The realism of the ECHAM5 models to simulate the hydrological cycle in the Arctic and North European area*

Klaus Arpe, Stefan Hagemann*, Daniela Jacob and Erich Roeckner

Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany
*Corresponding author. E-mail: *hagemann@dkrz.de*

Received November 2004; accepted in revised form 7 July 2005

Abstract A new version of the ECHAM model is investigated in respect of the hydrological cycle in the Arctic and North European area. Several horizontal and two vertical resolution versions are studied. The higher-resolution ECHAM5 models are, in many respects, superior to the lower-resolution versions of the same model family and the older ECHAM4 model. The vertical resolution has a decisive impact but also increased horizontal resolution leads mostly to improvements. Here T106 (about 110 km) often gives the best results. The summer maxima of precipitation, surface temperature and latent heat flux are simulated too early by about a month for several river catchment areas. This shift is strongest in the T106 and T159 models. Another problem with the annual cycle of precipitation is a relative minimum in August to October, especially in the low-resolution ECHAM5 models. The precipitation of the ECHAM5 simulations over the Arctic region exceeds all observational estimates by 5–15 mm/month, strongest in May–June. The latent heat flux over the river catchments has a clear trend towards increased fluxes with higher horizontal and vertical resolution, which seems to reach a maximum with T106. In the comparison of annual mean *P-E* (precipitation minus evaporation) with observed river discharge only the horizontal resolution seems to be important, again giving best results for the high-resolution models. The year-by-year variability of the simulations is too high, which is more pronounced for the higher-resolution versions. Especially strong impacts are found from the vertical resolution. The interannual variability of the latent heat flux is much smaller than that of precipitation and therefore the results shown for precipitation apply also for the simulated river discharge. Some forcing of ocean temperature anomalies on the precipitation over the Rhine, Kolyma and Indigirka catchment areas have been found, from the northeastern Atlantic and from the Pacific with developing El Niños. Despite the increased random variability in the higher-resolution models, the signal could be detected in almost all simulations. On the whole the higher-resolution (horizontal and vertical) ECHAM5 model simulations are quite improved compared to the low-resolution version of the same model and an older T42 model version. Increasing the vertical resolution from 19 to 31 levels is decisive for this better performance.

Keywords Baltex; ECHAM5; hydrological cycle; model validation; precipitation; resolution dependency; river discharge

Introduction

A new version of the ECHAM model (Roeckner *et al.* 2003) has become available. It is expected that many users will use this model extensively. It will be used in a coupled atmosphere–ocean system or for driving a limited area model by providing boundary values. It is worth investigating simulations with several versions of this model to document its merits and deficiencies for the support of other users. The study should help to select the optimal resolution for the given tasks. It has its focus on the hydrological cycle in the Arctic and North European area. Four versions of horizontal resolution and two versions of vertical resolution are considered, i.e. resolution dependency; 19 and 31 levels in the vertical and horizontal resolutions from 42 (T42) to 159 spherical harmonics (T159). The

*Paper presented at the 4th BALTEX Study Conference, Bornholm, Denmark, May 2004.

Nordic Hydrology **Vol 36 No 4–5 pp 349–367** © IWA Publishing 2005

comparison will be concentrated on the ECHAM5 model versions 5T159L31, 5T106L31, 5T63L31, 5T63L19 and T42L19 and on the former standard ECHAM4 model 4T42L19 (Roeckner *et al.* 1996). For brevity the specification of levels and model version will be dropped in the following unless needed, i.e. the short names will be: T159, T106, T63, T42 (for ECHAM5 with 31 levels), T106L, T63L, T42L (for ECHAM5 with 19 levels) and 4T42 (ECHAM4 with 19 levels). 5T106L19 and 5T42L31 simulations are available but are not likely to be used in future experiments and are mentioned only occasionally.

With the ECHAM5 model versions AMIP2-type simulations were conducted covering the time period 1979–1999. In order to pay regard to model spin-up, all simulations were started in 1978. The simulations were forced with AMIP2 sea surface temperatures (Taylor *et al.* 2000) over the ocean. The horizontal resolutions of T42, T63, T106 and T159 correspond to grid sizes of about 2.8°, 1.9°, 1.1° and 0.75° or rather 300 km, 200 km, 110 km and 80 km, respectively. For 4T42L19, an ensemble of 6 simulations is considered that cover the time period 1951–1993. These simulations were forced with SST from the GISST dataset (Rayner *et al.* 1996).

This is an accompanying report to Hagemann *et al.* (2005b) which gives an overview of the performance of these models from a global point of view while this report focuses on the hydrological cycle over the North European and Arctic area.

An overview of the performance on some dynamical quantities will be given in the section on dynamical quantities. Long-term means of precipitation will be discussed in the section on precipitation.

The geographical distribution of differences between simulations and observational estimates as well as their annual cycles will be shown. For investigating the latent heat flux (LHFX) over the continents in the corresponding section only the estimates by Mintz and Serafini (1992) and the values from ERA40 (ECMWF reanalysis project of 40 years (Simmons and Gibson 2000)) are available for comparison: both are not very reliable estimates of the truth. Therefore the difference between precipitation and evaporation, which eventually results in river discharge, will be studied in detail in the section called river discharge for several rivers in the Arctic–North European area. The available observed river discharge data extend hardly into the most recent time, seldom beyond 1984, which leaves only a short overlapping period between observations and simulations. For 4T42 much longer simulations are available, partly starting in 1903 or before. These experiments will therefore be used to bridge the gap in time between the two data sets. The interannual variability of precipitation is investigated in the following. Here impacts from the sea surface temperature (SST) on the precipitation variability are studied. The results are summed up in the last section: summary and conclusion.

Dynamical quantities

Figure 1 shows the mean ERA40 500 hPa height field in winter (DJF) overlayed with the error field of four model versions. One recognizes the typical error patterns with negative values over ridge areas, i.e. Iceland–Europe and Alaska–east Siberia and positive values over trough areas, i.e. America and west Siberia. This means a more zonal flow in the simulations. This is a well known error pattern, which was first documented in 1976 in the ECMWF model (Arpe *et al.* 1976; Arpe 1990). The strength of this pattern has been considerably reduced since then. In this comparison two classes in the error strength are found: larger errors in the 19 level versions (T42L and T63L, also T106L but not shown) and smaller errors in the simulations with 31 levels: only T63 is shown but it applies for all resolutions T42 to T159. Within the 31 level group version T159 seems to have larger errors than the others but even with a 20-year average the sampling error could play a role. Furthermore a large positive error near the North Pole can be found, which is less in T106

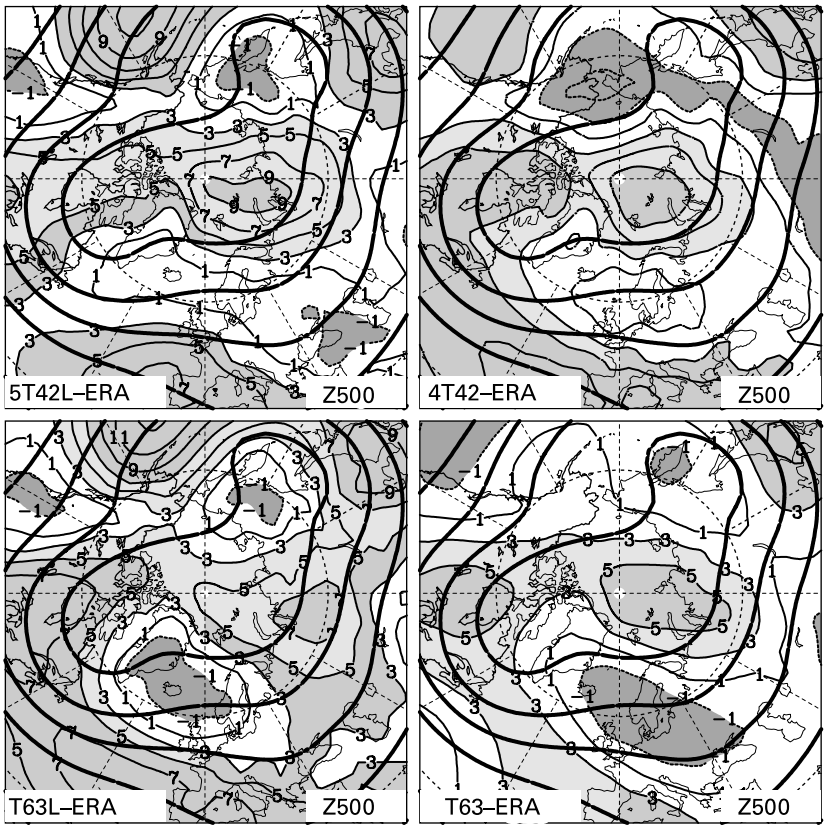

Figure 1 500 hPa height field error (model ERA40) overlayed by the height fields of ERA40 for DJF. Contours of errors at ± 1, 3, 5, 7, 9, 11 dam with shading for $>+3$ and <-3 dam, negative contours are dashed. Contours of ERA40 height fields every 16 dam (heavy lines)

and T63L. In some respects 4T42 has lower errors than T42L. The dipole patterns of the height errors in the 19 level simulations (T42L and T63L) along the American Pacific coast in Figure 1 is due to a westward shift of the Pacific ridge.

In summer (not shown) the error fields are dominated by a belt of more uniform negative errors around the world at about 60°N. Again the same division into two classes of simulations becomes obvious, i.e. with 19 levels and 31 levels. In this season T159 has the lowest errors. Again large positive errors near the North Pole can be found. 4T42 has lower errors than T42L.

The mean error in the height field is closely connected with an error pattern in the wind field.

Figure 2 shows cross sections of the zonal mean wind error overlayed with the position of the jet streams, which are indicated by shading areas with wind speeds of more than 25 m/s in the ERA40 analysis. Again a typical error field can be seen with positive values poleward and above the mean position of the jet stream and negative values below and equatorward of the jet stream. This can be interpreted as an upward and poleward shift of the mean jet stream. It is especially clear in the northern hemisphere in both seasons. As with the 500 hPa height field error, also for this feature two groups can be distinguished: larger errors in the simulations with 19 levels and smaller ones for 31 levels. Again 4T42 has less error than T42L.

A more detailed investigation shows, however, that the jet is not really shifted poleward but is strengthened in regions where it is positioned nearer to the poles and weakened where

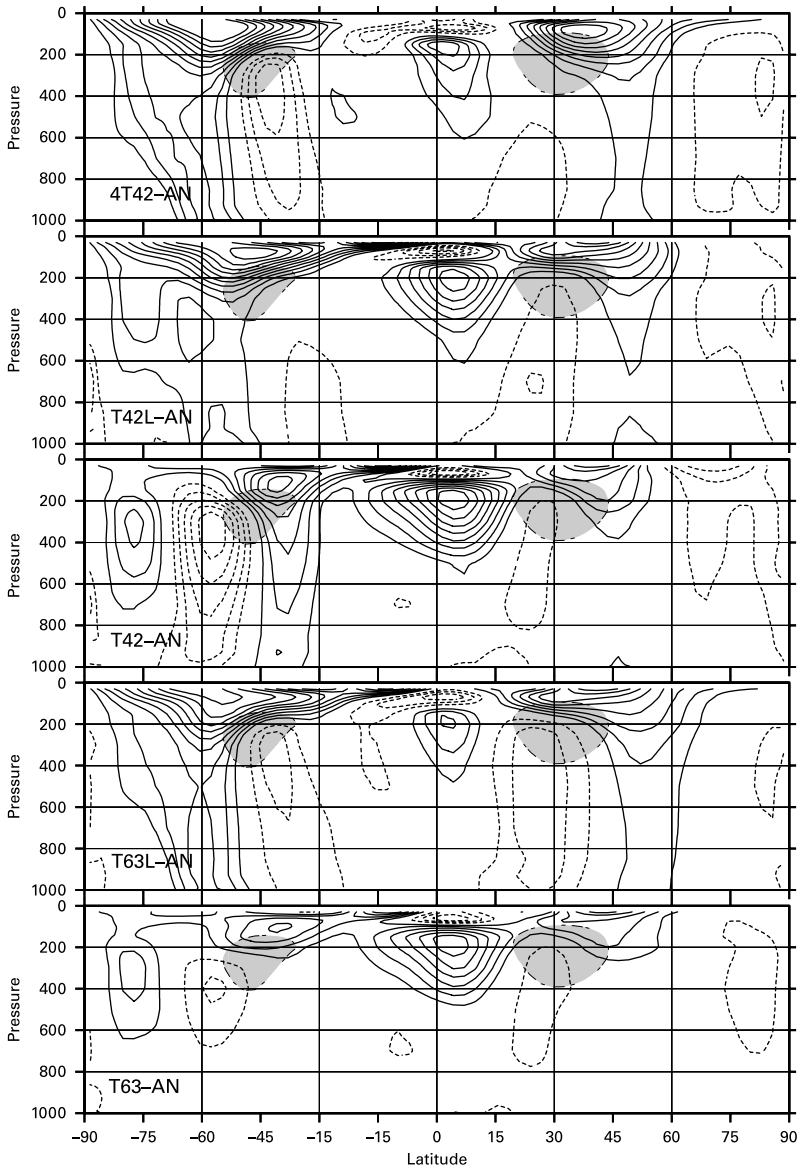

Figure 2 Cross sections of the zonal mean wind error for DJF overlayed with the position of the jet streams, which are indicated by shading the areas with wind speeds of more than 25 m/s in the ERA40 analysis. Contours at ± 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 m/s, negative contours are dashed

the jet is further away from the pole. This can be demonstrated by maps of winds at 300 hPa in the area of interest (not shown). The model simulations exhibit a polar jet over the Atlantic that extends further into Europe than the analyses: this extension is least in T159. Over the western Atlantic, where the jet stream has a position further south, the wind speeds of the analyses exceed those of the simulations. The same applies for the subtropical jet over Egypt with more than 40 m/s in the analyses and 2–5 m/s less wind speeds in the simulations. The two jet streams are separated over western Europe by a belt of lower wind speeds that is less well simulated by the models due to the eastward extension of the polar jet. On the whole, T106 is probably best but differences between models are small.

The jet is guiding, or at least linked with, the cyclone tracks and these are important for the precipitation, a main topic of this study. Errors in the jet will most likely be reflected by the precipitation, which will be investigated below.

A further systematic error in the wind field (Figure 2) can be found in the upper troposphere of the tropics. The easterlies in the analysis are almost reversed to westerlies in the model simulations. This leads to a very strong vertical wind shear at 100 hPa near the equator. Here the error is reduced with higher horizontal resolution in the case of 19 and 31 vertical levels but with less vertical levels the error is smaller as well almost zero for T106L.

The parametrisation of precipitation depends strongly on the relative humidity at all levels and on the vertical temperature profile (static stability). Figure 3 shows N–S sections of temperature differences between simulations and the ERA40 analysis for summer over the European sector (0–30°E) at 850 and 300 hPa. Positive deviations from ERA40 temperatures at 300 hPa or negative deviations at 850 hPa mean that the simulated stratifications are more stable. Especially north of 70°N the simulated stratifications are less stable than ERA40. 4T42 is extreme in this destabilisation. The 19 level models show a different class of errors compared to the 31 level models south of 55°N where the 19 level models are more stable. This is also shown in Table 1, where the stability is given as a temperature difference between 300 and 850 hPa. From this, more precipitation might be expected in the 31 level models than in the 19 level models.

For comparison, the temperature difference between both levels of a wet-adiabatic sounding is −62 K. More positive numbers mean a more stable stratification. The 19 level simulations are clearly more stable than the 31 level simulations for all horizontal resolutions, and therefore convection is less likely in 19 level simulations, which is indicated by less precipitation in the T42L–T42, T63L–T63 and T106L–T106 comparison of Table 1. There is also a dependency of precipitation on the horizontal resolution, with less precipitation in the coarser resolution models. This is accompanied by relative humidity in the opposite sense, i.e. more precipitation and lower relative humidity occur with higher horizontal resolution models.

In winter all simulations provide lower temperatures in the upper troposphere (not shown) than ERA40, with differences of around 5 K at 200 hPa. For the upper tropospheric levels in winter the T159 simulated temperatures are clearly different to the simulations with lower

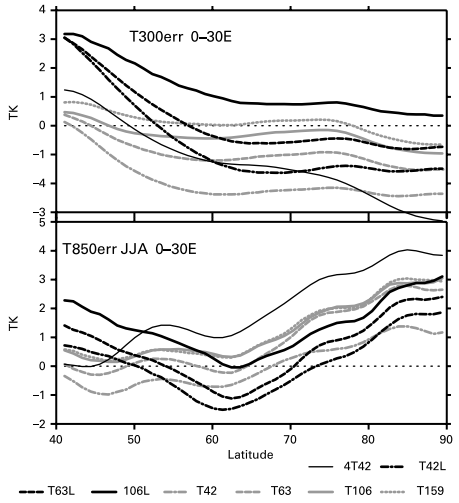

Figure 3 N–S section of temperature differences between simulations and ERA40, averaged 0–30°E at the 300 and 850 hPa levels

Table 1 Area means in JJA for 0–30°E/45–55°N (central Europe). Observed values for precipitation from GPCP (Huffman *et al.* 1996) and for dynamics from ERA40

	GPCP/ ERA40	T159	T106	T106L	T63	T63L	T42	T42L	Units
Precip.	78	71	69	44	62	44	61	39	mm/ month
T300–850	−51.7	−52.0	−52.4	−51.1	−52.6	−51.1	−52.7	−51.1	K
RH500	52	51	54	51	58	54	60	55	%
RH850	74	74	75	67	77	71	79	72	%

resolutions. The T159 also seems to be inferior to the lower-resolution models in some other respects. Some parametrisations depend on parameters, which are resolution-dependent. These parameters must be partly determined empirically. It is believed that the gravity wave drag is still not optimally tuned for the T159 model, which might have caused this temperature error.

Figure 4 investigates vertical profiles of relative humidity. A most prominent feature is the large difference between the older ERA15 (Gibson *et al.* 1997) and the recent ERA40 analyses; ERA15 is much dryer throughout the atmosphere. Comparisons with radiosonde observations suggest that ERA15 is more realistic but conventional humidity observations, especially in the upper troposphere, are not very reliable, which guided ECMWF in the beginning of their operations not to use any humidity observations above 300 hPa. Now they assume a large observational error for the analysis scheme. With this in mind it can be assumed that the model formulations, which are used in the analysis cycle, play an important role for the humidity of the analysis. The decrease of the humidity in the ECMWF operational analyses with the change of the convection scheme from Kuo (1974) to Tiedtke (1989) was nicely shown by Arpe (1990). In Figure 4 it can be seen that ERA15 has by far the lowest humidity, ERA40 is mostly below the values of the ECHAM simulations. Within the simulations there is a clear trend towards higher relative humidity with decreasing horizontal resolution, especially at higher levels (see also Table 1). One might expect more precipitation with higher atmospheric humidity but in Table 1 the opposite was found.

The convection scheme in the ECHAM models seems to be more efficient in extracting humidity from the atmosphere with higher resolution, which is reasonable because intuitively it is easier to fulfil the conditions of instability or reaching a critical level of relative humidity for a smaller column of atmosphere than with a larger one. A similar resolution dependency has been found for other models. This resolution effect may be important when using humidity data from a global model to feed a higher resolution limited area model.

Precipitation

In Figure 5 the winter precipitation fields of the models are compared with observational estimates. Probably the best estimate is the one by GPCP (Global Precipitation Climate Project (Huffman *et al.* 1996)). The difference over land to the estimates by CRU (Climate Research Unit at East Anglia University (New *et al.* 2000)) results from a better observational database by GPCP, especially with an enhanced correction of possible transmission errors and the application of a correction of precipitation amounts due to undercatchments in cases of snowfall. The latter leads to higher values, especially in winter in GPCP compared to CRU. The correction used by GPCP might, however, be too large (Rudolf, personal communication). In areas without any conventional precipitation observations GPCP makes use of IR, TOVS and SSMI observations by satellite while CRU

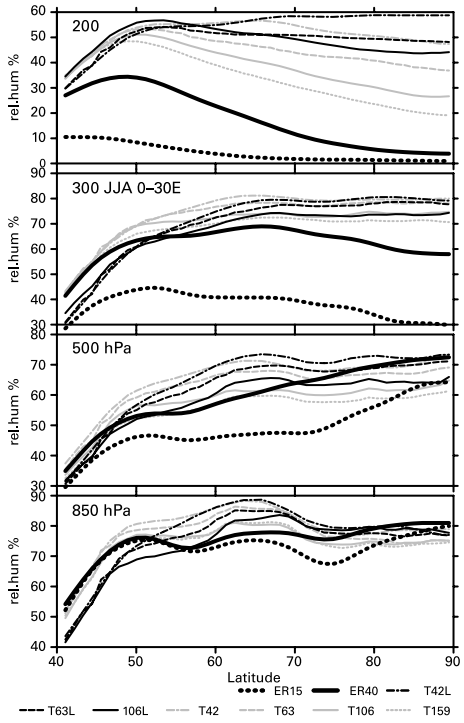

Figure 4 N–S section of relative humidity averaged 0–30°E at different levels

has to fall back on climatological means. The CRU data will be used here as well because it is available for a much longer period. The estimates of ERA40 are not shown here although this data set is used extensively for other quantities in this study but its quality with respect to precipitation was found to be low (Hagemann *et al.* 2005a). Generally the different precipitation patterns are very similar and it is hard to give an advantage to one or other model.

A wavy structure in the precipitation fields for T106 and T159 (not shown) is caused by the presentation of orography in spherical harmonics. This becomes especially obvious over the oceans where there are still orographic waves of the order of 20 m amplitudes. Over such artificial hills there is enhanced precipitation. This occurs not only in the T106 and T159 simulations but also in the lower-resolution models. In fact the amplitudes of precipitation and orography are even larger with T42 than with T106 but because of their longer wavelengths they are not so obvious in Figure 5.

In maps of precipitation possible errors are not so evident. Therefore, maps of differences between simulations and GPCP normalized by the maximum of either of them (relative error) are shown in Figure 6 for T63. There is too little precipitation in winter over Siberia and too much over northern America, the Arctic and China and less errors in L19 over North America. Relative errors of more than 70% occur only over areas with low precipitation amounts and therefore the absolute error values are small. The error fields of the other resolutions are very similar.

The differences between the model versions are made more obvious in Figure 7 (upper panel) where the precipitation difference between all model versions and GPCP are compared for the belt 40–60°N in DJF. Large errors can be found over the oceans. Here improvements with increased resolution in the horizontal and vertical are obvious.

For summer (Figure 6, left panel) the relative errors for T63 show too much precipitation for eastern Siberia and north-western America. Lower errors for the 31 level model can

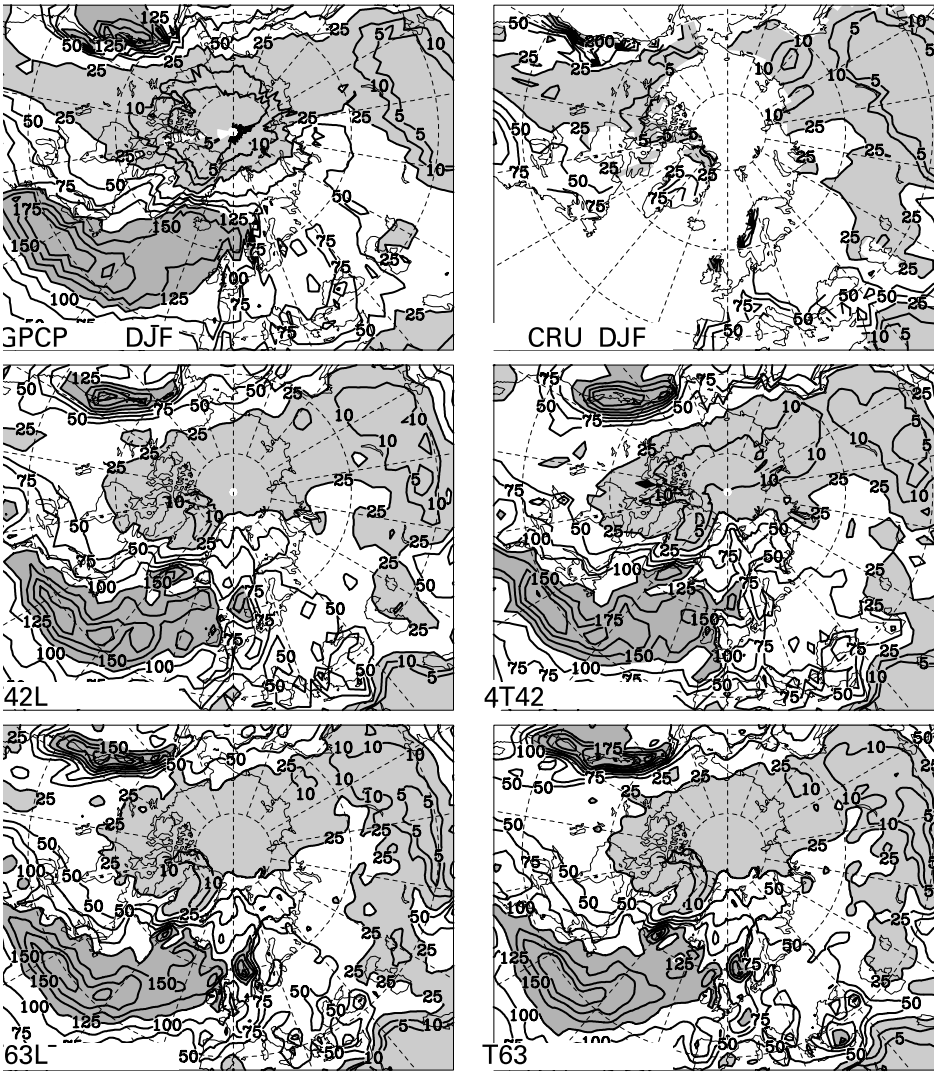

Figure 5 Precipitation in DJF. Contours at 5, 10, 25, 50, 75, 100, 125, 150, 175, 200 mm/month, shading for >125 mm/month and <25 mm/month

easily been recognized. In Figure 7 (lower panel) the errors are shown for all model versions for the belt 35–45°N. Within the ECHAM5 simulations a clear improvement with increasing resolution can be found: in particular the increase of vertical levels results in lower errors.

Earlier a relation between precipitation error and jet stream error was anticipated. The error pattern of the 500 hPa height field has often been brought in context with an under-representation of the occurrence of blockings in simulations. The daily data required to do a statistic of blockings were not available but in Figure 6 the effect of an underestimation of blocking events can be recognised. A typical European blocking consists of a Scandinavian anticyclone with cold air outbreaks to the west and east, e.g. over Spain and Turkey, where it will result in enhanced precipitation while there is reduced precipitation over Germany. Such a situation is typical for winter. In long-term means, missing blockings will result in enhanced precipitation over central Europe and reduced precipitation over the Mediterranean, such as found in Figure 6. An under-prediction of Scandinavian anticyclones results in a zonalisation of the circulation that was shown in Figure 1.

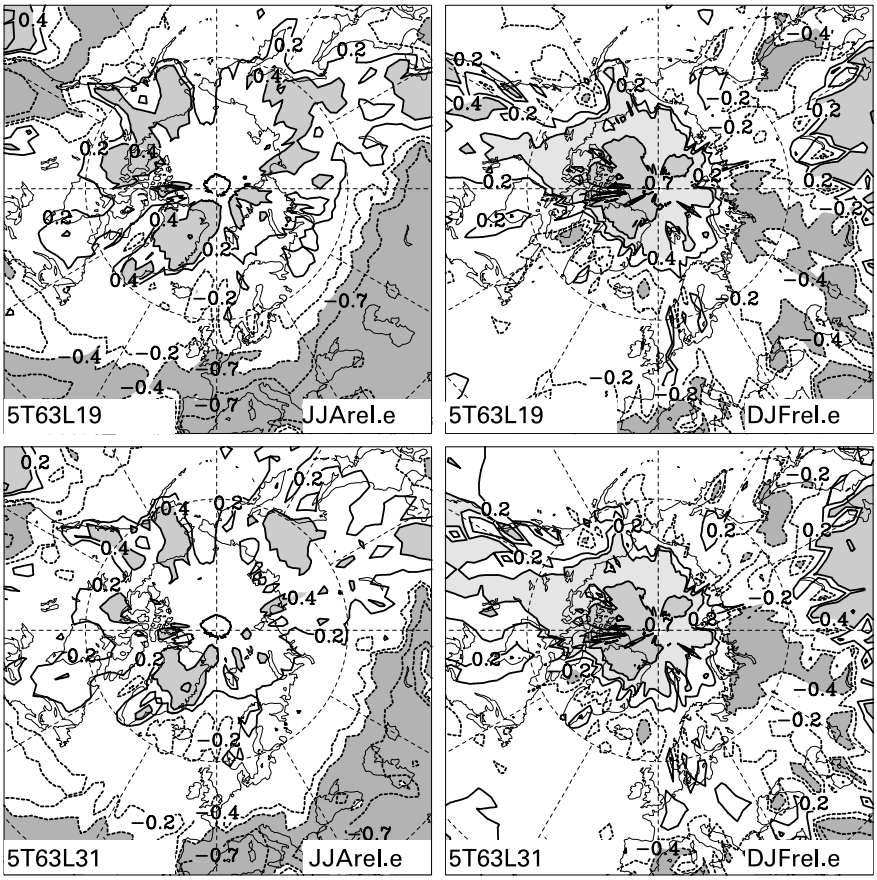

Figure 6 Precipitation errors (model GPCP), normalised by the maximum of both (relative errors). Contours at $\pm$ 0.2, 0.4, 0.7, shading for >0.4 and <-0.4, negative contours are dashed. Fields are smoothed to T21 resolution

In Figure 8 the annual cycles of precipitation averaged for selected river catchment areas are shown. The full list of rivers, which have been investigated, can be seen in Table 2. T106 and T159 are very similar and often nearest to GPCP, e.g. Baltic catchment, for northern Europe. Over North America (Mackenzie) T106 with both vertical resolutions is on the high side with respect to precipitation amounts. For eastern Siberia (Indigirka) all simulations give too much precipitation in summer with perhaps the best values from 4T42. T106 and T159 are best in the ECHAM5 family. Over western Siberia and eastern Europe (Ob) the 31 level model versions again provide the best estimates of the truth. For all regions the T42L and T63L simulations are mostly very similar and inferior to the others.

The summer maxima are simulated too early by a month for the Mackenzie, Churchill and most Siberian rivers in all simulations. For the latter the shift is strongest in T159 and T106. The enhanced precipitation in spring/early summer in the simulations go hand in hand with too early maxima of temperatures and LHFX (not shown). In particular the low resolution models have unrealistic relative minima in August–October for all rivers except Indigirka.

Over the whole Arctic region the simulations provide much more precipitation than GPCP (Figure 6). The differences exceed 70% of the model precipitation. The true precipitation in the Artic is very difficult to measure and there are hardly any stations. GPCP relies there on satellite observation but SSMI data are not usable over snow. TOVS need

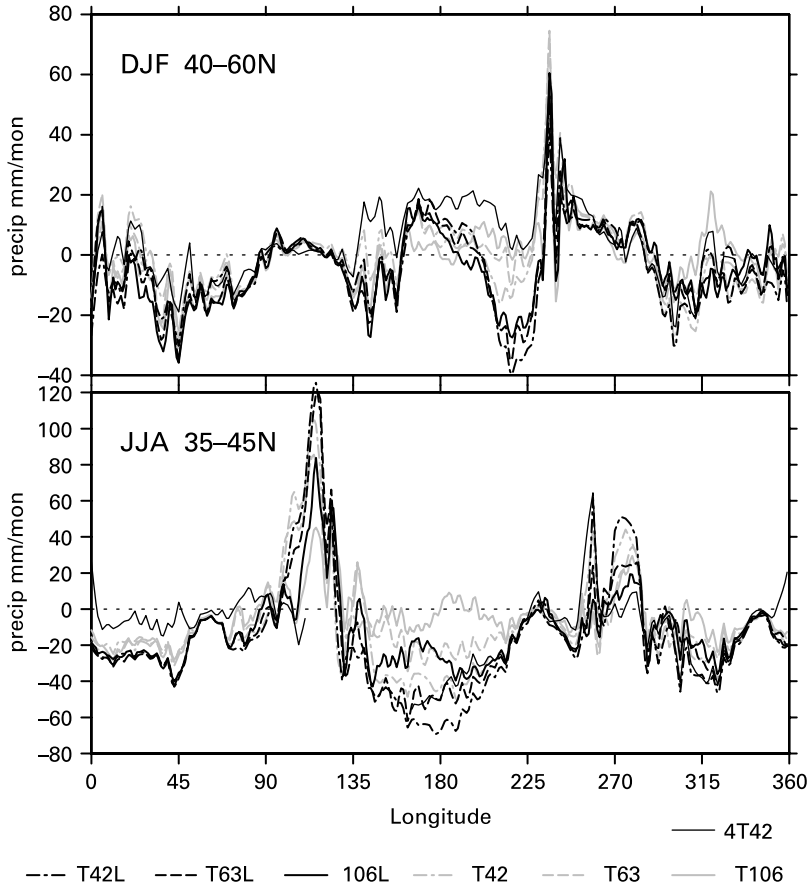

Figure 7 Precipitation errors (model GPCP) along a longitudinal belt

ground truth for calibration, which is not available. The only real ground observations are those published by Radionov *et al.* (1997) from some drifting ice sheet stations. These are probably the basis for other observational estimates although they cover a period before the GPCP analysis. All ECHAM5 simulations provide precipitation amounts exceeding all observational estimates throughout the year by 5–15 mm/month (Table 3). Especially large differences occur in May–June, perhaps connected with the same problem which leads to an early summer maximum over the river catchments. 4T42 agrees well with the observational estimates.

Latent heat flux

Estimates of LHFX from observational data are available from Mintz and Serafini (1992). These are partly based on the estimates of precipitation by Jaeger (1976). These estimates bear a large portion of uncertainty. In Figures 9 and 10 maps of LHFX in winter and summer are shown. For comparison the values from ERA40 are also presented. In both seasons the Mintz–Serafini and the 4T42 simulations provide the extremes, in winter Mintz–Serafini has the lowest and in summer the highest values while 4T42 shows just the opposite. In winter the ERA40 values are intermediate between the ECHAM5 model simulations and Mintz–Serafini and there is only little variability within the ECHAM5 values. In summer two classes are found over Europe/Siberia, the 19 level models give less LHFX than the 31 level models (T106, not shown, is similar to T63), recognisable from the extent of the

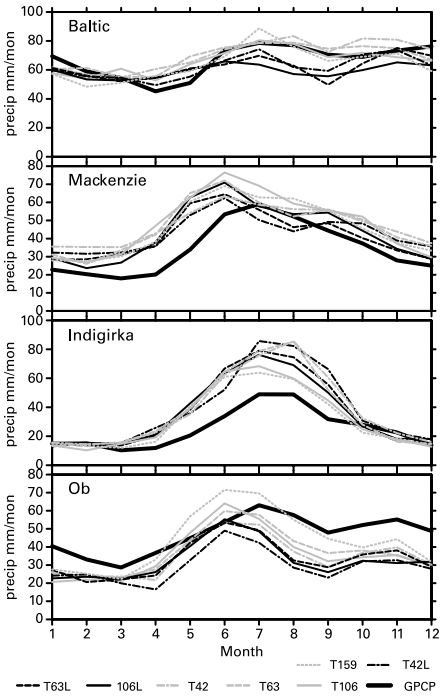

Figure 8 Annual cycles of precipitation for some selected river catchment areas

30 mm/month contour (light shaded area). The ERA40 values are nearer to the 19 level model simulations while Mintz–Serafini is nearer to the 31 level model values. In summer for eastern Canada the 19 level ECHAM5 simulations have LHFX values exceeding any other estimate. However, these differences are of the order of 15 mm/month, which are small compared to the differences which were found for the precipitation.

For selected river catchments the annual cycles of LHFX have been investigated (not shown). The somewhat surprising result is the similarity between all simulations and the ERA40 data. The Mintz and Serafini (1992) estimates show the strongest annual cycle, as already seen from Figures 9 and 10. Often 4T42 is nearest to ERA40 except over Siberia. Within the ECHAM5 simulations one finds mostly a stronger annual cycle in T106 and T159. The shift of the summer maximum towards spring, seen for the precipitation and surface temperature, can be found as well for the latent heat flux. There is a stronger annual cycle in T42L than in 4T42.

Earlier it was shown that the differences between the model simulations are small compared to the amplitude of the annual cycle. If, however, the annual means are considered (Table 2), differences of up to 20% between different ECHAM5 simulations are found, extremely for the Ob. The latent heat flux over the river catchments has a clear trend to increased fluxes with higher horizontal and vertical resolution, which seems to reach a maximum with T106 and a minimum with T42L. Because of the uncertainty in our knowledge of the truth it is not clear if this trend means also an improvement. For the same reason it is hard to judge which of the simulations is more realistic, except that the excessive values over E. Canada in summer in the 19 level models are less likely.

River discharge

For comparing model data of *P-E* (precipitation minus evaporation) and observations of river discharge we restrict ourselves to annual means because precipitation is stored in

Table 2 Area and time means of LHFX for selected river catchment areas in different ECHAM simulations, ERA40 and the Mintz and Serafini (1992) LHFX climatology. The period for averaging is quite inhomogenic as the ECHAM4 simulations cover 1951–1994, ECHAM4 and the ERA40 data used cover 1979–1999, and the Mintz–Serafini climatology includes long-term means older than 1990. Units: mm/month

	T159	T106	T106L	T63	T63L	T42	T42L	4T42	ERA	Mintz Seraf.
Baltic inflow	38	39	37	38	37	37	36	35	35	39
Rhine	48	49	46	48	47	48	45	51	44	52
Elbe	45	46	43	46	43	45	42	43	42	48
Vistula	46	46	43	47	44	45	42	41	44	43
Odra	44	46	42	45	42	44	38	40	44	45
Goeta–Vaenern	41	42	39	42	41	40	40	39	35	42
Mackenzie	31	32	30	31	29	29	27	33	31	28
Churchill	39	40	38	38	36	36	33	38	31	30
Nelson	43	42	40	42	41	42	39	41	38	36
Kolyma	21	21	20	20	18	19	18	18	22	22
Indigirka	18	18	18	17	16	16	16	14	20	21
Lena	29	29	28	27	27	26	25	24	24	28
Jenissei	31	32	30	30	29	27	26	29	29	31
Ob	32	29	26	29	26	26	23	30	32	31
N.Dvina	34	35	33	33	31	30	29	32	30	34
Neman	44	46	43	45	43	42	40	36	42	42
Mean	37	37	35	36	34	35	32	34	34	36

the ground as soil moisture and on the ground as snow and released at the mouth of the rivers with quite a delay. The model mostly already parametrises the processes leading to such delays but bears some uncertainty which we do not want to investigate here.

In Figure 11 selected time series of annual means of simulated *P-E* and observed river discharge (Dümenil Gates *et al.* 2000) are shown. The river discharge has been divided by the catchment area of the river to obtain the units of mm/month. There is only a short overlap in time between the observations and the ECHAM5 simulations. The 4T42 simulations are available for much longer times and will be used more extensively below. For 4T42 an average of 6 individual runs are shown here and should therefore be less variable in time. As all 6 runs were forced with the same SST, it could be assumed that the remaining variability would be a signal of SSTs on the variability of *P-E* over the single river catchment areas which will be discussed below.

Table 4 gives an overview of *P-E* long-term means for different catchments in both simulations and observations. The reasonably well-simulated values are marked with a (*), i.e. within 10% of the observed values. Those marked with (+) or (−) deviate widely (more than 50%) from the observations. The simulation T106L comes out best with 6 good marks

Table 3 Arctic precipitation north of 85°N. Units: mm/month

	Radionov	Legates	GPCP	T159	T106	T106L	T63	T63L	T42	T42L	4T42
DJF	10	5	4	11	12	12	12	12	12	13	8
MAM	8	11	7	15	16	16	15	16	14	15	9
JJA	21	19	23	33	33	34	31	31	28	24	17
SON	17	20	14	21	24	22	24	22	24	22	17

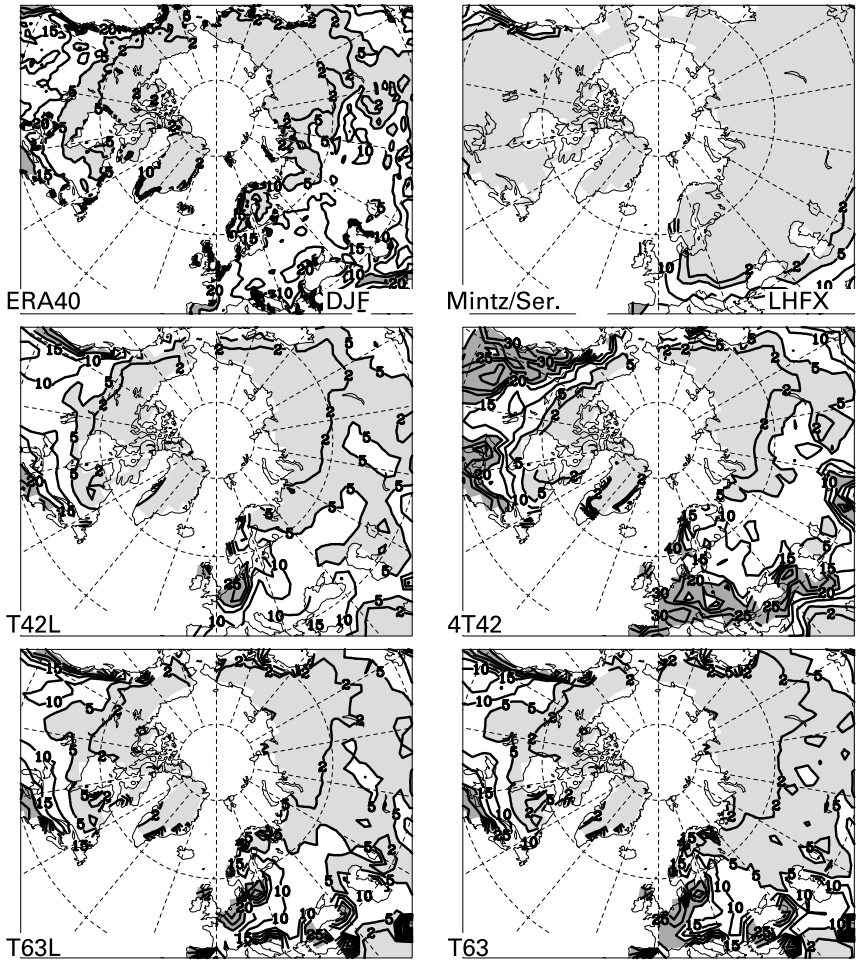

Figure 9 Latent heat flux (LHFX) over continents in DJF. Contours at 2, 5, 10, 15, 20, 25, 30, 40, 50 mm/month, shading for >20 mm/month and <5 mm/month

but one bad mark from 16 possible. 4T42 got 7 good marks but has also 3 bad ones. As in the previous discussion T42 and T42L got the worst scores.

For the markings in Table 4 one has to consider that observations and simulations often do not cover the same period. Trying to do a subjective judgment from plots similar to Figure 11 concentrating on the overlapping period, one might come to different conclusions. Churchill and Nelson have not been evaluated in this respect because the overlapping period is very short and the variabilities are very large.

Mostly the subjective judgement agrees with the objective markings in Table 4 only that the subjective method allows more good markings. For Elbe and Odra the T63 simulation was marked as extremely erroneous from the long-term means while the subjective judgement regards it as well done. The reason is an abrupt increase of the simulated P-E after the end of the river discharge observation.

The summing up of the marks at the end of Table 4 exhibits a clear trend in the ECHAM5 simulations with the best results for the high horizontal resolution simulations and the worst for the low-resolution ones, while the vertical resolution seems to be of less importance. The 4T42 simulations obtain quite good marks.

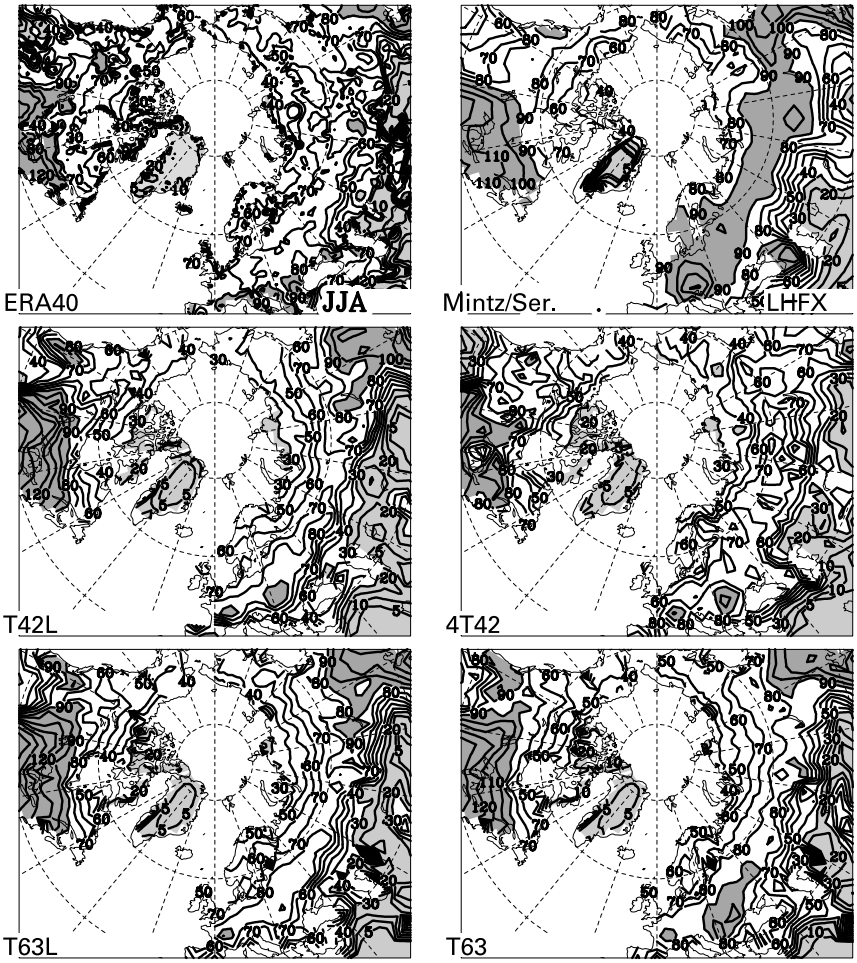

Figure 10 Latent heat flux (LHFX) over continents in JJA. Contours at 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 mm/month, shading for >90 mm/month and <30 mm/month

Interannual variability of precipitation

In Table 5 the variances of annual mean precipitation in the simulations and observations for different river catchment areas are presented. The variances of GPCP exceed mostly that of CRU. This might result from the analysis method used in the case of missing station observations of precipitation. Then the CRU analysis tends towards climatological values while GPCP uses, in such cases, estimates from satellite observations. The different sampling periods may be an issue as well. The variances of the mean of 6 single 4T42 experiments are shown. The mean of the 4T42 simulations provides clearly lower variabilities than the single simulations (not shown), as would be expected if there is only a weak common forcing which is the SST. A strong impact on the precipitation over the river catchment areas would lead to small differences between single simulations and the variance of the mean simulation could be, in the extreme case, the same variance as that of a single simulation. But all the variances of the mean simulation are clearly lower than that of the single simulation, mostly by a factor of 2, near to a theoretical value of $\sqrt{(6)} = 2.5$ in the case

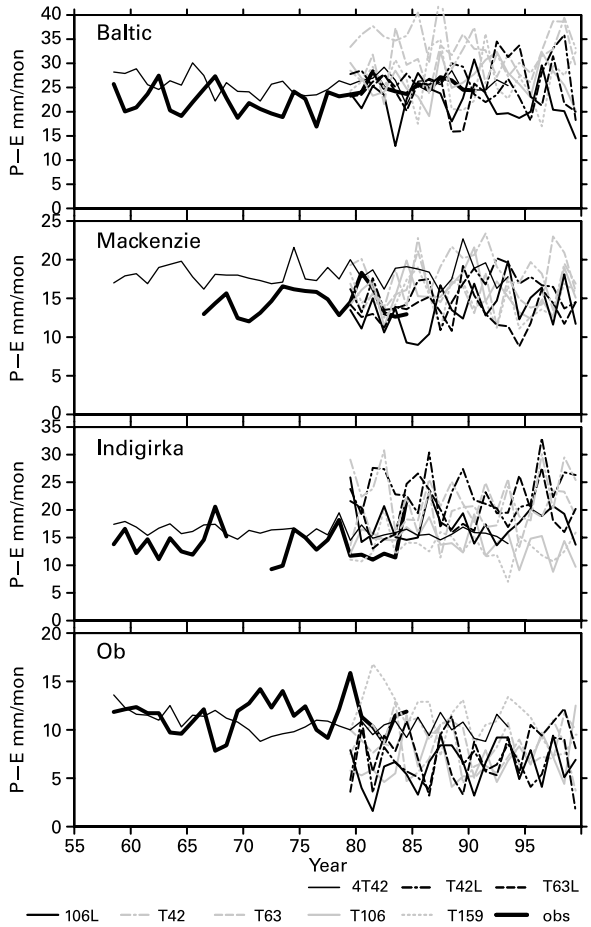

Figure 11 Time series of annual mean observed river discharge and simulated *P-E* for some selected river catchment areas. The river discharge has been divided by the catchment area to obtain the same units as *P-E*

of random variability. Accordingly only low correlations between the analysed and simulated precipitation are also found, which will be discussed below.

The variances of single simulations, except T42L for some rivers, clearly exceeds that of CRU or GPCP, showing that the models provide too much year-by-year variability, which was already obvious in Figure 11. There is a clear tendency to higher variability with increasing vertical resolution. The new features in the formulation of the ECHAM5 model compared to ECHAM4 seem to be less important for the variability as T42L lies mostly within the spread of the 4T42 simulations.

Correlations between analysed annual mean precipitation and the simulated values were carried out. Correlations done in the study are always anomaly correlations. We concentrate first on the mean 4T42 run because higher statistical significance can be expected. Higher values (>0.18) can be found for Rhine, Indigirka, N. Dvina and Neman. For these the correlations between simulation and analysis of seasonal means show that the higher correlation in the annual means results mostly only from one season.

The SST is the only common external forcing of the atmosphere in reality and simulation which can be responsible for the positive correlation between observations and simulations. For both precipitation data sets a strong signal over the Pacific is found with positive correlations indicating connections with developing El Niños. A further signal can be found

Table 4 Area and time means of *P-E* for selected river catchment areas in different ECHAM simulations and observations. LeSe results from the difference of the Legates and Willmott (1990) precipitation and the Mintz and Serafini (1992) LHFX climatology. The period for averaging is quite inhomogenic. Reasonably well simulated values are marked with a (*). Those marked with (+) or (−) deviate extremely. By (#) good performance from subjective impression for the overlapping period is indicated. Units: mm/month

	T159	T106	T106L	T63	T63L	T42	T42L	4T42	LeSe	Obs. river discharge
Baltic inflow	28 #	28 #	23*#	32	26 #	34 +	28 #	27#	24*	23
Rhine	47 #	42*#	26	49	32	49#	26	41*#	36*	39
Elbe	21 #	23 #	13	27 + #	16*	29 +	18	26 +	23	16
Vistula	22 #	22 #	10#	25	13 #	30 +	13 #	23 + #	20	15
Odra	20 #	19 #	10	21 + #	10	21 +	9	17 #	18	14
Goeta-V	32*#	23	19	36 #	25	43	33	32*#	22	29
Mackenzie	16 #	17 #	15*	17#	14*	20	17 #	19	4	14
Churchill	10*	12	9*	11*	8	13	7	10*	7	10
Nelson	8	8	5	7*	6	8	5	7*	6	7
Kolyma	19	21	24 +	27 +	28 +	34 +	35 +	21	8	15
Indigirka	14*#	15*#	18	20#	21 +	23 +	24 +	17#	2−	14
Lena	19*#	21*#	20*#	22 #	22 #	24	20*#	14	5−	19
Jenissei	23	22 #	19*#	20*#	19*#	19*#	15	18*#	7−	19
Ob	12*#	8	7	9#	9	10*	8	11*#	14	11
N.Dvina	29	26	25*#	30	33	30	34 +	25*#	19	23
Neman	21 #	23 #	13	25 +	15*	29 +	18 #	24 +	22	16
good (*)	4	3	6	3	4	2	1	7		
bad + or −	0	0	1	4	2	7	2	3		
good subj. #	11	10	5	8	4	2	6	9		

Table 5 Variances of annual mean precipitation over different catchments in simulations and analyses using maximum lengths of sampling, CRU and 4T42 1951–1993, GPCP and ECHAM5 simulations 1979–1999. For 4T42 the variance of the mean simulation and the extremes of the single run are given

	GPCP	T159	T106	T106L	T63	T63L	T42	T42L	CRU	Mean 4T42
Baltic	3.9	6.2	5.1	4.9	4.8	5.7	4.1	2.7	2.7	2.0
Rhine	8.5	10.5	11.3	7.0	15.8	12.4	8.2	5.5	6.1	3.8
Elbe	8.2	7.5	10.1	6.2	11.3	6.8	7.2	4.3	5.5	3.1
Vistula	6.1	10.0	8.4	7.1	5.7	5.8	6.6	4.3	4.0	3.6
Odra	7.2	8.4	8.8	7.6	7.8	6.0	5.8	3.9	5.0	3.0
Goeta-V.	5.5	9.6	6.7	5.9	11.6	11.5	6.8	7.9	4.9	3.5
Mackenzie	2.9	3.1	3.3	3.7	2.7	3.1	3.2	3.6	2.1	1.3
Churchill	2.5	5.3	6.0	6.9	3.7	4.5	5.7	4.2	2.4	2.0
Nelson	3.0	4.0	6.8	4.5	4.8	3.6	4.8	3.8	2.5	1.5
Kolyma	4.7	4.8	3.8	5.2	4.5	5.8	6.3	4.9	2.6	1.7
Indigirka	3.7	3.3	3.1	3.1	4.3	4.0	5.0	4.0	2.4	1.0
Lena	2.6	3.1	3.5	3.5	5.0	3.5	4.0	3.8	2.1	1.0
Jenissei	2.4	3.7	3.9	3.8	3.3	3.6	3.2	4.3	1.7	1.4
Ob	3.7	2.8	3.6	4.1	3.3	3.5	2.5	3.5	2.5	1.3
N.Dvina	5.6	5.9	7.2	7.9	5.1	6.4	6.1	5.1	2.7	2.2
Neman	5.8	9.1	7.8	8.9	7.6	8.4	6.3	5.7	4.1	2.9
mean	6.0	7.9	7.7	6.1	8.5	7.3	6.0	4.6	4.3	2.9

over the northern Atlantic especially along the European–African coast. These correlations are strongest when using a mean of 6 simulations of the 4T42 model.

Also the ECHAM5 simulations covering only the period 1979–1999 show the same signal. Statistical significance is obviously a problem in this investigation. From the fact that the same response pattern emerges from most of the simulations suggests that there is some forcing from the oceans on the precipitation over the Rhine, Indigirka, N. Dvina and Neman catchments. From the available data we cannot decide if the north-eastern Atlantic or the Pacific is dominating. The former connection is easier to understand and the latter agrees with other studies (Bengtsson *et al.* 1996; Fraedrich and Müller 1992) that demonstrated impacts from ENSO on Europe.

Summary and conclusion

The superiority of the higher-resolution ECHAM5 models has been shown in many respects. For dynamical quantities the vertical resolution has a decisive impact but also increased horizontal resolution leads mostly to improvements. However, T63, T106 and T159 of the 31 level model deviate only slightly from each other and T159 performs in few respects worse than the other high-resolution versions, possibly due to the gravity wave drag which still has to be optimally tuned for the extremely high resolution. There are also deviations from this general rule, e.g. the jet strength over Egypt and the zonal mean wind in the upper tropical troposphere are best in the 19 level models. 4T42 often has lower errors than T42L but larger ones than the 31 level and higher horizontal resolution versions of ECHAM5.

For precipitation the two 19 level model versions of ECHAM5 are clearly of lower quality than the other versions. T106 and T159 are very similar and mostly nearest to observations. The summer maxima of precipitation, surface temperature and latent heat flux are simulated for several river catchment areas too early by about one month. This shift is strongest in the T106 and T159 model. Another problem with the annual cycle of precipitation is an unrealistic indication of a relative minimum in August to October, especially in the

low-resolution ECHAM5 models. The precipitation of the ECHAM5 simulations over the Arctic region exceeds all observational estimates by 5–15 mm/month, strongest in May–June. But it has to be noted that in this region the uncertainties of the observational estimates are relatively large.

The latent heat flux over the river catchments has a clear trend to increased fluxes with higher horizontal and vertical resolution, which seems to reach a maximum with T106. Because of the uncertainty in our knowledge of the truth it is not clear if this trend also means an improvement. However, large fluxes in summer over eastern Canada in the 19 level ECHAM5 simulations seem to be excessive. In the comparison of *P-E* with observed river discharge only the horizontal resolution seems to be important for their realism, again giving best results for the high-resolution models. The 4T42 simulations obtain quite good ratings.

The year-by-year variability of the simulations is too high, which is more pronounced for the higher-resolution versions. Especially strong impacts are found from the vertical resolution. Differences in the parametrisation schemes of the ECHAM model versions are less important. The interannual variability of the LHFX is much smaller than that of precipitation and therefore the results shown for precipitation do apply also for the simulated river discharge. Some forcing of ocean temperature anomalies on the precipitation over the Rhine, Kolyma and Indigirka catchment areas has been found.

The relative humidity, especially at higher levels, is steadily dropping with higher resolution. This may create some problems for nesting high-resolution models into global models or for forcing limited area models with global model data because at the boundaries excessive precipitation can be expected.

The low-resolution ECHAM5 T42 L19 model is in many respects inferior to the ECHAM4 model with the same resolution, which is probably due to the fact that the latter has been tuned extensively over a long period of operation. There is hope also that ECHAM5 can be improved by small adjustments after it has been in service over a longer period.

Acknowledgements

This study would not been possible without the support by many colleagues at the Max Planck Institute for Meteorology and at the German Climate Computing Centre (DKRZ).

References

Arpe, K. (1990). Impacts of changes in the ECMWF analysis-forecasting scheme on the systematic error of the model. In: *ECMWF Seminar on 10 Years of Medium-Range Weather Forecasting, 4–8 September1989*, ECMWF, Reading, UK, pp. 69–114.

Arpe, K., Bengtsson, L., Hollingsworth, A. and Janjic, Z. (1976). A case study of a ten day prediction, *ECMWF Technical Report* 1, 105 S.

Bengtsson, L., Arpe, K., Roeckner, E. and Schulzweida, U. (1996). Climate predictability experiments with a general circulation model. *Climate Dyn.*, **12**, 261–278.

Dümenil Gates, L., Hagemann, S. and Golz, C. (2000). Observed historical discharge data from major rivers for climate model validation. *Max Planck Institute for Meteorology, Hamburg, Report* no. 307.

Fraedrich, K. and Müller, K. (1992). Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.*, **12**, 25–31.

Gibson, J.K., Kållberg, P., Uppala, S., Hernandez, A., Nomura, A. and Serrano, E. (1997). *ERA Description. ECMWF Reanal. Proj. Rep. Ser.* 1. European Centre for Medium-Range Weather Forecasting: Reading, UK.

Hagemann, S., Arpe, K. and Bengtsson, L. (2005a). Validation of the hydrological cycle of ERA40. *Reports on Earth System Science* 10, Max-Planck-Institut für Meteorol., Hamburg, Germany.

Hagemann, S., Arpe, E. and Roeckner, E. (2005b). Evaluation of the hydrological cycle in the ECHAM5 model. *J. Climate Special Issue on MPI-M models*, in press.

Huffman, G.J., Adler, R.F., Arkin, P.A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., Joyce, R.J., McNab, A., Rudolf, B., Schneider, U. and Xie, P. (1996). The Global Precipitation Climatology Project, (GPCP) combined precipitation data set. *Bull. Am. Meteor. Soc.*, **78**, 5–20.

Jaeger, L. (1976). *Monatskarten des Niederschlags fuer die ganze Erde.* Ber. d. Dt. Wetterdienstes 139, Offenbach, Germany.

Kuo, H.L. (1974). Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. Atmos. Sci.*, **31**, 1232–1240.

Legates, D.R. and Willmott, C.J. (1990). Mean seasonal and spacial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127.

Mintz, Y. and Serafini, Y.V. (1992). A global monthly climatology of soil moisture and water balance. *Climate Dyn.*, **8**, 13–27.

New, M., Hulme, M. and Jones, P. (2000). Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.

Radionov, V.F., Bryazgin N.N. and Alexandrov, E.I (1997). The snow cover over the Arctic Basin. *Technical Report* APL-UW TR 9701. Applied Physics Laboratory, University of Washington, Seattle, WA (published 1996 in Russian: Gidrometeoizdat, St. Petersburg).

Rayner, N.A., Hearten, E.B., Parker, D.E., Folland, C.K. and Hacked, R.B. (1996). Version 2.2 of the global sea-ice and sea surface temperature data set, 1903–1994. *Clim. Res. Tech. Note*, CRTN74, Bracknell, UK.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U. and Schulzweida, U. (1996). The atmospheric general circulation model ECHAM-4: model description and simulation of present-day climate. *Max Planck Institute for Meteorology, Hamburg, Report* no. 349.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U. and Tompkins, A. (2003). The atmospheric general circulation model ECHAM5, Part I: Model description. *Max Planck Institute for Meteorology, Hamburg, Report no.* 349.

Simmons, A.J. and Gibson, J.K. (2000). The ERA-40 project plan, *ERA-40 Project Report Series* 1. ECMWF: Shinfield Park, Reading, UK.

Taylor, K.E., Williamson, D. and Zwiers, F. (2000). The sea surface temperature and sea ice concentration boundary conditions for AMIP II simulations, *PCMDI Report No.* 60, LLNL, Livermore, CA.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large scale models. *Mon. Weather Rev.*, **117**, 1779–1800.