

# **Weather Derivatives Fallback Methodology Study for International Swaps and Derivatives Association**

**February 2003**

**Risk Management Solutions, Inc.**



## **Executive Summary**

- RMS was engaged by ISDA to undertake a comparative study of several proposed methods for calculating missing temperature values at a weather station using data from that station and a nearby fallback station. The ultimate objective of this work is to establish a standard missing value substitution methodology for the global weather market.
- RMS analyzed 5 possible methods by using each to estimate the temperature value for a series of historical dates with known values. From these analyses, RMS compiled a variety of statistics on the errors in the estimates produced by each methodology.
- Based on the results of these analyses, RMS recommends that ISDA adopt a methodology based on a comparison of values between a primary station and its fallback station over the 10 days before and after the date(s) with the missing value(s). Other similar approaches using different time windows produce comparable results but are not quite as accurate over the sample of stations and dates analyzed in this study. A method based on comparison of values around the date in question for the prior 10 years was clearly inferior to any of the other methods.

## Table of Contents

Section	Page
1. Analysis scope and approach	1
2. Results for filling one missing value	5
3. Results for filling fifteen consecutive missing values	21
4. Analysis of extreme prediction errors	24
5. Conclusions and recommendation	26

## **ANALYSIS SCOPE AND APPROACH**

Risk Management Solutions, Inc. (RMS) was retained by the International Swaps and Derivatives Association (ISDA) to analyze three proposed fallback methodologies for filling missing temperature values for the purpose of settling weather derivatives. The specific goals for this analysis are summarized below:

### **Fallback Methodologies**

ISDA requested that RMS evaluate three fallback methodologies, described below. In addition, RMS proposed and analyzed two additional methodologies (Methods A and C).

- **Method A:** On a given day where either the minimum and/or maximum temperature (Tmin or Tmax) is missing at a primary station, take the data from the fallback station 2 days before and 2 days after that day and compare these values to the values at the primary station. Calculate the mean difference between the two stations across these 4 days and apply that mean difference to the value reported at the fallback station on the date where data was missing for the primary station to estimate a value for the primary station's temperature.
- **Method B:** Same as Method A, except that 5 days before and after a missing value are used.
- **Method C:** Same as Method A, except that 10 days before and after a missing value are used.
- **Method D:** Same as Method A, except that 15 days before and after a missing value are used.
- **Method E:** On a given day where either Tmin or Tmax is missing at a primary station, replace it by adjusting the relevant temperature at the fallback station by the arithmetic mean of the daily differences between the relevant temperature at the fallback station and the primary station for the same day for each of the previous 10 years. If there is missing data for this day in any of the previous 10 years at either the primary station or the fallback station, go back until 10 years of data can be utilized to determine the adjustment.

## Stations

The analysis was conducted for a set of 11 primary stations in the U.S., Canada, Europe, Japan, and Australia that were specified by ISDA and for a supplemental set of 39 stations selected by RMS. The following list shows the main (first 11 stations) and supplemental stations used for the study.

Primary Station	Fallback Station	Primary Station	Fallback Station
Chicago O'Hare	Chicago Midway	Raleigh	Greensboro
Laguardia	NY Central Park	Casper	Laramie
Philadelphia	Allentown	Richmond	Washington Nat.
Phoenix	Tucson	Newark	NY Central Park
Atlanta	Montgomery	Olympia	Seattle
London Heathrow	Northolt	Kansas City	Topeka
Berlin-Dahlem	Berlin Tempelhof	Milwaukee	Madison
Oslo-Blindern	Gardermoen	Orlando	Daytona Beach
Melbourne R.O.	Melbourne Airport	Pittsburgh	Youngstown
Tokyo	Yokohama	Fargo	Grand Forks
Amsterdam Schiphol	DeBilt	Chattanooga	Knoxville
Baltimore	Washington Reagan	Bismarck	Fargo
Boston Logan	Providence	Charleston	Savannah
Covington	Colombus	Boise Air	Lewiston
Dallas Fort Worth	Dallas Love	Essen	Dusseldorf
Des Moines	Omaha	Rome	Grazzanise
Houston Bush	Houston William	Verona	Treviso
Indianapolis	Louisville	Toronto	London Airport
Minneapolis	St Cloud	Montreal	Ottawa
Nashville	Knoxville	Halifax	Charlottetown
Hartford	Providence	Belfast	Machrihanish
Providence	Boston Logan	Nottingham	Waddington
Washington Reagan	Washington Dulles	Osaka	Kobe
Seattle	Spokane	Fukuoka	Saga
Birmingham	Atlanta	Yokohama	Tokyo

### **Analysis approach**

For each day of the historical record back to 1960 (or to the start of the record if more recent than 1960) we predicted Tmin and Tmax at the primary station using each of the five fallback methodologies. In order to use the maximum amount of data for the different analyses, slightly different lengths of data were predicted in each case. For Methods A, B, C and D we predicted all days except for the first and last 2, 5, 10 and 15, respectively, of the data record. For Method E we predicted all days available except for the first 10 years.

Seasonal effects were analyzed using the same methods with data divided into four seasons: Dec-Feb, Mar-May, Jun-Aug, Sep-Nov.

In the case where 15 consecutive days were considered, missing data was divided into blocks of 15 days and every day in each block was predicted using data from neighboring blocks. This way all days except the first and last 15 were predicted.

### **Statistical measures**

From the prediction errors a range of statistics were calculated:

- “Mean absolute error” is the average of the absolute errors
- “SD error” is the standard deviation of the errors
- “Bias (% + errors)” is the percentage of over predictions (positive errors)
- “Maximum + error” is the maximum error observed
- “Maximum – error” is the maximum negative error observed

In addition, 7 quantiles were calculated to characterize the error distribution.

Average errors are often used as measures for prediction precision. However, we have not reported any average errors because these, by the design of the study, are forced to be very close to zero for Methods A-D, whatever the prediction of errors. As such, average error is not a suitable measure for a comparison of methods. It is possible to circumvent this problem, but due to time constraints we chose instead to focus on the percentage of positive errors and the median as measures for prediction precision.

Both “Mean absolute error” and “SD error” are measures of the spread of the error distribution and, whereas the mean absolute error is always smaller than the SD, they often result in a similar ranking of methods. Nevertheless, they are useful for different purposes. The SD is a standard statistical measure of distribution spread with which many are familiar. Mean absolute error, on the other hand, answers the question “How far off is the fallback prediction from the true value, on average”.

Results are in Fahrenheit for all U.S. stations and Celsius for all other stations.

**Skill scores**

In order to summarize the results for the 11 main stations, two types of skill scores were produced for each Method A-E.

The first skill score highlights the best and worst method by assigning 1 point to the method that scored best and -1 to the method that scored worst for each of four key statistics: mean absolute error, bias, the 5% quantile, and the 95% quantile. If a method was neither best nor worst it was assigned a zero. The assignment was done separately for Tmin and Tmax across all stations, and the individual scores were added up to provide a single skill score for each method.

The second skill score was calculated in the same way as the first but instead of assigning -1, 0 and 1, the best method was assigned 4, the next best 3, the third best 2 and the worst method 1 for each of the four error statistics. This skill score puts less emphasis on the best and the worst method than the first type of skill score and distinguishes more clearly between all methods.

## **RESULTS FOR FILLING ONE MISSING VALUE**

The following pages contain summaries of the results of the analyses undertaken by RMS:

- A. Individual statistics for 11 main stations
- B. Summary statistics for 11 main stations
- C. Summary statistics for 39 supplemental stations
- D. Skill score analysis
- E. Analysis of seasonality for 11 main stations

**A. INDIVIDUAL STATISTICS FOR THE 11 MAIN STATIONS**

The following tables compare the five methods for Tmin and Tmax separately. In each case, the result from the best method is shaded in green and the worst method in yellow. One table is included for each of the 11 main stations.

Chicago O'Hare	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
<b>Mean absolute error</b>	1.84	1.86	1.91	1.92	2.19	1.38	1.33	1.32	1.32	1.57
<b>SD error</b>	2.40	2.41	2.45	2.46	2.78	2.00	1.91	1.88	1.88	2.17
<b>Bias (% + errors)</b>	46%	46%	45%	46%	45%	45%	47%	47%	48%	48%
<b>Maximum + error</b>	17.75	17.80	18.50	18.47	18.10	22.25	18.20	18.90	18.87	19.20
<b>Maximum - error</b>	-15.00	-14.80	-14.95	-15.30	-14.30	-29.25	-23.50	-23.10	-23.37	-23.30
<b>5% quantile</b>	-3.75	-3.60	-3.60	-3.57	-4.20	-2.75	-2.70	-2.65	-2.67	-3.10
<b>10% quantile</b>	-2.75	-2.80	-2.80	-2.83	-3.40	-2.00	-2.00	-2.00	-2.00	-2.30
<b>25% quantile</b>	-1.50	-1.60	-1.65	-1.63	-2.00	-1.00	-1.10	-1.05	-1.03	-1.20
<b>50% quantile</b>	0.00	-0.20	-0.25	-0.27	-0.30	0.00	-0.10	-0.05	-0.07	0.00
<b>75% quantile</b>	1.50	1.40	1.50	1.47	1.60	1.00	0.90	0.95	0.93	1.20
<b>90% quantile</b>	3.00	3.10	3.20	3.23	3.60	2.00	2.10	2.05	2.03	2.50
<b>95% quantile</b>	4.00	4.20	4.30	4.37	4.80	3.00	2.90	2.87	2.87	3.40

LaGuardia	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	1.09	1.09	1.09	1.10	1.40	1.32	1.27	1.26	1.26	1.54
SD error	1.47	1.46	1.47	1.47	1.82	1.75	1.69	1.67	1.67	1.98
Bias (% + errors)	44%	46%	45%	45%	45%	47%	49%	49%	49%	47%
Maximum + error	20.25	20.00	20.20	20.17	13.30	27.75	20.20	18.00	17.03	11.60
Maximum - error	-21.50	-18.50	-18.75	-19.40	-7.20	-25.00	-24.50	-24.45	-24.27	-12.50
5% quantile	-2.25	-2.10	-2.10	-2.10	-2.80	-2.75	-2.60	-2.60	-2.60	-3.10
10% quantile	-1.75	-1.70	-1.65	-1.63	-2.20	-2.00	-2.00	-1.95	-2.00	-2.40
25% quantile	-1.00	-0.90	-0.90	-0.93	-1.20	-1.00	-1.00	-1.05	-1.03	-1.30
50% quantile	0.00	-0.10	-0.10	-0.13	-0.10	0.00	0.00	0.00	-0.03	-0.10
75% quantile	0.75	0.80	0.80	0.77	1.00	1.00	1.00	1.00	0.97	1.20
90% quantile	1.75	1.70	1.75	1.77	2.20	2.00	2.00	1.95	1.97	2.50
95% quantile	2.50	2.40	2.50	2.47	3.10	2.75	2.70	2.65	2.67	3.30

Philadelphia	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.30	2.30	2.32	2.32	2.76	2.20	2.12	2.11	2.11	2.42
SD error	2.94	2.93	2.96	2.96	3.49	2.97	2.89	2.88	2.88	3.24
Bias (% + errors)	49%	50%	51%	51%	48%	52%	54%	54%	55%	52%
Maximum + error	13.75	13.20	14.40	14.17	14.70	17.50	17.60	19.00	19.00	20.00
Maximum - error	-20.75	-18.70	-18.20	-17.67	-17.30	-22.25	-22.40	-23.20	-22.83	-23.30
5% quantile	-5.00	-4.90	-4.90	-4.93	-6.10	-5.00	-5.00	-5.00	-5.00	-5.50
10% quantile	-3.75	-3.70	-3.75	-3.77	-4.70	-3.50	-3.40	-3.40	-3.40	-3.90
25% quantile	-1.75	-1.90	-1.85	-1.90	-2.50	-1.50	-1.40	-1.40	-1.40	-1.80
50% quantile	0.00	0.10	0.10	0.10	-0.10	0.25	0.30	0.30	0.30	0.20
75% quantile	2.00	1.90	1.95	1.97	2.10	1.75	1.80	1.75	1.77	2.00
90% quantile	3.50	3.60	3.60	3.60	4.10	3.25	3.20	3.15	3.13	3.60
95% quantile	4.75	4.60	4.65	4.63	5.40	4.50	4.10	4.05	4.00	4.70

Phoenix	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.44	2.47	2.48	2.49	3.06	2.18	2.20	2.23	2.25	2.75
SD error	3.18	3.20	3.21	3.22	3.85	2.96	2.96	2.98	3.01	3.55
Bias (% + errors)	48%	48%	48%	49%	42%	47%	48%	49%	49%	47%
Maximum + error	20.00	20.80	20.35	19.97	20.60	16.50	18.10	17.05	16.77	17.30
Maximum - error	-14.25	-13.80	-13.65	-14.03	-18.10	-16.75	-17.80	-17.75	-17.63	-18.90
5% quantile	-5.00	-5.10	-5.01	-5.07	-6.70	-4.75	-4.70	-4.75	-4.77	-5.80
10% quantile	-4.00	-3.90	-3.90	-3.87	-5.40	-3.50	-3.50	-3.50	-3.50	-4.40
25% quantile	-2.00	-2.00	-2.05	-2.03	-3.10	-1.75	-1.70	-1.70	-1.77	-2.40
50% quantile	0.00	-0.10	-0.10	-0.07	-0.70	0.00	0.00	-0.05	-0.07	-0.20
75% quantile	2.00	1.90	1.95	1.93	1.90	1.50	1.60	1.70	1.70	2.10
90% quantile	4.00	3.90	4.00	4.03	4.30	3.50	3.50	3.55	3.60	4.30
95% quantile	5.25	5.40	5.40	5.43	5.90	5.00	4.90	4.90	4.93	5.70

Atlanta	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.57	2.65	2.70	2.73	3.21	2.96	2.90	2.87	2.86	3.29
SD error	3.37	3.47	3.54	3.57	4.16	4.13	4.05	4.03	4.03	4.46
Bias (% + errors)	48%	48%	48%	49%	48%	47%	47%	47%	47%	47%
Maximum + error	20.25	19.40	19.45	20.53	21.70	28.50	27.40	28.85	29.67	30.50
Maximum - error	-23.25	-20.30	-19.65	-20.10	-21.50	-30.25	-26.40	-26.70	-26.53	-27.40
5% quantile	-5.50	-5.60	-5.60	-5.67	-6.93	-6.25	-6.00	-5.85	-5.80	-6.70
10% quantile	-4.00	-4.20	-4.30	-4.37	-5.30	-4.50	-4.40	-4.30	-4.27	-5.00
25% quantile	-2.00	-2.10	-2.20	-2.20	-2.80	-2.25	-2.20	-2.25	-2.27	-2.70
50% quantile	0.00	-0.10	-0.10	-0.07	-0.20	0.00	-0.20	-0.25	-0.23	-0.20
75% quantile	2.00	2.00	2.04	2.07	2.40	2.00	2.00	1.90	1.87	2.30
90% quantile	4.00	4.30	4.35	4.40	4.90	4.50	4.50	4.40	4.43	5.20
95% quantile	5.50	5.80	5.85	5.93	6.70	6.75	6.70	6.75	6.83	7.40

London	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
<b>Mean absolute error</b>	0.69	0.71	0.71	0.71		0.35	0.34	0.34	0.34	
<b>SD error</b>	0.90	0.92	0.92	0.92		0.46	0.45	0.45	0.45	
<b>Bias (% + errors)</b>	54%	58%	60%	60%		48%	47%	46%	46%	
<b>Maximum + error</b>	3.38	3.43	3.71	3.80		2.48	2.49	2.55	2.49	
<b>Maximum - error</b>	-3.93	-4.00	-3.89	-4.09		-2.40	-2.10	-1.97	-2.00	
<b>5% quantile</b>	-1.70	-1.79	-1.81	-1.78		-0.70	-0.67	-0.67	-0.67	
<b>10% quantile</b>	-1.25	-1.33	-1.34	-1.36		-0.55	-0.52	-0.51	-0.50	
<b>25% quantile</b>	-0.48	-0.48	-0.47	-0.47		-0.30	-0.29	-0.28	-0.28	
<b>50% quantile</b>	0.10	0.16	0.20	0.21		-0.02	-0.03	-0.04	-0.03	
<b>75% quantile</b>	0.57	0.60	0.63	0.61		0.28	0.27	0.26	0.26	
<b>90% quantile</b>	1.03	1.02	0.96	0.95		0.55	0.57	0.55	0.55	
<b>95% quantile</b>	1.35	1.27	1.22	1.17		0.75	0.75	0.76	0.77	

Note that Method E could not be used for London Heathrow due to data at the fallback station Northolt only being available from 2/15/1995.

Berlin-Dahlem	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	0.63	0.64	0.65	0.65	0.71	0.36	0.35	0.35	0.35	0.39
SD error	0.84	0.84	0.84	0.85	0.91	0.50	0.48	0.47	0.47	0.52
Bias (% + errors)	47%	46%	46%	46%	47%	48%	48%	48%	48%	48%
Maximum + error	5.78	5.22	5.50	5.25	4.28	4.93	4.71	4.77	4.73	4.16
Maximum - error	-4.35	-4.35	-4.15	-4.06	-4.16	-2.93	-3.02	-3.14	-3.22	-3.21
5% quantile	-1.33	-1.29	-1.30	-1.30	-1.47	-0.77	-0.73	-0.72	-0.72	-0.81
10% quantile	-1.00	-0.97	-0.98	-0.99	-1.13	-0.57	-0.54	-0.54	-0.54	-0.62
25% quantile	-0.50	-0.51	-0.52	-0.53	-0.60	-0.28	-0.28	-0.28	-0.28	-0.33
50% quantile	-0.05	-0.07	-0.08	-0.08	-0.07	0.00	-0.01	-0.02	-0.01	-0.03
75% quantile	0.45	0.46	0.48	0.49	0.53	0.27	0.26	0.25	0.26	0.28
90% quantile	1.08	1.11	1.12	1.12	1.18	0.57	0.53	0.54	0.53	0.61
95% quantile	1.50	1.52	1.55	1.52	1.57	0.78	0.76	0.74	0.74	0.82

Oslo-Blindern	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	1.24	1.27	1.30	1.32	1.48	0.81	0.80	0.80	0.80	0.88
SD error	1.63	1.66	1.69	1.71	1.89	1.12	1.11	1.12	1.12	1.21
Bias (% + errors)	52%	54%	54%	55%	57%	50%	51%	51%	51%	52%
Maximum + error	8.08	7.95	7.18	7.65	6.92	9.68	9.63	9.58	9.60	10.03
Maximum - error	-8.15	-8.00	-8.32	-8.26	-9.39	-7.10	-7.52	-8.54	-8.79	-9.22
5% quantile	-2.78	-2.92	-3.02	-3.07	-3.23	-1.78	-1.76	-1.76	-1.77	-1.86
10% quantile	-2.05	-2.18	-2.20	-2.26	-2.39	-1.28	-1.24	-1.27	-1.25	-1.32
25% quantile	-0.95	-0.97	-1.00	-1.02	-0.97	-0.58	-0.58	-0.58	-0.58	-0.62
50% quantile	0.08	0.13	0.17	0.19	0.31	0.02	0.01	0.03	0.03	0.05
75% quantile	0.98	1.03	1.07	1.09	1.37	0.63	0.60	0.60	0.60	0.71
90% quantile	1.90	1.94	1.97	1.99	2.35	1.25	1.24	1.22	1.21	1.36
95% quantile	2.55	2.54	2.53	2.56	2.98	1.73	1.70	1.68	1.68	1.86

Melbourne	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	0.87	0.83	0.82	0.81	0.87	0.73	0.70	0.69	0.69	0.76
SD error	1.13	1.09	1.07	1.07	1.14	0.98	0.95	0.93	0.93	1.00
Bias (% + errors)	50%	49%	50%	50%	50%	51%	52%	53%	53%	58%
Maximum + error	6.55	6.51	6.64	6.67	4.76	4.70	4.84	4.98	4.95	4.27
Maximum - error	-5.80	-5.49	-5.31	-5.39	-5.25	-7.70	-7.81	-7.84	-7.99	-7.88
5% quantile	-1.80	-1.73	-1.69	-1.66	-1.79	-1.63	-1.58	-1.54	-1.53	-1.51
10% quantile	-1.35	-1.31	-1.27	-1.26	-1.33	-1.15	-1.10	-1.09	-1.07	-1.03
25% quantile	-0.70	-0.66	-0.65	-0.64	-0.69	-0.53	-0.50	-0.49	-0.49	-0.42
50% quantile	0.00	-0.01	-0.01	-0.01	0.00	0.03	0.04	0.05	0.06	0.17
75% quantile	0.70	0.67	0.64	0.64	0.72	0.60	0.58	0.57	0.58	0.74
90% quantile	1.40	1.34	1.30	1.30	1.45	1.15	1.09	1.08	1.05	1.29
95% quantile	1.85	1.78	1.75	1.74	1.89	1.50	1.43	1.37	1.37	1.63

Tokyo	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	0.53	0.51	0.50	0.50	0.54	0.70	0.66	0.65	0.64	0.68
SD error	0.73	0.70	0.69	0.69	0.72	0.93	0.88	0.87	0.86	0.90
Bias (% + errors)	48%	49%	49%	49%	49%	49%	49%	49%	49%	47%
Maximum + error	6.20	5.89	6.18	6.16	5.89	6.30	6.19	6.20	6.32	6.78
Maximum - error	-4.68	-4.33	-4.48	-4.46	-4.02	-4.83	-4.19	-3.77	-3.66	-4.04
5% quantile	-1.13	-1.05	-1.03	-1.03	-1.14	-1.43	-1.38	-1.36	-1.35	-1.46
10% quantile	-0.83	-0.79	-0.78	-0.78	-0.86	-1.10	-1.05	-1.03	-1.02	-1.12
25% quantile	-0.43	-0.41	-0.41	-0.40	-0.44	-0.57	-0.54	-0.53	-0.51	-0.58
50% quantile	0.00	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.02	-0.05
75% quantile	0.40	0.38	0.38	0.38	0.41	0.55	0.50	0.49	0.49	0.48
90% quantile	0.83	0.79	0.78	0.77	0.84	1.08	1.02	0.99	0.98	1.00
95% quantile	1.15	1.08	1.08	1.06	1.11	1.48	1.40	1.38	1.36	1.37

	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Amsterdam Schiphol										
Mean absolute error	0.84	0.85	0.86	0.86	0.93	0.62	0.60	0.59	0.59	0.66
SD error	1.11	1.12	1.12	1.12	1.21	0.81	0.79	0.79	0.79	0.87
Bias (% + errors)	50%	52%	52%	53%	53%	49%	49%	49%	50%	49%
Maximum + error	7.78	7.94	8.62	8.44	8.11	5.73	5.64	5.74	5.61	5.64
Maximum - error	-6.63	-6.40	-6.95	-7.08	-7.25	-5.68	-5.81	-5.48	-5.43	-5.18
5% quantile	-1.85	-1.89	-1.92	-1.93	-2.06	-1.28	-1.26	-1.25	-1.25	-1.44
10% quantile	-1.38	-1.41	-1.42	-1.43	-1.52	-0.98	-0.94	-0.93	-0.92	-1.07
25% quantile	-0.65	-0.66	-0.66	-0.66	-0.69	-0.50	-0.47	-0.47	-0.46	-0.55
50% quantile	0.02	0.05	0.06	0.07	0.09	0.00	-0.01	-0.01	-0.01	-0.02
75% quantile	0.67	0.68	0.68	0.69	0.78	0.48	0.47	0.47	0.46	0.50
90% quantile	1.33	1.31	1.31	1.29	1.43	0.97	0.93	0.93	0.93	1.01
95% quantile	1.79	1.75	1.75	1.73	1.89	1.30	1.27	1.26	1.26	1.37

**B. SUMMARY STATISTICS FOR 11 MAIN STATIONS**

	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
5 Fahrenheit stations										
Mean absolute error	2.05	2.08	2.11	2.12	2.54	2.03	1.99	1.98	1.98	2.34
SD error	2.77	2.80	2.83	2.85	3.35	2.91	2.85	2.85	2.85	3.25
Bias (% + errors)	47%	48%	48%	48%	45%	48%	49%	49%	49%	48%
Maximum + error	20.25	20.80	20.35	20.53	21.70	28.50	27.40	28.85	29.67	30.50
Maximum - error	-23.25	-20.30	-19.65	-20.10	-21.50	-30.25	-26.40	-26.70	-26.53	-27.40
5% quantile	-4.50	-4.50	-4.50	-4.53	-5.70	-4.50	-4.40	-4.35	-4.33	-5.10
10% quantile	-3.25	-3.30	-3.30	-3.33	-4.30	-3.00	-3.00	-3.00	-3.03	-3.70
25% quantile	-1.50	-1.60	-1.60	-1.60	-2.20	-1.50	-1.40	-1.40	-1.40	-1.80
50% quantile	0.00	-0.10	-0.10	-0.10	-0.30	0.00	0.00	0.00	0.00	-0.10
75% quantile	1.50	1.50	1.55	1.57	1.70	1.50	1.40	1.40	1.40	1.70
90% quantile	3.25	3.40	3.45	3.47	3.80	3.00	3.00	3.00	3.00	3.60
95% quantile	4.50	4.60	4.70	4.73	5.30	4.50	4.30	4.25	4.27	5.00

	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
6 Celsius stations										
Mean absolute error	0.82	0.82	0.83	0.83	0.91	0.63	0.61	0.60	0.60	0.67
SD error	1.12	1.13	1.13	1.14	1.25	0.87	0.85	0.85	0.85	0.92
Bias (% + errors)	50%	50%	51%	51%	51%	49%	50%	50%	50%	50%
Maximum + error	8.08	7.95	8.62	8.44	8.11	9.68	9.63	9.58	9.60	10.03
Maximum - error	-8.15	-8.00	-8.32	-8.26	-9.39	-7.70	-7.81	-8.54	-8.79	-9.22
5% quantile	-1.83	-1.84	-1.86	-1.87	-1.99	-1.38	-1.32	-1.32	-1.31	-1.43
10% quantile	-1.30	-1.30	-1.30	-1.30	-1.40	-0.98	-0.94	-0.94	-0.93	-1.03
25% quantile	-0.60	-0.58	-0.59	-0.58	-0.63	-0.45	-0.44	-0.44	-0.43	-0.49
50% quantile	0.00	0.01	0.01	0.02	0.04	0.00	0.00	0.00	0.00	0.01
75% quantile	0.60	0.62	0.63	0.63	0.72	0.47	0.45	0.45	0.44	0.51
90% quantile	1.30	1.29	1.30	1.29	1.48	0.98	0.95	0.94	0.93	1.05
95% quantile	1.80	1.79	1.79	1.80	2.01	1.38	1.33	1.31	1.30	1.44

**C. SUMMARY STATISTICS FOR THE 39 SUPPLEMENTARY STATIONS**

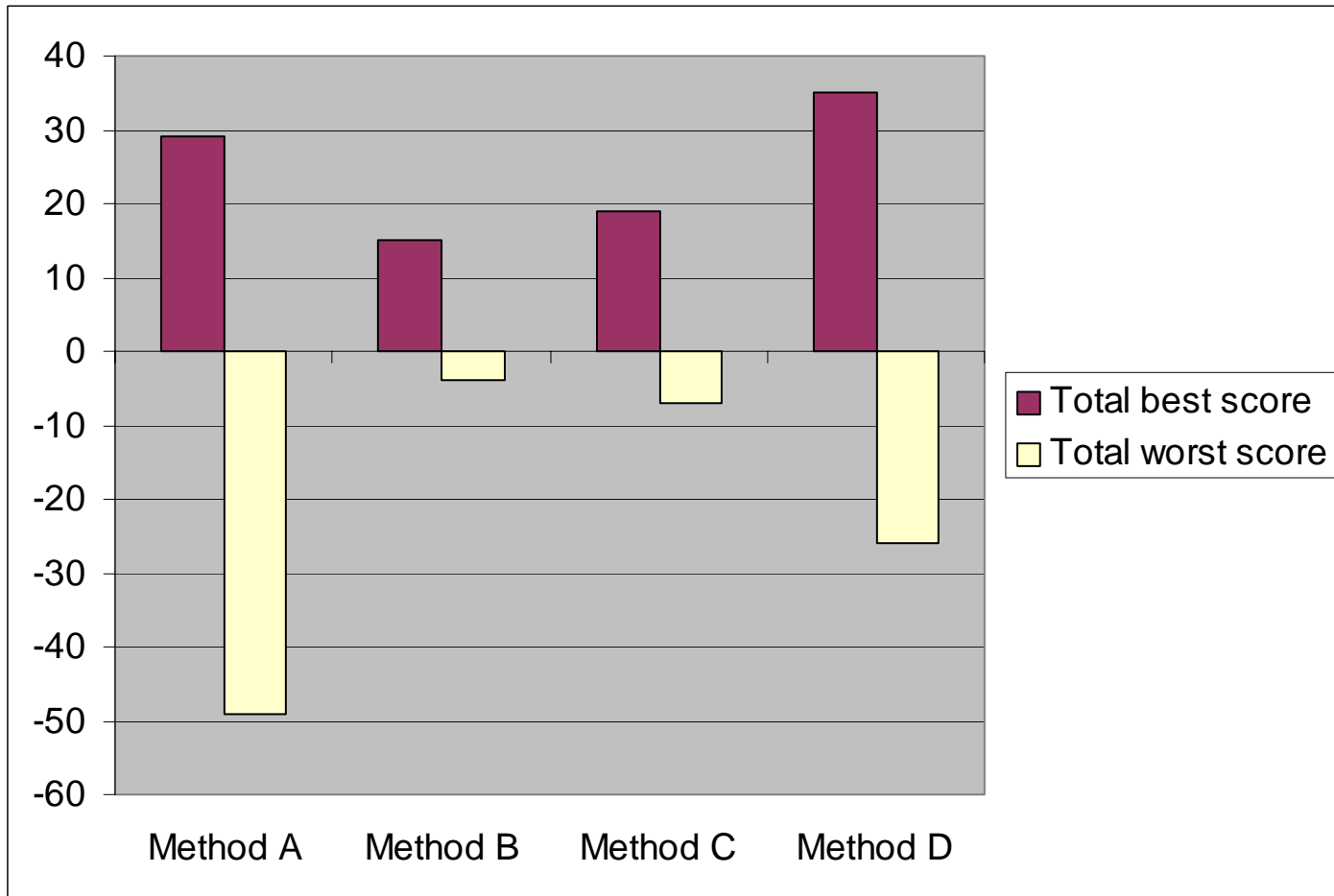
28 Fahrenheit stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.64	2.69	2.71	2.72	3.08	2.61	2.58	2.57	2.58	2.92
SD error	3.58	3.65	3.68	3.71	4.17	3.70	3.68	3.69	3.70	4.10
Bias (% + errors)	48%	49%	49%	50%	50%	48%	49%	50%	50%	49%
Maximum + error	38.75	39.30	42.30	40.97	45.60	42.00	45.50	46.40	45.83	51.10
Maximum - error	-34.75	-39.20	-37.55	-37.83	-44.80	-33.25	-32.60	-36.45	-37.77	-37.70
5% quantile	-5.75	-5.80	-5.80	-5.83	-6.50	-6.00	-5.80	-5.80	-5.80	-6.50
10% quantile	-4.25	-4.20	-4.20	-4.23	-4.80	-4.00	-4.00	-4.00	-4.00	-4.60
25% quantile	-2.00	-2.00	-2.05	-2.03	-2.30	-1.75	-1.80	-1.80	-1.80	-2.10
50% quantile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75% quantile	2.00	2.10	2.10	2.10	2.40	1.75	1.80	1.80	1.80	2.10
90% quantile	4.25	4.20	4.25	4.27	4.90	4.00	4.00	3.95	3.97	4.50
95% quantile	5.75	5.70	5.75	5.77	6.50	5.75	5.70	5.75	5.73	6.40

11 Celsius stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	1.12	1.10	1.10	1.10	1.20	1.11	1.08	1.07	1.07	1.17
SD error	1.53	1.51	1.52	1.52	1.64	1.56	1.52	1.51	1.51	1.63
Bias (% + errors)	50%	51%	51%	51%	49%	50%	50%	50%	50%	51%
Maximum + error	15.95	17.39	16.60	16.12	17.44	16.18	13.85	13.30	13.26	13.35
Maximum - error	-18.93	-18.94	-18.71	-17.96	-16.35	-19.63	-20.43	-20.71	-20.95	-20.78
5% quantile	-2.48	-2.46	-2.47	-2.47	-2.77	-2.48	-2.40	-2.38	-2.38	-2.57
10% quantile	-1.80	-1.79	-1.79	-1.78	-2.02	-1.75	-1.70	-1.70	-1.69	-1.83
25% quantile	-0.83	-0.81	-0.81	-0.81	-0.95	-0.80	-0.77	-0.77	-0.77	-0.84
50% quantile	0.03	0.02	0.03	0.03	-0.02	0.00	0.01	0.01	0.01	0.03
75% quantile	0.83	0.82	0.82	0.82	0.84	0.80	0.77	0.76	0.76	0.86
90% quantile	1.75	1.73	1.72	1.71	1.80	1.73	1.66	1.64	1.64	1.83
95% quantile	2.45	2.41	2.40	2.39	2.51	2.45	2.37	2.35	2.35	2.57

#### D. SKILL SCORE ANALYSIS

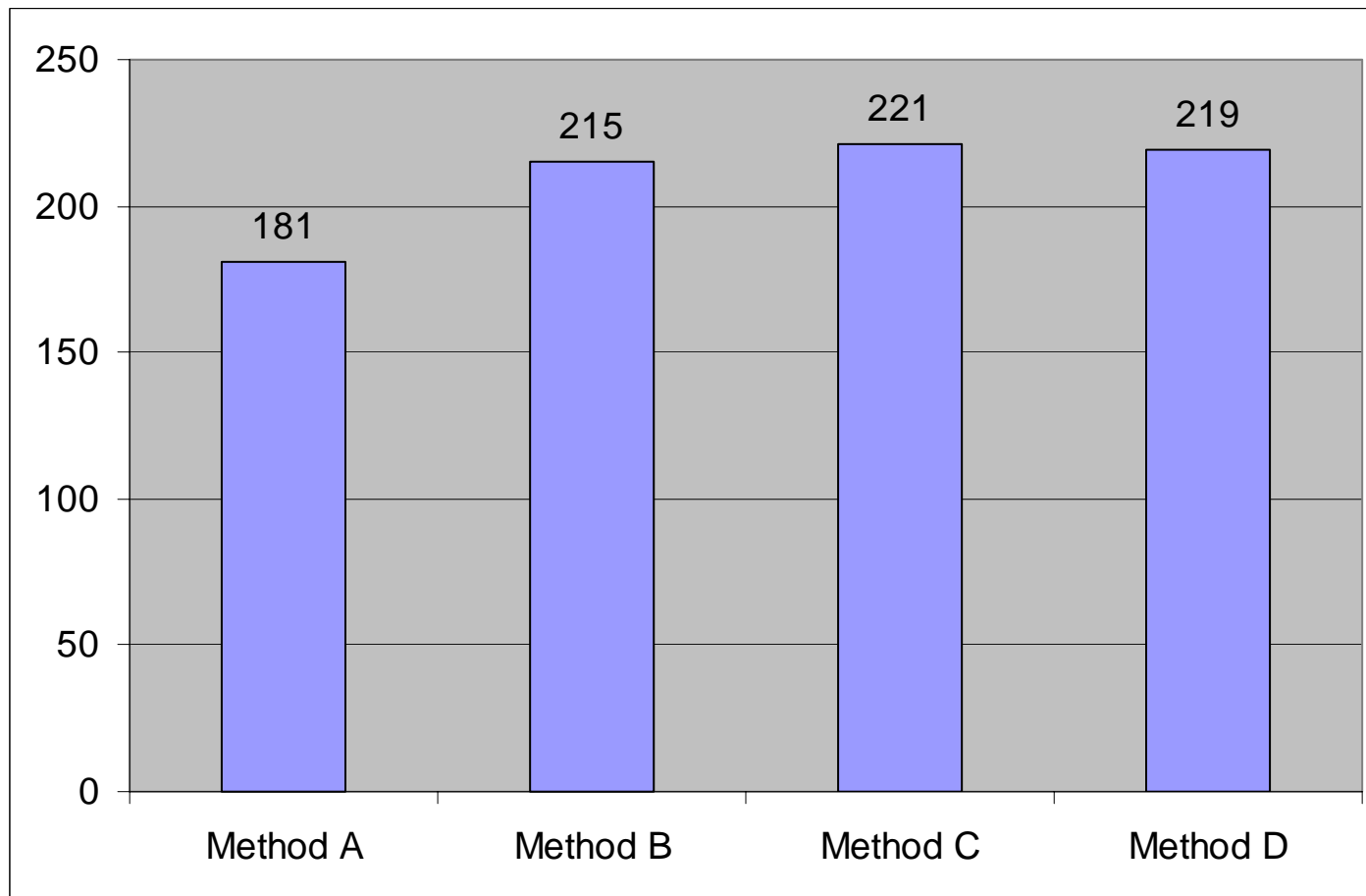
Because Method E turned out to be the worst method judged by almost all error statistics, we have excluded it from the skill score analyses. This allows for a more focused analysis of the remaining four methods.

The following graph shows the first type of skill score, i.e. the total number of times each of Methods A, B, C and D gave the best result and the worst result for four error statistics on Tmin and Tmax at the 11 main stations. Though Methods A and D show the highest occurrences of getting the best result, they also show the highest occurrences of the worst result.



This graph shows the second type of skill score for each of Methods A, B, C and D. This skill score is calculated by ranking each method from worst to best for each station and for each of Tmin and Tmax across four error statistics. A value of 4 is given to the best method, 3 for the next best, 2 for the third best and 1 for the worst method.

The score is summed up for all cases to give a total value for each model. The graph shows that Methods B, C and D are roughly equally good and slightly better than Method A.



### E. ANALYSIS OF SEASONALITY FOR THE 11 MAIN STATIONS

We analyzed the effects of seasonality on the results and have found that no strong seasonal pattern exists which would warrant different approaches for different seasons. The following tables summarize our findings.

#### December – January

	Tmin	Tmin	Tmin	Tmin	Tmin	Tmax	Tmax	Tmax	Tmax	Tmax
5 Fahrenheit stations	Method A	Method B	Method C	Method D	Method E	Method A	Method B	Method C	Method D	Method E
Mean absolute error	2.19	2.21	2.24	2.25	2.61	2.28	2.20	2.20	2.19	2.49
SD error	3.01	3.04	3.09	3.10	3.52	3.34	3.23	3.22	3.22	3.54
Bias (% + errors)	47%	47%	47%	47%	45%	47%	49%	49%	50%	48%
Maximum + error	23.25	21.60	21.35	20.17	21.70	28.50	27.40	28.35	27.77	30.50
Maximum - error	-23.25	-20.30	-20.20	-20.10	-21.50	-25.00	-24.50	-24.45	-24.27	-21.90
5% quantile	-4.75	-4.90	-4.95	-4.93	-6.00	-5.25	-5.10	-5.05	-5.07	-5.70
10% quantile	-3.50	-3.50	-3.50	-3.53	-4.40	-3.50	-3.50	-3.45	-3.40	-3.90
25% quantile	-1.50	-1.60	-1.65	-1.63	-2.10	-1.50	-1.50	-1.50	-1.47	-1.80
50% quantile	0.00	-0.10	-0.15	-0.17	-0.30	0.00	0.00	0.00	0.00	-0.10
75% quantile	1.50	1.60	1.60	1.63	1.70	1.50	1.40	1.45	1.47	1.70
90% quantile	3.50	3.60	3.70	3.67	4.00	3.50	3.40	3.35	3.30	3.80
95% quantile	5.00	5.00	5.15	5.10	5.60	5.00	4.90	4.92	4.90	5.40

	Tmin	Tmin	Tmin	Tmin	Tmin	Tmax	Tmax	Tmax	Tmax	Tmax
6 Celcius stations	Method A	Method B	Method C	Method D	Method E	Method A	Method B	Method C	Method D	Method E
Mean absolute error	0.84	0.86	0.87	0.88	0.97	0.67	0.66	0.66	0.66	0.73
SD error	1.20	1.23	1.25	1.27	1.38	0.99	0.98	0.98	0.98	1.07
Bias (% + errors)	50%	50%	50%	51%	53%	50%	51%	51%	51%	53%
Maximum + error	8.08	7.95	8.62	8.44	8.11	9.68	9.63	9.58	9.60	10.03
Maximum - error	-8.15	-8.00	-8.32	-8.26	-9.39	-7.10	-7.52	-8.29	-8.59	-9.22
5% quantile	-1.90	-1.99	-2.08	-2.13	-2.20	-1.58	-1.55	-1.56	-1.57	-1.62
10% quantile	-1.33	-1.34	-1.35	-1.36	-1.41	-1.05	-1.02	-1.02	-1.02	-1.09
25% quantile	-0.58	-0.56	-0.56	-0.56	-0.59	-0.43	-0.42	-0.41	-0.41	-0.45
50% quantile	0.00	0.01	0.01	0.01	0.06	0.02	0.01	0.02	0.02	0.05
75% quantile	0.58	0.59	0.61	0.61	0.72	0.48	0.46	0.47	0.47	0.56
90% quantile	1.30	1.34	1.37	1.37	1.59	1.05	1.02	1.01	1.01	1.16
95% quantile	1.90	1.98	2.01	2.03	2.32	1.50	1.46	1.42	1.41	1.58

March – May

5 Fahrenheit stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.09	2.10	2.12	2.13	2.52	2.15	2.11	2.10	2.11	2.44
SD error	2.80	2.81	2.83	2.85	3.34	3.05	3.00	2.99	3.00	3.37
Bias (% + errors)	47%	47%	47%	48%	46%	48%	49%	49%	49%	49%
Maximum + error	17.25	19.40	20.35	21.13	17.40	23.75	25.90	27.70	29.73	29.00
Maximum - error	-16.25	-17.20	-17.05	-16.90	-16.80	-30.25	-25.90	-23.90	-23.97	-23.30
5% quantile	-4.50	-4.50	-4.50	-4.50	-5.60	-4.75	-4.50	-4.42	-4.43	-5.20
10% quantile	-3.25	-3.30	-3.35	-3.33	-4.20	-3.25	-3.20	-3.20	-3.20	-3.80
25% quantile	-1.50	-1.60	-1.60	-1.63	-2.10	-1.50	-1.60	-1.55	-1.57	-1.80
50% quantile	0.00	-0.10	-0.10	-0.13	-0.20	0.00	0.00	-0.05	-0.07	0.00
75% quantile	1.50	1.60	1.60	1.60	1.70	1.50	1.50	1.50	1.50	1.80
90% quantile	3.25	3.40	3.40	3.43	3.90	3.25	3.20	3.20	3.23	3.80
95% quantile	4.75	4.70	4.75	4.73	5.40	4.75	4.60	4.50	4.57	5.30

6 Celcius stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	0.83	0.83	0.84	0.85	0.92	0.63	0.60	0.60	0.60	0.65
SD error	1.14	1.14	1.15	1.16	1.25	0.86	0.83	0.82	0.82	0.88
Bias (% + errors)	50%	50%	50%	51%	51%	49%	49%	49%	49%	49%
Maximum + error	6.95	6.51	6.64	6.67	5.45	6.05	5.76	5.87	5.94	6.60
Maximum - error	-7.13	-6.83	-7.65	-7.75	-8.20	-6.83	-6.95	-7.32	-7.36	-7.88
5% quantile	-1.85	-1.88	-1.89	-1.89	-1.98	-1.35	-1.28	-1.28	-1.27	-1.38
10% quantile	-1.33	-1.31	-1.33	-1.33	-1.44	-1.00	-0.94	-0.94	-0.93	-1.02
25% quantile	-0.60	-0.59	-0.59	-0.59	-0.63	-0.48	-0.46	-0.46	-0.46	-0.51
50% quantile	0.00	0.01	0.01	0.01	0.03	0.00	-0.02	-0.01	-0.01	-0.02
75% quantile	0.63	0.63	0.65	0.65	0.75	0.48	0.44	0.44	0.44	0.49
90% quantile	1.33	1.32	1.34	1.35	1.52	1.00	0.94	0.95	0.94	1.04
95% quantile	1.83	1.81	1.81	1.82	2.01	1.38	1.33	1.33	1.32	1.43

June- August

	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
5 Fahrenheit stations										
Mean absolute error	1.89	1.92	1.95	1.97	2.42	1.80	1.77	1.79	1.81	2.23
SD error	2.51	2.54	2.58	2.60	3.14	2.53	2.48	2.50	2.52	3.00
Bias (% + errors)	47%	48%	48%	49%	45%	48%	49%	50%	50%	48%
Maximum + error	16.50	16.40	16.90	17.13	18.00	20.25	17.50	18.20	18.23	17.80
Maximum - error	-14.50	-12.80	-11.90	-13.00	-18.10	-29.25	-23.50	-23.10	-23.37	-23.30
5% quantile	-4.00	-4.10	-4.15	-4.17	-5.40	-4.00	-3.80	-3.90	-3.87	-4.80
10% quantile	-3.00	-3.10	-3.05	-3.07	-4.10	-2.75	-2.70	-2.75	-2.77	-3.50
25% quantile	-1.50	-1.50	-1.55	-1.53	-2.20	-1.25	-1.30	-1.30	-1.33	-1.80
50% quantile	0.00	0.00	-0.05	-0.07	-0.30	0.00	0.00	0.00	0.03	0.00
75% quantile	1.50	1.50	1.50	1.50	1.70	1.25	1.30	1.30	1.33	1.60
90% quantile	3.00	3.10	3.15	3.17	3.60	2.75	2.70	2.70	2.73	3.40
95% quantile	4.25	4.20	4.27	4.33	4.85	3.75	3.70	3.70	3.77	4.60

	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
6 Celcius stations										
Mean absolute error	0.81	0.80	0.81	0.81	0.89	0.62	0.60	0.59	0.59	0.66
SD error	1.08	1.06	1.06	1.06	1.17	0.82	0.80	0.79	0.79	0.87
Bias (% + errors)	50%	50%	51%	51%	51%	49%	49%	49%	50%	50%
Maximum + error	5.28	5.19	5.29	5.27	5.82	4.95	4.67	4.54	4.42	4.95
Maximum - error	-7.85	-6.92	-6.67	-6.70	-5.90	-4.70	-5.01	-4.89	-4.86	-4.67
5% quantile	-1.80	-1.78	-1.79	-1.79	-1.91	-1.33	-1.29	-1.28	-1.28	-1.42
10% quantile	-1.30	-1.29	-1.28	-1.28	-1.41	-0.98	-0.94	-0.94	-0.93	-1.05
25% quantile	-0.60	-0.61	-0.62	-0.62	-0.66	-0.48	-0.47	-0.46	-0.46	-0.52
50% quantile	0.00	0.01	0.02	0.02	0.03	0.00	-0.01	-0.01	0.00	0.01
75% quantile	0.63	0.64	0.64	0.65	0.73	0.48	0.46	0.45	0.45	0.51
90% quantile	1.33	1.29	1.28	1.26	1.43	1.00	0.95	0.94	0.93	1.03
95% quantile	1.78	1.72	1.70	1.69	1.89	1.38	1.32	1.29	1.31	1.42

September – November

5 Fahrenheit stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	2.08	2.12	2.16	2.18	2.61	1.92	1.89	1.88	1.90	2.22
SD error	2.78	2.83	2.87	2.90	3.42	2.71	2.68	2.67	2.70	3.06
Bias (% + errors)	47%	48%	48%	49%	46%	48%	49%	49%	49%	48%
Maximum + error	20.00	20.80	20.35	19.97	20.60	21.50	23.00	23.65	23.50	22.20
Maximum - error	-15.00	-14.80	-14.95	-15.30	-16.80	-26.25	-24.80	-24.90	-25.10	-27.40
5% quantile	-4.50	-4.60	-4.55	-4.60	-5.90	-4.25	-4.20	-4.10	-4.10	-4.80
10% quantile	-3.25	-3.40	-3.40	-3.47	-4.40	-3.00	-2.90	-2.85	-2.93	-3.50
25% quantile	-1.50	-1.70	-1.70	-1.70	-2.30	-1.50	-1.40	-1.35	-1.37	-1.70
50% quantile	0.00	-0.10	-0.05	-0.07	-0.30	0.00	0.00	-0.05	-0.03	-0.10
75% quantile	1.50	1.60	1.60	1.60	1.80	1.50	1.40	1.35	1.37	1.60
90% quantile	3.25	3.50	3.50	3.57	3.90	3.00	2.90	2.90	2.93	3.50
95% quantile	4.50	4.80	4.85	4.87	5.40	4.25	4.10	4.05	4.10	4.80

6 Celsius stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmin Method E	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D	Tmax Method E
Mean absolute error	0.80	0.80	0.80	0.80	0.87	0.59	0.57	0.57	0.57	0.63
SD error	1.09	1.08	1.09	1.09	1.18	0.82	0.80	0.79	0.79	0.86
Bias (% + errors)	49%	50%	51%	51%	51%	50%	49%	50%	50%	49%
Maximum + error	6.85	5.77	6.23	5.90	6.30	7.85	7.97	7.77	7.64	8.28
Maximum - error	-7.10	-6.57	-6.84	-7.06	-6.94	-7.70	-8.37	-8.29	-8.31	-7.45
5% quantile	-1.75	-1.75	-1.78	-1.77	-1.93	-1.28	-1.23	-1.23	-1.23	-1.36
10% quantile	-1.28	-1.27	-1.26	-1.26	-1.32	-0.93	-0.89	-0.87	-0.87	-0.98
25% quantile	-0.60	-0.58	-0.58	-0.58	-0.61	-0.43	-0.43	-0.41	-0.41	-0.48
50% quantile	0.00	0.00	0.02	0.02	0.03	0.00	0.00	0.00	0.00	-0.01
75% quantile	0.60	0.61	0.62	0.62	0.69	0.45	0.43	0.43	0.43	0.47
90% quantile	1.28	1.25	1.25	1.25	1.41	0.93	0.91	0.89	0.88	0.98
95% quantile	1.75	1.75	1.72	1.71	1.90	1.28	1.23	1.22	1.22	1.34

## RESULTS FOR FILLING FIFTEEN CONSECUTIVE MISSING VALUES

We repeated the entire analysis for missing blocks of 15 consecutive days for the 11 main stations and the 39 supplemental stations. The statistics shown are for the daily errors when 15 days are missing so that they can be compared to the previous study. Method E will give the same daily error when filling 15 consecutive days as when filling one day. Because of this and because it produces the worst results, we have not included Method E in this section. Summary results for Methods A – D are shown on the following pages.

### Summary statistics for the 11 main stations

Fahrenheit Stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D
<b>Mean absolute error</b>	2.40	2.26	2.21	2.19	2.26	2.12	2.07	2.05
<b>SD error</b>	3.22	3.04	2.97	2.95	3.20	3.02	2.95	2.93
<b>Bias (% + errors)</b>	47%	48%	48%	48%	48%	49%	49%	49%
<b>Maximum + error</b>	21.75	20.70	21.60	22.17	30.50	28.80	30.40	29.80
<b>Maximum - error</b>	-20.25	-20.20	-19.55	-19.43	-28.75	-27.80	-27.55	-27.07
<b>5% quantile</b>	-5.25	-4.90	-4.75	-4.73	-5.00	-4.60	-4.50	-4.47
<b>10% quantile</b>	-3.75	-3.60	-3.45	-3.47	-3.50	-3.30	-3.15	-3.13
<b>25% quantile</b>	-1.75	-1.70	-1.70	-1.67	-1.75	-1.50	-1.50	-1.50
<b>50% quantile</b>	0.00	-0.10	-0.10	-0.10	0.00	0.00	0.00	-0.03
<b>75% quantile</b>	1.75	1.70	1.65	1.63	1.50	1.50	1.45	1.47
<b>90% quantile</b>	4.00	3.70	3.60	3.57	3.50	3.20	3.15	3.10
<b>95% quantile</b>	5.25	5.10	4.95	4.90	5.00	4.60	4.45	4.43

Celsius Stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D
<b>Mean absolute error</b>	0.94	0.89	0.87	0.86	0.68	0.64	0.62	0.62
<b>SD error</b>	1.28	1.22	1.19	1.19	0.95	0.89	0.88	0.87
<b>Bias (% + errors)</b>	50%	51%	51%	51%	50%	50%	50%	50%
<b>Maximum + error</b>	9.93	8.95	8.59	8.52	9.68	9.56	9.48	9.60
<b>Maximum - error</b>	-10.05	-9.25	-9.27	-9.30	-9.65	-10.25	-9.58	-9.50
<b>5% quantile</b>	-2.05	-1.97	-1.95	-1.93	-1.48	-1.39	-1.36	-1.34
<b>10% quantile</b>	-1.45	-1.39	-1.36	-1.36	-1.05	-0.99	-0.96	-0.95
<b>25% quantile</b>	-0.68	-0.63	-0.62	-0.61	-0.50	-0.47	-0.45	-0.45
<b>50% quantile</b>	0.03	0.02	0.02	0.02	0.00	0.00	0.00	0.00
<b>75% quantile</b>	0.73	0.67	0.66	0.65	0.50	0.47	0.46	0.46
<b>90% quantile</b>	1.48	1.39	1.35	1.35	1.08	0.99	0.97	0.96
<b>95% quantile</b>	2.03	1.92	1.86	1.85	1.48	1.39	1.35	1.33

Summary statistics for the 39 supplemental stations

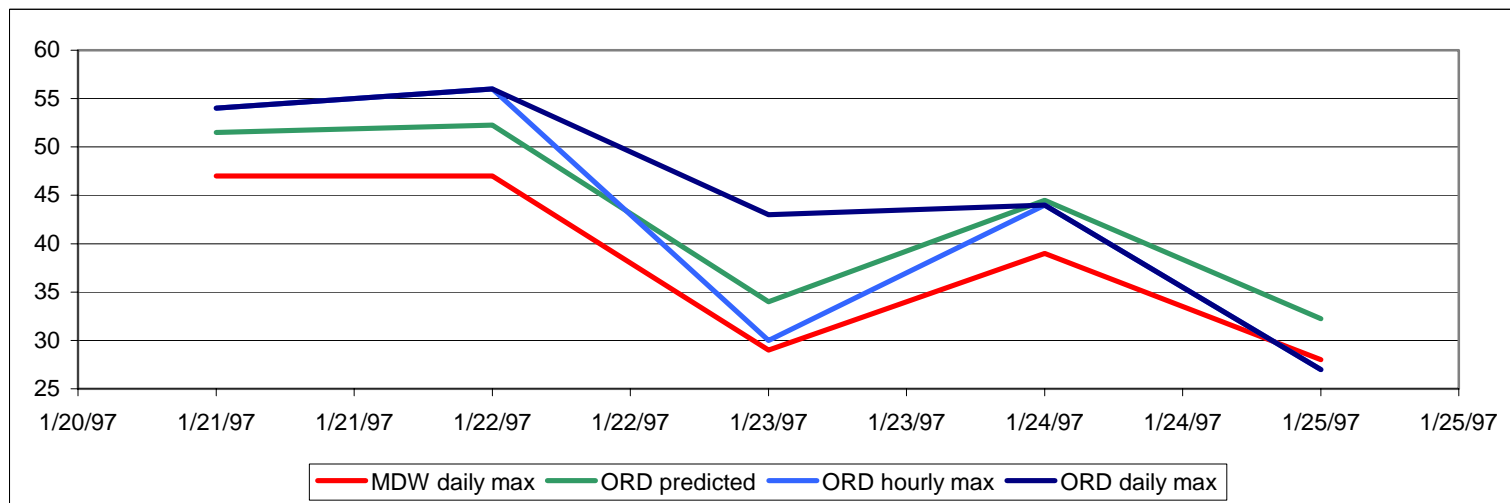
28 Fahrenheit stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D
Mean absolute error	3.09	2.92	2.85	2.82	2.94	2.77	2.69	2.67
SD error	4.20	3.98	3.89	3.86	4.17	3.94	3.85	3.83
Bias (% + errors)	48%	49%	50%	50%	48%	49%	50%	50%
Maximum + error	48.25	47.90	48.00	45.63	49.00	45.30	46.30	46.57
Maximum - error	-44.00	-43.80	-40.90	-38.27	-40.25	-39.40	-40.35	-38.60
5% quantile	-6.75	-6.30	-6.10	-6.03	-6.50	-6.20	-6.05	-6.00
10% quantile	-4.75	-4.50	-4.40	-4.37	-4.50	-4.30	-4.15	-4.13
25% quantile	-2.25	-2.20	-2.15	-2.10	-2.00	-1.90	-1.90	-1.87
50% quantile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75% quantile	2.25	2.20	2.20	2.17	2.00	1.90	1.85	1.87
90% quantile	4.75	4.60	4.45	4.40	4.50	4.30	4.15	4.07
95% quantile	6.50	6.20	6.05	6.00	6.50	6.20	6.00	5.93

11 Celsius stations	Tmin Method A	Tmin Method B	Tmin Method C	Tmin Method D	Tmax Method A	Tmax Method B	Tmax Method C	Tmax Method D
Mean absolute error	1.25	1.18	1.15	1.14	1.21	1.14	1.11	1.10
SD error	1.71	1.61	1.58	1.56	1.69	1.60	1.56	1.55
Bias (% + errors)	51%	51%	51%	51%	50%	50%	50%	50%
Maximum + error	16.70	17.42	17.02	16.68	14.95	14.72	14.16	14.09
Maximum - error	-19.63	-17.72	-17.27	-17.38	-22.13	-21.18	-21.20	-21.30
5% quantile	-2.75	-2.64	-2.57	-2.55	-2.70	-2.53	-2.47	-2.45
10% quantile	-2.00	-1.91	-1.86	-1.83	-1.90	-1.80	-1.76	-1.74
25% quantile	-0.90	-0.87	-0.84	-0.83	-0.88	-0.82	-0.80	-0.79
50% quantile	0.03	0.02	0.04	0.03	0.00	0.01	0.02	0.01
75% quantile	0.95	0.88	0.86	0.84	0.85	0.81	0.79	0.77
90% quantile	1.98	1.83	1.79	1.76	1.88	1.75	1.71	1.68
95% quantile	2.75	2.55	2.50	2.47	2.65	2.49	2.44	2.42

## ANALYSIS OF EXTREME PREDICTION ERRORS

The maximum positive and negative errors observed for all of the methods are surprisingly high, ranging from  $-44.8$  to as high as  $51.1$  degrees F. For European stations the range is less pronounced, ranging from  $-9.39$  to  $10.03$  degrees C, but Celsius stations taken as a whole also produce very large errors, ranging from  $-18.94$  to  $20.95$ .

In order to better understand why such large errors occur, we looked into a number of cases for U.S. stations. For all of the cases we examined, it seems that the large errors are a result of erroneous values in the historical records for either the primary station or the fallback station. These erroneous values can be far off of the value that would be estimated with a substitution methodology and can also influence estimated values for surrounding days when that day is included for estimating the average difference between a primary station and its fallback station. For example, on January 23, 1997 at Chicago, O'Hare (ORD) the final edited value for Tmax was 43 degrees F and 29 degrees F at Chicago Midway (MDW), and the various fallback methodologies all significantly underestimate the ORD Tmax value for the day. However, the highest hourly value recorded at ORD on that day was only 30 degrees, suggesting that the 43 degree value is erroneous and that the fallback methodologies are not actually producing significant errors in comparison to the true values. This is illustrated on the graph below, which shows the recorded daily maximum temperature at ORD and MDW together with the prediction from Method A. The line labeled "ORD hourly max" is the daily ORD values with the value on January 23 replaced by the hourly maximum for that day.



We looked at each of the five occurrences where the prediction error at ORD was greater than ten degrees F since 1997 and found that only one of these was the result of an apparently bad edited value at the primary station, ORD. The other four came from bad data at the fallback station, MDW, for which edited data is unavailable. One way to minimize such occurrences would be to select high-quality fallback stations for which edited data is available.

It is reasonable to assume that data quality problems which would cause such large errors are less likely to occur today than in the past, given improved quality control procedures in place at most national meteorological offices.

It is worth noting that none of the methods examined in this study could deal with these types of situations and that a study using data from the primary station (such as hourly data) would be needed instead.

## **CONCLUSIONS AND RECOMMENDATION**

### **Analysis Conclusions**

- Method E is consistently inferior to Methods A – D and we recommend that it be excluded from consideration. This finding is not surprising, since the presence of factors such as trends and measurement discontinuities can affect two neighboring stations differently and can therefore make historical data less reliable than recent data for analyses of this type. Furthermore, because of the persistence of weather patterns, one would expect a method that uses local data to be more accurate than a method that relies on climate statistics.
- Methods A-D all produce very similar measures of mean absolute error, the measure of the magnitude of the average difference between an estimated value and the actual value. Furthermore, these values are generally in range of 2 to 3 degrees F or 0.6 to 1.2 degrees C, which seems to be an acceptable range given the simplicity of the methods.
- The bias statistics suggest that none of the Methods A – D are significantly biased.
- The quantile statistics for the 11 main stations indicate that Methods A and B are less likely to generate significantly erroneous values than Methods C or D for Tmin, although all of the four methods are sometimes ranked highest for specific quantiles at individual stations. This pattern is reversed for Tmax. However, when the supplementary 39 stations are included in the analysis these differences are less pronounced.
- The maximum positive and negative errors observed for all of the methods are surprisingly high, although they occur on a very small number of the total dates analyzed and appear to be caused by data quality issues and not methodology flaws. Given current quality control procedures at most national meteorological offices, it is unlikely that such large errors will be a common occurrence in the future. However, fallback stations with reliable edited data should be used wherever possible to minimize this risk.
- The skill scores presented in Section 3 provide another perspective on the relative accuracy of Methods A – D and demonstrate that Method C is marginally the most accurate across the 11 primary stations.
- By analyzing each season separately for each of the 11 primary stations, we conclude that there is no significant difference in the accuracy of Methods A – D between seasons.

## Recommendations

- RMS believes that any of Methods A – D would be suitable for use in the weather market. All methods provide reasonable accuracy, and the differences between them are minimal.
- Of the 4 methods, the RMS analyses show that for the 11 main stations, Methods A and B are generally superior for Tmin and that Methods C and D are generally superior for Tmax. However, we do not believe this is significant given the very small differences in error between the approaches.
- We recommend that ISDA adopt Method C. In our study, Methods A and D produce the highest frequencies of best values, but they carry the added risk of having the highest frequencies of worst values too. Methods B and C give the best combined results across both Tmin and Tmax, and the skill score analysis reveals Method C to be slightly superior to B across the 11 main stations of interest.
- Our analysis of Methods A – D for filling 15-day blocks of missing data shows that Method D is superior. However, given that the difference in accuracy between methods C and D is small, we do not recommend using a different methodology for filling large blocks than for filling smaller blocks of missing data.