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.

PROFESSOR EMERITUS

PRESIDENT

APPLIED CLIMATOLOGISTS, INC.

ADAM J. KALKSTEIN, Ph.D.

CLIMATOLOGIST AND ASSOCIATE PROFESSOR

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL ENGINEERING

U.S. MILITARY ACADEMY

WEST POINT, NY

DECEMBER, 2018

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 125 three month forecast periods, from January/February/March, 2008 through May/June/July, 2018. The data used for these comparisons were provided by WTI.

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 three month 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. 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 by WTI 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:

HSS (%) = 100 * (H - E) / (T - E)

where H = number of correct forecasts, E = expected number of correct forecasts (1/3 of total, described in the next paragraph), 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 utilized by CPC in forecasting: above average, below average, and average (the CPC defines these as "above normal", "normal", and "below normal" (CPC, 2016; http://www.cpc.ncep.noaa.gov/products/predictions/long_range/tools.html). CPC virtually never issues a "normal" forecast; they have historically only issued above and below normal forecasts. CPC also uses an "equal chance" (EC) term in regions where forecasters cannot determine whether conditions will favor any of the three categories; thus, the chance of any of these three categories is defined to be 33.3% each, labeling these areas EC. With this in mind, we have instead utilized a 2-class system of above and below normal, eliminating the EC and normal categories; this two-class system matches the above and below normal forecasts issued by WTI. Consequently, the area in white within Figure 1 is not included in the comparison, since there is a 33.3%-33.3%-33.3% chance of each of the three categories in that area. In this manner, both the WTI and CPC forecast evaluations are based on a two-category system, above and below normal, to keep the forecasts parallel.

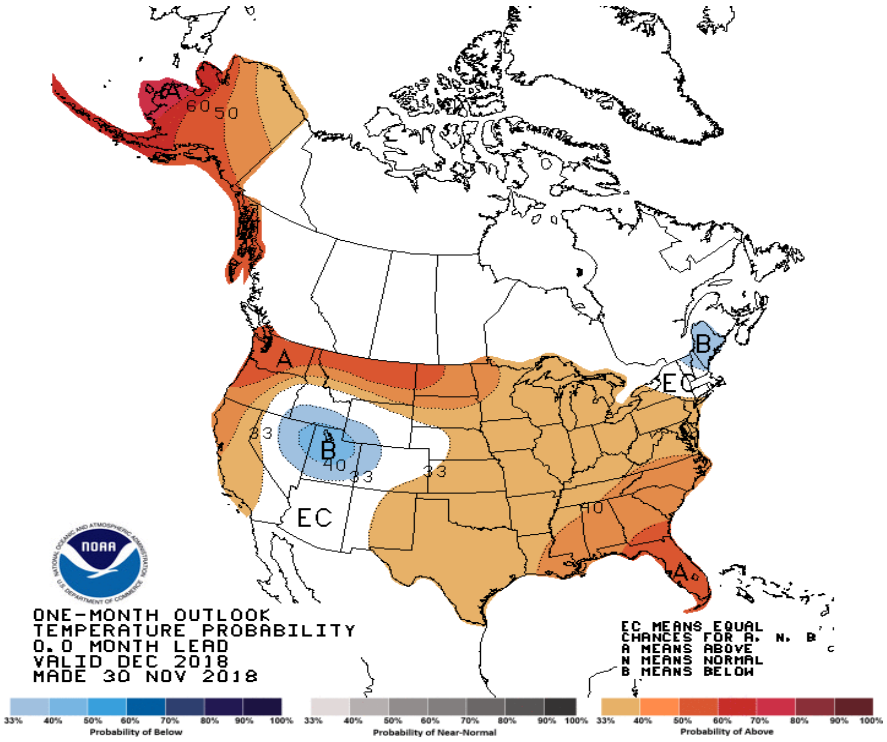

Figure 1. CPC one month forecast for December, 2018, constructed 30 November 2018. Note the EC areas, where there is an equal chance for above normal, normal, or below normal temperature conditions (white area between the 33 percent isolines).

Table 1 shows the comparison of skill scores for part of the data extending from the beginning of 2008 to the end of 2009 (the full 125 month forecast dataset is found in **Appendix 1**). Hit rates are color-coded; the darker the green color, the better the accuracy. For the total period, WTI skill scores exceed the CPC scores in 104 of the 125 months, or 83 percent of the total forecasts issued during this time period. Thus, on average during this period, WTI bested the CPC on about 5 out of every 6 forecasts, a highly statistically significant difference in forecast accuracy.

Start Year	Months	Hits (out of 260)	Hit Rate (%)	WTI Skill Score	CPC Skill Score
2008	JFM	147	56.53%	34.81%	-4.74%
2008	FMA	149	57.30%	35.96%	-6.68%
2008	MAM	173	66.53%	49.81%	-6.03%
2008	AMJ	194	74.61%	61.92%	10.78%
2008	MJJ	183	70.38%	55.58%	5.82%
2008	JJA	190	73.07%	59.62%	9.91%
2008	JAS	173	66.53%	49.81%	11.42%
2008	ASO	166	63.84%	45.77%	-12.50%
2008	SON	168	64.61%	46.92%	-6.90%
2008	OND	144	55.38%	33.08%	2.80%
2008	NDJ	181	69.61%	54.42%	-1.72%
2008	DJF	134	51.53%	27.31%	8.41%
2009	JFM	160	61.53%	42.31%	10.13%
2009	FMA	200	76.92%	65.38%	26.08%
2009	MAM	198	76.15%	64.23%	12.50%
2009	AMJ	199	76.53%	64.81%	9.70%
2009	MJJ	170	65.38%	48.08%	13.15%
2009	JJA	137	52.69%	29.04%	9.91%
2009	JAS	129	49.61%	24.42%	27.37%
2009	ASO	146	56.15%	34.23%	4.31%
2009	SON	195	75.00%	62.50%	3.45%
2009	OND	117	45.00%	17.50%	-21.98%
2009	NDJ	146	56.15%	34.23%	7.97%
2009	DJF	75	28.84%	-6.73%	10.78%

Table 1. Skill scores for a portion of the evaluation period. The number of "hits" is a proportion of WTI correct forecasts out of 260 cities. Hit rates below 45% are colored brown, between 45% and 55%, they are not colored, and above 55% they are colored green, with the intensities increasing as the rates go down/up. Skill scores are based upon the Heidke algorithm found above. For the skill score columns, light green colored skill scores designate the winning 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 125 month forecast period, WTI registered only 3 missed forecasts as defined by Heidke skill scores, (Figure 2; red and yellow quadrats), which represents an over 95% percent accuracy rate based upon skill scores, all the more impressive considering that these forecasts were developed 11 months prior to the actual forecast date. CPC posted 24 missed forecasts (green and yellow quadrats), which is an approximately 81 percent accuracy rate. It also represents approximately 8 times more missed forecasts than WTI, a significantly higher rate. The diagonal line in Figure 2 represents equal forecast accuracy by the two groups; dots to the left of that line represent a better skill score (forecast accuracy) 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 21 of 125 forecast periods, a 16.8

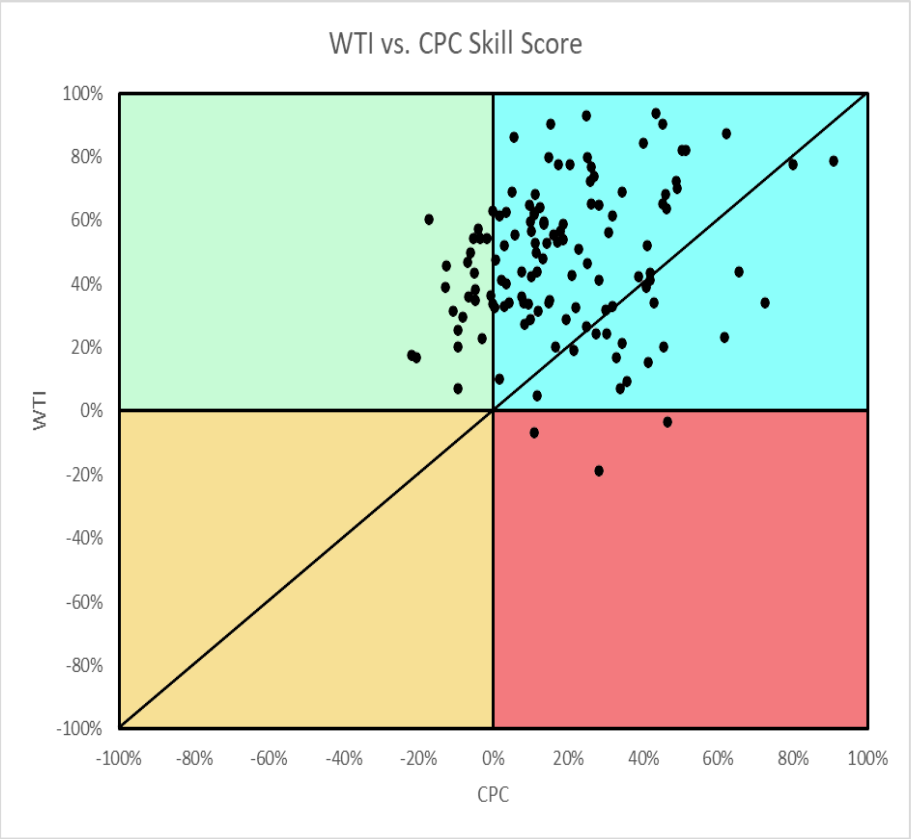

Figure 2. A comparison of forecast "hits" vs. "misses" based upon Heidke skill scores. The blue quadrat includes forecast hits by both WTI and CPC. The green quadrat represents hits by WTI and misses by CPC. The red quadrat represents hits by CPC and misses by WTI. The yellow quadrat represents misses by both groups.

percent rate (this can also be seen in the full dataset, Appendix 1). Thus, WTI's forecasts were more accurate than CPC's in 5 of every 6 forecasts during the evaluation period.

A temporal analysis of the skill scores is instructive to see when forecast differentials were greatest (Figure 3). 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 and end 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 28 consecutive month time span from 2012 to 2014.

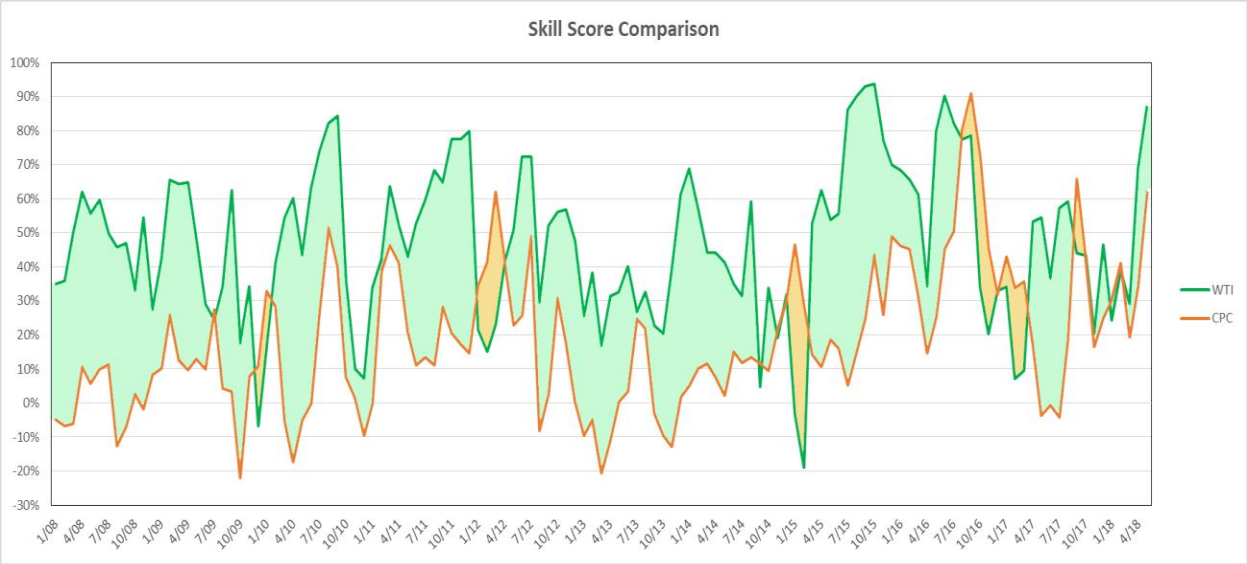

Figure 3. Skill score comparison through the evaluation period. Green areas represent periods when WTI's accuracy exceeded CPC's; orange areas show periods when CPC's skill scores were superior.

We also evaluated the magnitude of the skill scores using a series of intervals (Figure 4). 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 .30 (30%) 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.

Although the CPC results are clearly inferior in terms of accuracy to the WTI scores, it is at least encouraging to see that CPC accuracy rates have been improving (Figure 5). Through the 125 month evaluation period, the CPC skill scores have increased from an average of near 0 at the beginning of the period to about 35 percent toward the end of the evaluation period. This is an encouraging trend, although even toward the end of the period CPC's scores remain significantly lower than the WTI values.

The evaluation strongly suggests that WTI has offered much more accurate forecasts than CPC over the 10+ year evaluation period. 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. The bottom line indicates that WTI issues a more accurate forecast about 83 percent of the time, CPC's forecasts are associated with considerably more negative skill scores, and WTI forecasts are associated with higher skill scores than CPC's in long consecutive stretches of months throughout the evaluation period.

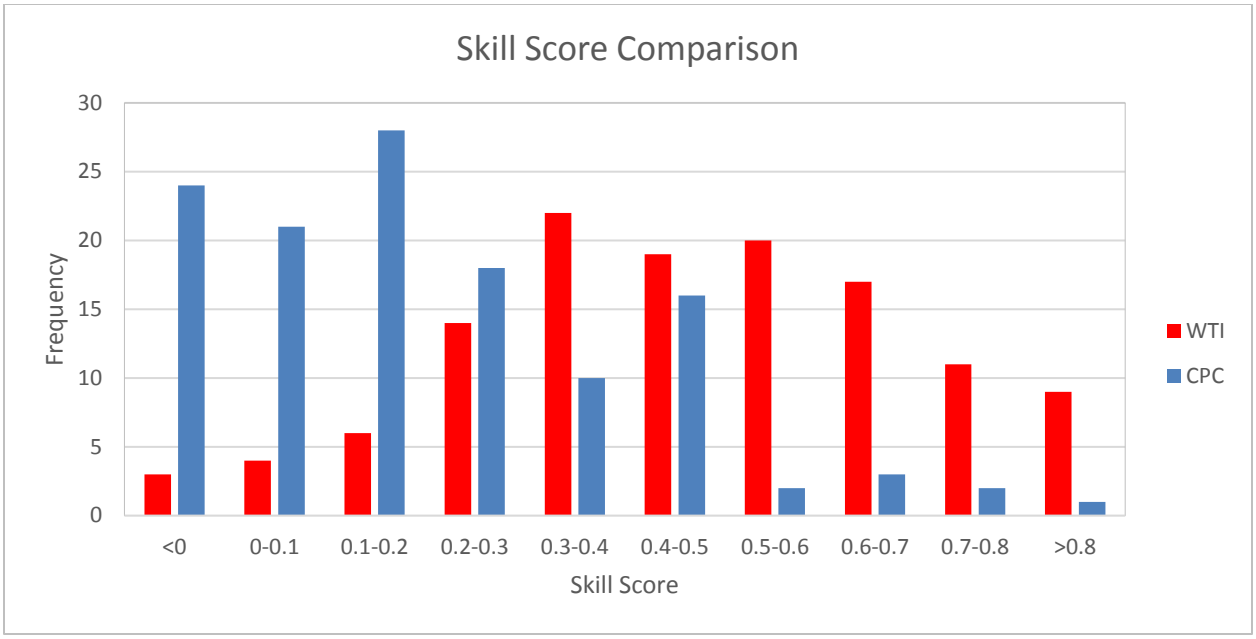

Figure 4. A comparison of skill score intervals, shown in tenths. The skill scores become more accurate as you move to the right.

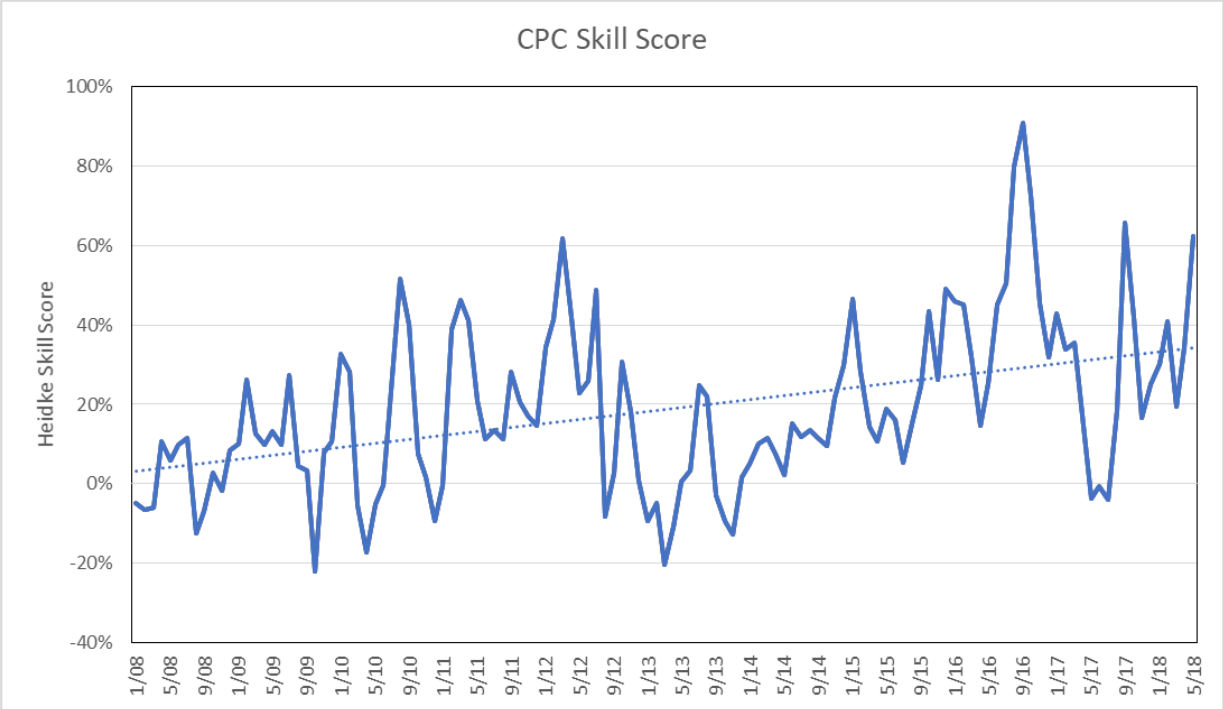

Figure 5. CPC temporal evaluation of skill scores. Y-axis represents Heidke skill score value, x-axis represents month. Dotted line represents trend.

APPENDIX 1

SKILL SCORE VALUES FOR THE ENTIRE EVALUATION PERIOD

This is the 10 year dataset constructed by WTI. Hit rates represent WTI hits; they are color-coded and the darker the yellow/brown color, the worse the hit rate. The darker the green color, the better the hit rate. Values within 45-55% are not colored. For the WTI and CPC skill score data, light green represents the higher skill score value between CPC and WTI.

Start Year	Months	Hits (out of 260)	Hit Rate (%)	WTI Skill Score	CPC Skill Score
2008	JFM	147	56.53%	34.81%	-4.74%
2008	FMA	149	57.30%	35.96%	-6.68%
2008	MAM	173	66.53%	49.81%	-6.03%
2008	AMJ	194	74.61%	61.92%	10.78%
2008	MJJ	183	70.38%	55.58%	5.82%
2008	JJA	190	73.07%	59.62%	9.91%
2008	JAS	173	66.53%	49.81%	11.42%
2008	ASO	166	63.84%	45.77%	-12.50%
2008	SON	168	64.61%	46.92%	-6.90%
2008	OND	144	55.38%	33.08%	2.80%
2008	NDJ	181	69.61%	54.42%	-1.72%
2008	DJF	134	51.53%	27.31%	8.41%
2009	JFM	160	61.53%	42.31%	10.13%
2009	FMA	200	76.92%	65.38%	26.08%
2009	MAM	198	76.15%	64.23%	12.50%
2009	AMJ	199	76.53%	64.81%	9.70%
2009	MJJ	170	65.38%	48.08%	13.15%
2009	JJA	137	52.69%	29.04%	9.91%
2009	JAS	129	49.61%	24.42%	27.37%
2009	ASO	146	56.15%	34.23%	4.31%
2009	SON	195	75.00%	62.50%	3.45%
2009	OND	117	45.00%	17.50%	-21.98%
2009	NDJ	146	56.15%	34.23%	7.97%
2009	DJF	75	28.84%	-6.73%	10.78%
2010	JFM	116	44.61%	16.92%	32.76%
2010	FMA	158	60.76%	41.15%	28.23%
2010	MAM	181	69.61%	54.42%	-5.39%
2010	AMJ	191	73.46%	60.19%	-17.24%
2010	MJJ	162	62.30%	43.46%	-5.17%
2010	JJA	196	75.38%	63.08%	-0.22%
2010	JAS	215	82.69%	74.04%	26.94%
2010	ASO	229	88.07%	82.12%	51.51%
2010	SON	233	89.61%	84.42%	40.09%

2010	OND	149	57.30%	35.96%	7.54%
2010	NDJ	104	40.00%	10.00%	1.51%
2010	DJF	99	38.07%	7.12%	-9.48%
2011	JFM	145	55.76%	33.65%	-0.22%
2011	FMA	160	61.53%	42.31%	38.79%
2011	MAM	197	75.76%	63.65%	46.34%
2011	AMJ	177	68.07%	52.12%	41.16%
2011	MJJ	161	61.92%	42.88%	20.91%
2011	JJA	178	68.46%	52.69%	11.21%
2011	JAS	190	73.07%	59.62%	13.58%
2011	ASO	205	78.84%	68.27%	11.21%
2011	SON	199	76.53%	64.81%	28.23%
2011	OND	221	85.00%	77.50%	20.47%
2011	NDJ	221	85.00%	77.50%	17.24%
2011	DJF	225	86.53%	79.81%	14.66%
2012	JFM	124	47.69%	21.54%	34.48%
2012	FMA	113	43.46%	15.19%	41.38%
2012	MAM	127	48.84%	23.27%	61.85%
2012	AMJ	158	60.76%	41.15%	41.81%
2012	MJJ	175	67.30%	50.96%	22.84%
2012	JJA	212	81.53%	72.31%	25.86%
2012	JAS	212	81.53%	72.31%	48.92%
2012	ASO	138	53.07%	29.62%	-8.19%
2012	SON	177	68.07%	52.12%	2.80%
2012	OND	184	70.76%	56.15%	30.82%
2012	NDJ	185	71.15%	56.73%	17.89%
2012	DJF	169	65.00%	47.50%	0.65%
2013	JFM	131	50.38%	25.58%	-9.48%
2013	FMA	153	58.84%	38.27%	-4.74%
2013	MAM	116	44.61%	16.92%	-20.47%
2013	AMJ	141	54.23%	31.35%	-10.78%
2013	MJJ	143	55.00%	32.50%	0.43%
2013	JJA	156	60.00%	40.00%	3.45%
2013	JAS	133	51.15%	26.73%	24.78%
2013	ASO	143	55.00%	32.50%	21.98%
2013	SON	126	48.46%	22.69%	-3.02%
2013	OND	122	46.92%	20.38%	-9.48%
2013	NDJ	154	59.23%	38.85%	-12.72%
2013	DJF	193	74.23%	61.35%	1.72%
2014	JFM	206	79.23%	68.85%	4.96%
2014	FMA	185	71.15%	56.73%	10.13%
2014	MAM	163	62.69%	44.04%	11.64%
2014	AMJ	163	62.69%	44.04%	7.54%
2014	MJJ	158	60.76%	41.15%	2.16%
2014	JJA	147	56.53%	34.81%	15.09%

2014	JAS	141	54.23%	31.35%	11.85%
2014	ASO	189	72.69%	59.04%	13.58%
2014	SON	95	36.53%	4.81%	11.64%
2014	OND	145	55.76%	33.65%	9.48%
2014	NDJ	120	46.15%	19.23%	21.55%
2014	DJF	142	54.61%	31.92%	29.96%
2015	JFM	81	31.15%	-3.27%	46.55%
2015	FMA	54	20.76%	-18.85%	28.23%
2015	MAM	178	68.46%	52.69%	14.22%
2015	AMJ	195	75.00%	62.50%	10.78%
2015	MJJ	180	69.23%	53.85%	18.75%
2015	JJA	183	70.38%	55.58%	16.16%
2015	JAS	236	90.76%	86.15%	5.39%
2015	ASO	243	93.46%	90.19%	15.30%
2015	SON	248	95.38%	93.08%	24.78%
2015	OND	249	95.76%	93.65%	43.53%
2015	NDJ	220	84.61%	76.92%	26.08%
2015	DJF	208	80.00%	70.00%	49.14%
2016	JFM	205	78.84%	68.27%	46.12%
2016	FMA	200	76.92%	65.38%	45.26%
2016	MAM	193	74.23%	61.35%	31.90%
2016	AMJ	146	56.15%	34.23%	14.66%
2016	MJJ	225	86.53%	79.81%	25.22%
2016	JJA	243	93.46%	90.19%	45.26%
2016	JAS	229	88.07%	82.12%	50.43%
2016	ASO	221	85.00%	77.50%	79.96%
2016	SON	223	85.76%	78.65%	90.95%
2016	OND	146	56.15%	34.23%	72.63%
2016	NDJ	122	46.92%	20.38%	45.47%
2016	DJF	144	55.38%	33.08%	31.90%
2017	JFM	146	56.15%	34.23%	42.89%
2017	FMA	99	38.07%	7.12%	33.84%
2017	MAM	103	39.61%	9.42%	35.56%
2017	AMJ	179	68.84%	53.27%	17.03%
2017	MJJ	181	69.61%	54.42%	-3.66%
2017	JJA	150	57.69%	36.54%	-0.65%
2017	JAS	186	71.53%	57.31%	-4.09%
2017	ASO	189	72.69%	59.04%	18.53%
2017	SON	163	62.69%	44.04%	65.73%
2017	OND	162	62.30%	43.46%	41.81%
2017	NDJ	122	46.92%	20.38%	16.59%
2017	DJF	167	64.23%	46.35%	25.00%
2018	JFM	129	49.61%	24.42%	30.17%
2018	FMA	154	59.23%	38.85%	40.95%
2018	MAM	137	52.69%	29.04%	19.40%

2018	AMJ	206	79.23%	68.85%	34.48%
2018	MJJ	238	91.53%	87.31%	62.28%