

**A COMPARISON OF FORECASTING SKILL BETWEEN
WEATHER TRENDS INTERNATIONAL (WTI) AND
THE CLIMATE PREDICTION CENTER (CPC) AT
THE NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA)**

LAURENCE S. KALKSTEIN, Ph.D.

CLIMATOLOGIST AND PROFESSOR

DEPARTMENT OF PUBLIC HEALTH SCIENCES

MILLER SCHOOL OF MEDICINE

UNIVERSITY OF MIAMI

ADAM J. KALKSTEIN, Ph.D.

CLIMATOLOGIST AND ASSISTANT PROFESSOR

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL ENGINEERING

U.S. MILITARY ACADEMY

WEST POINT

MAY, 2014

The purpose of this analysis is to compare the accuracy of three-month forecasts that are issued by the Climate Prediction Center (CPC) against the accuracy of similar forecasts issued by Weather Trends International (WTI). The comparison extends for 77 three month forecast periods, from October/November/December, 2007 through February/March/April, 2014.

These forecasts were developed for 260 cities around the country, and are issued by WTI at the end of the calendar month for the time period that is valid 11 months into the future. For example, the May, 2010 forecast was issued by WTI in late June, 2009. They are issued one month at a time, and three months were combined for the purposes of this analysis and to match up to CPC's forecast period. Once the forecasts are issued, they are not updated by WTI.

The CPC's forecasts are constructed differently; they are developed for three combined month intervals, but in this analysis, the CPC forecast developed one month prior to the three month period was used in the comparison. Thus, for example, the May, 2010 CPC forecast was utilized in this comparison to evaluate their June, July, and August, 2010 forecast. Unlike the WTI forecasts, this permits the CPC to alter their forecasts through the time period, right up to the month before. All things being equal, this would offer a large advantage to the accuracy of the CPC forecast as compared to WTI's.

"Skill scores" were calculated for both sets of forecasts for each three month period. The CPC has historically utilized the Heidke Skill Score (http://www.cpc.ncep.noaa.gov/products/Soilmst_Monitoring/US/Outlook/cas_score.shtml), which is defined as follows:

$$\text{HSS (\%)} = 100 * (H - E) / (T - E)$$

where H = number of correct forecasts, E = expected number of correct forecasts (1/3 of total), and T = total number of valid forecast-observation pairs. An HSS above 0 translates to more "hits" than "misses" among the 260 forecasts for each three month period; conversely, an HSS below 0 indicates more misses than hits. An HSS of 0 would be, in the long term, the random result if there was no forecasting skill. There are three categories in forecasting: above average, below average, and average. The CPC virtually never issues an "average" forecast; instead, for large segments of the country, their forecast for a three month period is "equal chance", which translates to no forecast at all, as it assumes an equal probability of above or below average temperatures. WTI does not issue an "equal chance" forecast. So this forecasting evaluation is based largely on a two category system, above and below average, to keep the forecasts parallel.

Table 1 shows the comparison of skill scores for part of the data extending from 2012 to the beginning of 2014 (the full 77 month forecast dataset is found in **Appendix 1**). CPC scores are divided into two types; one includes the "equal chance" probability that is utilized by the CPC and the other uses only non-equal chance forecasts. In either case, the WTI skill scores exceed the CPC scores in 65 of the 77 months, or 84.4 percent of the total forecasts issued during this time period. Thus, on average during this period, WTI bested the CPC on 6 out of every 7 forecasts, a highly statistically significant difference in forecast accuracy.

Year Start	Season	Hits	Hit Rate	WTI Skill Score	CPC Skill Score (All Forecasts)	CPC Skill Score (non EC Forecasts)
2012	JFM	185	71.15%	42.31%	34.48%	54.79%
	FMA	149	57.31%	14.62%	41.38%	76.19%
	MAM	162	62.31%	24.62%	61.85%	86.45%
	AMJ	206	79.23%	58.46%	41.81%	76.38%
	MJJ	203	78.08%	56.15%	22.84%	42.40%
	JJA	227	87.31%	74.62%	25.86%	45.45%
	JAS	232	89.23%	78.46%	48.92%	58.81%
	ASO	183	70.38%	40.77%	-8.19%	-11.59%
	SON	172	66.15%	32.31%	2.80%	3.89%
	OND	218	83.85%	67.69%	30.82%	48.31%
	NDJ	219	84.23%	68.46%	17.89%	45.60%
	DJF	201	77.31%	54.62%	0.65%	1.43%
2013	JFM	147	56.54%	13.08%	-9.48%	-20.56%
	FMA	143	55.00%	10.00%	-4.74%	-9.57%
	MAM	118	45.38%	-9.23%	-20.47%	-24.48%
	AMJ	177	68.08%	36.15%	-10.78%	-13.51%
	MJJ	171	65.77%	31.54%	0.43%	0.60%
	JJA	190	73.08%	46.15%	3.45%	4.71%
	JAS	177	68.08%	36.15%	24.78%	43.89%
	ASO	193	74.23%	48.46%	21.98%	36.96%
	SON	190	73.08%	46.15%	-3.02%	-8.75%
	OND	124	47.69%	-4.62%	-9.48%	-23.16%
	NDJ	89	34.23%	-31.54%	-12.72%	-22.01%
	DJF	189	72.69%	45.38%	1.72%	5.06%
2014	JFM	179	68.85%	37.69%	4.96%	11.86%
	FMA	172	66.15%	32.31%	10.13%	18.08%
Average				36.57%	12.23%	20.28%

Table 1. Skill scores for a portion of the evaluation period. The number of “hits” is a proportion of correct forecasts out of 260 cities. Skill scores are based upon the Heidke algorithm found above. Green colored skill scores designate the best score for either WTI or CPC.

In a vast majority of the three month forecast periods, WTI scored positive Heidke skill scores, indicating many more forecast “hits” than “misses”. During the 77 month forecast period, WTI registered only 8 missed forecasts (Figure 1; red and yellow quadrats), which represents an almost 90 percent accuracy rate, all the more impressive considering that these forecasts were developed 11 months prior to the actual forecast date. CPC posted 21 missed forecasts (green and yellow quadrats), which is approximately a 73 percent accuracy rate. This represents approximately 2.6 times more missed forecasts than WTI, a significantly higher rate. The diagonal line in Figure 1 represents equal forecast accuracy by the two groups; dots to the left of that line represent a better skill score (forecast accuracy)

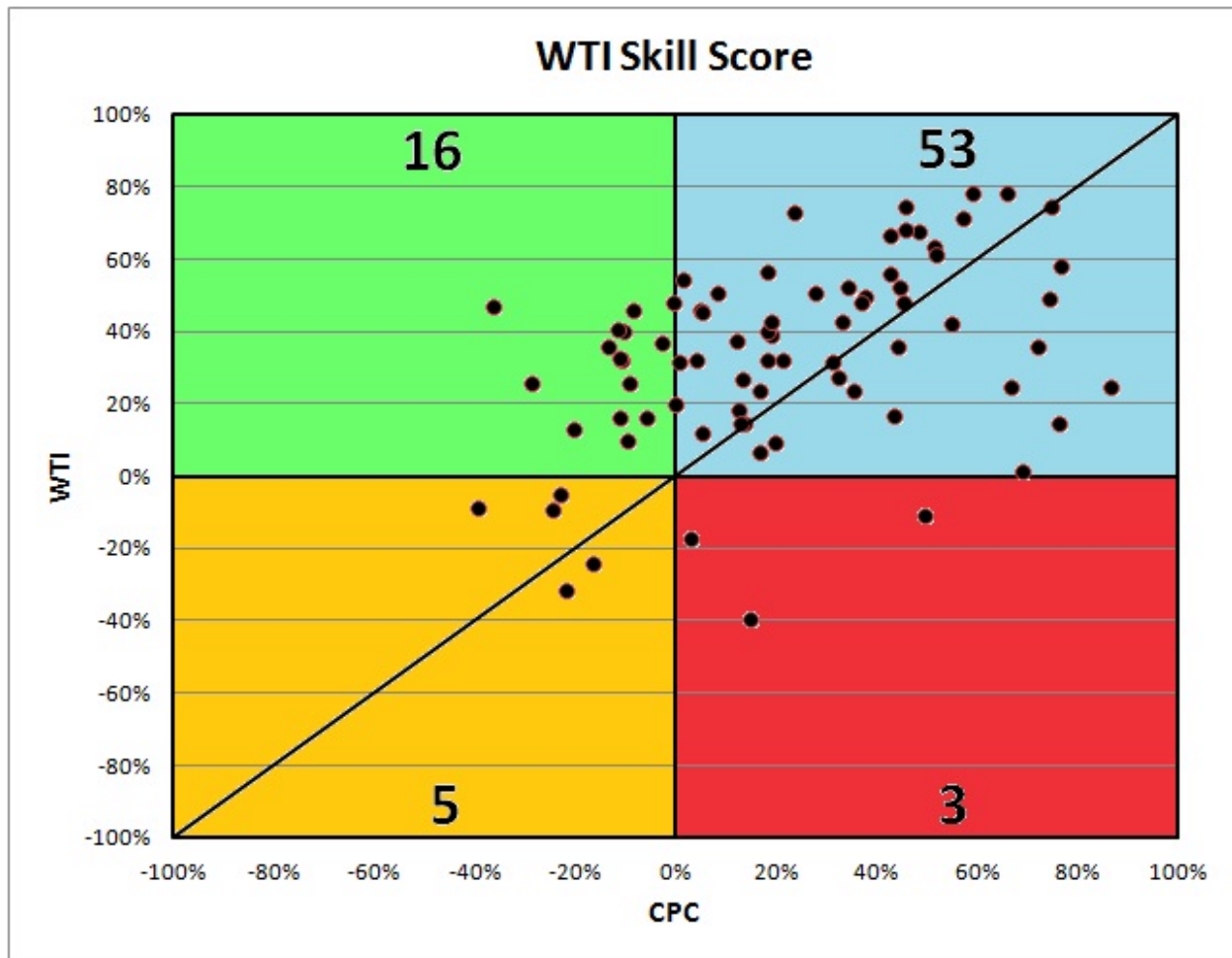


Figure 1. A comparison of forecast “hits” vs. “misses” based upon Heidke skill scores. The blue quadrat includes forecast hits by both WTI and CPC. The green quadrat represents hits by WTI and misses by CPC. The red quadrat represents hits by CPC and misses by WTI. The yellow quadrat represents misses by both groups.

by WTI; dots to the right represent a better accuracy by CPC. It is clear that the dots are strongly skewed to the left, indicating the superior skill scores by WTI. In fact, CPC had a superior skill score to WTI in just 19 of 77 forecast periods, a 24.6 percent rate. Thus, WTI’s forecasts were more accurate than CPC’s in 3 of every 4 forecasts during the evaluation period.

A temporal analysis of the skill scores is instructive to see when forecast differentials were greatest (Figure 2). It is clear that WTI had greater forecast accuracy during virtually all the time span, with the exception of some short periods in the middle of the forecast period. At no point did CPC exceed WTI’s skill score for more than three consecutive months, yet WTI’s skill scores were higher for a 20 consecutive month period at the start of the time span, and a 19 consecutive month time span during 2012 and 2013.

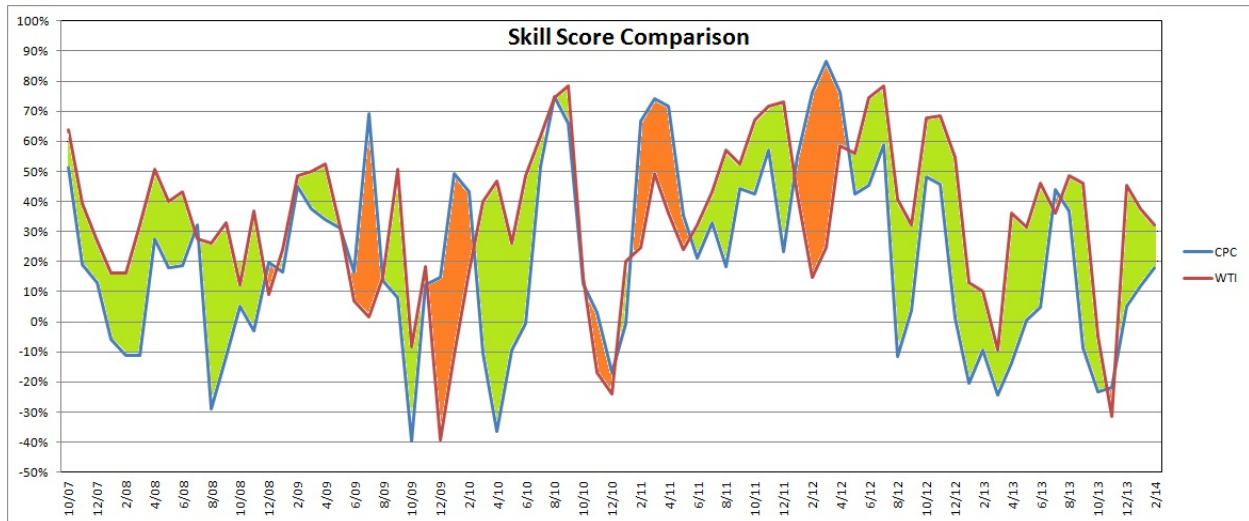


Figure 2. Skill score comparison through the evaluation period. Green areas represent periods when WTI’s accuracy exceeded CPC’s; orange areas show periods when CPC’s skill scores were superior.

We also evaluated the magnitude of the skill scores using a series of intervals (Figure 3). Once again, WTI shows a clearly superior set of scores, translating into more accurate forecasts. CPC’s more negative scores, and slightly positive scores, stands in contrast to the larger number of highly positive scores that are indicative of the WTI forecasts. The switch occurs at the .20 (20%) Heidke skill score; all intervals higher than that show WTI forecasts as more dominant; lower intervals are more populated by CPC forecasts, and by a sizable margin.

The evaluation strongly suggests that WTI has offered much more accurate forecasts than CPC over the last 7 years. This is in spite of the fact that the WTI forecasts are issued 11 months prior to the actual forecast period, while CPC’s forecasts are issued a month before. Whether the “equal chance” forecast is included for CPC or if it is excluded, the WTI forecasts are largely superior. The bottom line indicates that WTI issues an accurate forecast almost 90 percent of the time, CPC’s forecasts are associated with negative skill score values at a rate 2.6 times higher than WTI’s, and WTI forecasts are associated with higher skill scores than CPC’s in 3 of every 4 months, on average, during the evaluation period.

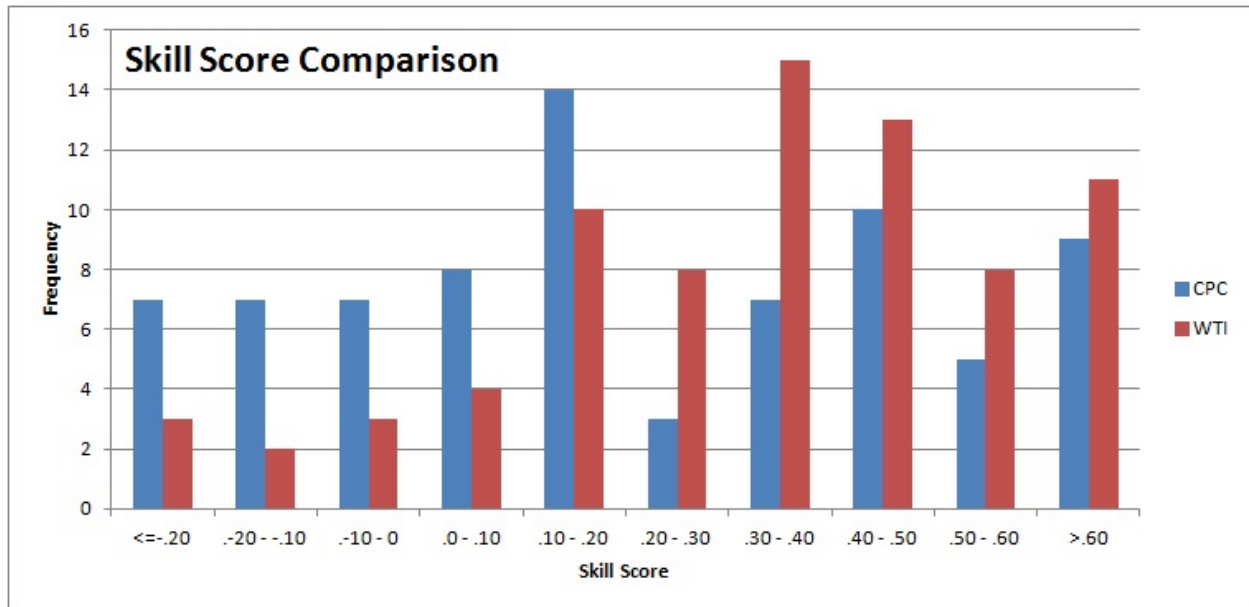


Figure 3. A comparison of skill score intervals, shown in tenths. Negative skill scores are to the left; the most accurate skill scores are to the right.

APPENDIX 1

SKILL SCORE VALUES FOR THE ENTIRE EVALUATION PERIOD

Year Start	Season	WTI City Hits	WTI Hit %	WTI Skill Score	CPC NOAA Skill Score (All Forecasts)	
2007	OND	213	81.92%	63.85%	40.30%	1
	NDJ	181	69.62%	39.23%	15.09%	2
	DJF	165	63.46%	26.92%	10.13%	3
2008	JFM	151	58.08%	16.15%	-4.74%	4
	FMA	151	58.08%	16.15%	-6.68%	5
	MAM	172	66.15%	32.31%	-6.03%	6
	AMJ	196	75.38%	50.77%	10.78%	7
	MJJ	182	70.00%	40.00%	5.82%	8
	JJA	186	71.54%	43.08%	9.91%	9
	JAS	166	63.85%	27.69%	11.42%	10
	ASO	164	63.08%	26.15%	-12.50%	11
	SON	173	66.54%	33.08%	-6.90%	12
	OND	146	56.15%	12.31%	2.80%	13
	NDJ	178	68.46%	36.92%	-1.72%	14
	DJF	142	54.62%	9.23%	8.41%	15
2009	JFM	161	61.92%	23.85%	10.13%	16
	FMA	193	74.23%	48.46%	26.08%	17
	MAM	195	75.00%	50.00%	12.50%	18
	AMJ	198	76.15%	52.31%	9.70%	19
	MJJ	171	65.77%	31.54%	13.15%	20
	JJA	139	53.46%	6.92%	9.91%	21
	JAS	132	50.77%	1.54%	27.37%	22
	ASO	149	57.31%	14.62%	4.31%	23
	SON	196	75.38%	50.77%	3.45%	24
	OND	119	45.77%	-8.46%	-21.98%	25
	NDJ	154	59.23%	18.46%	7.97%	26
	DJF	79	30.38%	-39.23%	10.78%	27
2010	JFM	116	44.62%	-10.77%	32.76%	28
	FMA	152	58.46%	16.92%	28.23%	29
	MAM	182	70.00%	40.00%	-5.39%	30
	AMJ	191	73.46%	46.92%	-17.24%	31
	MJJ	164	63.08%	26.15%	-5.17%	32
	JJA	193	74.23%	48.46%	-0.22%	33

	JAS	210	80.77%	61.54%	26.94%	34
	ASO	227	87.31%	74.62%	51.51%	35
	SON	232	89.23%	78.46%	40.09%	36
	OND	149	57.31%	14.62%	7.54%	37
	NDJ	108	41.54%	-16.92%	1.51%	38
	DJF	99	38.08%	-23.85%	-9.48%	39
2011	JFM	156	60.00%	20.00%	-0.22%	40
	FMA	162	62.31%	24.62%	38.79%	41
	MAM	194	74.62%	49.23%	46.34%	42
	AMJ	177	68.08%	36.15%	41.16%	43
	MJJ	161	61.92%	23.85%	20.91%	44
	JJA	172	66.15%	32.31%	11.21%	45
	JAS	186	71.54%	43.08%	13.58%	46
	ASO	204	78.46%	56.92%	11.21%	47
	SON	198	76.15%	52.31%	28.23%	48
	OND	217	83.46%	66.92%	20.47%	49
	NDJ	223	85.77%	71.54%	17.24%	50
	DJF	225	86.54%	73.08%	14.66%	51
2012	JFM	185	71.15%	42.31%	34.48%	52
	FMA	149	57.31%	14.62%	41.38%	53
	MAM	162	62.31%	24.62%	61.85%	54
	AMJ	206	79.23%	58.46%	41.81%	55
	MJJ	203	78.08%	56.15%	22.84%	56
	JJA	227	87.31%	74.62%	25.86%	57
	JAS	232	89.23%	78.46%	48.92%	58
	ASO	183	70.38%	40.77%	-8.19%	59
	SON	172	66.15%	32.31%	2.80%	60
	OND	218	83.85%	67.69%	30.82%	61
	NDJ	219	84.23%	68.46%	17.89%	62
	DJF	201	77.31%	54.62%	0.65%	63
2013	JFM	147	56.54%	13.08%	-9.48%	64
	FMA	143	55.00%	10.00%	-4.74%	65
	MAM	118	45.38%	-9.23%	-20.47%	66
	AMJ	177	68.08%	36.15%	-10.78%	67
	MJJ	171	65.77%	31.54%	0.43%	68
	JJA	190	73.08%	46.15%	3.45%	69
	JAS	177	68.08%	36.15%	24.78%	70
	ASO	193	74.23%	48.46%	21.98%	71
	SON	190	73.08%	46.15%	-3.02%	72
	OND	124	47.69%	-4.62%	-9.48%	73
	NDJ	89	34.23%	-31.54%	-12.72%	74

	DJF	189	72.69%	45.38%	1.72%	75
2014	JFM	179	68.85%	37.69%	4.96%	76
	FMA	172	66.15%	32.31%	10.13%	77
Average			66.76%	33.53%	11.97%	