Skillful seasonal predictions of boreal winter accumulated heating degree days and relevance for the weather derivative market

S. Brands *

Instituto de Física de Cantabria, IFCA (CSIC-UC), Santander, Spain.

---

* *Corresponding author address:* Instituto de Física de Cantabria, IFCA (CSIC-UC), Santander, Spain

E-mail: brandssf@unican.es

ABSTRACT

It is demonstrated that boreal winter accumulated heating degree days, a weather derivative product frequently demanded by energy suppliers (among others), can be skillfully predicted with a lead time of one month, i.e. at the beginning of the previous November, for many regions of the Northern Hemisphere extratropcis. This finding contradicts the assumption of poor seasonal predictability for this variable. The present short note is meant to properly inform the participants of the weather derivative market in order to assure market transparency and to foster the scientific discussion on how to disseminate with this formerly unknown expert knowledge.

1. Introduction

The seasonal predictability of the boreal winter (DJF) climate in the Northern Hemisphere (NH) extratropics is generally assumed to be poor (Kushnir et al. 2006; Folland et al. 2012; Kim et al. 2012). However, it was recently demonstrated that DJF-mean 2m air temperatures over a large fraction of the NH can be predicted from Eurasian snow cover increase during the previous October, as described by the Snow Advance Index (SAI) (Cohen and Jones 2011; Cohen et al. 2012). In a follow-up study, it was shown that this index can be equally used for predicting DJF-precipitation totals in Europe, obtaining significant forecast skill over the Iberian Peninsula and southern Norway (Brands et al. 2012).

The physical basis for these empirical findings has been provided by a large number of observational and idealized general circulation model (GCM) studies (Cohen et al. 2001; Gong et al. 2003; Cohen et al. 2007; Hardiman et al. 2008; Fletcher et al. 2009; Kolstad and Charlton-Perez 2011; Smith et al. 2011). Following the conceptual model in Cohen et al. (2007), an above-normal Eurasian snow cover extend in October enhances the upward wave activity flux, which, in turn, weakens the stratospheric polar vortex. Due to the relatively long decorrelation time (or memory) of the stratosphere, this weakening tends to persist and propagate downward into the troposphere (Baldwin et al. 2003), thereby favoring the negative phase of the Arctic Oscillation during the following winter season. Although some aspects of this teleconnection pathway are not yet fully understood, it is increasingly recognized that the predictability of the boreal winter climate might be much larger than previously thought and that replicating the above mentioned lagged teleconnection in operative dynamical seasonal forecasting models could be key for improving their forecast skill (Cohen and Jones 2011; Peings et al. 2012).

The present study assesses the relevance of this formerly unknown expert knowledge for the weather derivative market and discusses some implications. This is done for the example of accumulated heating degree days (HDD), a common weather derivative product (Zeng 2000) which, for example, is traded at the Chicago Mercantile Exchange (CME) GLOBEX electronic platform (`http://www.cmegroup.com/trading/weather/`). Before the start of the winter season, energy suppliers (among others) usually make contracts based on HDD in order to insure against a possible loss of sales that would be caused by an anomalously warm winter (Quayle and Diaz 1980; Sailor and Munoz 1997; Timmer and Lamb 2007). In this study, it will be shown that winter accumulated heating degree days (hereafter 'DJF-HDD') can be skillfully forecasted in many regions of the Northern Hemisphere extratropcis, among them the southern United States, by using a statistical forecasting method based on simple linear regression. Since the standard deviation with respect to the climatological mean is usually taken as reference for defining the 'strike', i.e. the pre-negotiated threshold value above/below which the energy supplier/buyer will be paid out or not (Zeng 2000; Jewson and Brix 2005; Considine 2012), it constitutes the forecast variable in the present study. The added value/skill of the proposed forecasting method will be measured in terms of the percentage with which the purely climatological forecast (i.e. always predicting the climatological mean/standard deviation = 0) is outperformed. Having available a skillful prediction for the DJF-HDD on the 1st of November, i.e. one month ahead, is a kind of expert knowledge which should be made available in the field of weather derivative trading in order to assure market transparency.

2. Data and Methods

Accumulated heating degree days are commonly defined as follows:

$$HDD = \sum_{i=1}^{n} (T_{base} - T_{mean})^{+} \tag{1}$$

where n is the number of days the HDD is calculated for, i.e. n=90 for the DJF-season focussed on in this study (note that the 29th of February is ignored in case of leap years), T_{base} is the baseline temperature which is set at the commonly used value of $18°C$ or $65°Fahrenheit$, valid for a typical uninsolated building, and T_{mean} is the daily-mean 2m air temperature at a given location. The plus sign indicates that only positive values are accumulated (Buyukalaca et al. 2001).

T_{mean} are calculated upon 6-hourly reanalysis data from ERA-Interim (Dee et al. 2011), which are publicly available at the European Center of Medium-Range Weather Forecasts (`http://data-portal.ecmwf.int/data/d/interim_daily/`). To confirm the results obtained with this gridded dataset for selected cities of the United States, in-situ station data from the United States National Climatic Data Center, which are freely available at the Average Daily Temperature Archive of the University of Dayton (`http://academic.udayton.edu/kissock/http/Weather/default.htm`), are used in addition. Note that some station time series suffer from a very small number ($< 0.5\%$ in any case) of large outlier values which are first set to missing values and then are linearly interpolated (Jewson and Brix 2005). Thereafter, for each DJF-season between 1997/98 and 2011/12 (n=15) and for each reanalysis grid-box and station location, the HDD values are calculated as described in Eq. 1 and represent the values to be predicted (the predictands).

The increase of Eurasian snow cover in October is calculated using daily satellite retrievals from NOAA (Ramsay 1998), which are stored at the National Snow and Ice Data Center (`ftp://sidads.colorado.edu/pub/DATASETS/NOAA/G02156/24km/`). For each October between 1997 and 2012, the daily snow cover over a spatial domain covering $25-60°N$ and $0-180°E$ is calculated and a time series of d = 31 days of square kilometer values is obtained (Cohen and Jones 2011). Then, the robust linear regression approach described in Street et al. (1988) is applied to this time series. For this specific problem, robust regression is preferable to ordinary regression since it avoids the fitted straight line from being 'moved' towards outlier values that do occur in the daily snow cover time series. Consequently, robust regression is resistant to outlier values and essentially removes one uncertainty source. The slope of the robust regression line (i.e. the regression coefficient) is then taken as index value (Cohen and Jones 2011), which represents the rate of advance of the snow cover extend in October (and not the mean value). This index value is available on November 1st and is applied as the only predictor from which DJF-HDDs are predicted.

As a statistical forecasting method, simple linear regression is applied in a one-year out cross-validation framework (Michaelsen 1987). The objective of this approach is to construct a loop in which the i^{th} of $n = 15$ DJF-HDD values (expressed in standardized anomalies) is forecasted from the regression parameters obtained from regressing the remaining $n - 1$ HDD values against its corresponding predictor values. In order to avoid artificial skill (von Storch and Zwiers 1999), the i^{th} of the $n = 15$ DJF-HDD values is forecasted from the standardized anomalies of the remaining $n - 1$ HDD values, i.e. the calculation of the mean and standard deviation is repeated in each step of the cross validation. Note that the temporal autocorrelation of the DJF-HDD time series was found to be spurious in those regions where the forecasting method is successful (see below for a definition of 'successful'). This is in qualitative agreement with Madden (1977) and Kushnir et al. (2006) and justifies the assumption of data independence the cross-validation approach relies on.

To measure the forecast/out-of-sample skill of the proposed forecasting scheme, the time series of 15 independend predictions is compared to its observed counterpart. The first measure of forecast skill applied here is the root mean square error skill score ($rmsess$) defined as follows (Jolliffe and Stephenson 2003):

$$rmsess = (1 - \frac{rmse}{rmse_{ref}}) \times 100 \tag{2}$$

where $rmse$ is the root mean square error of the time series predicted by the statistical forecasting method and $rmse_{ref}$ is the $rmse$ obtained from always predicting the climatological mean, the latter being 0, since standardized anomalies/zero-mean data are calculated in each step of the cross-validation. Note at this point that calculating the climatological mean upon 15 or even 10 values only is a standard procedure in weather derivative trading (Considine 2012). Thus, $rmsess$ gives the percentage with which the error committed by assuming 'normal' winters is decreased by using the statistical forecasting method proposed here.

In addition to $rmsess$, the Pearson correlation coefficient (r) between the observed and forecasted DJF-HDD values is applied and its significance is computed using a two-sided t-test (null hypothesis $H_0 : r = 0$), which is a standard procedure for assessing the skill of seasonal forecasts (Kim et al. 2012).

The above mentioned one-year-out cross-validation is applied separately for each grid-box and station. Due to the massive repetition of the method to thousands of grid boxes covering the NH, significant local out-of-sample skill ($\alpha_{local} = 0.05$) over a areal fraction of 5% would be expected even if the method had no predictive capability (assuming zero spatial autocorrelation). Consequently, it must be assured that the areal fraction obtained by the forecasting approach is very unlikely to be a product of chance. To this end, and similar to the method applied in DelSole and Shukla (2009), the cross-validation approach is repeated 1000 times, using the randomly re-shuffled predictor time series in each step. The 99th percentile of the resulting 1000 areal fractions of locally significant forecast skill arising from chance is then assumed to represent the critical value above which global/field significance (α_{global}) is obtained. Note that the spatial autocorrelation of the predictand is taken into account by this technique.

To additionally demonstrate that the results are not sensitive to possible trends in the observed DJF-HDD time series, the statistical forecasting approach was also applied to linearly de-trended time series (the detrending was repeated in each step of the cross-validation in order to avoid artificial skill). The corresponding results were found to closely match those obtained from undetrended time series. Since the method-related uncertainties of de-trending short-time series are manyfold (Jewson and Brix 2005), and since one-year-out cross-validation disrupts the temporal sequence (i.e. the independent variable of the trend), the results for the raw time series will be shown in the next section.

3. Results

Fig. 1a shows the forecast skill in terms of *rmsess* for the DJF-HDD values obtained from gridded ERA-Interim 2m air temperatures (see Eq. 1). Areas exhibiting an *rmsess* $> 20\%$ are found for the following regions (in alphabetic order):

- Central America
- Eastern Mediterranean
- Greenland
- North Africa
- Norway
- Siberia
- Southeast Asia
- Southern Alaska
- **Southern U.S.A**

the latter being most relevant for weather derivative trading. Note that the areal fraction of locally significant correlations ($\alpha_{local} = 0.05$) between observed and forecasted values — which are not shown since they roughly correspond to *rmsess* values $> 20\%$— was found to be highly significant ($\alpha_{global} = 0.01$), i.e. field/global significance was obtained.

Since HDD values for the major U.S. cities are directly traded at the CME GLOBEX electronic platform, the statistical forecasting and cross-validation procedure is explicitly applied to the respective station data. In other words, the weather derivative product 'heating degree days for U.S. cities' is directly forecasted in this step of the study. As visualized in Fig. 1b, the station-specific results are in close agreement with those obtained from reanalysis data. With a *rmsess* of up to 47% and an r of up to 0.85 for the case of the

DJF-HDD in Houston (Texas), considerable forecast skill is yielded for the southern U.S. cities. Note that r is significant at a test level of at least 1% for each of the 6 major cities of the southern United States (exception: Jacksonville, where r is significant at a test level of 5%). An overview of the validation results for the Southern cities, including $rmsess$, r and the $p - value$ is given in Tab. 1. As an example, observed and forecasted time series are shown in Fig. 1c for Houston (Texas).

4. Discussion

The statistical forecasting method proposed in this study is expected to be of interest for both the buyers and sellers of the weather derivative market, in order to decide on whether to close a contract or not. For instance, on the basis of the forecast for DJF 2011/12 (see Fig. 1c), which was available on the 1st of November 2011 and turned out to be reasonably correct, an energy supplier fearing financial loss in case the standard deviation for DJF-HDD is largely negative might have insured against this risk. In contrast, on the basis of the forecast for DJF 2009/10, which also turned out to be reasonably correct, he probably would not have done so. In turn, if the forecast for DJF 2011/12 would have been available for the insurer, he/she probably would have raised the premium/price for the above mentioned insurance.

In swap contracts, the 'fair strike' (i.e. the strike for which the expected pay-off is zero) is usually defined as the unconditional/climatological mean of the observed time series of DJF-HDD (Jewson and Brix 2005). However, in those regions where DJF-HDD are predictable (see Fig. 1a), this definition should be reconsidered since the expected value for the DJF-HDD following a large (small) Eurasian snow cover increase in October is larger (smaller) than the unconditional mean. As a possible solution for this problem, the 'fair strike' could be defined as a the mean value conditioned to the index for October Eurasian snow cover increase. However, the sample size of this index is small to-date (n = 15, due to the limited availability of daily snow cover data), which, presently, would be one argument against the calculation of conditional means.

Finally, it should be noted that the inter-annual variability of the boreal winter climate is not solely driven by Eurasian snow cover (Kushnir et al. 2006; Fletcher and Saunders 2006; Folland et al. 2012). Moreover, due to the short sample size available to date, possible non-stationarities in the described teleconnection cannot be currently assessed. Consequently, the proposed statistical forecasting scheme should be re-evaluated in the future.

5. Conclusions

This study has shown that DJF accumulated heating degree days over a large and highly significant fraction of the Northern Hemisphere extratropics can be skillfully predicted from Eurasian snow cover increase in the previous October. For this purpose, simple linear regression has been applied as a statistical forecasting method which has been validated in a one-year-out cross-validation framework. With a root mean square error skill score of up to 47% and a Pearson correlation coefficient of up 0.85, the forecast skill at 6 major cities of the southern United States is significant at a test level of up to 0.1%, and, thus, is very unlikely a product of chance. Since the latter variables are directly traded as weather derivative products, it is important to communicate this expert knowledge to the respective sellers and buyers (i.e. from the energy sector) in order to guarantee fair trading.

Acknowledgments.

This study was developed in the framework of the SPECS project, financed by the European Commission through the Seventh Framework Programme for Research, Grant Agreement 3038378. The author would like to thank the 'Consejo Superior de Investigaciones Científicas JAE-PREDOC' programm for financial support and acknowledges the ECMWF ERA-Interim, NSIDC and University of Dayton Average Daily Temperature

Archive datasets. He would also like to thank the three anonymous reviewers for their helpful comments on the manuscript. A special thanks goes to Dr. José Manuel Gutiérrez for supervising this work and to Dr. J.J. Taboada (MeteoGalicia) for discussing the initial idea.

REFERENCES

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301 (5633)**, 636–640, doi:{10.1126/science.1087143}.

Brands, S., R. Manzanas, J. M. Gutierrez, and J. Cohen, 2012: Seasonal Predictability of Wintertime Precipitation in Europe Using the Snow Advance Index. *J. Climate*, **25 (12)**, 4023–4028, doi:{10.1175/JCLI-D-12-00083.1}.

Buyukalaca, O., H. Bulut, and T. Yilmaz, 2001: Analysis of variable-base heating and cooling degree-days for Turkey. *Applied Energy*, **69 (4)**, 269–283, doi:{10.1016/S0306-2619(01) 00017-4}.

Cohen, J., M. Barlow, P. J. Kushner, and K. Saito, 2007: Stratosphere-troposphere coupling and links with Eurasian land surface variability. *J. Climate*, **20 (21)**, 5335–5343, doi: {10.1175/2007JCLI1725.1}.

Cohen, J. and J. Jones, 2011: A new index for more accurate winter predictions. *Geophys. Res. Lett.*, **38**, doi:{10.1029/2011GL049626}.

Cohen, J., K. Saito, and D. Entekhabi, 2001: The role of the Siberian high in Northern Hemisphere climate variability. *Geophys. Res. Lett.*, **28 (2)**, 299–302, doi:{10.1029/ 2000GL011927}.

Cohen, J. L., J. C. Furtado, M. A. Barlow, V. A. Alexeev, and J. E. Cherry, 2012: Arctic warming, increasing snow cover and widespread boreal winter cooling. *Environ. Res. Lett.*, **7 (1)**, doi:{10.1088/1748-9326/7/1/014007}.

Considine, G., 2012: Introduction to weather derivatives. Tech. rep., Weather Derivatives Group, Aquila Energy, Chicago.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137 (656, Part a)**, 553–597, doi:{10.1002/qj.828}.

DelSole, T. and J. Shukla, 2009: Artificial Skill due to Predictor Screening. *J. Climate*, **22 (2)**, 331–345, doi:{10.1175/2008JCLI2414.1}.

Fletcher, C. G., S. C. Hardiman, P. J. Kushner, and J. Cohen, 2009: The Dynamical Response to Snow Cover Perturbations in a Large Ensemble of Atmospheric GCM Integrations. *J. Climate*, **22 (5)**, 1208–1222, doi:{10.1175/2008JCLI2505.1}.

Fletcher, C. G. and M. A. Saunders, 2006: Winter North Atlantic Oscillation hindcast skill: 1900-2001. *J. Climate*, **19 (22)**, 5762–5776, doi:{10.1175/JCLI3949.1}.

Folland, C. K., A. A. Scaife, J. Lindesay, and D. B. Stephenson, 2012: How potentially predictable is northern European winter climate a season ahead? *Int. J. Climatol.*, **32 (6)**, 801–818, doi:{10.1002/joc.2314}.

Gong, G., D. Entekhabi, and J. Cohen, 2003: Modeled Northern Hemisphere winter climate response to realistic Siberian snow anomalies. *J. Climate*, **16 (23)**, 3917–3931, doi:{10.1175/1520-0442(2003)016⟨3917:MNHWCR⟩2.0.CO;2}.

Hardiman, S. C., P. J. Kushner, and J. Cohen, 2008: Investigating the ability of general circulation models to capture the effects of Eurasian snow cover on winter climate. *J. Geophys. Res*, **113 (D21)**, doi:{10.1029/2008JD010623}.

Jewson, S. and A. Brix, 2005: *Weather derivative valuation. The meteorological, statistical, financial and mathematical foundations*. Cambridge University Press, Cambridge.

Jolliffe, I. and D. Stephenson, 2003: *Forecast Verification*. Wiley, Chichester.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dyn.*, **39 (12)**, 2957–2973, doi:{10.1007/s00382-012-1364-6}.

Kolstad, E. W. and A. J. Charlton-Perez, 2011: Observed and simulated precursors of stratospheric polar vortex anomalies in the Northern Hemisphere. *Climate Dyn.*, **37 (7-8)**, 1443–1456, doi:{10.1007/s00382-010-0919-7}.

Kushnir, Y., W. A. Robinson, P. Chang, and A. W. Robertson, 2006: The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability. *J. Climate*, **19 (23)**, 5949–5970, doi:{10.1175/JCLI3943.1}.

Madden, R., 1977: Estimates of autocorrelations and spectra of seasonal mean temperatures over North-America. *Mon. Wea. Rev.*, **105 (1)**, 9–18, doi:{10.1175/1520-0493(1977)105⟨0009:EOTAAS⟩2.0.CO;2}.

Michaelsen, J., 1987: Cross-validation in statitical climate forecast models. *J. Climate Appl. Meteor.*, **26 (11)**, 1589–1600, doi:{10.1175/1520-0450(1987)026⟨1589:CVISCF⟩2.0.CO;2}.

Peings, Y., D. Saint-Martin, and H. Douville, 2012: A Numerical Sensitivity Study of the Influence of Siberian Snow on the Northern Annular Mode. *J. Climate*, **25 (2)**, 592–607, doi:{10.1175/JCLI-D-11-00038.1}.

Quayle, R. and H. Diaz, 1980: Heating degree-day data applied to residential heating energy-consumption. *J. Appl. Meteorol.*, **19 (3)**, 241–246, doi:{10.1175/1520-0450(1980)019⟨0241:HDDDAT⟩2.0.CO;2}.

Ramsay, B. H., 1998: The interactive multisensor snow and ice mapping system. *Hydrol. Processes*, **12 (10-11)**, 1537–1546, doi:{10.1002/(SICI)1099-1085(199808/09)12:10/11⟨1537::AID-HYP679⟩3.0.CO;2-A}.

Sailor, D. and J. Munoz, 1997: Sensitivity of electricity and natural gas consumption to climate in the USA - Methodology and results for eight states. *Energy*, **22 (10)**, 987–998, doi:{10.1016/S0360-5442(97)00034-0}.

Smith, K. L., P. J. Kushner, and J. Cohen, 2011: The Role of Linear Interference in Northern Annular Mode Variability Associated with Eurasian Snow Cover Extent. *J. Climate*, **24 (23)**, 6185–6202, doi:{10.1175/JCLI-D-11-00055.1}.

Street, J., R. Carroll, and D. Ruppert, 1988: A note on computing robust regression estimates via iteratively reweighted least-squares. *Am. Stat.*, **42 (2)**, 152–154, doi:{10.2307/2684491}.

Timmer, R. P. and P. J. Lamb, 2007: Relations between temperature and residential natural gas consumption in the central and eastern United States. *J. Appl. Meteorol. Climatol.*, **46 (11)**, 1993–2013, doi:{10.1175/2007JAMC1552.1}.

von Storch, H. and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.

Zeng, L., 2000: Weather derivatives and weather insurance: Concept, application, and analysis. *Bull. Amer. Meteor. Soc.*, **81 (9)**, 2075–2082, doi:{10.1175/1520-0477(2000)081⟨2075:WDAWIC⟩2.3.CO;2}.

List of Tables

TABLE 1. Forecast skill in terms of the root mean square error skill score (*rmsess*, in %), the Pearson correlation coefficient (r) and the p-value of the Pearson correlation coefficient (*pval*, in %). Also given are the Weather Bureau Army Navy (WBAN) station numbers.

Name	***Station number***	***rmsess***	***r***	***pval in %***
Atlanta	13874	36.9	0.78	0.07
Dallas/Fort Worth	03927	31.8	0.73	0.19
Houston	12960	47.0	0.85	0.01
Little Rock	13963	24.3	0.65	0.82
Jacksonville	13889	19.9	0.60	1.78
Raleigh/Durham	13722	28.4	0.70	0.36

List of Figures

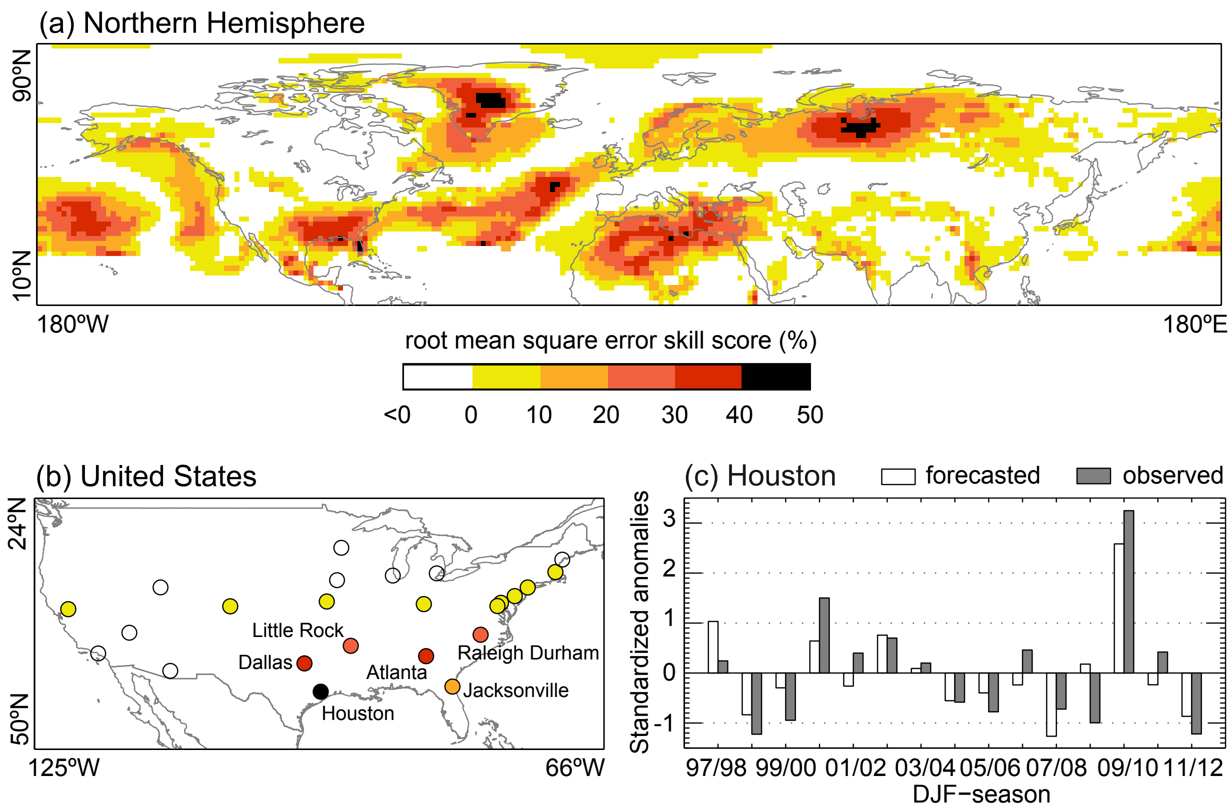

Fig. 1. **Panel (a):** One-month lead forecast skill for DJF accumulated heating degree days, calculated upon 2m air temperatures from ERA-Interim, in terms of the root mean square error skill score, the latter calculated with respect to the climatological mean. Results are from a one year-out cross-validation for the DJF-seasons between 1997/98 and 2011/12 (n=15). **Panel (b):** as (a) but using in-situ station data. The forecast skill for the six named stations is significant in terms of the p-value of the Pearson correlation coefficient (see Tab. 1 for more details. **Panel (c):** Forecasted vs. observed DJF-HDD for Houston (Texas).