Evaluation of WRF-Forecasts Over Siberia: Air Mass Formation, Clouds and Precipitation

Debasish PaiMazumder[1], David Henderson[2] and Nicole Mölders[*,2]

[1]*National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA*

[2]*University of Alaska Fairbanks, College of Natural Science and Mathematics, Department of Atmospheric Sciences, and Geophysical Institute, 903 Koyukuk Drive, P.O. Box 757320, Fairbanks, AK 99775-7320, USA*

Abstract: The Weather Research and Forecasting (WRF) model was run as a regional model without data assimilation or nudging (31 36h-simulations) for July and December 2005 over a limited area domain covering Siberia to examine weather formation in an air-mass source region. The WRF-results were compared to NCEP1/NCAR-reanalysis, International Satellite Cloud Climatology Project, Global Precipitation Climatology Centre and Canadian Meteorological Centre data to assess model performance and identify shortcomings. WRF is capable of predicting air-mass formation. Simulation errors are within the error range of other models. The timing of best/worst agreement differs among quantities depending on their sensitivity to systematic (model deficiencies) and/or unsystematic errors (e.g. initial conditions). Overall, the WRF-results agree better with reanalysis for July than December. WRF-results and reanalysis agree best under persistent high pressure and worst during frontal passages and transition from one pressure regime to another. In July, WRF provides smaller diurnal amplitudes of 2m-temperature with up to 5.4 K lower, and 3.5 K higher values at 0000 and 1200 UTC than the reanalysis. In December, WRF overestimates 2m-temperature by 1.4 K. WRF-temperatures excellently agree with the reanalysis from 700 hPa to 300 hPa. Except during frontal passages, wind-speed shows positive bias. Typically root-mean-square errors and standard deviation of errors of wind-speed (temperature) increase (decrease) with height. In December, WRF has difficulty predicting the position and strength of the polar jet. WRF underestimates cloudiness and snow-depth, but overestimates precipitation. In July, predicted convective precipitation is related strongly to boundaries between different land-cover. WRF-predicted snow-depth strongly correlates with terrain and misses the observed fine features.

1. INTRODUCTION

The scientific community has developed the next generation numerical weather prediction model called the Weather Research and Forecasting model (WRF [1]) for applications ranging from meters to thousands of kilometers based on the best scientific knowledge and experience gained from WRF's predecessor models. A useful forecast model should be able to provide results that agree well with known analytical solutions, results of other independently developed models, reanalysis, analysis and observational data; it must conserve energy, mass and moisture; and its code and model description must be freely available [2, 3]. WRF fulfills these criteria [4-9].

Any model intended to be a community model, in addition, needs to be applicable in the range of its intended use, i.e. it is important to know the application limits. The model must provide reliable results in other regions than it was developed for, no matter of climate, terrain and prevailing ecosystems. General applicability does not

contradict efforts to extend a model for special purposes (e.g., as a regional climate model) or to optimize a model for a region, as is done, for instance, in case of the Antarctic Mesoscale Prediction System [10].

First steps to examine WRF's transferability to other regions showed that when applied to the Baiu front over Japan WRF-simulated and observed rainfall distribution agreed broadly, and that WRF better captured the position of heavy rainfall than another well-evaluated mesoscale model [11]. Simulated and observed precipitation and cloud distributions agreed well when applying WRF for simulating a pre-monsoon thunderstorm over the east coast of India and a weak cyclonic circulation associated with low-pressure over south India [12].

The goals of this paper are to examine WRF's ability to forecast the formation of air-masses, cloudiness, precipitation, frontal structure, and to identify need for future model improvement. Calm winds with stagnant, extremely stable conditions in winter or calm winds in summer are prerequisites to produce an air-mass of relative horizontally homogeneous characteristics by prolonged contact (days to weeks) with the underlying surface. Such conditions are extremely challenging to forecast for any limited area model, when these stagnant winter highs or summer lows have to develop in the model domain itself and are not advected into the model domain over the lateral boundaries. The ability to

*Address correspondence to this author at the University of Alaska Fairbanks, College of Natural Science and Mathematics, Department of Atmospheric Sciences, and Geophysical Institute, 903 Koyukuk Drive, P.O. Box 757320, Fairbanks, AK 99775-7320, USA; Tel: +1-907-474-7910; Fax: +1-907-474-7379; E-mails: molders@gi.alaska.edu, cmoelders@alaska.edu

forecast adequately the formation of air-masses is required for any regional meteorological weather-forecast model or the meteorological part of any regional climate or Earth System model that is to be applied over air-mass source regions like Canada or Siberia.

For our study, we performed WRF-simulations without nudging or data assimilation over Siberia for July 2005 and December 2005 (i.e. 31 36h-simulations for each month). We assessed WRF's performance in simulating air mass formation, clouds and precipitation by means of the National Center for Environmental Prediction and National Center for Atmospheric Research reanalysis version 1 known as NCEP1/NCAR-reanalysis (NNR; [13,14]), International Satellite Cloud Climatology Project (ISCCP), Global Precipitation Climatology Centre (GPCC) and Canadian Meteorological Centre (CMC) data. The NNR-data were chosen over analysis data as they use all available quality controlled and quality assured (QA/QC) observational data. Using reanalysis and climatology data for model assessment is common practice in modeling studies of the mesoscale-α or larger [15,16].

Siberia was chosen as a test-bed as it is one of the world's high-latitude air-mass source regions for continental polar (cP) air masses. Siberia's hydro-climatic and insolation conditions differ strongly from those in the mid-latitude US. Broadleaf and coniferous forests dominate; wetlands and tundra cover Siberia that is widely underlain by warm permafrost soils. Siberia faces calm winds with stagnant, strongly stable conditions in December and free convection in July. The radiation and energy-budget conditions of the Siberian dark days and white nights are quite different from those in mid-latitudes. In December, northern and southern Siberia experience 0-5 h and 5-9 h of daylight, which is at least 8 h less daylight than in July.

2. MODEL SET-UP AND INITIALIZATION

We used eta-coordinates ($\eta=(p-p_T)/(p_s-p_T)$ where p is the pressure at a given atmospheric level, p_s is the surface pressure, and p_T is the pressure at the top of the model atmosphere). The eta-levels were at 0.9965, 0.9865, 0.973, 0.958, 0.9415, 0.923, 0.9025, 0.8805, 0.8565, 0.83, 0.801, 0.7695, 0.7355, 0.699, 0.6595, 0.6175, 0.573, 0.5255, 0.476, 0.4245, 0.3715, 0.3175, 0.263, 0.212, 0.1665, 0.1265, 0.0915, 0.0605, 0.0335, and 0.0105, i.e. the 31 layers increased with height. The model domain was centered over 60 ^{o}N, 90 ^{o}E and encompassed the atmosphere over Siberia from the surface to 50 hPa with 70×150 grid-points of 50 km grid-increment (Fig. **1**). This grid-increment is of the order of the grid-increment for a potential regional Earth System Model framework for Arctic applications. This magnitude of grid-increment was also chosen in many downscaling applications in climate research [17,18] and as an intermediate parent grid-increment for further dynamic downscaling to grid-increments even smaller than 10 km [19]. This magnitude of grid-increment has been used for regional reanalysis for the parent domain that holds child-domain(s) with smaller grid-increments [20]. Since the quality of forecasts affects the quality of any smaller scale model driven with the forecasts, it is important to have high accuracy at the parent-scale as well.

Cloud and precipitation-formation processes on the resolvable scale were determined with a six-water class microphysics scheme [21]. This scheme predicts the mixing ratios of water vapor, cloud-water, rainwater, ice, snow, and graupel at each time step. It considers, among other things, sedimentation of ice, co-existence of ice and super-cooled water, and gradual melting of falling snow. Since with respect to the grid-increment cumulus convection is of subgrid-scale, cumulus clouds were considered by a cumulus-ensemble-mean approach [22]. This approach determines the mean impact from a simulated cumulus cloud ensemble and permits subsidence effects to spread to neighboring grid columns. The calculation of long-wave radiation considered multiple bands, trace gases, and microphysics species [23]. Shortwave radiation was calculated by the Goddard scheme that considers 11 spectral bands for both diffuse and direct solar radiation [24]. Atmospheric boundary layer (ABL) physics were dealt with by the Mellor-Yamada-Janjić scheme in combination with the Eta-model surface-layer scheme that is based on similarity hypotheses, where the parameterization of the viscous sub-layer considers variable roughness height for temperature and humidity. A Mellor-Yamada level 2.5 turbulence-closure model served to simulate the full range of atmospheric turbulent regimes of the ABL and the free atmosphere; the differential equation for turbulent kinetic energy production/dissipation is solved iteratively [25]. The exchange of heat and moisture at the land-atmosphere interface was simulated by a modified version of the Rapid Refresh land-surface model [26]. The fluxes at the atmosphere-surface interface were determined for the dominant land-cover and soil type within a grid-cell assuming that the dominant land-cover and soil type are representative for the grid-cell (strategy of dominant land-cover, soil-type). The soil model has six layers with upper boundaries at 0, 0.05, 0.2, 0.4, 1.6, and 3 m depth. It solves the heat-diffusion and Richards' moisture-transfer equations and considers frozen soil physics in the prediction of soil-moisture and temperature states. The multi-layer snow model considers changing snow density, refreezing of liquid water percolating through the snowpack, and snow-depth. Snow albedo depends on both snow-depth and temperature. Melting of snow can occur at the snow-atmosphere and snow-soil interface. Grid-cells can be partially snow-covered if the grid-averaged skin-temperature is above freezing. In this case, calculations are made separately for a snow-covered and snow-free part. The resulting fluxes and state variables will be averaged according to their fractional coverage of the grid-cell.

The 1 km resolution state soil geographic database and 10 min resolution terrain and vegetation data were applied for soil-texture, land-cover type and terrain elevation, respectively. The land-cover and soil-type datasets distinguish 24 land-cover and 16 soil-types, respectively.

Initial and boundary conditions stemmed from the $1^{o}\times1^{o}$, 6 h resolution global final analyses (FNL) that are sponsored by NOAA/NWS within the framework of the Global Forecast System. No data assimilation was applied for initialization. No nudging was applied during the simulation because nudging "pulls" the model towards the observed situation, and would make it difficult to assess
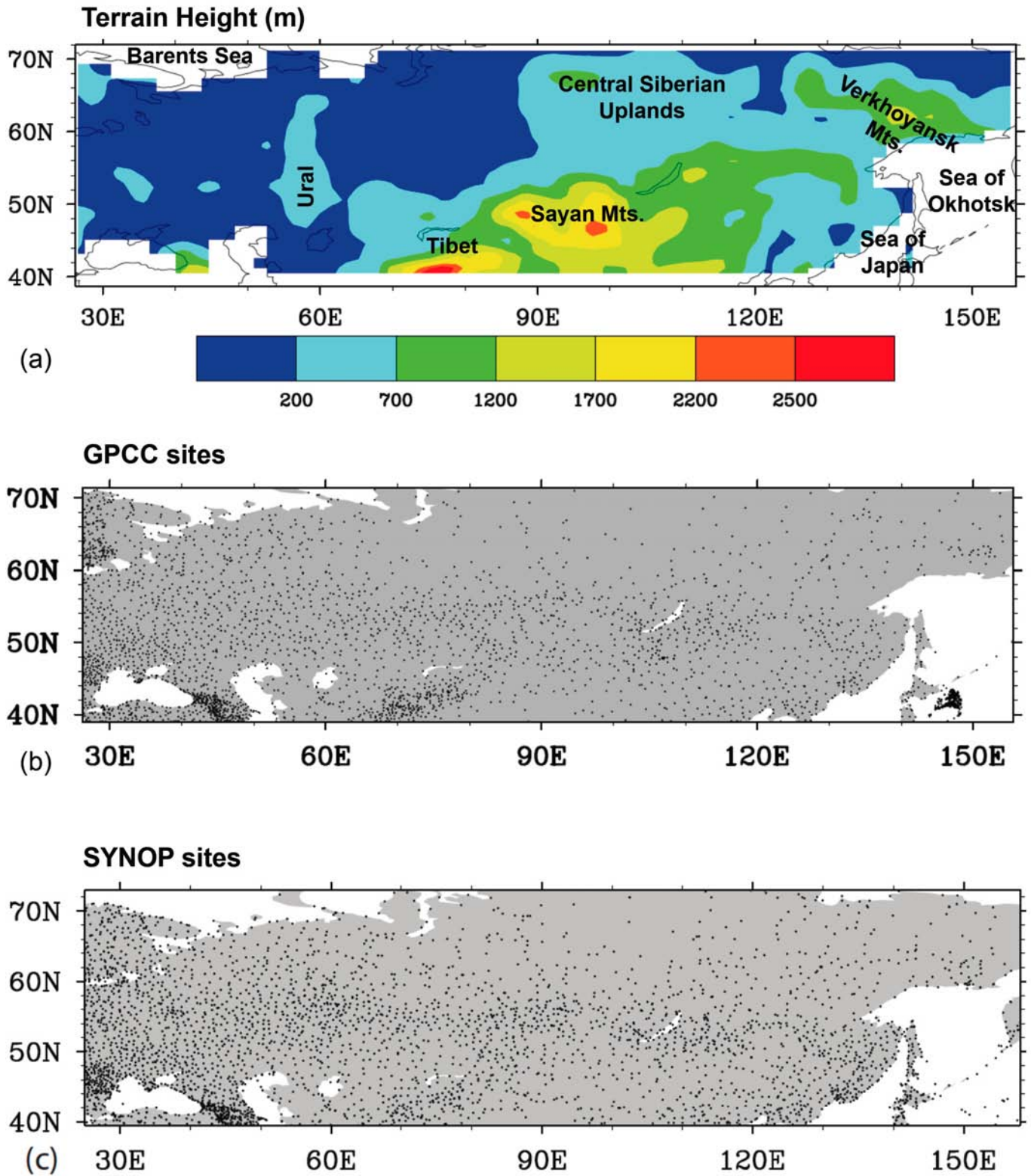
Terrain Height (m)

(a)

GPCC sites

(b)

SYNOP sites

(c)

Fig. (1). (**a**) Model domain and terrain elevation, (**b**) location of precipitation and (**c**) synoptic (SYNOP) observational sites. Names refer to locations mentioned in the text.

whether WRF is able to describe features like air-mass formation that build within the domain.

The FNL-data provided the initial snow-depths, soil temperatures and moisture conditions. WRF was run without any nesting for each day of July and December 2005 for 30 h. The first 6 h were discarded for spin-up, and the 6–30 hour periods were used in the comparison with the NNR, ISCCP, GPCC and CMC data.

3. DATA

Various centers (e.g., European Center for Medium Range Weather Forecast [ECMWF], NCAR, NCEP, Japan Meteorological Agency) have produced reanalysis data [13, 27-29]. In reanalysis, quality assessed/quality controlled observational data from various sources (e.g., synoptic stations, radiosondes, aircrafts, satellites) provided by different countries and organizations are assimilated by a modern analysis/forecast system to obtain a physically and thermodynamically consistent global reanalysis of atmospheric fields (See the aforementioned literature for details).

The NNR-data effort involved the recovery of land-surface, ship, rawinsonde, aircraft, satellite, and other observations. All observations were run through a QA/QC before being assimilated with a data assimilation system kept unchanged over the reanalysis period [13]. The 3D-VAR assimilation system has a triangular truncation of 62 waves corresponding to about 210 km horizontal resolution. It consists of the NCEP global spectral model operational in 1995 with 28 vertical sigma levels. The assimilation procedure provided a gridded dataset based on all available high-quality observations of temperature, horizontal wind and specific humidity from upper air rawinsondes, operational TIROS Operational Vertical Sounder vertical temperature soundings over ocean, temperature soundings over land, cloud-tracked winds derived from geostationary satellite observations, aircraft observations of wind and temperature, and surface reports of surface pressure, temperature, horizontal wind, and specific humidity (SYNOP data).

The main advantages of reanalysis data is that they are gridded products available every 6 h and include more quality-assured/quality-controlled observations than individual observational datasets or analysis. Reanalysis data fill in large data voids (e.g. those in the SYNOP-station network; Fig. **1**). The temporal resolution exceeds the frequency of many routine observations in Polar Regions, especially during winter.

The various reanalysis products differ only marginally for pressure, temperature, relative humidity and wind-speed [28]. The short-term variability among ECMWF's ERA40, Climatic Research Unit (CRU) [30] and NNR data is very similar [27]. ERA40 and CRU data show high correlation and no statistically significant differences [31]. Correlation between NNR and ERA40 data is ~0.99 globally and for Siberia and differences are statistically non-significant. In comparison with CRU, ERA40 near-surface temperatures show positive bias over the Northern Hemisphere in winter. On average, bias is less than 3 K over Siberia, and little in summer; the bias between NNR and ERA40 near-surface temperatures is -0.16 K for Siberia, i.e. NNR and ERA40 provide similar results [32]. Comparison of the ERA40, NNR, NCEP2 and JRA-25 global reanalysis in Polar Regions showed large differences for hydrological components and cloud variability. NCEP1 and NCEP2 data differ mainly with respect to snow-cover [33].

Due to the non-significant differences among reanalysis products for pressure, temperature, and wind-speed over Siberia, we restricted our discussion to a comparison of

WRF-simulated pressure, temperature, and wind-speed with NNR data of these quantities. Due to the large differences among reanalysis products for hydrological components, we assessed WRF-simulated clouds, precipitation and snow-depth by the ISCCP, GPCC and CMC data, respectively. The ISCCP-data provide monthly cloud-cover fraction on a $2.5^o \times 2.5^o$ grid. ISCCP total cloud-cover fraction is the percentage of pixels identified as cloudy within $2.5^o \times 2.5^o$ [34]. We used the $0.5^o \times 0.5^o$ resolution GPCC monthly precipitation data that is based on quality-controlled and homogenized time-series from 50721 stations worldwide [35], 3309 of them are in the model domain (Fig. **1**). This number of sites is sufficient to provide reliable gridded averages [36]. The CMC-data provide monthly snow-depth on a $0.5^o \times 0.5^o$ resolution.

4. ANALYSIS

Each NNR grid-cell encompasses various WRF grid-cells and more grid-cells in the South than North of the model domain. We aggregated the WRF-data of temperature, pressure, and wind-speed to the $2.5^o \times 2.5^o$ resolution of the NNR-data. In doing so, we averaged all WRF-simulated values that fall into a given NNR grid-cell and applied a Cressman-type method as described in [37]. WRF grid-cells that partly fall into more than one NNR grid-cell were considered according to their area-weight. Since the NNR-data are available at 0000, 0600, 1200, and 1800 UTC, we used the WRF-output at these times. The aggregated values of WRF-simulated sea-level pressure (SLP), wind-speed and temperature were compared with the NNR-values. In both WRF and the reanalysis, the barometric equation serves to determine SLP. We aggregated the WRF-precipitation and snow-depth on the $0.5^o \times 0.5^o$-grid as described above and determined the monthly accumulated value for each $0.5^o \times 0.5^o$-grid for comparison with the GPCC- and CMC-data. The snow-depth values obtained by WRF are the initial snow-depth plus the snow accumulated by WRF-simulated snowfall. We calculated the monthly average snow-depth using the daily snow-depth values obtained by WRF.

None of WRF's cloud parameterizations and cumulus schemes predicts fractional cloud-cover [1]. Therefore, we compared WRF-simulated cloud existence and ISCCP cloud fraction following [38]. If WRF-simulated cloud and/or ice-mixing ratios exceeded 0.001 g/kg for any grid-cell of a grid-column, clouds were assumed to exist in the grid-column. This threshold was determined from independent sensitivity studies that compared cloud existence determined with various thresholds to Moderate Resolution Imaging Spectroradiometer (MODIS) data [39]. We determined cloud-fraction for each $2.5^o \times 2.5^o$-grid-cell by considering the fractional area within each $2.5^o \times 2.5^o$-grid-cell that was identified to have cloud-existence in WRF, dividing it by the total $2.5^o \times 2.5^o$-grid-cell area and multiplying by 100. Monthly cloud fraction was determined by averaging the WRF-derived cloud fractions. We compared the so-determined WRF-cloud fraction to the ISCCP-data.

Theoretically, systematic and unsystematic errors can contribute to discrepancies between WRF and the data to

which we compared them. Gridded observational data only approximates actual conditions. Discrepancies between simulations and such observations are partially due to different spatial scale representations. Thus, if the time-series of simulated daily quantities is highly correlated with an observational estimate, the model performance is impressive, even when there are magnitude errors. For instance, forecasting the transitions between clear and cloudy days would be enough of an achievement.

To assess potential sources for discrepancies we calculated performance skills following [40]. In our experimental design, bias quantifies systematic discrepancies from consistent differences in numerical, geometrical, and/or physical factors. In the comparison of WRF and reanalysis data, such discrepancies can result from differences of terrain height, discretization, grid-resolution, parameterizations, land-cover and soil parameters, and the aggregation method (for a discussion see [36]). The standard deviation of error (SDE) quantifies unsystematic errors from uncertainty in initial and boundary conditions and/or measurement errors [41]. Since any quantity calculated with observations is "error"-burdened too [42], errors can be inherent in the reanalysis due to the assimilation of observations, in the ISCCP, GPCC and CMC-data from interpolation of observations and in the WRF-data from boundary and/or initial conditions (e.g., soil temperature and moisture, sea-surface temperatures, snow-cover). However, observation errors are likely marginal. Secondary errors result from propagation of primary errors into other quantities and contribute to the overall error that was assessed by the root-mean-square error (RMSE).

Since diurnally varying differences may cancel each other out [43], we determined skill-scores for 0000, 0600, 1200, and 1800 UTC each. Siberia covers several time zones. Thus, these times correspond to 0200-0900, 0900-1600, 1500-2200, and 2100-0400 local time, respectively. We used UTC for both the simulations and the reanalysis for consistency of time.

We tested the hypothesis that WRF-data do not differ significantly (at the 95% or higher confidence level) from the NNR, GPCC, ISCCP and CMC-data according to Student t-tests. Herein, the t-value is given by $t = (X - \mu)/(\frac{s}{\sqrt{n}})$ where X, s and n are the mean, standard deviation and number of data. Once the t-value is calculated, a p-value can be found using a table of values from Student's t-distribution. In the following, the term "*significant*" is only used when the p-value is below 0.05, i.e. the computed probability that data are significantly correlated exceeds 95%.

5. SURFACE PRESSURE

Real forecasts would have GCM-predictions as lateral boundary data. Within the framework of a limited area Earth System Model, the lateral boundary conditions would stem from a General Circulation or other global model. In our study, WRF is driven by the FNL-analysis, which is an observational dataset. Thus, WRF downscales the FNL-data from 1°×1° to ~0.5°×0.5°, which is one aspect of why the forecasts are highly correlated with the reanalysis. However,

WRF needs to have realistic thermodynamics to create the highs and lows within the domain.

WRF-simulated and NNR-pressure differ non-significantly. The same is true for WRF-simulated and analysis-pressure. No significant effect of the forecast length can be detected for all pressure skills. Our finding is consistent with the findings by [44]. This author used over 30 120 h WRF-simulations and observations and found only marginal, but non-significant decrease in performance with increasing forecast length.

Synoptic-scale changes that are not imported over the lateral boundaries cannot be simulated unless they have their source within the domain. WRF is able to predict the formation of the high and low-pressure cell in the domain in December and July, respectively. In December, the source of high-pressure systems is subsidence with calm wind, stagnant, stable, cool and dry conditions over the Siberian snow-cover in response to low insolation and a negative radiation balance. In July, the lows form in response to long insolation, calm wind over homogenous flat surface that allow for warm, moist conditions and free convection.

Compared to reanalysis, WRF, on average, overestimated SLP with slightly higher bias in July than in December. In both months, WRF captured the distribution and trends of SLP well (Fig. **2**). In July, the WRF and reanalysis data correlate between 0.92 and 0.93 with 0.92 on average. In December, correlation is slightly less (0.9-0.93, 0.91 on average). In July (December), the diurnal cycle caused pressure-bias to vary between 3.8 and 6 hPa (3.9-6.8 hPa) each day with 5.1 hPa (5.5 hPa) on average. Errors in temperature (see section 7) can have notable impact on pressure *via* the equation of state, while errors in specific humidity affect pressure marginally. Especially over mountain chains, some pressure-bias results from the differences of terrain height and complexity between WRF and the reanalysis.

For July (December), RMSEs and SDEs between WRF and reanalysis-surface pressure ranged between 22.5 and 23.2 hPa (24.9-26.0 hPa) and 21.8 and 22.7 hPa (24.3-25.3 hPa), respectively. The July (December) mean RMSE and SDE were 22.7 (25.3) and 22.2 (24.8) hPa, respectively (Fig. **2**). RMSEs and SDEs and their extremes were higher for December than July because high pressure dominated the synoptic situation in December. In December, the Siberian High is very strong and gradients between the high and the adjacent lows are large. In the simulations, these lows enter the model domain over the lateral boundaries, while WRF has to forecast the high in response to the regional forcing occurring within the domain.

Peaks in bias, RMSE and SDE exist during transitions or frontal passages (Fig. **2**). The SDE is about four times higher than the bias which indicates that WRF and reanalysis-SLP differ partly due to unsystematic errors. Unsystematic errors stem from measurement errors and hence are inherent in the FNL- and reanalysis-data. However, the bias and RMSE suggest additional error due to model deficits and discretization.

WRF-derived monthly average snow-depth and extent differed notably from the CMC observations (Fig. **3**). The too small (initial) snow-cover extent explains some of the
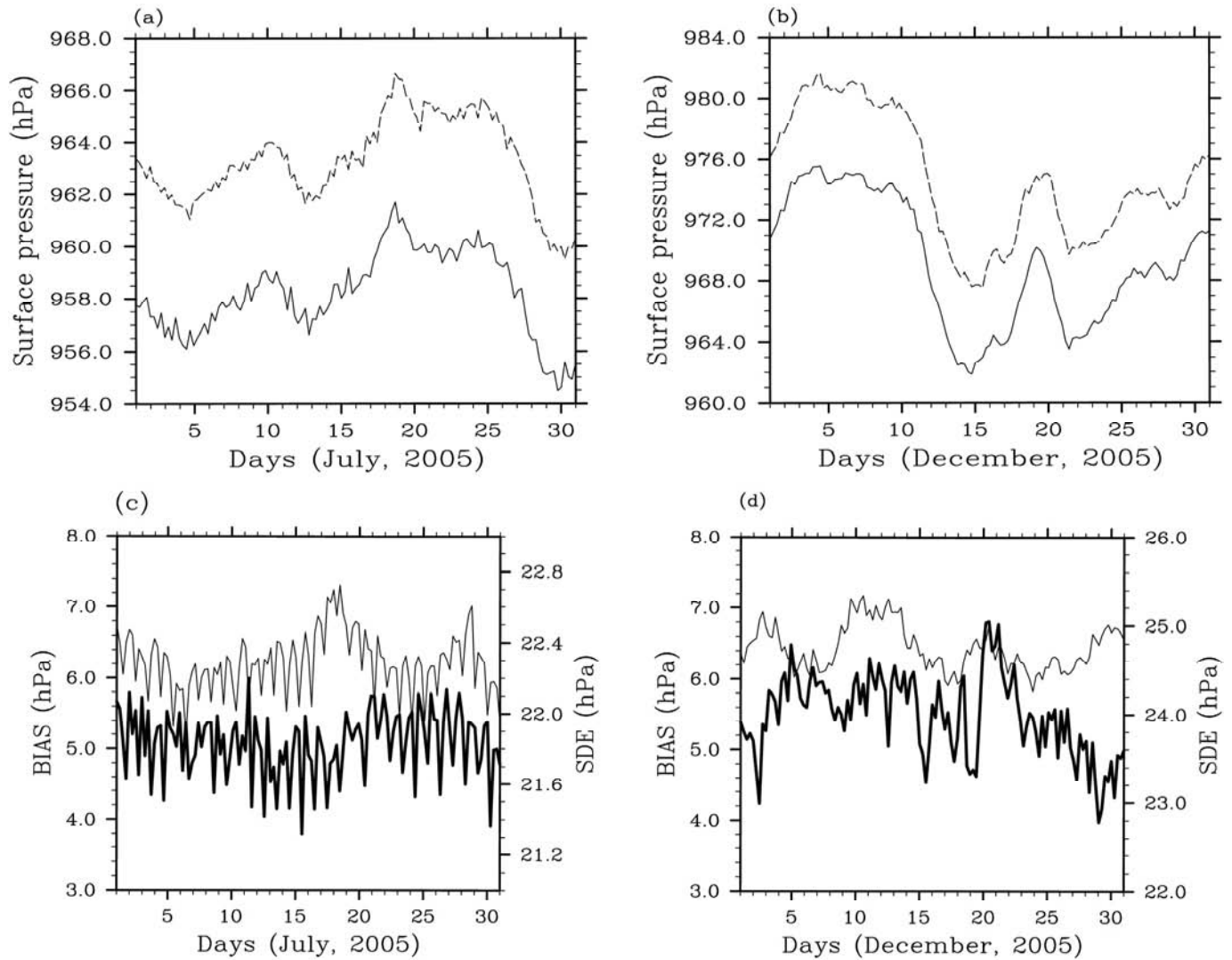
Fig. (2). Temporal evolution of domain-averaged surface pressure, bias and SDE (hPa) for July (**a**), (**c**), and December (**b**), (**d**). In (**a**) and (**b**) the dashed and bold lines represent the WRF-simulated and reanalysis values, respectively. In (**c**) and (**d**) the bold and thin lines represent bias and SDE between WRF-derived and reanalysis values. Bias and SDE have to be read at the left and right Y-axis, respectively.

about 5.5 hPa discrepancy between WRF and the reanalysis SLP in December. MM5-simulations performed for Alaska without and with consideration of snow showed that changes in atmospheric humidity and temperature caused by snow propagate into the pressure field; the presence of snow may reduce SLP by more than 1 hPa; the effect of permafrost increases the pressure difference to 1.2 hPa for the permafrost-snow system compared to no-snow-no-permafrost [45]. In a statistical study, [46] identified comparable effects of snow on East Coast cyclones as found in our study. The enhanced December-RMSE and SDE over the mountain chains result from errors in initialization of snow and simulated snow-depth (Fig. **3**; see section 3.5). Errors in initialization of permafrost and active layer depth may contribute to some of the 5.1 hPa overestimate in July when snow only exists at high elevation. Unfortunately, the Siberian soil-temperature network encompassed only agriculturally used areas [36] and currently gridded soil-temperature data derived from this network exist only from the 1800s through 1990 [47]. Thus, this hypothesis cannot be tested for our simulations.

The land-sea masks of WRF and the reanalysis differ. Since water and land temperature differ more strongly in December than July, RMSE and SDE are higher for December than July along the Barents Sea coast west of the Ural.

6. WIND

In December, a persistent high pressure system led to calm winds, i.e. less variability than in July when convection initiates gusts (Fig. **4**). WRF-simulated wind-speeds differ non-significantly from the NNR wind-speeds or analysis. Again, no significant effect of the forecast length on the performance skills exists. In both months, the temporal evolution of WRF and reanalysis agree well for 10 m wind-speeds and acceptably for wind-speeds at various pressure levels. While WRF captures the trend of 10 m wind-speed better in December (r=0.78) than in July (r=0.60; Fig. **4**) the opposite is true at the various pressure levels (over all pressure levels r=0.57 in December, r=0.82 in July). During low pressure or frontal passages WRF-simulated 10 m wind-
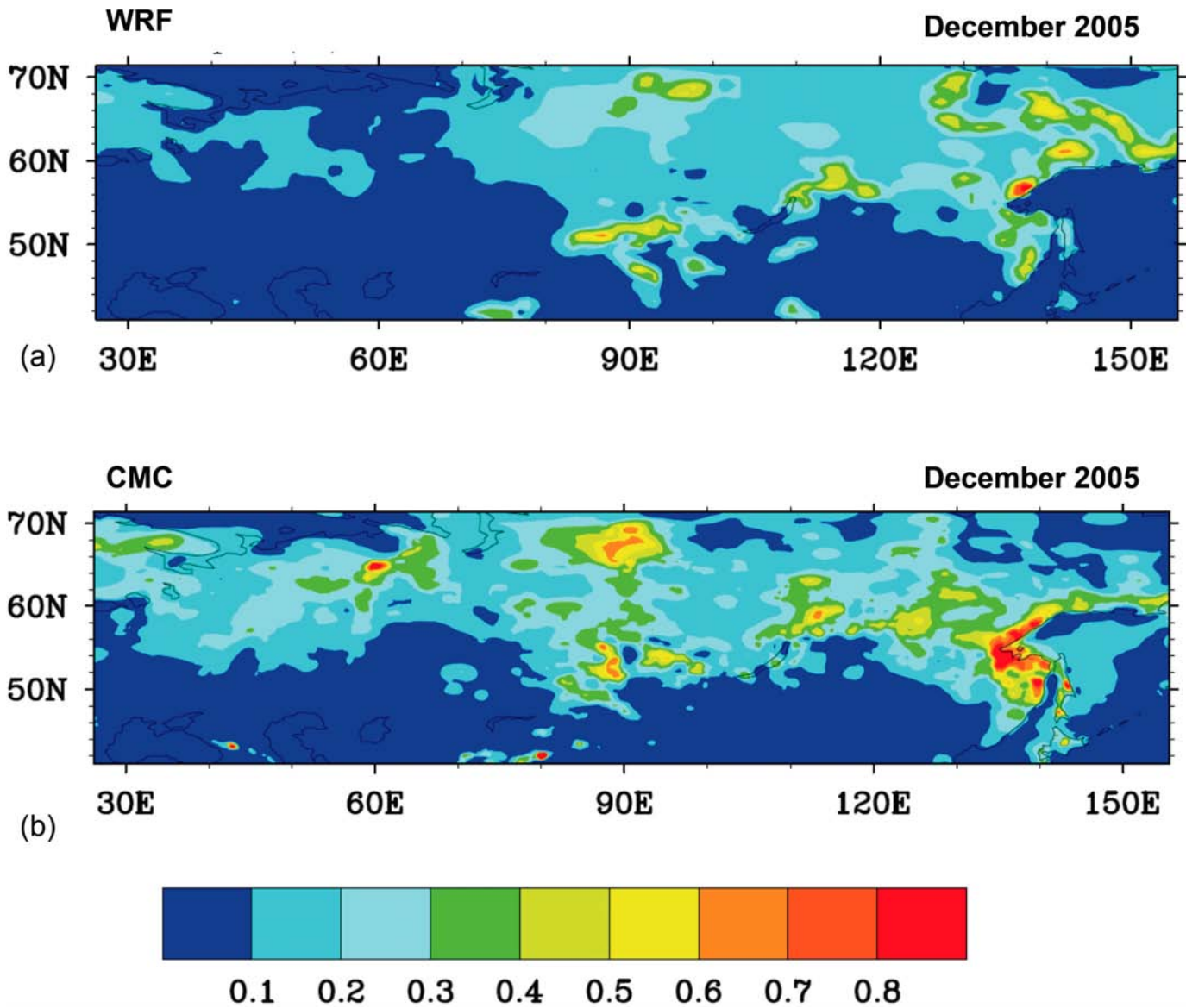
Fig. (3). Monthly average snow-depth as obtained from (**a**) WRF and (**b**) the CMC data.

speeds exceed those of the reanalysis leading to low correlation, especially when timing is slightly off. Some error results from the aggregation. When a front runs ahead/behind, aggregation to a coarser grid can place that front into a neighbored $2.5^o \times 2.5^o$ grid-cell producing seemingly large errors, especially in the southern part of the domain. Here, 2.5^o cover a larger west-east distance than in the North.

During frontal passages in December, huge differences in wind-speed occur at all pressure levels (Fig. **5**) resulting in daily correlation-skill scores between 0.09 and 0.51. Wind-speed is overall positively biased in both months compared to reanalysis. However, during strong high pressure wind-speed is biased negatively for December. All these discrepancies indicate that WRF has difficulty to capture the pressure gradients.

WRF overestimates 10 m wind-speed at 0600, 1200, and 1800 UTC (mean bias < 0.36 ms^{-1}), while it often underestimates 10 m wind-speed slightly at 0000 UTC (e.g. Fig. **6**). WRF's 10 m wind-speed biases for Siberia compare well with other models and fall within the performance range of WRF over the US. When evaluating WRF over the western US, [6], for instance, found a tendency to overestimate 10 m wind-speed by 0.5 ms^{-1} during summer, and a diurnal variation of bias with afternoon maxima and nighttime minima. A MM5-evaluation for the Great Lakes region, for instance, showed mean bias of 0.36 and 0.50 ms^{-1} in summer and winter, respectively [43]. The biases of a 33 h simulations with the 29 km Eta model for Florida for the warm (cold) season [48] are higher than our biases of ±1 ms^{-1} ($\pm 2\ ms^{-1}$).

The quality of wind simulations decreases with increasing stability or free convection. In the ABL, WRF and reanalysis wind-speeds correlate highest at 0000 and 0600 UTC, and lowest at 1200 and 1800 UTC. Errors in WRF-simulated convection cause the high errors at 1200 and 1800
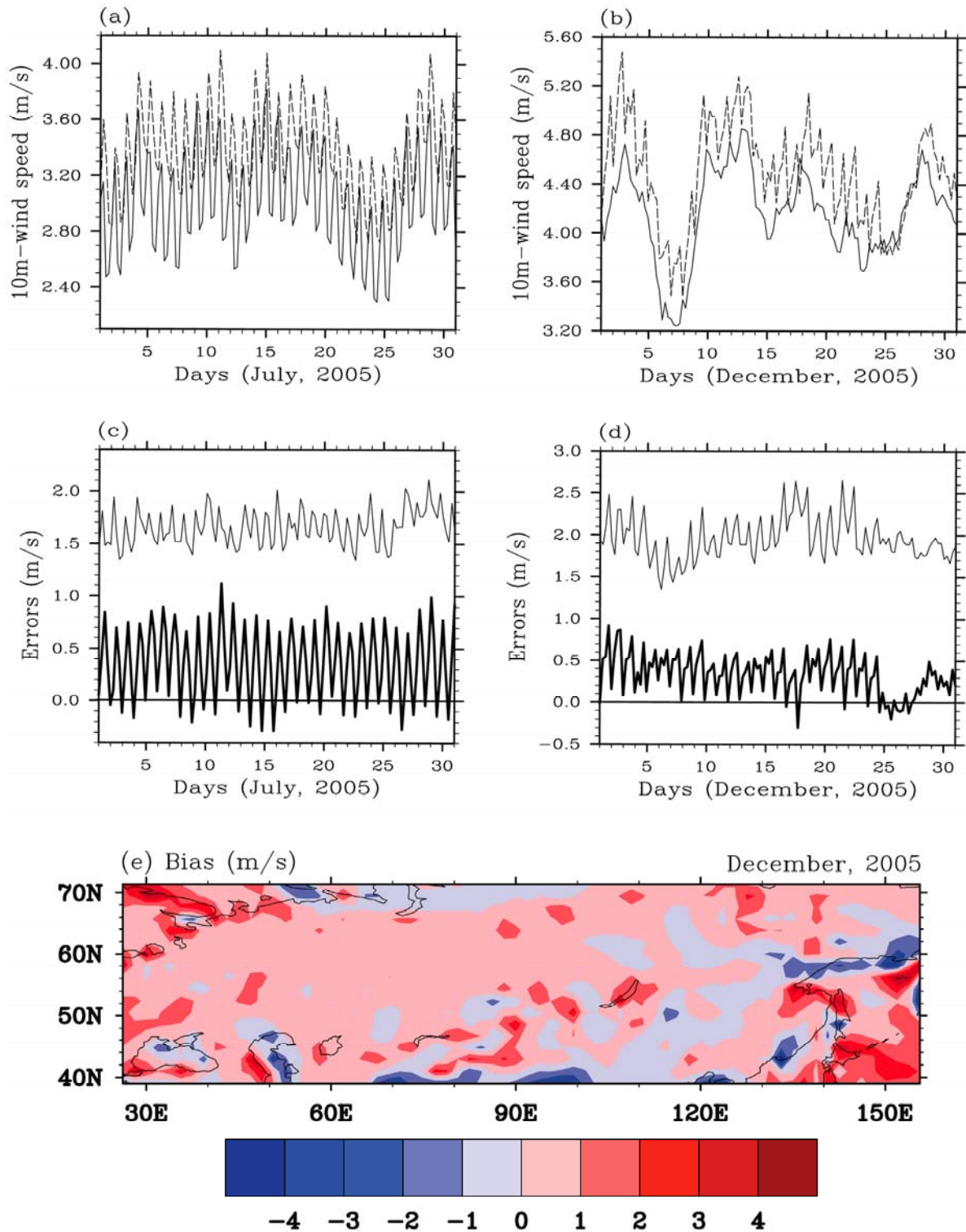
Fig. (4). Temporal evolution of domain-averaged 10 m wind-speed, bias, and SDE (ms^{-1}) for July (**a**), (**c**), and December (**b**), (**d**), and horizontal distribution of bias for December (**e**). In (**a**) and (**b**), the dashed and bold lines represent the WRF-derived and reanalysis values, respectively. In (**c**) and (**d**), the bold and thin lines represent bias and SDE between WRF-derived and reanalysis values. The zero-line is indicated for readability. Horizontal panels for RMSE and SDE look similar to those of bias (therefore not shown). Legends and Y-axes differ among panels.

UTC. As discussed in section 8, the distribution of land-cover that looks like being downscaled without any nature justified procedure (Fig. **7**), yields locally incorrect surface forcing. Convection simulated in the wrong places affects the wind field below 700 hPa. Nevertheless, wind-speed biases fall in the lower range of those reported by [49] for their January-February-March evaluation of the Eta-model, NGM, MRF, MESO, MM5, and ULAM.
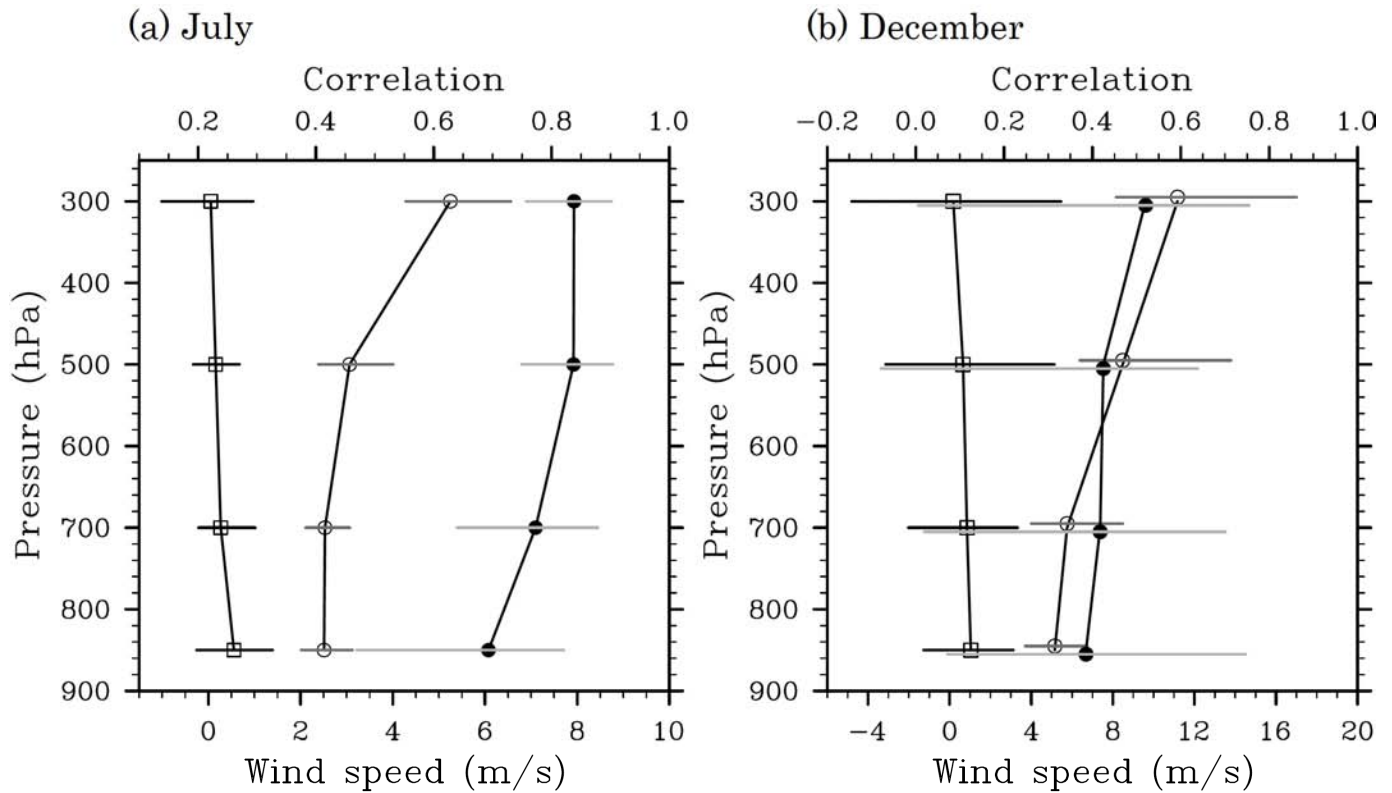
Fig. (5). Bias (line with squares), SDE (line with circles) and correlation (line with filled dots) between WRF-derived and reanalysis wind-speed at 850, 700, 500 and 300 hPa and their ranges shown as horizontal black, dark and light gray bars, respectively for (**a**) July and (**b**) December. Correlation has to be read at the top X-axis, bias and SDE at the bottom X-axis.
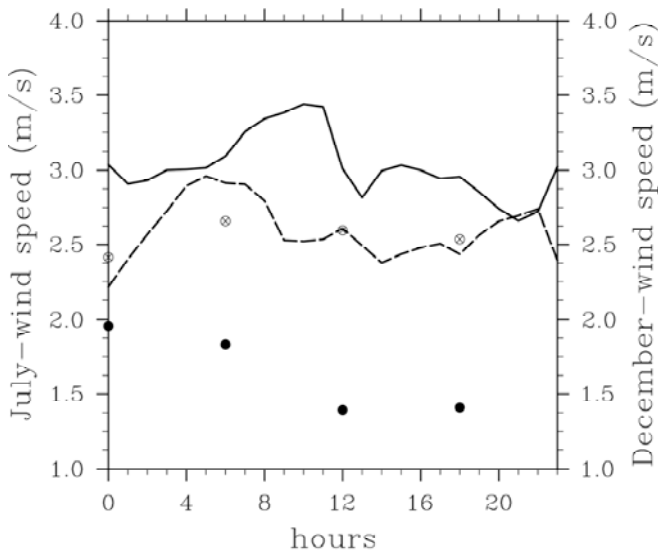

Fig. (6). WRF-derived monthly average diurnal cycle of 10m wind-speed (ms^{-1}) with respect to UTC for July (solid line) and December (dashed line) for the grid-cell centered around 58 N and 94 E. July and December averaged NNR-data are only available at 0000, 0600, 1200 and 1800 UTC and are given by the filled circle and circle with cross, respectively.

In both months, positive bias exists (up to 4 ms^{-1} in December) over the Barents Sea, Sea of Okhotsk and mountain chains. Comparison with the topography, land-cover and land-sea line and to published land-cover classifications based on remote sensing data analysis [50, 51]

identifies differences in landscape representation as a cause (Figs. **1**, **4**, **7**). Assumed surface roughness, resolution and the surface-layer parameterization can cause such discrepancies [52]. The surface parameters used in WRF and for creating the NNR-data differ. Surface roughness and turbulence are less in WRF than required to capture the friction and reduction of wind-speed towards the surface [53].

Throughout the atmosphere, mean RMSEs and SDEs were smaller in July than in December (Fig. **5**) and daily RMSEs and SDEs were lower during high than low pressure. For 10 m wind-speed SDEs reached up to 2.11 ms^{-1} (2.64 ms^{-1}) in July (December) during frontal passages, while SDEs decreased to 1.34 ms^{-1} during high-pressure episodes (Table **1**; Fig. **4**). Mean values of 10 m wind-speed RMSE and SDE (Table **1**) compared well to the summer and winter RMSEs of 1.75 and 1.93 ms^{-1} and July and December SDEs of 1.57 and 1.72 ms^{-1} reported for MM5 by [43].

In December, WRF simulated the polar-front position and strength incorrectly (Figs. **8**, **9**) which caused the larger RMSEs and SDEs in the mid and upper troposphere than in the lower troposphere (Fig. **5**). Related to the errors in position, fronts run ahead/behind and cyclones were misplaced slightly, that caused errors in temperature, relative humidity, cloudiness and precipitation. The comparison of the NNR and WRF-data showed that in December, the conditions at the northern and eastern lateral boundary differed notably. The lower wind-speeds at these boundaries contributed to the underestimation of the polar-jet strength. The failure to capture the strong horizontal pressure
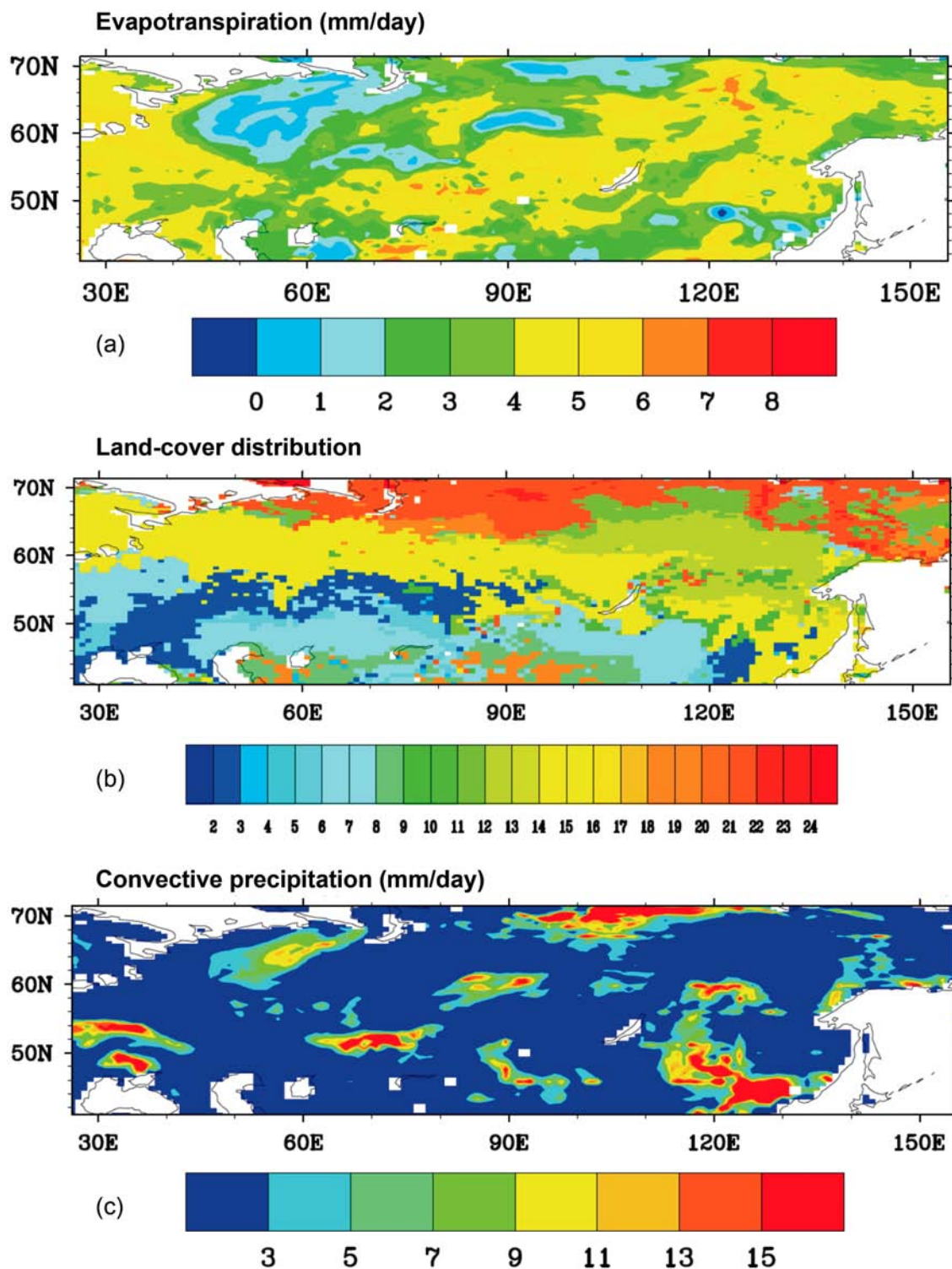
Evapotranspiration (mm/day)

Land-cover distribution

Convective precipitation (mm/day)

Fig. (7). (**a**) Evapotranspiration, (**b**) land-cover distribution and (**c**) convective precipitation as obtained for July 8, 2005. Distribution of evapotranspiration and convective precipitation on other convective July days look similar (therefore not shown). The coding for the land-cover categories is as follows: 1 urban and built-up land, 2 dryland, cropland and pasture, 3 irrigated cropland and pasture, 4 mixed dryland/irrigated cropland and pasture, 5 cropland/grassland mosaic, 6 cropland/woodland mosaic, 7 grassland, 8 shrubland, 9 mixed shrubland/grassland, 10 savanna, 11 deciduous broadleaf forest, 12 deciduous needleleaf forest, 13 evergreen broadleaf, 14 evergreen needleleaf, 15 mixed forest, 16 water bodies, 17 herbaceous wetland, 18 wooden wetland, 19 barren or sparsely vegetated, 20 herbaceous tundra, 21 wooded tundra, 22 mixed tundra, 23 bare ground tundra, 24 snow or ice.

gradients and full strength of the Siberian High contributed to the incorrect positioning of the jet in December. In July, the lateral boundary data of NNR and WRF agreed well and the strength of the polar jet was predicted well (Fig. **8**).

Table 1.　**Mean and Standard Deviation (±StDev) Given in Parenthesis, and Ranges of Bias, Root-Mean-Square Errors (RMSE), Standard Deviation of Error (SDE), and Correlation-Skill Score and the Mean of these Errors Given in Parenthesis for 10m Wind-Speed (v_{10}), and 2m-Temperature (T_2) for July and December 2005. The Mean Values Refer to Statistics Over the Entire Month while the Range Refers to the Statistics for the Individual 6-30h Simulation Data**

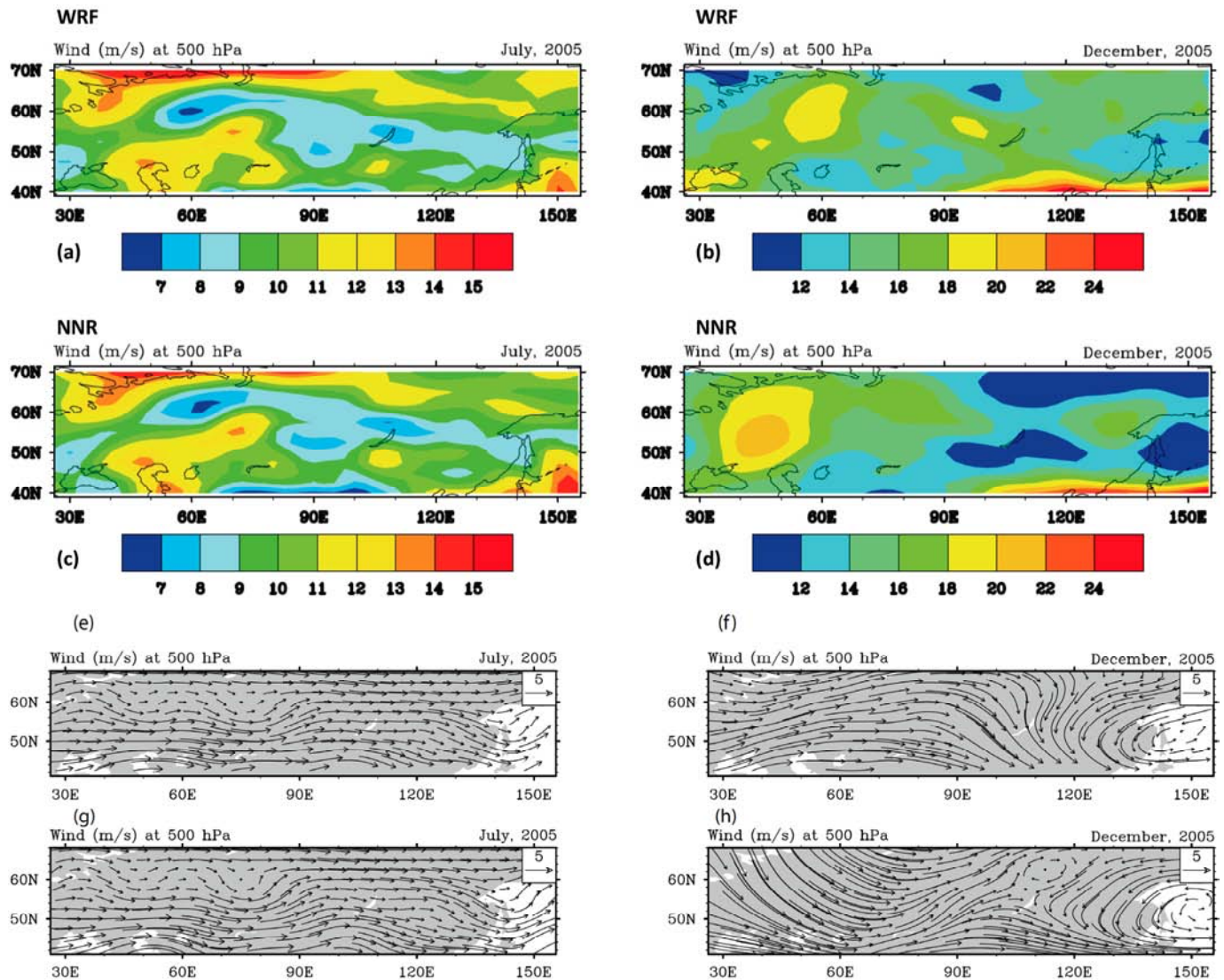|		Mean (StDev)	Bias Range (Mean)	RMSE Range (Mean)	SDE Range (Mean)	Correlation-Skill Score Range (Mean)
v_{10} (ms^{-1})	July	3.02 (±0.37)	-0.28 to 1.12 (0.36)	1.35 to 2.15 (1.73)	1.34 to 2.11 (1.66)	0.38 to 0.74 (0.60)
	December	4.36 (±0.37)	-0.29 to 0.91 (0.33)	1.44 to 2.73 (1.99)	1.34 to 2.64 (1.95)	0.58 to 0.90 (0.78)
T_2 (K)	July	290.6 (±2.7)	-5.4 to 3.6 (-0.5)	3.7 to 6.9 (4.9)	3.6 to 5.0 (4.1)	0.69 to 0.89 (0.81)
	December	258.9 (±1.8)	-0.8 to 3.5 (1.4)	4.0 to 6.3 (4.9)	4.0 to 5.0 (4.1)	0.89 to 0.95 (0.92)

Fig. (8). Monthly average wind-speed at 500 hPa as obtained by WRF (**a**), (**b**), and NNR (**d**), (**e**) for July (**a**), (**c**), and December (**b**), (**d**) and wind field as obtained by WRF (**e**), (**f**), and NNR (**g**), (**h**) for July (**e**), (**g**), and December (**f**), (**h**).

The increase of model-layer thickness with height caused the increase of RMSE with height, and exists for other models. Our upper level July-RMSEs were appreciably lower than those reported in [49] study, while our December-RMSEs were similar to their findings. July-RMSEs fell

within the 2 to 4 ms^{-1} range reported by [48] for the 29 km Eta model.

In December, SDE and RMSE, and in July, SDE of 10 m wind-speed were the smallest over central Siberia (not shown). December-RMSEs and SDEs and July-SDE were
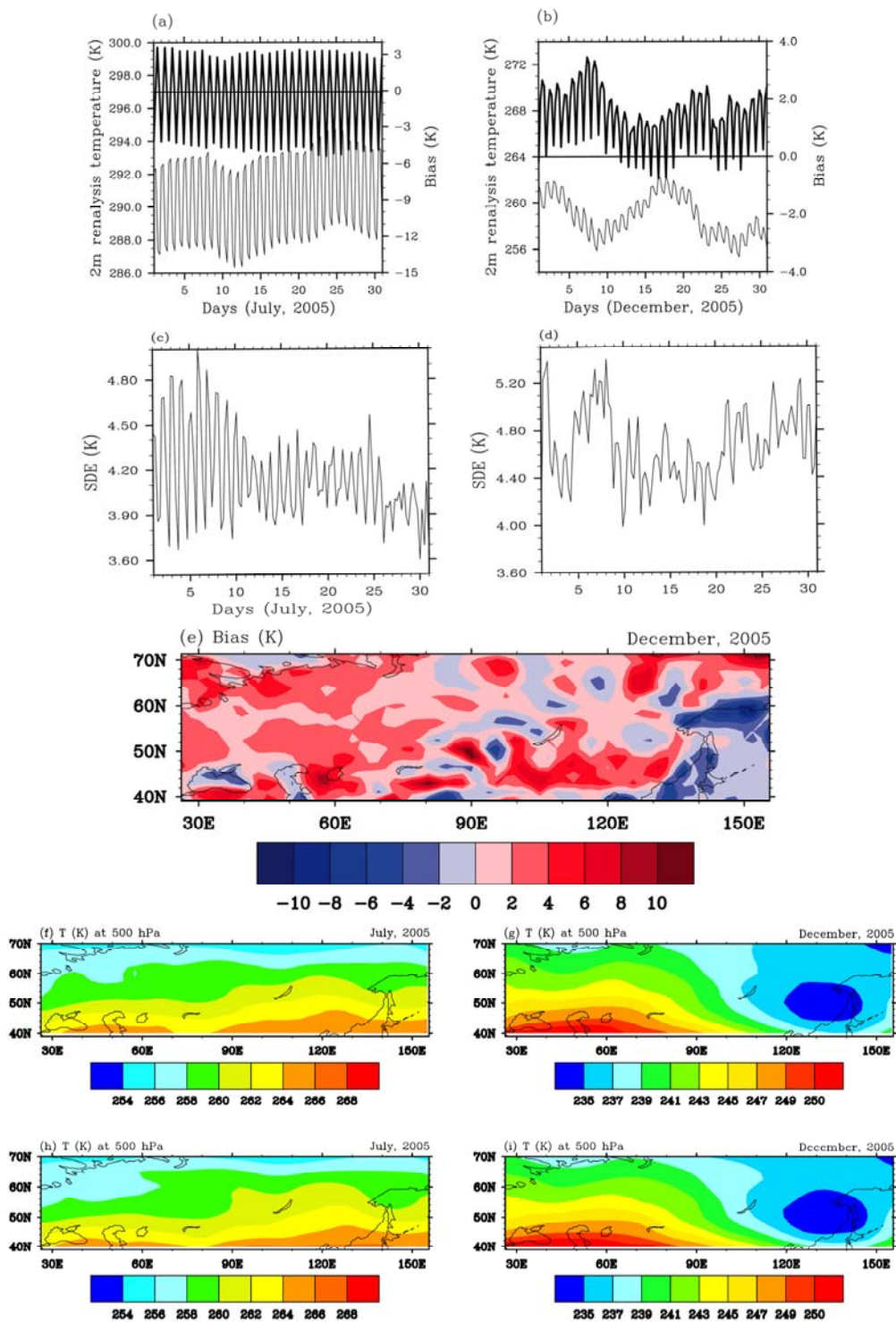
Fig. (9). Temporal evolution of domain-average reanalysis 2m-air temperature, bias, and SDE for 2 m air temperature (K) and SDE for July (**a**), (**c**), and December (**b**), (**d**), and horizontal distribution of bias for December (**e**), and of monthly average temperatures at 500 hPa as obtained by WRF (**f**), (**g**), and NNR (**h**), (**i**) for July (**f**), (**h**), and December (**g**), (**i**). Horizontal plots for RMSE look similar to those of bias and SDE (therefore not shown). In (**a**) and (**b**), the thin and bold lines represent the temporal evolution of domain-average reanalysis 2 m air temperature and bias. The zero-bias-line is superimposed. Domain-average reanalysis 2 m temperature and bias have to be read at the left and right Y-axis, respectively.

the greatest over the Sea of Okhotsk. At all pressure levels, skill-scores of wind-speeds showed no obvious horizontal pattern in July. However, in December, strong negative bias existed west of the Ural at all levels and decreased towards the surface. West of the Ural, RMSEs between WRF and reanalysis wind-speed were extremely high at all levels and reach about 17.5 ms^{-1} at 300 hPa. SDEs exceeded the absolute value of the biases up to four times in this region. The low skill in this area resulted from the incorrect position of the polar jet in WRF (Fig. **8**).

The low biases and comparatively great SDEs indicated that much of the total RMSEs of wind-speed resulted from unsystematic errors (e.g. observational uncertainty, initial and boundary conditions). Inhomogeneity exists in radiosonde-wind data due to different wind-finding systems. Differences were in the order of 0.5 ms^{-1} but can reach up to 3 ms^{-1} in remote areas like the Kara Sea [54]. Intercomparison of aerosonde and radiosonde wind-data showed mean wind-magnitude differences ranging from 0.02 to 1.7 ms^{-1} [55]. These errors are smaller than the SDEs of upper level wind-speed. Thus, observational errors alone do not explain the large errors in December. The boundary condition used by WRF affected strongly the result. In December, the FNL and NNR winds differed strongly at the boundaries of the model domain at upper levels with consequences for the forecasted position of the polar-jet location (Figs. **8**, **9**).

7. TEMPERATURE

WRF-simulated and NNR-temperature differed non-significantly from each other. The same is true for WRF-simulated and FNL-analysis temperature. Like for pressure and wind, no significant effect of the forecast length can be detected for all temperature skills. On average, WRF and reanalysis agreed acceptably for 2m-temperatures and well at all pressure levels in both months (Table **1**; Figs. **9**, **10**). Correlation between WRF and reanalysis 2 m temperatures was stronger in December (r=0.92) than in July (r=0.81) because of the relative homogeneous temperature distributions during subsidence inversions under persistent high pressure (Fig. **9**). In July, convection led to high variability in temperatures leading to higher bias than in December.

Despite WRF overestimated daily temperatures on average, it underestimated the amplitude of the diurnal cycle in most places most of the time (e.g. Fig. **11**). Consequently, WRF failed to capture the diurnal temperature extremes to

their full extent. In July, WRF and reanalysis agreed typically better at local morning and evening (e.g. around 0000 and 1200 UTC in the example of Fig. **11**) when atmospheric conditions were close to neutral. WRF underestimated 2 m temperatures up to 5 K around local noon and overestimated them up to 3.5 K at local night. The errors in simulated convection due to the erroneous land-cover distribution (see section 8) explain the discrepancies in the afternoon, when free-convection occurred frequently. The dampening of the diurnal cycle in winter was due to erroneous snow-depth/coverage. Some of these snow-depth/coverage errors already stem from the initialization. WRF's 2 m temperature performance in our study was similar to that of MM5 in a summer study over the Great Lakes region [43]; the daily mean bias was small as day and nighttime errors canceled each other out.

Bias between WRF and reanalysis 2m-temperature was mostly positive throughout December and 1.4 K on average (Table **1**, Fig. **9**). In December, biases were lowest at local morning (e.g. Fig. **11**). Obviously, WRF tended to underestimate temperature under stable conditions. Between December 12 and 17, for instance, temperature inversions occurred due to radiative cooling. Note that the radiation balance is negative in December. The magnitude of bias found in our study is similar to that reported by [48].

In the ABL, the quality of temperature simulations decreased with increasing stability or free convection. At all pressure levels, temperature-biases were within the range of observational uncertainty. July-biases were overall slightly negative and decreased with increasing height (Fig. **10**). On the contrary, in December, biases were higher at 850 and 500 hPa than at 700 and 300 hPa. WRF mostly underestimated December-temperatures at 850 hPa. The biases fall within the range of [49] and [48] studies. The higher biases near the surface and 850 hPa than at 700 to 300 hPa result from errors in the surface-energy budget. In the surface-energy budget, the partitioning of available energy into sensible and latent
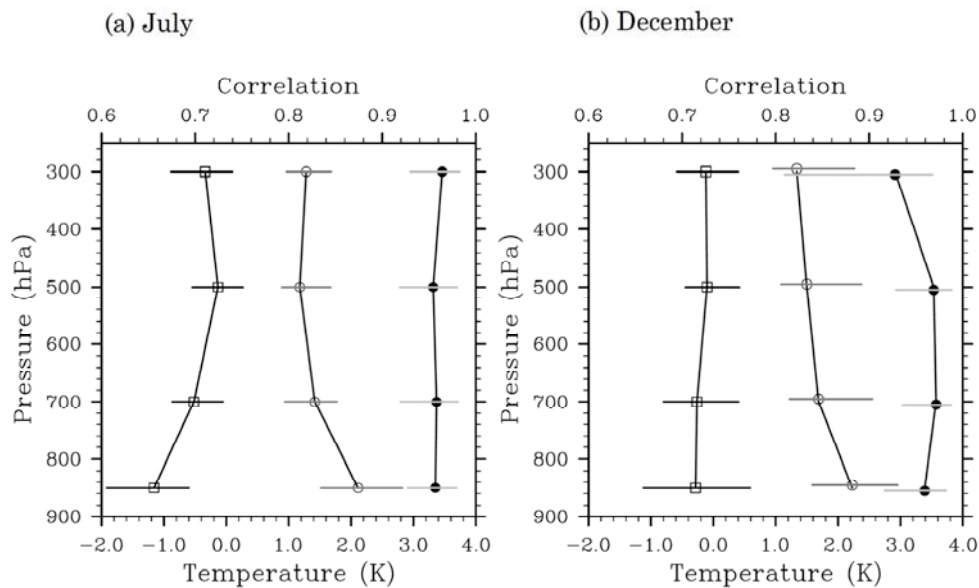

Fig. (10). Bias (line with squares), SDE (line with circles) and correlation (line with filled dots) between WRF-derived and reanalysis temperature (K) at 850, 700, 500 and 300 hPa and their ranges shown as horizontal black, dark and light gray bars, respectively for (**a**) July and (**b**) December. Correlation has to be read at the top X-axis, bias and SDE at the bottom X-axis.

heat fluxes and reflection into space strongly depends on the land-cover. The latent heat fluxes, for instance, among other things strongly depend on stomatal resistance. Albedo affects net radiation and increases (decreases) in albedo decrease (increase) the surface moist static energy [56]. Thus, errors in land-cover distribution affect the surface energy budget and hence, the temperature and moisture structure in the lower atmosphere. The nighttime cold bias can be attributed to errors in cloudiness (see section 8) that caused errors in the amount of long-wave radiation and nighttime cooling; simulating too much clouds hinders the atmosphere below the clouds to cool appropriately, while simulating too few clouds leads to too strong radiative cooling. Errors in 2 m temperatures and in the ABL propagated into discrepancies in relative humidity, with further impacts on simulated cloudiness.
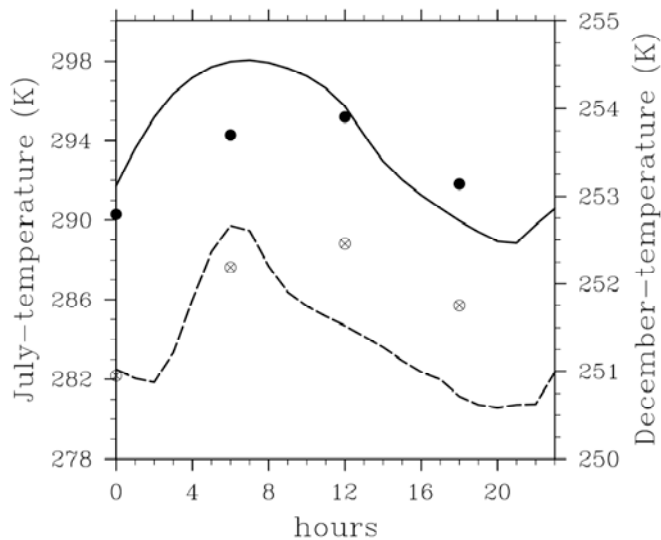

Fig. (11). WRF-derived monthly average diurnal cycle of temperature (K) with respect to UTC for July (solid line) and December (dashed line) for the grid-cell centered around 58 N and 94 E. July and December averaged NNR-data are only available at 0000, 0600, 1200 and 1800 UTC and are given by the filled circle and circle with cross, respectively.

The spatial distribution of bias notably matched the land-water and elevated terrain distributions (Figs. **1**, **9**) indicating differences in terrain height and the representation of the coastline between WRF and the reanalysis as major reasons. This pattern was more obvious for December because of the then greater contrast between land and ocean temperatures than in July.

In July (December), 2 m temperature RMSEs and SDEs between WRF and reanalysis amounted 4.9 K (4.9 K) and 4.1 K (4.7 K), respectively (Table **1**). The range of July-RMSEs was about twice that of the warm season RMSE-range of the 29 km Eta model in [48]. December-RMSEs were within the range reported by these authors for their cold season study. Approximately 2 K lower RMSEs and SDEs were found for MM5 over the Great Lake region [43] than we obtain for the comparison of WRF with reanalysis for Siberia.

WRF had difficulty in describing the near-surface temperature conditions over this permafrost region in July. A main reason was the erroneous land-cover distribution (Fig.

7b); soil heats differently under different land-cover with consequences for soil temperature and moisture, sensible and latent heat fluxes.

Temperature RMSEs and SDEs decreased with height in both months except at 300 hPa in July (Fig. **10**) and during frontal passages. The higher December than July RMSEs and SDEs can be related to unsystematic errors in the initial snow-distribution *via* the surface-energy budget and snow-albedo feedback [56]. Comparison of simulated monthly average snow-depth with the monthly average snow-depth of the CMC data showed huge discrepancies (Fig. **3**). Locally, too large (small) snow-cover extent led to incorrect atmospheric cooling (warming).

Our December-RMSE were on average smaller than the RMSEs reported by [49] for their comparison of six models; our July-RMSEs fall in the range of the best performances in their study. Our findings indicated that WRF-temperature simulations were very reliable above the ABL except for difficulty at the polar-jet level that were related to errors in the simulated position of the jet discussed before.

At upper levels, monthly averaged temperature bias, RMSEs and SDEs were less than ±2 K everywhere over Siberia in both months (therefore not shown). Like for 2 m temperature, at lower levels, bias, RMSEs and SDEs between WRF and reanalysis were several degrees higher over Tibet and its adjacent mountain chains than elsewhere in the domain because of differences in terrain height and snow-distribution.

The much smaller bias than SDE indicates that unsystematic errors in the reanalysis-temperature distribution and in initial and boundary conditions in WRF mainly contributed to the RMSE. Such unsystematic errors stem from the observations used in the assimilation to produce the NNR and FNL-data and the interpolation used to create the initial and boundary fields for WRF. Note that the FNL-data are analysis data. Analysis data can have errors due to observations. These errors are imported into the model over the lateral boundaries like model errors from a parent model are imported over the lateral boundaries into the child domain in a nested simulation. Typically, GPS-equipped radiosondes show bias of 0.25 K and 1 K at 100 hPa and 10 hPa, respectively [57]. Cloud-top temperatures that are derived from satellite data can have uncertainty of about 1 to 2 K. Thus, observational errors alone do not explain the discrepancies found.

8. CLOUD FRACTION

WRF-derived and ISCCP cloud-fraction distribution differed appreciably (Fig. **12**). WRF tended to forecast a sort of north-south gradient in cloud fraction for the entire domain. However, such a gradient only existed west of 70 °E in July and east of 70 °E to 100 °E in December. In both cases, the observed gradient was fuzzier and less distinct than simulated by WRF. In July, the observed gradient was stronger than simulated. In December, a maximum was simulated between 50 and 60 °N that was not observed.

WRF, on average, underestimated ISCCP-cloud fraction with more negative bias in July (-20%) than in December (-9%) due to misrepresentation of convective clouds that existed frequently in July. In July, bias reached up to -40%
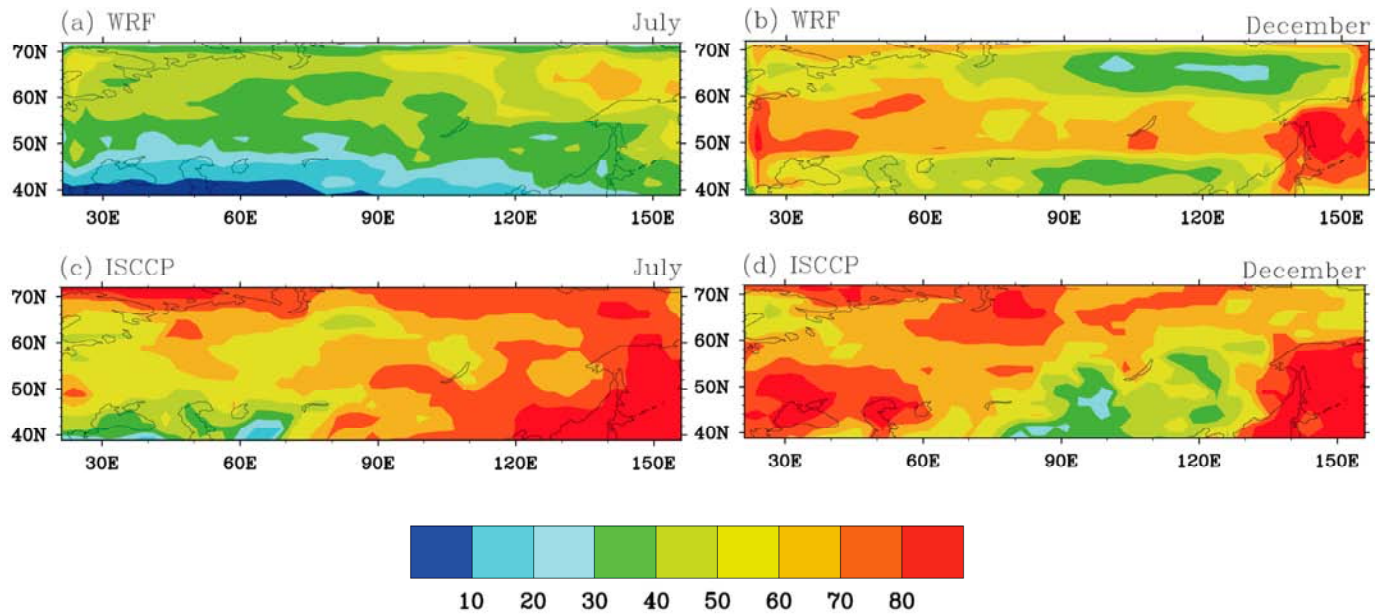
Fig. (12). Monthly average cloud fraction as derived from WRF for (**a**) July and (**b**) December, and the ISCCP for (**c**) July and (**d**) December.

over southern Siberia, the Barents Sea, Sea of Okhotsk and Sea of Japan. In December, bias reached up to -30% over the Central Siberian Uplands, Verkhoyansk Mts. and southwestern Siberia whereas it reached up to 30% over the Sayan Mts. Underestimation of cloud fraction enhanced incoming solar radiation and hence surface heating and heating of near-surface air in July. In July, too early onset of convective clouds contributed to the on average too low daily maximum temperatures. Too early onset of cloudiness means a too high cloud fraction at that time, less insolation, decreased net radiation, and reduced warming. In December, in the few areas with overestimation of cloudiness downward long-wave radiation increased and caused warming. Thus, some of the temperature-difference discussed above resulted from errors in simulated cloudiness.

In July, some error in cloudiness resulted from the land-cover data and land-cover treatment used in WRF. In nature, land-cover typically varies less in mountainous terrain than in flat or moderate terrain [51]. This is not reflected in the land-cover data (Fig. **7**). Furthermore, WRF assumed the dominant land-cover within a grid-cell as the representative one for the exchange of heat and moisture at the earth-atmosphere interface [1]. This procedure led to a more homogeneous land-cover distribution than in nature.

WRF-monthly accumulated evapotranspiration showed similarity with the dominant land-cover pattern in July (Fig. **7a**, **b**) that was higher on individual convective days than on monthly average. On convective days, clouds formed frequently along boundaries of land-cover, and were often accompanied by convective precipitation in their downwind. This means the strategy of dominant land-cover type affects the water supply to the atmosphere, convection, cloudiness and precipitation.

9. PRECIPITATION AND SNOW-DEPTH

Compared to the GPCC-data, WRF, on average, overestimated precipitation with slightly higher bias in

December (24.4 mm/mon) than in July (19.2 mm/mon). The fact that WRF underestimated cloudiness, but overestimates precipitation means that the cloud-schemes used in our simulations seem to convert too quickly or much cloud-water/ice into rainwater/snow. Such behavior can occur when the actual shapes of the ice crystals differ from those assumed in the cloud-microphysical scheme and hence actual terminal velocities are less than those simulated [58]. Other reasons can be the assumptions on conversion of cloud to rainwater [58]. Nevertheless, in both months, WRF captured the spatial distribution of precipitation reasonably well (r>0.70). In July, precipitation biases were greatest in central Siberia, over mountain chains located in the southern part of the domain and along the coast.

Errors in WRF-snow-depth stem from both incorrect initial snow-depth distribution and incorrectly predicted snowfall. However, in long-term simulations, the combined error affects the exchange of fluxes at the surface-atmosphere interface. WRF underestimated snow-depth on average, and failed to capture the fine structure in the distribution (Fig. **3**). WRF-provided snow-depth followed terrain height in most of the domain and missed the increased snow-depth south of the Barents Sea and along the coasts of the Sea of Japan and Okhotsk (Figs. **1**, **3**) that were due to local snowfall enhancement due the relatively warmer water than air.

The cloud microphysical scheme used a prescribed aerosol distribution. This assumption caused some errors, as aerosols strongly affected the efficiency of precipitation formation and amount [21, 58, 59]. Some error in predicted precipitation was related to the erroneous land-cover distribution (Fig. **7b**, **c**). In December, biases were highest along the coasts indicating differences in landscape as causes for differences between WRF and GPCC-precipitation. The errors in predicted precipitation along the coast explain the errors in predicted snow-depth in these areas.

Errors in simulated precipitation also resulted from misrepresentation of the partitioning between solid and liquid precipitation, assumptions of ice-crystal types, terminal velocity, vertical grid-resolution and inaccurate simulation of atmospheric moisture transport related to errors in the position of the polar front and polar jet (Figs. **8**, **9**). Interpolation of the observed precipitation to regional averages, and network density and design contributed to the errors [36]. For solid precipitation, some of the discrepancies can be due to catch deficiencies that for high wind-speeds may be as large as 30% [60] despite the gridding method used for the GPCC-data corrects for such kind of errors [35].

The enhanced biases of temperature during frontal passages, discussed before, indicate misinterpretation of cloud processes, assumptions on cloud parameters and vertical mixing as sources for discrepancies. Terminal velocity of ice-crystals, for instance, strongly depends on ice-crystal type, size and shape [61]. The microphysical scheme assumed hexagonal graupel-like snow. However, in nature, a huge variety of ice-crystal types may occur that have different terminal velocities than that of graupel-like snow. Consequently, WRF over (under)estimates the sedimentation of ice crystals depending on whether the terminal velocities of the natural ice crystals are smaller (larger) than those assumed by WRF. Thus, where WRF over(under)estimates sedimentation of ice-crystals, precipitation and snow-depth can be over(under)estimated.

CONCLUSIONS

The goals of our study were to examine WRF's performance in simulating the formation of air-masses, clouds and precipitation and to identify model shortcomings and need for future model improvement. This assessment sought to examine WRF's performance for and transferability to a hydro-climatically different region, and to assess its suitability as meteorological part in a regional Earth System Model for the Arctic. Since data access and spatial resolution is very limited for this region and biased low density network data may provide large errors when used to assess regional conditions [36], we used coarse-resolution gridded observational data despite such data may be less confident than any comparison with single point observation. We compared the 6-30h of WRF-simulations, initialized at 0000 UT on each day in July and December 2005 over Siberia for July and December 2005 with FNL, NNR, GPCC, ISCCP and CMC data.

Skill-scores of all examined quantities fell within the range found by other authors in their WRF-evaluation case studies over the US, in other model comparisons and for other mesoscale model case studies. We conclude that WRF produced reliable results with similar quality for Siberia for our July and December 2005 cases than for the region for which WRF was developed. The scientific question whether WRF is suitable for some applications as the meteorological part of a future Arctic system model can be answered positively, despite some work is required as identified in the following.

WRF-simulated pressure, temperature, and wind-speed differed non-significantly (at the 95% or higher confidence level) from the FNL and NNR-values. Compared with NNR, overall WRF simulated wind-speeds and temperatures well

in both months. Except for pressure, WRF and NNR agreed best under persistent high pressure and worst during frontal passages and transition from high to low pressure or vice versa. WRF-simulated and reanalysis of pressure agreed better for low than high pressure. WRF overestimated SLP compared to reanalysis and had some difficulty in capturing the positions of fronts. Some of the systematic differences in pressure and wind related to misinterpretation of terrain height and complexity. In WRF like in any numerical weather prediction model, terrain is "smoothed" because an average of the terrain height within a grid-cell is assumed to represent its topographic conditions. However, the reanalysis inherit the effect of the correct terrain height due to the assimilation of observations. On average, WRF simulations agreed stronger with reanalysis in July than December. Biases, RMSEs, and SDEs between WRF and reanalysis were lower for pressure and wind-speed in July than in December, while temperature-RMSEs and SDEs were similar in both months. The results suggested including parameterizations that consider subgrid-scale terrain impacts on the atmosphere when WRF is to be used within the framework of an Arctic system model.

Some of the temperature errors are due to errors in simulated cloudiness. The general underestimation of cloudiness and huge discrepancies in cloudiness suggest that testing other cloud parameterizations and/or improving the parameterization of cloud processes may be a promising path for better capturing the amplitude of the diurnal cycle. Improved cloud microphysical parameterizations also could lead to better precipitation and snow-depth predictions.

Our findings indicated that WRF-temperature simulations were very reliable above the ABL except for difficulty at the polar-jet level in December that was related to errors in the positioning of and strength of the polar jet (Fig. **8**). These errors can be partly attributed to the lateral boundary conditions of the wind field and the failure to capture the strength of the Siberian High to its full extent. Obviously, WRF had difficulty in producing the strong pressure gradients to the adjacent lows. The errors in predicting the cloud and snow-distributions were among the reasons for this failure and were caused partly by incorrect land-cover data and the smoothed terrain.

In July, WRF-results for the ABL suffered notably from the inaccurate land-cover data. Land-cover affects the energy fluxes [56]. Consequently, the monthly-accumulated evapotranspiration and on many days near-surface air temperatures well reflected the distribution of land-cover. Thus, we have to suspect that errors in land-cover distribution can propagate *via* evapotranspiration to errors in cloud and precipitation formation.

Incorrect land-cover and initial snow-cover/depth data also are also suspected to have contributed to the dampening of the diurnal cycle found for both months. Since wind-speed, temperature and humidity depend on the fluxes of momentum, and sensible and latent heat, errors propagated into these quantities and again affected the fluxes. Secondary differences resulted from error propagation as the incorrect fluxes and state variables affected simulated cloud and precipitation formation; cloudiness again affected the radiation processes. The accumulated convective precipitation documented well this error propagation and the

strong impact of land-cover as accumulated convective precipitation was correlated well with the boundaries of assumed land-cover. WRF strongly responded to land-cover and land-surface conditions under weak synoptic forcing conditions. The fact that model performance was suspect to the accuracy of the land-cover data means that accurate land-cover data are an urgent need for all WRF-applications over land and urges for improved and continuous updates of land-cover datasets for all atmospheric modeling purposes. The underestimation of cloudiness was suspected to be partly a result of the errors in land-cover and snow distribution and assumption in the microphysical schemes. Based on this study we have to suspect that in Arctic System modeling applications errors in predicted land-cover and/or snow-depth distribution will contribute to errors in fluxes and state variables at least in the ABL with consequences for simulated regional near-surface climate.

Although the discrepancies between the model results and observations are to a strong degree unsystematic according to the relative values of bias and SDE, we have to expect the conclusion can change for simulations with different choice of physical schemes, especially for cloud and precipitation. This is due to the fact that there is still systematic error due to model deficits.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

None declared.

REFERENCES

[1] Skamarock WC, Klemp JB, Dudhia J, *et al*. A description of the advanced research WRF version 2. NCAR/TN–468+STR, NCAR Technical Note 2005.
[2] Pielke RA. Mesoscale meteorological modeling. Academic Press: New York 2002.
[3] Schlünzen KH. On the validation of high-resolution atmospheric mesoscale models. J Wind Eng Ind Aerodyn 1997; 67-68: 479-92.
[4] Done J, Davis CA, Weisman M. The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model. Atmos Sci Lett 2004; 52: 110-7.
[5] Knievel JC, Ahijevych DA, Manning KW. Using temporal modes of rainfall to evaluate the performance of a numerical weather prediction model. Mon Wea Rev 2004; 132: 2995-3009.
[6] Cheng WYY, Steenburgh WJ. Evaluation of surface sensible weather forecasts by the WRF and the Eta models over the western United States. Wea Forecast 2005; 20: 812-21.
[7] Kain JS, Weiss SJ, Levit JJ, Baldwin ME, Brigh DR. Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: the SPC/NSSL spring program 2004. Wea Forecast 2006; 21: 167-81.
[8] Klemp JB, Skamarock WC, Dudhia J. Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. Mon Wea Rev 2007; 135: 2897-913.
[9] Trentmann JKC, Salzmann M, Barthlott C, *et al*. Multi-model simulations of a convective situation in low-mountain terrain in central Europe. Meteorol Atmos Phys 2009; 103: 95-103.
[10] Bromwich DH, Monaghan AJ, Manning KW, Powers JG. Real-time forecasting for the Antarctic: an evaluation of the Antarctic

[11] Kusaka H, Crook A, Dudhia J, Wada K. Comparison of the WRF and MM5 models for simulation of heavy rainfall along the Baiu front. SOLA 2005; 1: 197-200.
[12] Vaidya SS. Simulation of weather systems over Indian region using mesoscale models. Meteorol Atmos Phys 2007; 95: 15-26.
[13] Kalnay E, Kanamitsu M, Kistler R, *et al*. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 1996; 77: 437-71.
[14] Kanamitsu M, Ebisuzaki W, Woollen J, *et al*. NCEP–DOE AMIP-II reanalysis (R-2). Bull Am Meteorol Soc 2002; 83: 1631-43.
[15] Furevik T, Bentsen M, Drange H, Kindem IKT, Kvamstø NG, Sorteberg A. Description and evaluation of the Bergen climate model: ARPEGE coupled with MICOM. Climate Dyn 2004; 21: 27-51.
[16] Frankignoul C, Kestenare E, Botzet M, *et al*. An intercomparison between the surface heat flux feedback in five coupled models, COADS and the NCEP reanalysis. Climate Dyn 2004; 22: 373-88.
[17] Oh J-H, Kim T, Kim MK, Lee S-H, Min S-K, Kwon W-T. Regional climate simulation for Korea using dynamic downscaling and statistical adjustment. J Meteorol Soc Jpn 2004; 82: 1629-43.
[18] Semmler T, Jacob D, Schlünzen KH, Podzun R. The water and energy budget of the arctic atmosphere. J Clim 2005; 18: 2515-30.
[19] Zhang J, Bhatt US, Tangborn VW, Lingle CS. Climate downscaling for estimating glacier mass balances in northwestern North America: validation with a USGS benchmark glacier. Geophys Res Lett 2007; 34: L21505.
[20] Stein J, Richard E, Lafore JP, Pinty JP, Asencio N, Cosma S. High-resolution non-hydro-static simulations of flash-flood episodes with grid-nesting and ice-phase parameterization. Meteorol Atmos Phys 2000; 72: 203-21.
[21] Thompson G, Rasmussen RM, Manning K. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: description and sensitivity analysis. Mon Wea Rev 2004; 132: 519-42.
[22] Grell GA, Dévényi D. A generalized approach to parameterizing convection. Geophys Res Lett 2002; 29: 4.
[23] Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res 1997; 102D: 16663-82.
[24] Chou M-D, Suarez MJ. An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech Memo 104606 1994; pp. 85.
[25] Janjić ZI. Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. NCEP Office Note 437 2002; p. 61.
[26] Smirnova TG, Brown JM, Benjamin SG, Kim D. Parameterization of cold season processes in the MAPS land-surface scheme. J Geophys Res 2000; 105: D4077-86.
[27] Simmons AJ, Jones PD, da Costa Bechtold V, *et al*. Comparison of the trends and low frequency variability in CRU, ERA-40 and NCEP/NCAR analysis of surface air temperature. J Geophys Res 2004; 109: D24115.
[28] Onogi K, Koide H, Sakamoto M, *et al*. JRA-25: Japanese 25-year re-analysis project-progress and status. Quart J R Meteorol Soc 2005; 131: 3259-68.
[29] Uppala SM, Simmons AJ, Andrae U, *et al*. The ERA-40 reanalysis. Quart J R Meteorol Soc 2005; 131: 2961-3012.
[30] Mitchell TD, Carter TR, Jones PD, Hulme M, New M. A comprensive set of high resolution grids of monthly climate for Europe and Globe: and 16 scenarios (2001-2100), Tyndall, Centre for Climate Change Research 2004; p. 55.
[31] Drobot SMJ, Herzfeld UC, Fowler C. Uncertainty in temperature and precipitation datasets over terrestrial regions of western Arctic. Earth Interaction 2006; 10: 1-17.
[32] Hagemann S, Arpe K, Bengtson L. Validation of the hydrological cycle of ERA40, ECMWF ERA-40 project report. Series 24, Reading, UK 2005; p. 42.
[33] Bromwich DH, Fogt RL, Hodges KI, Walsh JE. A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. J Geophys Res 2007; 112: D10111.
[34] Rossow WB, Schiffer RA. Advances in understanding clouds from ISCCP. Bull Am Meteorol Soc 1999; 80: 2261-88.
[35] Schneider U, Fuchs T, Meyer-Christoffer A, Rudolf B. Global precipitation analysis products of the GPCC. Global Precipitation

Climatology Centre (GPCC), Deutscher Wetterdienst, Offenbach, Germany 2008; p. 12.

[36] PaiMazumder D, Mölders N. Theoretical assessment of uncertainty in regional averages due to network density and design. J Appl Meteor Clim 2009; 48: 1643-66.

[37] PaiMazumder D, Miller J, Mölders N, *et al.* Evaluation of community climate system model soil temperatures using observations from Russia. Theor Appl Climatol 2008; 94: 187-213.

[38] Yarker MB, PaiMazumder D, Cahill CF, Dehn J, Prakash A, Mölders N. Theoretical investigations on potential impacts of high-latitude volcanic emissions of heat, aerosols and water vapor and their interactions on clouds and precipitation. Open Atmos Sci J 2010; 4: 24-44.

[39] Beres J, Prakash A, Gens R, Mölders N. Comparison of WRF model outputs and MODIS image products for cloud presence: a case study. ARSC NSF-REU report 2009; p. 22.

[40] von Storch H, Zwiers FW. Statistical analysis in climate research. Cambridge University Press: UK 1999.

[41] Chang JC, Hanna SR. Air quality model performance evaluation. Meteorol Atmos Phys 2004; 87: 167-96.

[42] Mölders N. Plant and soil-parameter-caused uncertainty of predicted surface fluxes. Mon Wea Rev 2005; 133: 3498-516.

[43] Zhong S, In H, Bian X, Charney J, Heilman WBP. Evaluation of real-time high resolution MM5 predictions over the Great Lakes region. Wea Forecast 2005; 20: 63-81.

[44] Mölders N. Suitability of the weather research and forecasting (WRF) model to predict the June 2005 fire weather for Interior Alaska. Wea Forecast 2008; 23: 953-73.

[45] Mölders N, Walsh JE. Atmospheric response to soil-frost and snow in Alaska in March. Theor Appl Climatol 2004; 77: 77-105.

[46] Ross B, Walsh JE. Synoptic-scale influences of snow cover and sea ice. Mon Wea Rev 1986; 114: 1795-810.

[47] Zhang T, Barry RG, Gilichinsky D. Russian historical soil temperature data, Boulder, CO: National snow and ice data center. digital media 2006. Available at: http://nsidc.org/data/arcss078.html.

[48] Nutter PA, Manobianco J. Evaluation of the 29km-Eta model part I: objective verification at three selected stations. Wea Forecast 1999; 14: 5-17.

[49] White BG, Paegle J, Steenburgh WJ, *et al*. Short-term forecast validation of six models. Wea Forecast 1999; 14: 84-108.

[50] Gerlach R, Skinner L, Luckman A, Schmullius C. Deriving land cover information over Siberia: comparing results from MERIS and MODIS. Proceedings of the MERIS (A)ATSR Workshop 2005 (ESA SP-597). 26-30 September 2005 ESRIN, Frascati, Italy. Ed, H Lacoste. published on CDROM 2005; p. 46.1.

[51] Bergen KM, Zhao TT, Kharuk V, Blam Y, Brown DG, Peterson L, Miller N. Changing regimes: forested land-cover dynamics in central Siberia 1974–2001. Photogr Engineer Remote Sens 2008; 74: 787-98.

[52] Hines KM, Bromwich DH. Development and testing of polar weather research and forecasting (WRF) model. part I: greenland ice sheet meteorology. Mon Wea Rev 2008; 136: 1971-89.

[53] Mölders N, Kramm G. A case study on wintertime inversions in interior Alaska with WRF. Atmos Res 2010; 95: 314-32.

[54] Gruber C, Haimberger L. On the homogeneity of radiosonde wind time series. Meteorol Zeitschr 2008; 172: 631-43.

[55] Soddell JR, McGuffie K, Holland GJ. Intercomparison of atmospheric soundings from the aerosonde and radiosonde. J Appl Meteorol 2004; 43: 1260-9.

[56] Mölders N. Land-use and land-cover changes, atmospheric and oceanographic sciences Library 44, DOI 10.1007/978-94-007-1527-1 3, Springer Science+Business Media B.V 2011.

[57] Inai Y, Hasebe F, Shimizu K, Fujiwara M. Correction of radiosonde pressure and temperature measurements using simultaneous GPS height data. SOLA 2009; 5: 109-12.

[58] Mölders N, Kramm G, Laube M, Raabe A. On the influence of bulk parameterization schemes of cloud relevant microphysics on the predicted water cycle relevant quantities - a case study. Meteorol Zeitschr 1997; 6: 21-32.

[59] Muhlbauer A, Lohmann U. Sensitivity studies of the role of aerosols in warm-phase orographic precipitation in different dynamical flow regimes. J Atmos Sci 2008; 65: 2522-42.

[60] Yang D, Woo MK. Representativeness of local snow data for large scale hydrologic in-vestigations. Hydrol Process 1999; 131: 1977-88.

[61] Locatelli JD, Hobbs P. Fall speeds and masses of solid precipitation particles. J Geophys Res 1974; 79: 2185-97.