Comparison of precipitation and temperature fields in different data sets used for evaluating Regional Climate Models at the Hungarian Meteorological Service

1. Introduction

To be acquainted with the climate or simply the climate change of our future we must know what is hidden in the past. Thanks to our computerized world we have more and more available data sets.

Many studies were made comparing reanalysis and gridded data sets for the whole globe or for specified regions. We have a reanalysis data at the European Center for Medium range Weather Forecasting, ECMWF (ERA-40), different spatially-gridded data sets from Climatic Research Unit, CRU (CRU TS 1.2 and 2.1), a Hungarian gridded data set (HUGRID) derived from observed data and a raw, scattered station data. We make a comparative discussion related to cumulated precipitation and temperature at 2 meters.

In the first section we summarize the details, content of our available data sets. After that we compare them systematically for two regions: the smaller one is a rectangle region of the Hungarian territory and the larger one is an extended region including the Alps, the Carpathians and the Italian Peninsula. First the ERA-40, CRU TS 2.1 and the CRU TS 1.2 databases will be compared in the bigger region. Secondly the smaller region will be examined using the CRU TS 1.2, the HUGRID and the station data. Finally we draw the consequence. Our main goal after this study is to disapprove the worst ones and start working with the best ones for our smaller and bigger region in the Regional Climate Models (RCMs) applied at the Hungarian Meteorology Service (HMS).

2.1 About the ERA-40 reanalysis data set

The most important purpose of the ERA-40 project was to create high-quality global atmospheric analysis. Observations from September 1957 to August 2002 were analyzed using a version of the ECMWF data assimilation system. The databases are archived in the original T159 resolution or on the corresponding 'N80' reduced Gaussian grid. For plotting purposes the fields were interpolated onto a regular $1.125^{o}X1.125^{o}$ grid (Kallberg et al., 2005, Simmons et al., 2004). An overall report of ERA-40, with extended references to further documentation of the project, is given by Uppala et al. (2005).

Generally publicists go in for temperature trends at two meters. What we have as raw database is a set of analyses of temperature at a height of two meters for the observing times 00, 06, 12 and 18 UTC stored in Kelvin in the format of gridded-binary (grib) files.

"*Precipitation is not part of the primary variables that are reanalyzed in the ERA-40 data set, the precipitation was extracted from the ERA-40 prediction runs that are started at 00:00 and 12:00 GMT each day from the reanalyzed state of the atmosphere at this point in time. In order to reduce spin-up effects, the initial 12 hours of each prediction run were discarded and a sequence of 12 hourly accumulated precipitation fields were generated from the interval 12 to 24 hours after the start of each prediction run.*" (Crochet, 2004). We have this derived precipitation in different assortments of it, like convective, background precipitation and snow also stored in grib files for the times 00, 06, 12 and 18 UTC.

2.2 About CRU TS 1.2 and CRU TS 2.1 data sets

Time-series up to 1995 (1998 for temperature and precipitation) was already available at 0.5^{o} resolution (the CRU TS 1.0 and 1.1 data sets). These existing grids were extended to 2000 for Europe only, by using the methodology of New et al. (2000), then smoothed onto a 10' resolution for Europe (Mitchell et al., 2004). So the CRU TS 1.2 data set is based on previously constructed grids.

The creation of the CRU TS 2.1 data set (Mitchell and Jones, 2005) is rather different: it is derived directly from revised monthly station data and uses all available monthly station average of mean temperature and precipitation from land regions. Full details of the sources are given as for temperature in Jones and Moberg (2003) and as for precipitation in New et al. (2002). This data set replaces the previous versions (namely CRU TS 1.0 and 1.1) and employs a similar method to New et al. (2000), and its renamed to CRU TS 2.1. If we were going to use CRU TS 1.2 instead of CRU TS 2.1, then we should not disregard the facts of the poorer version of observed record and the higher resolution in space.

The CRU TS 2.1 database covers global land surface at half degree resolution while the CRU TS 1.2 only covers Europe, exactly: 34.0^{o}N:72.0^{o}N and 11.0^{o}W:32.0^{o}E, at 10' resolution. Both data sets comprise 1200 monthly grids of observed climate between 1901 and 2000 for five variables: cloud cover, diurnal temperature range, precipitation, temperature and vapor pressure. For now on call the CRU TS 1.2 as CRU10' and the CRU TS 2.1 as CRU0.5.

2.3 about the Hungarian gridded data set (HUGRID)

The HUGRID was created by the Meteorological Interpolation based on Surface Homogenized Data Basis (MISH) method, which was developed at the HMS for the spatial interpolation of different observed surface meteorological elements (Szentimrey et al., 2005). This method needs a homogenized data set, which is achieved by a method called Multiple Analysis of Series for Homogenization (MASH) describing by Szentimrey (1999).

Our disquisition will focus on monthly-accumulated precipitation and monthly average temperature at two meters. Each month is stored in a format of separated ASCII files. The data set contains 71 points in longitude-way starting at 16.0^{o}N and finishing at 23.0^{o}N and 30 points as for latitudes from 45.7^{o}N to 48.6^{o}N with the resolution of 0.1^{o}. The units of variables are mm and ^{o}C and the available time series extends from 1961 to 2000.

2.4 About the Hungarian observed database (STATION)

This raw database consists of 20 observations scattered within the region of the Hungarian border. Each file contains the exact coordinate of the stations (which is changeable due to war, inactivity), the daily average temperature depending on 8 measurements, or if it does not exist then the maximum and minimum temperature, as well as the daily precipitation for the period of 1961 and 2000. Neither data is homogenized.

Only six stations are complete for the full period looking at both variables: Budapest, Debrecen, Miskolc, Pécs, Szeged and Szombathely. The others including the six complete stations will be analyzed in virtue of precipitation and temperature in a shorter period, from 1971: Békéscsaba, Nagykanizsa, Nyíregyháza, Siófok, Szolnok and a station from the height around 1000 meters, called Kékestető. Sopron, Keszthely and Baja will only count as stations observing temperature. The average temperature of the last two stations with Győr will be calculated as the average of the maximum and minimum observed value of the day. Hence for the period of 1961-2000 six while for the period of 1971-2000 sixteen stations will be plotted and as for precipitation the numbers will be six and thirteen.

2.5 Summary of the data sets

Table 1. Short of available data sets at the HMS (only the used fields are shown!)

	Format	Period	Resolution	Variables
STATION	ASCII 1 station/file	daily data 1961-2000	scattered, 20 stations exact coordinates	t2m, t2max, t2min (C) precip (mm/day)
HUGRID	ASCII 1 month/file	monthly data 1961-2000	0.1^o in 16.0^o:23.0^o E 45.7:48.6 N	t2m (C) precip (mm/month)
ERA-40	GRIB 5 years/file	4 data daily 1957.09.01-2002.08.31	1^o global but 18^o:90^o N	t2m (K)
ERA-40	GRIB 1 month/file	4 data monthly 1961-2000	1^o global but 18^o:90^o N	convective, large-scale precip, snow (m/day)
CRU0.5	NetCDF	monthly data 1901-2002	0.5^o global (no ocean)	t2m (K) precip (mm/day)
CRU10'	ASCII 10 years/file	monthly data 1901-2000	10' Europe	t2m (tenth C) precip (tenth mm/month)

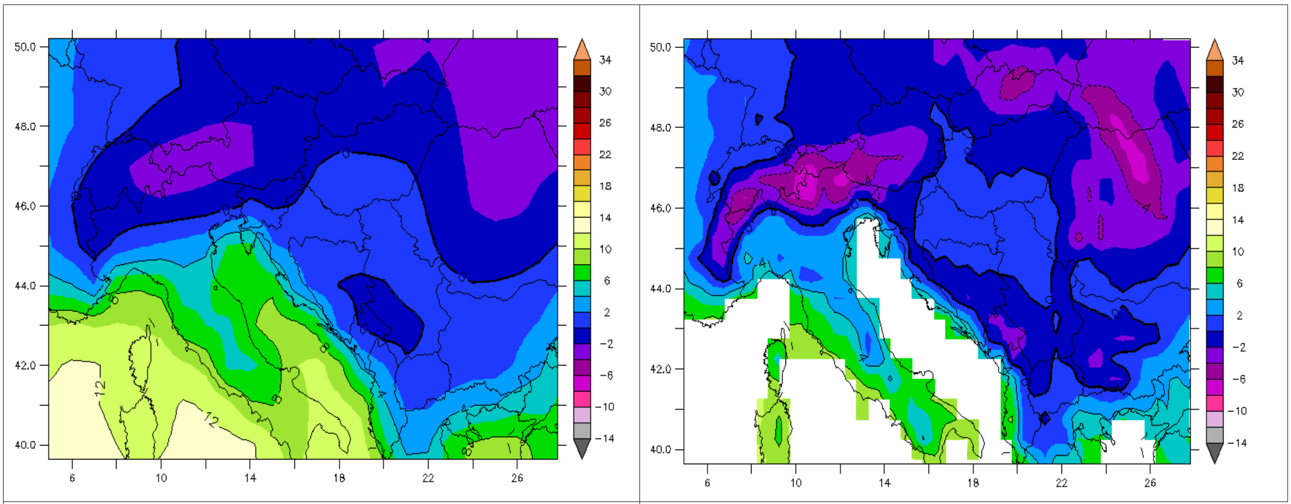
To make a statement we have to find the common climatological axis, the common file format and the same spatial resolution differently for the two chosen regions. As for the time resolution we produce a monthly climatology for decadal periods starting from 1961. Then we put thirty or forty years of climatology to use. Put in context, on the bigger region we have to interpolate spatially the databases onto at least the resolution of CRU10' while on the HUN-region we smoothen onto at least 0.1^o as its stored in the highest resolution database. The final short of the available data sets is shown in Table 1.

For the implementation we convert all the databases which are not already in to NetCDF (which stands for Network Common Data Form) file formats, which are machine-independent, self-describing, binary data formats standard for exchanging scientific data. The project homepage is hosted by UCAR (University Corporation for Atmospheric Research). The computations, analysis and visualizations are achieved in a program language, called FERRET provided by NOAA (National Oceanic and Atmospheric Administration). Further information can be reached on this page: http://ferret.pmel.noaa.gov .

3.1 Comparison of ERA-40, CRU0.5, CRU10 data sets

3.1.1 Absolute fields

All the plots that appears in the discussion of the bigger region will be represented within the region of the same used in climatological models applied at the HMS that means 4.94^o E : 27.82^o E and 39.64^o N : 50.20^o N. The territory for demonstration was chosen by Szepszo et al. (2007) for the purpose of selecting a region, which is suitable for all the models. The scale values vary between 0 mm and 400 mm as for the precipitation and in connection with temperature it changes from -14 °C to 34 °C. The range of colors goes all through the rainbow meaning proper values in the extremes.



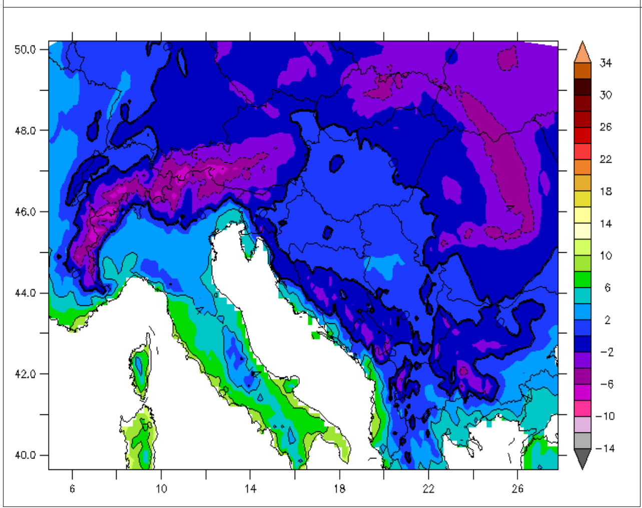
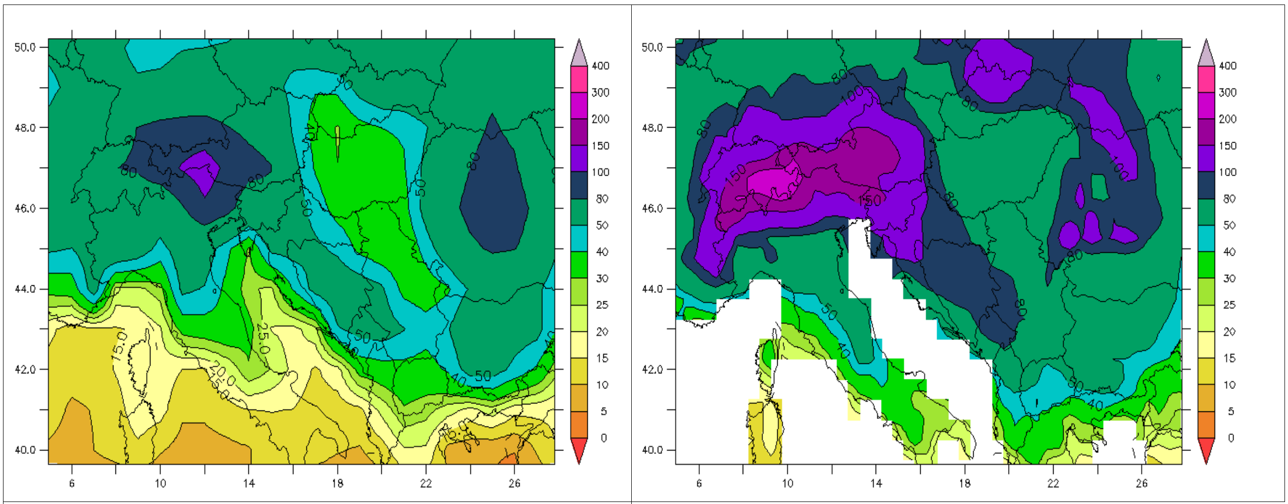


Figure 1.: Mean temperature in winter in the period of 1961-2000 in ERA-40 (top left), in CRU0.5 (top right) and in CRU10' (bottom) [°C]

As Figure 1. (mean temperature in winter) shows ERA-40 is not created for regional examination, its objectives were to create high-quality global analysis and to be used in studies of the general circulation, global change (Kallberg et al. (2005)). A blurred, but still recognizable difference can be found between the CRU0.5 and CRU10' beyond the resolution-related fact (the details are more remarkable on CRU10) is that the negative values are more existing on the fields of CRU0.5 mostly in position of Alps and Carpathians. This trend cannot be seen on plots of other seasons: negative extreme values (especially in the Alps and Carpathians) are discernable on CRU10'.

The monthly precipitation fields (in this case for summer) such as ERA-40, CRU0.5, CRU10' are shown on Figure 2. The point is that precipitation fields are scattered and the ERA-40 data set may not be accurate enough for such an analysis (Fil et al., 2005). Those points that are validated over a plain (like in Hungary) or over the region of the Alps represent lower precipitation values than the other two data sets. Still the relatively maximum values of ERA-40 are not positioned where the CRU databases are. However it is strange that in winter mostly we can find higher values in ERA-40 rather than in CRU data sets. Mainly the biggest source of error in ERA-40 is in connection with the orographic augmentation and rain-shadow effects, which are not well captured by the ECMWF model, may be due to its rough spatial resolution (Crochet (2004)).



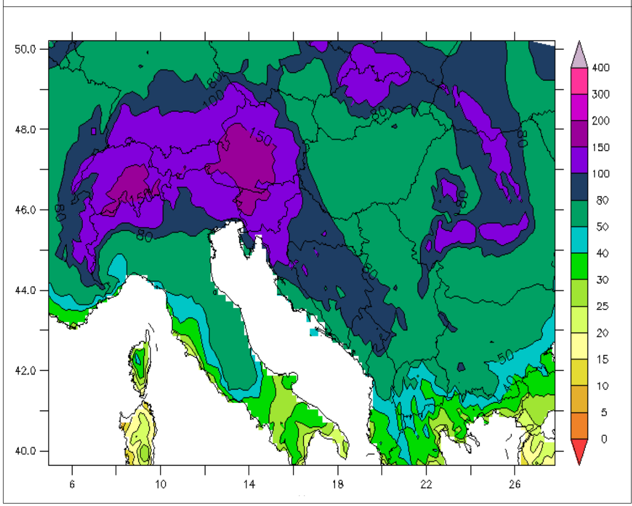


Figure 2.: Precipitation in summer in the period of 1961-2000 in ERA-40 (top left), in CRU0.5 (top right) and in CRU10' (bottom) [mm/month]

Above Hungary the fields are almost the same in the case of comparing the CRU0.5 and CRU10', however especially for the Alpine region the maximum rates are lower and less spacious in the more detailed CRU10' consistent with Mitchell et al. (2004). Similar, but less noticeably difference can be seen on the other seasons' fields (they are not shown in this study).

3.1.2 Difference fields

The difference plots were made by interpolating them mostly onto the resolution of 0.05^o in order to compare the data sets and to make the lines smoother. This does not count as a value changing operation in FERRET. We should point out that all the plots, which are mentioned here, were produced by changing the grid numbers both in longitude and latitude direction just to have nicer and smoother plots. Table 2. shows what kind of grid resolution the absolute and relative plots were calculated and then plotted on. Take a notice that only the CRU10' on the bigger region were calculated (and plotted) on its original resolution.

Table 2. Plotting resolution of the fields

Fields	*Resolution*
ERA40, CRU0.5, ERA40-CRU0.5, CRU0.5-CRU10'	0.1^o
CRU10' (CRU10' for HUN region)	10' (5')
HUGRID, HUGRID-CRU10'	0.05^o

As we saw in the last chapter that the ERA-40 especially in the Alpine, Carpathian region and near shores gives higher, or appropriately not as low values as the CRU0.5 does, can be noticed as difference fields on Figure 3. This occurs mostly because of the lower resolution. Finest differences can be found in the fields of autumn and winter (they are not shown here). On Figure 4. we can see different fields for precipitation: ERA-40 mostly gives lower values than CRU0.5, while in summer in the Eastern part of the region and in winter in the Western region of Hungary CRU0.5 is wetter than CRU10'.

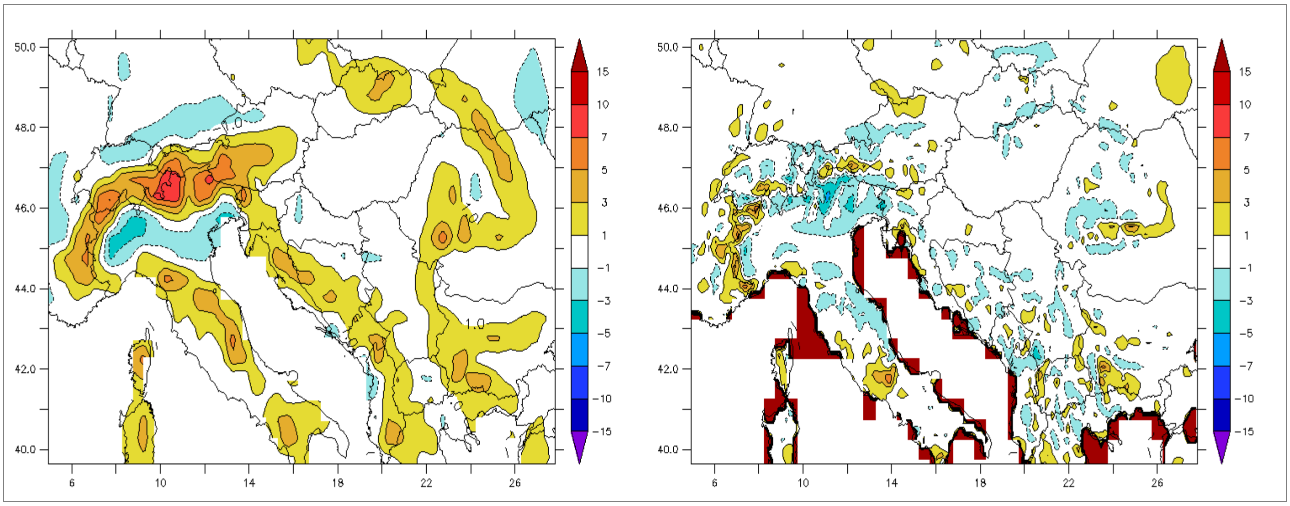


Figure 3.: Difference of temperature in spring in the period of 1961-2000 between ERA40 and CRU0.5 (left), and CRU0.5 and CRU10' (right) [^{o}C]

The problematic territories of the difference of CRU0.5 and CRU10' are also the higher elevated regions and the near-shores ones in the case of temperature. It's worth mentioning while in connection with the temperature the difference maps in different seasons are similar for the mountains, they are distinct in the case of precipitation (Figure 4.): in the Alpine region CRU0.5 is wetter, while near the shores and the Eastern Alps are drier. The Carpathians are ambiguous. We cannot say there are any tendencies in the plains either. Principally it may depend on the density of the used observations.

Examining Hungarian region we can declare that the differences of temperature are around zero in both cases, but in Northern Hungary CRU10' is warmer in all the seasons. The precipitation differences of ERA-40 and CRU0.5 except winter are always negative.

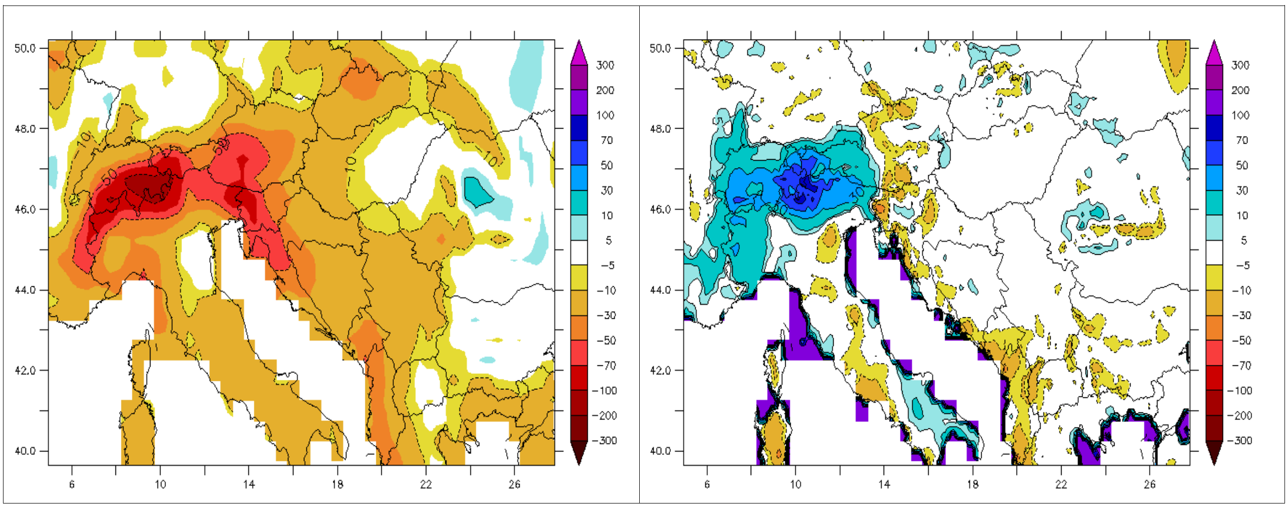


Figure 4.: Difference of annual precipitation in the period of 1961-2000 between ERA40 and CRU0.5 (left), and CRU0.5 and CRU10' (right) [mm/month]

3.2 Comparison of CRU10', HUGRID, STATIONS data sets

All the fields that appears in the smaller region will be represented in 16.0° E : 23.0° E and 45.7° N : 48.6° N. If we have a look at Figure 5., we can see that there is a slight difference for temperature between the CRU10' and the HUGRID data set. We can realize that negative values on the plot of difference only at higher elevations may due to the higher resolution of HUGRID. The other, more like positive values occur in the mountainous regions, which can be related to valleys. Mostly the difference is around zero.

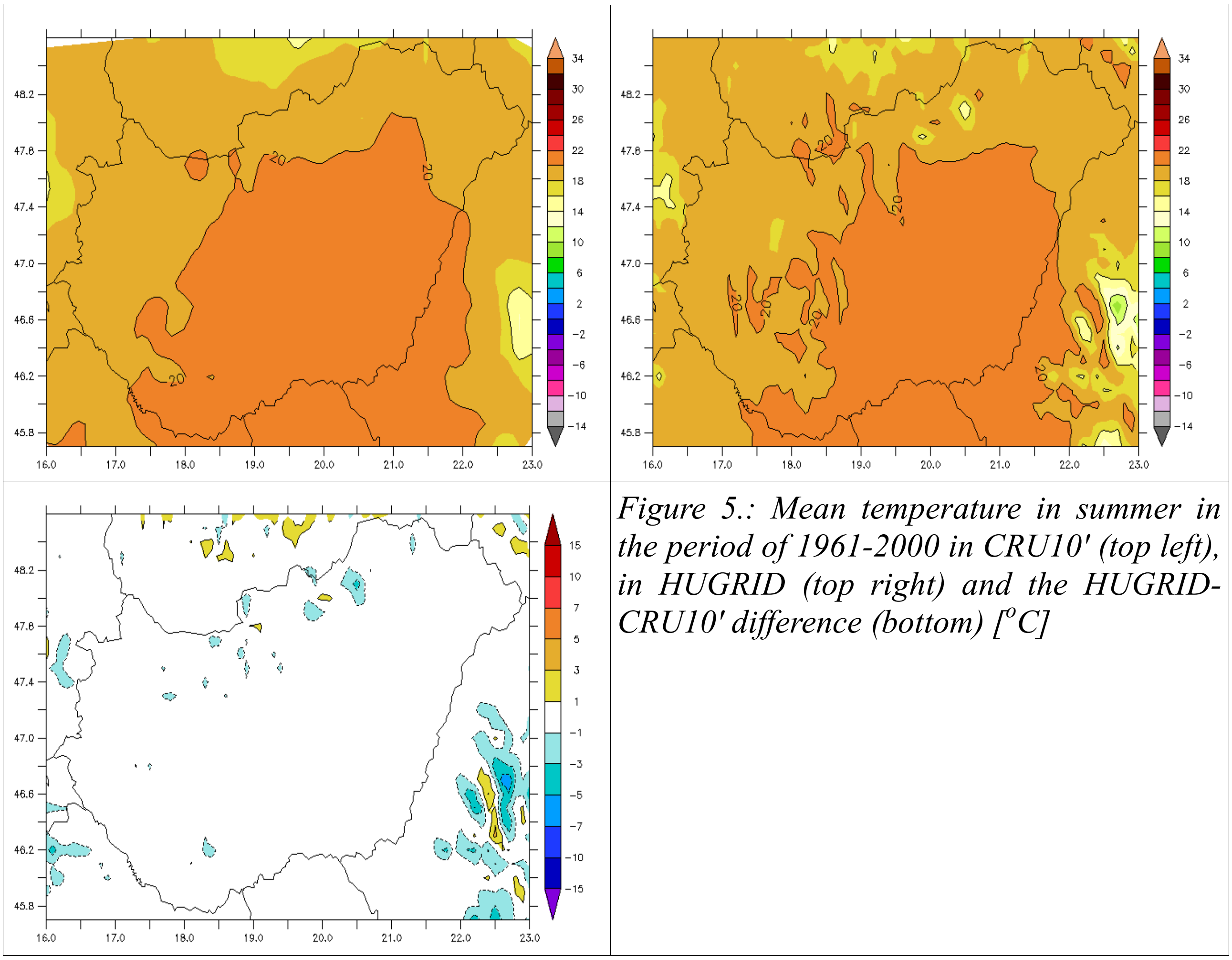
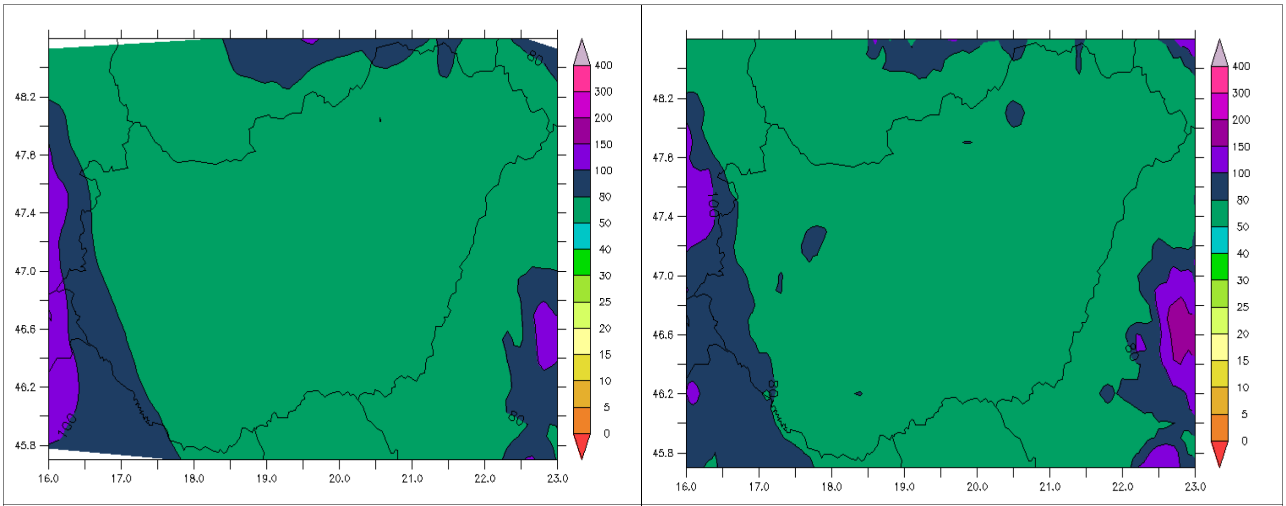


Figure 5.: Mean temperature in summer in the period of 1961-2000 in CRU10' (top left), in HUGRID (top right) and the HUGRID-CRU10' difference (bottom) [°C]

The precipitation has complicated difference field, even though on Figure 6. it is not that obvious and not expected due to the scale classification. The values on the absolute plots show that HUGRID has higher and maybe more real values due to the better original spatial resolution. Although they did not use any observations over the Hungarian borders through the assimilation process. Therefore it is raised the question whether the positive difference values in the Bihar-mountain are relevant and real.

On the difference plot negative values can be seen in the in northeastern and southwestern part of the region. The HUGRID gives more precipitation than the CRU10' for the mentioned Bihari-region and the Hungarian mountains relatively highly elevated. No apparent comments could be mentioned regarding the plain regions.



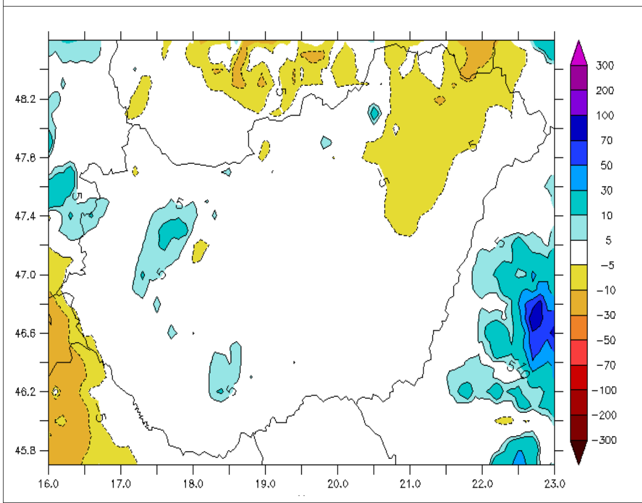


Figure 6.: Precipitation in summer in the period of 1961-2000 in CRU10' (top left), in HUGRID (top right) and the HUGRID-CRU10' difference (bottom) [mm/month]

3.3 HUGRID and STATION data sets

The last section of this study will focus on the Hungarian raw database (STATION) and the HUGRID one. One only has six/sixteen and six/thirteen points in space with valid values for the period of 1961/1971-2000 while the HUGRID has much more with the resolution of 10'. In order to compare these two databases we had to find the station's coordinates. Although it is changeable, we only used one with the precision of tenth degree as a station's location. As mentioned above neither temperature data is not homogenized.

Only one of the stations is at a higher elevation, so it's values differ much more from the HUGRID database than the other locations. Several colors and different-sized circles were made to show the difference values on Figure 7., which only shows the annual monthly precipitation and temperature difference for the period of 1971-2000. Except southwestern and northwestern Hungary the positive values are relevant (HUGRID is lower) if we have a look at the temperature plot. Only Kékestető, the mountainous observation contrasts much more.

The picture of precipitation is a bit complicated: near zero values can be seen except Szombathely, where the monthly precipitation is higher by 5-10 mm, and Kékestető, where it is lower by 5-10 mm for HUGRID. Higher differences can be seen in the field of summer (not shown here). The difference in winter tends to be positive.

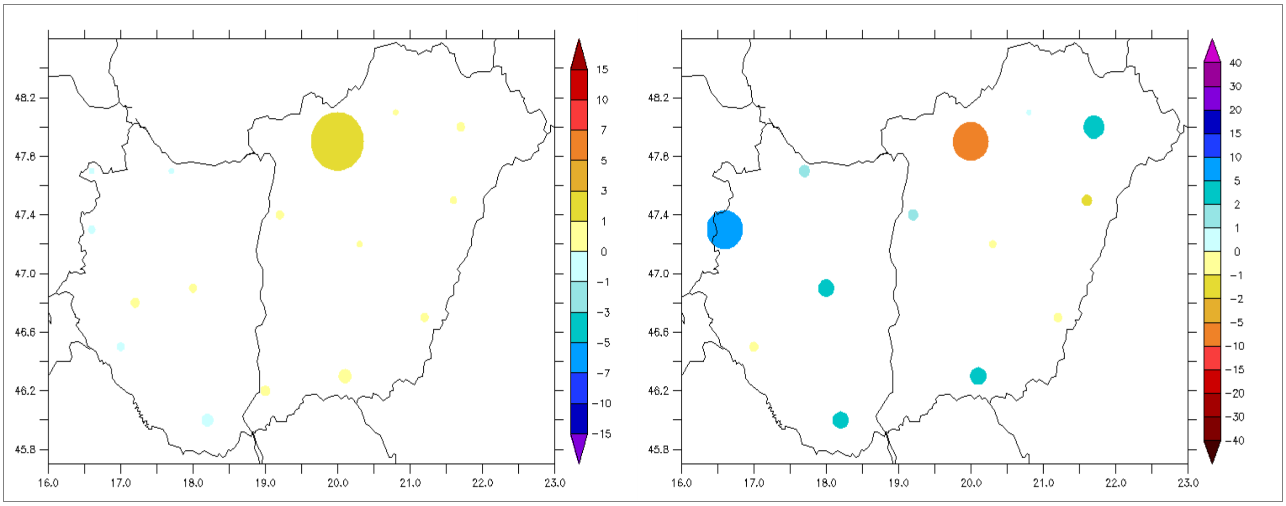


Figure 7.: Annual difference of HUGRID and STATIONS in temperature [°C] (left) and in precipitation [mm/month] in the period of 1971-2000 (right)

4. Conclusions

In this study we made a short comparative discussion of different data sets for evaluating RCMs at HMS. The below notes are our short conclusions:

- The precipitation and temperature of the ERA-40 reanalysis data is slubbered. We are not surprised because it is correlating with the rough resolution of ERA-40. The extreme values are lower and we suppose that the maximums are at their false position. If we have a look at the precipitation fields for the Hungarian territory we find the lowest values in autumn and summer, while the biggest values in winter.
- The smoother spatially gridded CRU gives lower extreme values both in temperature and precipitation, though its reverse was expected. Though the making process of the data sets is the same, we can interpret this with the different sources. The CRU TS 2.1, namely CRU0.5 was created later and uses more observations then CRU TS 1.2, namely CRU10'.
- If we compare at higher elevations the CRU10' and HUGRID, we will see that the latter data set gives lower values for temperature and higher values for precipitation. We can accept the first fact as the CRU10' is worse spatially gridded, but it is not that obvious in the case of precipitation as we saw in the case of the CRU0.5 and CRU10'. If we accept that CRU0.5 is closer to the reality and CRU10' gives false lower values consequently, than the HUGRID is more closer to reality also. To have a comparison of CRU0.5 and HUGRID would be irrelevant due to the considerable difference in spatial resolution.
- Kékestető, the only mountainous observation in Hungary, and Szombathely differ quite much from the equivalent point's value in the HUNGRID data set. We do not know its reason, but could be due to homogenization process of making temperature for HUGRID or due to the used orography in the interpolating process.

We also plan to have research in connection with some observations at higher altitudes outside of Hungary, for example in the Alps, in French or Czech territory because the most problematic territories are near mountains in the analyzed data sets.

References

Crochet, P., 2004: Comparison between ERA-40 derived precipitation and measured precipitation in Iceland. *Veðurstofa Íslands* - Greinargerð 04022: 71 p.

Fil, C., Dubus, L., 2005: Winter climate regimes over the North Atlantic and -European region in ERA40 reanalysis and Demeter seasonal hindcasts. *Tellus* 57: 290-307

Jones, P. D., and Moberg, A., 2003: Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001. *J. Clim.* 16: 206–223

Kallberg P., Berrisford P., Hoskins B., Simmons A., Uppala S., Lamy-Thepaut S., Hine R., 2005: ERA-40 Atlas. *R*eading, UK, ECMWF Re-Analysis Project

Mitchell, T. D., and Jones, P. D., 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Clim.* 25: 693-712

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., New, M., 2004: A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100). *Working Paper 55*. Tyndall Centre for Climate Change Research, University of East Anglia, Norwich, UK.

New, M., Hulme, M., Jones, P. D., 2000: Representing twentieth century space-time climate variability. Part 2: development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* 13: 2217-2238

New, M., Lister, D., Hulme, M., Makin, I., 2002: A high-resolution data set of surface climate over global land areas. *Climate Research* 21: 1–25

Simmons, A. J., Jones, P. D., da Costa Bechtold, V., 2004: Comparison of trends and variability in CRU, ERA-40 and NCEP/NCAR analysis of monthly-mean surface air temperatures. *ERA-40 Project Report Series 18. R*eading, UK, ECMWF Re-Analysis Project. 38p

Szentimrey, T., 1999: "Multiple Analysis of Series for Homogenization (MASH)". Proceedings of the Second Seminar for Homogenization of Surface Climatological Data. Budapest, Hungary; WMO, WCDMP 41: 27-46

Szentimrey, T., Bihari, Z., Szalai, S., 2005: Meteorological Interpolation based on Surface Homogenized Data Basis (MISH). *Geophysical Research* Abstracts 7, 2005

Szépszó, G., Csima, G., Pieczka, I., 2007: Regionális klímamodellek szisztematikus összehasonlítása. 3A/082/2004 NKFP-projekt, beszámoló.

Uppala, S. M.; Kallberg, P. W.; Simmons, A. J.; Andrae, U.; Bechtold, V. D.; Fiorino, M.; Gibson, J. K.; Haseler, J.; Hernandez, A.; Kelly, G. A.; Li, X.; Onogi, K.; Saarinen, S.; Sokka, N.; Allan, R. P.; Andersson, E.; Arpe, K.; Balmaseda, M. A.; Beljaars, A. C. M.; Van De Berg, L.; Bidlot, J.; Bormann, N.; Caires, S.; Chevallier, F.; Dethof, A.; Dragosavac, M.; Fisher, M.; Fuentes, M.; Hagemann, S.; Holm, E.; Hoskins, B. J.; Isaksen, L.; Janssen, P. A. E. M.; Jenne, R.; McNally, A. P.; Mahfouf, J. F.; Morcrette, J. J.; Rayner, N. A.; Saunders, R. W.; Simon, P.; Sterl, A.; Trenberth, K. E.; Untch, A.; Vasiljevic, D.; Viterbo, P.; Woollen, J., 2005: The ERA-40 re-analysis, *Q.J.R. Meteorol. Soc. 131*: 2961-3012.