1. INTRODUCTION

Climate forcing, including drivers of climate change, are parameterized in all climate models. There appears to be a controversy in climate modeling if the so called "model physics" has anything to do with actual physics of real processes, or if it is just a package of tunable statistical relationships of more obscure nature. Given how climate is generated in a climate model, it is exceedingly clear to us that unless "model physics" at least attempt to model the actual physics in a realistic way, climate modeling is not meaningful.

The Arctic is one of the most sensitive areas in the World to climate change. On average in 19 CMIP (Meehl et al. 2000) climate change simulations, the Arctic warms 2.5 times more the global average (Räisänen 2001). We see already today signs that global warming has started to impact the Arctic (Serreze et al. 2000, Comiso 2002). Still, the inter-model spread in the CMIP ensemble is largest in the Arctic (Räisänen 2001) and current GCM have problems reproducing today's Arctic climate (Walsh et al. 2002).

The large Arctic climate sensitivity is due in no small part to strong positive feedback mechanisms; the ice/snow-albedo feedback is probably the strongest. An adequate description of the fluxes of heat and momentum at the ice surface lay at the heart of a proper representation of this feedback. An evaluation of these in models has been difficult, due to lack of adequate data. The Surface Heat Budget of the Arctic Ocean (SHEBA, Uttal et al. 2002) experiment now makes this possible. The aim of the Arctic Regional Climate Model Intercomparison (ARCMIP, Curry and Lynch 2002) project is to improve climate models for the Arctic, by comparing models to SHEBA data and to each other.
2. MODELS

The models included in this study are state-of-the-art regional-scale climate models: ARCSym (Lynch et al., 1995), COAMPS\textsuperscript{TM} (Hodur 1997), HIRHAM (Christensen et al., 1996), Polar MM5 (Bromwich et al., 2001), RCA (Jones et al. 2004) and REMO (Jacob, 2001), see acronyms in Fig. 2. All models were set up with the same resolution on a common domain, centered on the SHEBA ice-drift track (Fig. 1). They all used the same 6-hourly lateral boundary conditions from taken ECMWF analyses. Sea- and ice-surface temperatures and ice fraction were prescribed from satellite observations. The models were run 13 months, from 1 September 1997. The experiment is described in detail in Tjernström et al. (2004) and Rinke et al. (2004).

3. RESULTS

In general, the relatively small domain ensures that the larger-scale dynamics in the models adhere to that of the driving analyses, although some differences between the models occur even on the synoptic scale (Rinke et al 2004). Fig. 2 (top panel) shows weekly averaged 2-meter air temperature from all the models for a few winter months. While the temperature of the ice surface was prescribed, the models are expected to follow the observations. It is therefore somewhat surprising to find some rather large differences between models and observations. During some cold periods in December 1997, many models are ~ 10 °C too warm, even in the weekly averages. The coldest period, around 1 January 1998, is, however, well captured by all models. In summer the differences are smaller, but with a systematic disparity between some models closer to ~ 0 °C, the melting point of fresh water, and others closer to ~ -1.8 °C, the melting point of salty ocean water.

Figure 3. Seasonal averages of temperature bias profiles. Lines show: Fall - black, winter – blue, spring – green and summer – yellow.

Seasonally averaged profiles of temperature bias are shown in Fig. 3. Two things are obvious in this figure. First, the biases are much larger and more variable below ~ 1 km. The larger biases closer to the surface indicate deficiencies in the boundary-layer parameterizations, probably also related to errors in the formation of low-level clouds. Note the summer low-level cold bias in practically all models, possibly due to an overestimation of cloud-top cooling. Second, different models behave very differently also in the free troposphere. Some models have a consistent bias through the year while others are very variable between the seasons and in general, the errors do not seem to tend to zero with height, in spite of the strong constraint by the lateral boundary conditions even on such a small domain.

Near-surface wind speeds follow the observed temporal variability well in all models, but with systematic biases (Figure 4). Annually averaged biases range from ~ -1 ms\textsuperscript{-1} in RCA to ~ 1.5 ms\textsuperscript{-1} in Polar MM5. In some cases, this is consistent with biases in friction velocity (Figure 5, top panel), for example the high $u^*$ bias in RCA is consistent with its low wind-speed bias. In Figure 5 (bottom panel), friction velocity is plotted against wind speed for
each model, meaning that the slope of a regression line through the data represents the square root of the drag coefficient. It is clear that the modeled momentum fluxes deviate from the observations (gray), sometimes significantly. The friction velocity is too high in ARCSym and RCA, in the latter possibly explaining the low bias in wind speed, while Polar MM5 show a hard lower limit beyond which $u^*$ is set constant.

Given the difficulties to model clouds, the surface radiation fluxes are relatively accurate in most models. While some models have biases of typically ± 30 W m$^{-2}$, the overall results are somewhat promising. The largest concern here is that the net error is often due to compensating problems in different fluxes, so that the net errors become sensitive to different processes.

Figure 6. Time series of weekly averaged sensible (top) and latent (bottom) heat flux for the whole year. Legends are as in Figure 2.

In a direct comparison of the turbulent heat fluxes, all models fail badly (Figure 6). None of the models is similar to any of the other models, and neither model shows significant similarity to the observations. The correlation coefficients between modeled and observed fluxes are consistently below 0.3 and modeled biases of especially latent heat is relatively large. Still, plotting the sensible heat flux scaled by the wind-speed against the low-level temperature difference most models follow the observed structure well, with the exception of the most stable conditions. Again, the slope of the almost linear dependences reflects the value of the heat transfer coefficient, which is too large in Polar MM5 but only slightly low in some other models. The region on the stable side, where the measurements indicate a strong dependence in the heat transfer coefficient on stability, is absent in all models except COAMPS$^\text{Tm}$. For the latent heat flux the situation is worse; the scaled dependences are very different between the models, in particular in RCA and REMO.

4. Discussion

In this study, the boundary conditions were constrained both by analyzed lateral boundaries and by prescribing the surface temperature over the ocean. In this sense it reflects a “best case scenario”: this is how good – or or bad – these models are when larges scale dynamics and surface temperatures are reasonably well known. The study reveal some systematic errors in temperature and wind speed and heat fluxes that have a reasonable functional dependence, but with time series that has almost no correlation what so ever to the observations.

It is our belief that this a result of model tuning. The description of the turbulent friction in these models was tuned to optimize the surface pressure development, to ensure reasonable development of synoptic systems. In this process, the actual friction was of less importance, as long as cyclones and anti-cyclones obtained the correct spin-up and spin-down. The modeled turbulence
then has to “pick up the slack” from other unknown deficiencies in the models. Non-linear feedbacks between the wind speed, the static stability and the turbulence then adjust to a new unrealistic balance, thus disrupting all the turbulent fluxes.

The results are often superficially nice representations of Arctic mean climate, often for the wrong reason. If these models, on the other hand, were to be coupled to an ocean model including sea-ice, we suggest that the end-results may easily become a quite poor representations of current conditions. We leave the consequences for the reliability of Arctic climate change simulations for the reader to ponder upon.

Acknowledgements
MT, MZ and GS were supported by the Swedish Science Council under contract 2002-5610 and by the Swedish climate-modelling program SWECLIM. They are grateful to the COAMPS-team at the Naval Research Laboratory in Monterey, CA. JC acknowledge the support from National Science Foundation. KD, AR, CJ and KW were supported by the European Union Framework 5 project GLIMPSE (EVK2-CT-2002-00164). KD and AR are also grateful to Jens Christenssen at the Danish Meteorological Institute. SP and TS would like to thank Daniela Jacob, for providing the REMO model and the opportunity to participate in ARCMIP, and Ralf Podzun, for technical support. CJ and KW are grateful for the assistance by Anders Ullerstig and Ulrika Willén. All authors acknowledge the invaluable work invested in setting up the ARCMIP program, and in the generation of the boundary and forcing fields for this experiment. In particular, we are grateful to Judy Curry, Amanda Lynch, Liz Cassano and Jeff Key. The authors are also grateful to the SHEBA Atmospheric Flux Group, mainly Ola Persson and Chris Fairall, for access to the observations and also to Chris Bretherton and Stefan de Roode for the compilation of the SHEBA soundings.

References


Serreze, M., and coauthors, 2000, Observational evidence of recent change in the northern high-latitude environment. *Climate Change*, **46**, 159-207.

