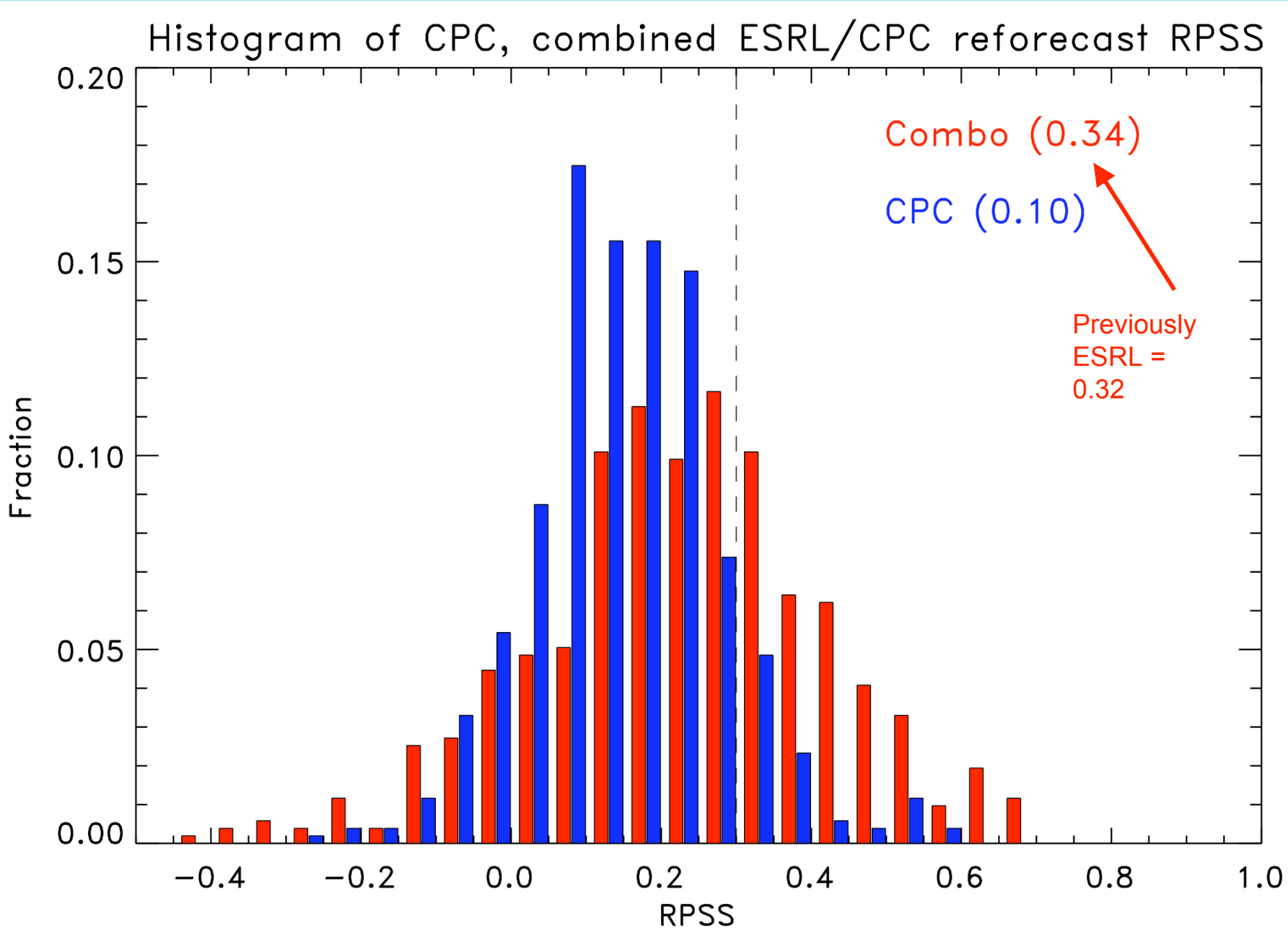
Adjusted probabilities using both ESRL, CPC

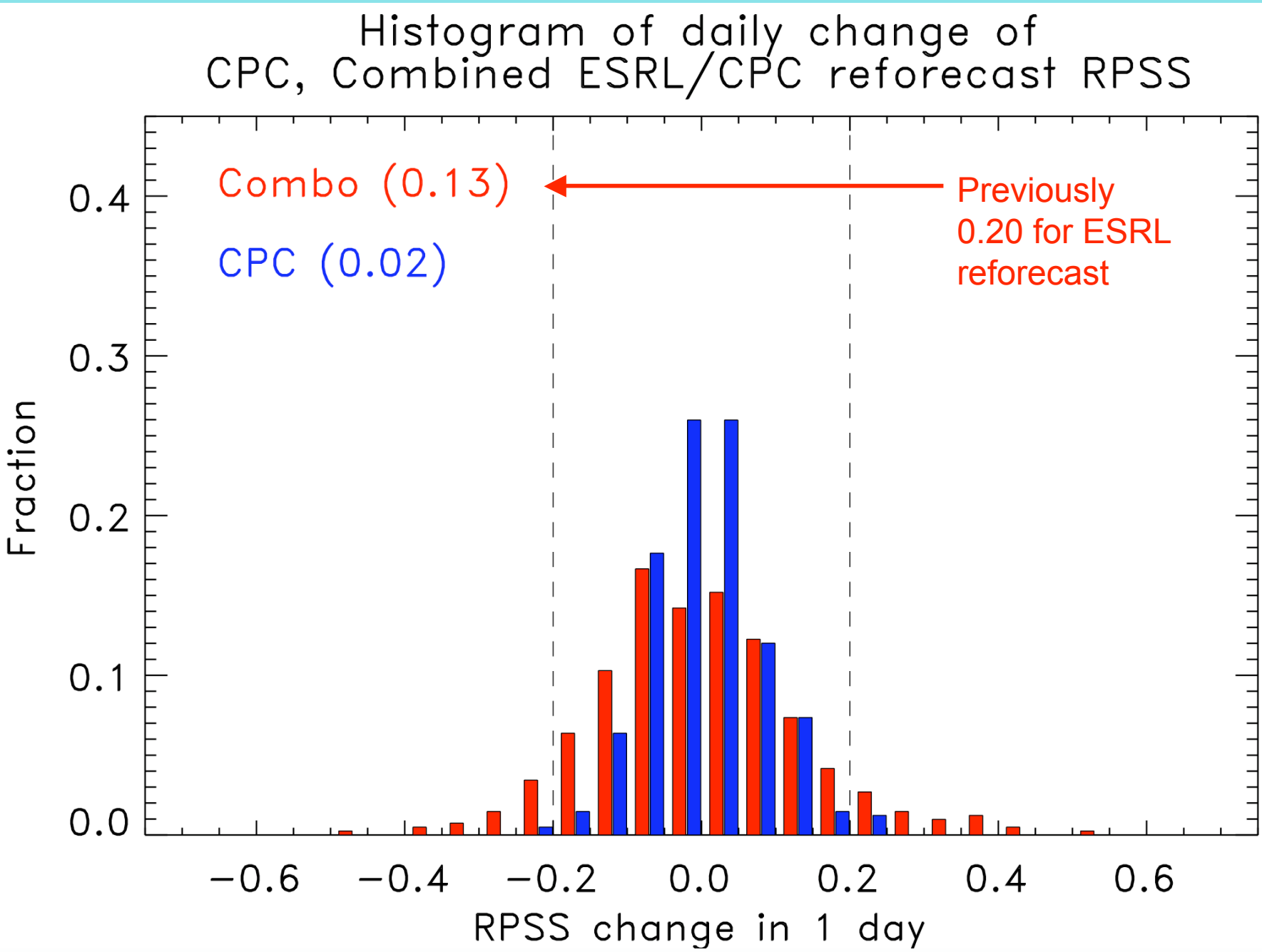
CPC \ ESRL	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-00
90-**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
80-90	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	1.000
70-80	0.000	1.000	0.857	0.000	0.500	0.889	1.000	0.817	0.898	0.827
60-70	0.000	0.385	0.769	0.806	0.728	0.861	0.917	0.922	0.909	0.961
50-60	0.444	0.412	0.631	0.626	0.677	0.696	0.764	0.825	0.893	0.961
40-50	0.259	0.436	0.465	0.537	0.600	0.634	0.715	0.736	0.831	0.917
30-40	0.127	0.220	0.335	0.423	0.471	0.519	0.599	0.662	0.721	0.938
20-30	0.087	0.145	0.237	0.317	0.383	0.440	0.513	0.586	0.799	0.814
10-20	0.037	0.060	0.099	0.110	0.293	0.256	0.245	0.467	0.250	1.000
0-10	0.017	0.012	0.019	0.043	0.171	0.000	0.000	0.000	0.000	0.000

0.000 or 1.000 = Untrustworthy because of small sample size


6–10 Day ESRL, CPC, combined RPSS
ESRL avg. RPSS = 0.197
CPC avg. RPSS = 0.144
Combo avg. RPSS = 0.221
Ranked Probability Skill Score (RPSS)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
Jan 04
Feb 04
Mar 04
Apr 04
May 04
Jun 04
Jul 04
Aug 04
Sep 04
Oct 04
Nov 04
Dec 04
Jan 05
Feb 05
Mar 05
Apr 05
May 05
Jun 05
Jul 05
Aug 05
Sep 05
Oct 05
Nov 05
Dec 05


Histogram of CPC, combined ESRL/CPC reforecast RPSS
Previously ESRL = 0.18
Combo (0.13)
CPC (0.11)
Fraction
0.20
0.15
0.10
0.05
0.00
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
RPSS


Histogram of CPC, combined ESRL/CPC reforecast RPSS
Combo (0.34)
CPC (0.10)
Previously
ESRL =
0.32
Fraction
0.20
0.15
0.10
0.05
0.00
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
RPSS


Histogram of daily change of
CPC, Combined ESRL/CPC reforecast RPSS
Combo (0.13)
CPC (0.02)
Previously
0.20 for ESRL
reforecast
Fraction
0.0
0.1
0.2
0.3
0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
RPSS change in 1 day

Conclusions

- Ambiguity in definition of 6-10 day (and week 2) and in observations. Let's agree upon a standard, use common data set.
- ESRL reforecasts (given caveats) appear more skillful on average, but more variable.
- There are ways to combine CPC and ESRL forecasts to, on average, do better than either individually.

What IS 6-10 day forecast?

Typically use Tmax, Tmin for 24-h period starting
At 1200 UTC. Ambiguity…do we thus use OR THIS?