Enabling Climate Information Services for Europe

DELIVERABLE 4.6
Report on future evolution of Eastern Sicily heavy precipitation events and their spatial distribution

Activity:	*WP4 – Cities*
Activity number:	*Task 4.4 - Flood risk assessment in cities of Eastern Sicily*
Deliverable:	*Report on future evolution of Eastern Sicily heavy precipitation events and their spatial distribution*
Deliverable number:	*4.6*
Authors:	*Andrea BERTOLINI, Michele BRUNETTI, Mistral GARZOGLIO, Claudia SIMOLO - ISAC-CNR; Maurizio MAUGERI – University of Milan*

The work leading to this publication has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 265240.

Summary

This report describes the activities performed by ISAC-CNR to produce Deliverable 4.6 of the ECLISE Project – Report on future evolution of Eastern Sicily heavy precipitation events and their spatial distribution. These activities were performed within task 4.4 - Flood risk assessment in cities of Eastern Sicily that also allowed producing deliverable 4.5 – Report on past heavy precipitation events for the eastern part of Sicily. Deliverable 4.6 was produced in order to fulfil SIAS (Servizio Informativo Agrometeorologico Siciliano, Sicily Regional Administration) needs. SIAS is one of the users of the ECLISE Project.

The activities we developed in the frame of task 4.4 went far beyond the planned goal of the project, which focused only on intense precipitation in eastern Sicily: we extended in fact the activities to the whole Sicily region.

1. Introduction

The general goal of ECLISE Project task 4.4 has already been outlined in Deliverable 4.5. Here we only underline once more that high frequency of heavy precipitation, elevated population density, complex geography and the lack, especially in the past, of an adequate urbanisation policy, make Sicily highly vulnerable to risks connected with heavy precipitation events.

Within this context it is very important to get a better knowledge of the spatial distribution of the probabilities of occurrence of these events and to assess whether climate change may affect such probabilities. We considered both these issues: the first results concerning the former issue were provided in Deliverable 4.5. This report aims both at integrating these results with new analyses performed in the second part of the ECLISE Project and at discussing the latter issue based on future climate scenario data.

2. Observational data, quality checks and homogenisation procedures

The network of observational daily precipitation records used for the analyses presented in this report has extensively been discussed in Deliverable 4.5, which gives detailed information on quality check and homogenisation of these records too.

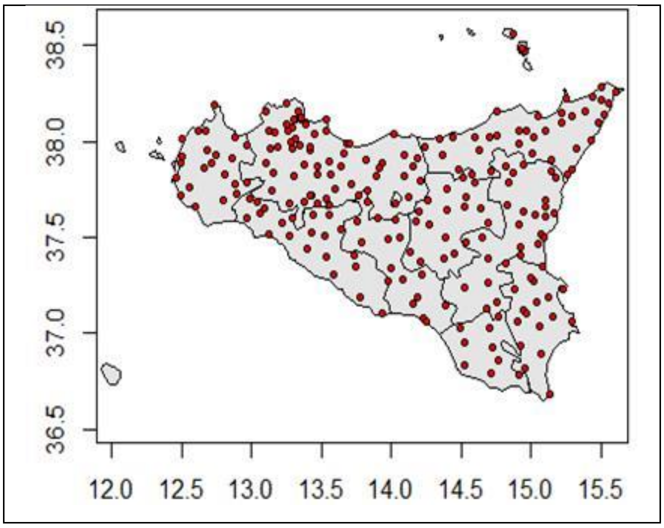

Figure 1. Position of the 231 stations of the rain-gauge network we used for the analyses presented in this report

In the first part of the ECLISE Project (see Deliverable 4.5) we selected a subset of 53 stations for the analyses and focused only on the 1951-2005 period. In the second part of the Project, thanks to the application of the homogenisation procedure to all the stations of the dataset, it was possible to extend the analyses to a much wider number of stations and to consider the whole 1921-2005 period. In particular, we selected 231 stations out of 325. The stations that were not considered are those with a too low number of data (we set a threshold of 24 years) and those that showed, after the homogenisation procedure, inhomogeneities in the number of rainy days.

3. Analysis of the station records

In Deliverable 4.5 we presented the results of Generalised Extreme Value Distribution (GEV) and Generalized Pareto Distribution (GPD) applied to subset of 53 Sicily stations. The results highlighted an interesting spatial pattern that allowed presenting spatial maps of the parameters and return levels. They highlighted however also very large uncertainties.

Within this context, we continued the analyses on the observational records in the second part of the ECLISE project.

The first step consisted, as already observed, in extending the number of stations; the second step consisted in adopting an index flood - Regional Frequency Analysis approach; the third step consisted in improving the method used in order to project the station results onto a high-resolution grid.

3.1 Index flood

Different index floods (see e.g. Hosking and Wallis, 1997) were used for GEV and GPD. In particular, for GEV we used the median of the daily annual maxima of each station, whereas for GPD we used percentile-85 of the daily station precipitation values above 1 mm. As far as GPD is concerned, we decided also to focus only on defined periods of the year. Specifically we studied a defined day of the year by considering the data of a window centred over that day and having a wide of 93 days. This time window is large enough to have a sufficient number of data to perform the analyses and permits to consider a rather homogeneous period in terms of precipitation regime (amount and frequency). Then we iterated the analyses moving the considered day through the whole year. This allowed us also identifying the periods of the year with highest risk of heavy precipitation.

3.2 Regional Frequency Analysis

After normalising all precipitation data by means of the index flood, we adopted a Regional Frequency Analysis approach (see e.g. Hosking and Wallis, 1997). Specifically, we first selected all the stations within 30 km from the point to investigate and merged then their data by means of the station-year method (see e.g. Reed et al., 1999). It simply consists in obtaining one single record considering all the data of the selected stations together. This merged record was then analysed to get GEV and GPD parameters and errors. They were obtained by means of both maximum likelihood and linear moments methods.

3.3 From station to gridded analyses

The analyses presented at sections 3.1 and 3.2 were performed not only for the station points, but also considering the grid-points of a 30-arc-second-resolution digital elevation model (DEM) over Sicily (we used the GTOPO30 DEM provided by the United States Geological Survey (USGS, 1996)). The main problem for this grid-point approach consisted in extending the index-flood values from the stations to the grid-points. This extension was performed searching for a linear regression between the station index flood values and the correspondent precipitation climatological normals, which were available from the ECLISE Project for all GTOPO30 grid-points (see Deliverable 6.4). This linear regression was performed within the same area considered for the station clustering of the regional frequency analysis. The final result consists of grid-point coefficients that allow estimating, from the precipitation normal of each grid-point, the corresponding median of the annual 1-day precipitation maxima and the percentile-85 of the 1-day precipitation values above 1 mm.

Once these values were available, it was easy to transform the grid-points estimations from results expressed in relative values (with reference to the corresponding index flood) into absolute precipitation values.

3.4 Grid-point uncertainties

The GPD results we obtained with the parameters estimated by means of the maximum likelihood method have also been used to get an estimation of the uncertainties of the grid-point return levels. These uncertainties have been estimated applying error propagation to all the steps that allowed obtaining the grid-point return levels. The highest contribution to the error turns out to be that arising from the estimation of the grid-point index flood from the corresponding precipitation climatology. This result is not surprising as the precipitation climatologies have themselves significant errors (about 15%), whereas the GEV and GPD parameters, thanks to the large data availability and to the regional frequency approach, have rather small errors.

3.5 Using an alternative approach

The use of the station-year approach is actually controversial and not all the authors like it (see e.g. Reed et al., 1999): the main benefit is that it allows an easy estimation of the GEV and GPD parameter errors; the main deficit is that it may be not completely correct if the station data are not independent. In Sicily, actually, the problem of the dependence of the station records is much lower than in many other areas as station correlation decreases very quickly with distance. However, in spite of this low precipitation spatial coherence, we decided to perform all analyses also with an alternative approach that consists in calculating the GEV and GPD parameters of each grid-point simply by means of a weighted average of the corresponding parameters of the stations within 30 km from it (Fowler and Kilsby, 2003). The weights are defined here by means of the number of available years of each station.

The results obtained with this alternative approach is not significantly different from those obtained with the station-year approach.

3.6 Present-day Sicily return levels

Most of the analyses presented in sections 3.1-3.5 are discussed in a degree thesis performed at Milan University (Garzoglio, 2014); the results we got within this thesis, together with some other results we got after its conclusion, will be presented and discussed in a paper that we plan to submitted to a scientific journal in the next future.

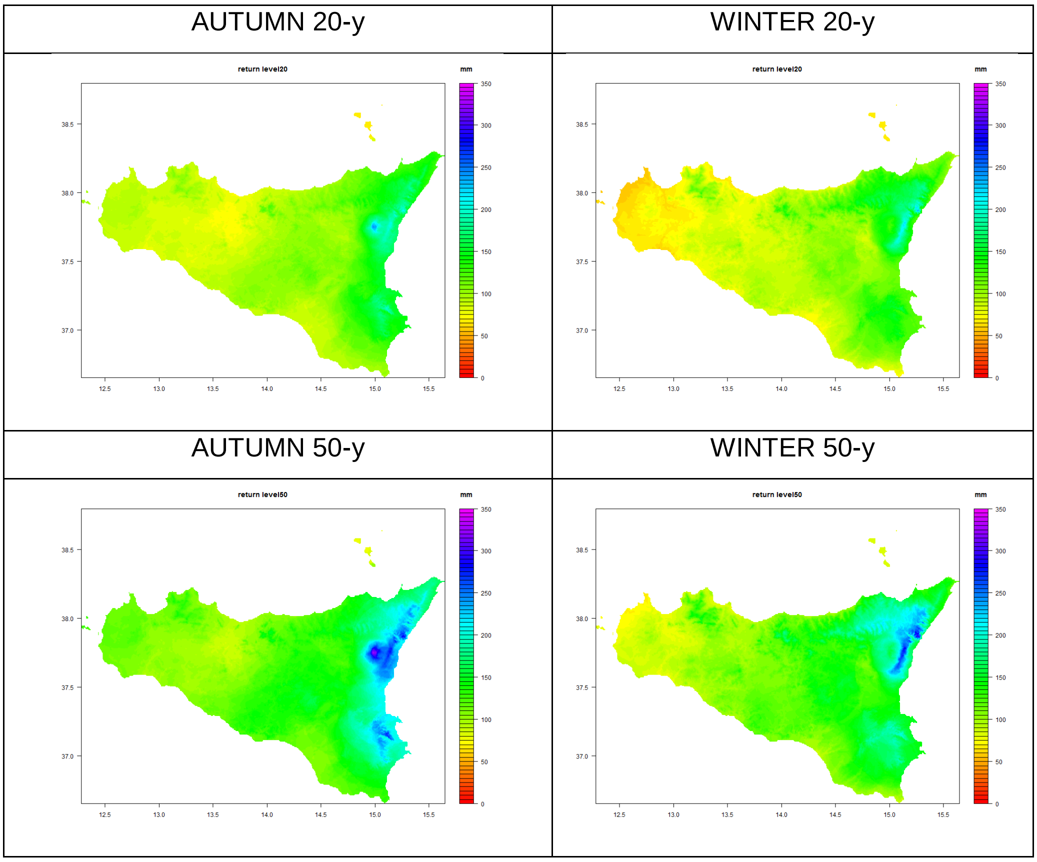

Figure 2. 20-year and 50-year return levels for 93-day windows centred on October 6th (left column graphs) and January 2nd (right column graphs).

Figures 2 and 3 show 20-year and 50-year return levels for 93-day windows centred on October 6th (corresponding to autumn season) and January 2nd (corresponding to winter season), obtained by means of GPD applied with a threshold corresponding to percentile-85 of 1-day precipitation values of the days with more than 1 mm of rain. GPD parameters were estimated by means of maximum likelihood, pooling stations within about 30-km distance and using index-flood – regional frequency analysis approach. The same maps can be obtained for windows centred on any day of the year and for any time length. The map corresponding to October 6th gives an indication for the season with the highest risk in some areas as, e.g., eastern Sicily. The map corresponding to January 2nd is representative of the season with the highest risk in some other areas as, e.g., the Madonie.

An estimation of the uncertainty level for the 6th October map is shown in figure 3. The error is within 30% for most Sicily grid-points.

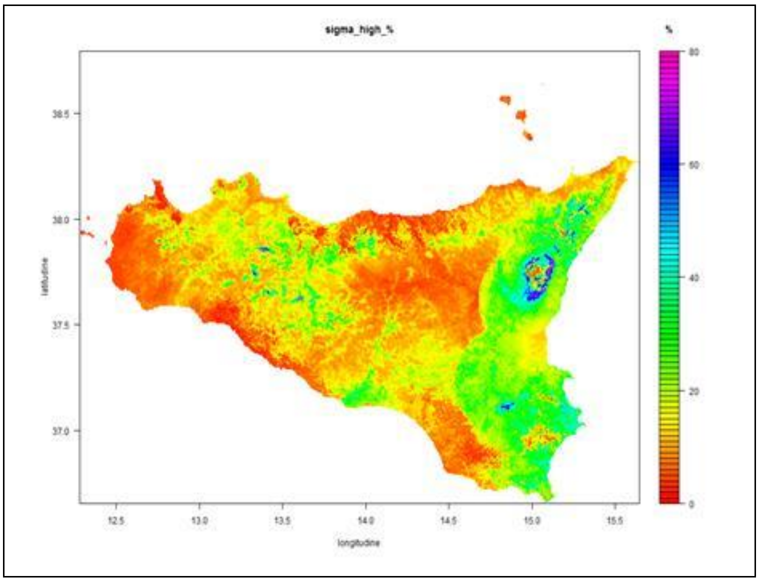

Figure 3. The figure shows an example of the uncertainty estimation for the return levels. Specifically the relative errors of the 50-year return levels for the 93-day window centred on October 6th (see figure 2) are shown here.

4. RCM precipitation scenario for the XXI century

XXI century precipitation values for Sicily were provided by the ENSEMBLES Project. Five RCM-GCM combinations were taken into account (table 1).

Institute	Scenario	Driving GCM	Regional Model	Resolution
KNMI	A1B	ECHAM5-r3	RACMO	25 km
SMHI	A1B	ECHAM5-r3	RCA	25 km
MPI	A1B	ECHAM5-r3	REMO	25 km
ETHZ	A1B	HadCM3Q0	CLM	25 km
HC	A1B	HadCM3Q0	HadRM3Q0	25 km

Table 1. List of RCM-GCM combinations taken into account in the present deliverable.

We considered the historical run of the models forced by GCM and their future projections under the A1B scenario. We considered also the historical run forced by reanalyses, but only in order to validate the downscaling procedure we present in the next paragraph.

Unlike GCMs, thanks to their higher spatial resolution, RCMs provide a better description of orographic effects, land-sea contrast and land-surface characteristics. They also give an improved treatment of fine scale physical and dynamical processes, and they are able to generate realistic mesoscale circulation patterns which are absent from GCMs. However, also RCMs have significant problems in correctly capturing the atmospheric processes leading to extreme precipitation events as highlighted by figure 4. This figure shows the 50-year return levels we get for the 5 RCMs of table 1 for the 1951-2000 period by means of the analysis of the annual maxima with the GEV technique. The results, obtained with the approach described in sections 3.1 and 3.2, are expressed only for the model grid-points (the grid has a resolution of 25 km). Figures 5 and 6 show the same results for the 2001-2050 and 2051-2100 periods, applied to the raw model output (without any downscaling) to test their performances.

These figures give evidence of a large spread among the return levels of the RCM-GCM combinations. Moreover, they show significant differences between the observations in the 1951-2000 period and the model results in the same period.

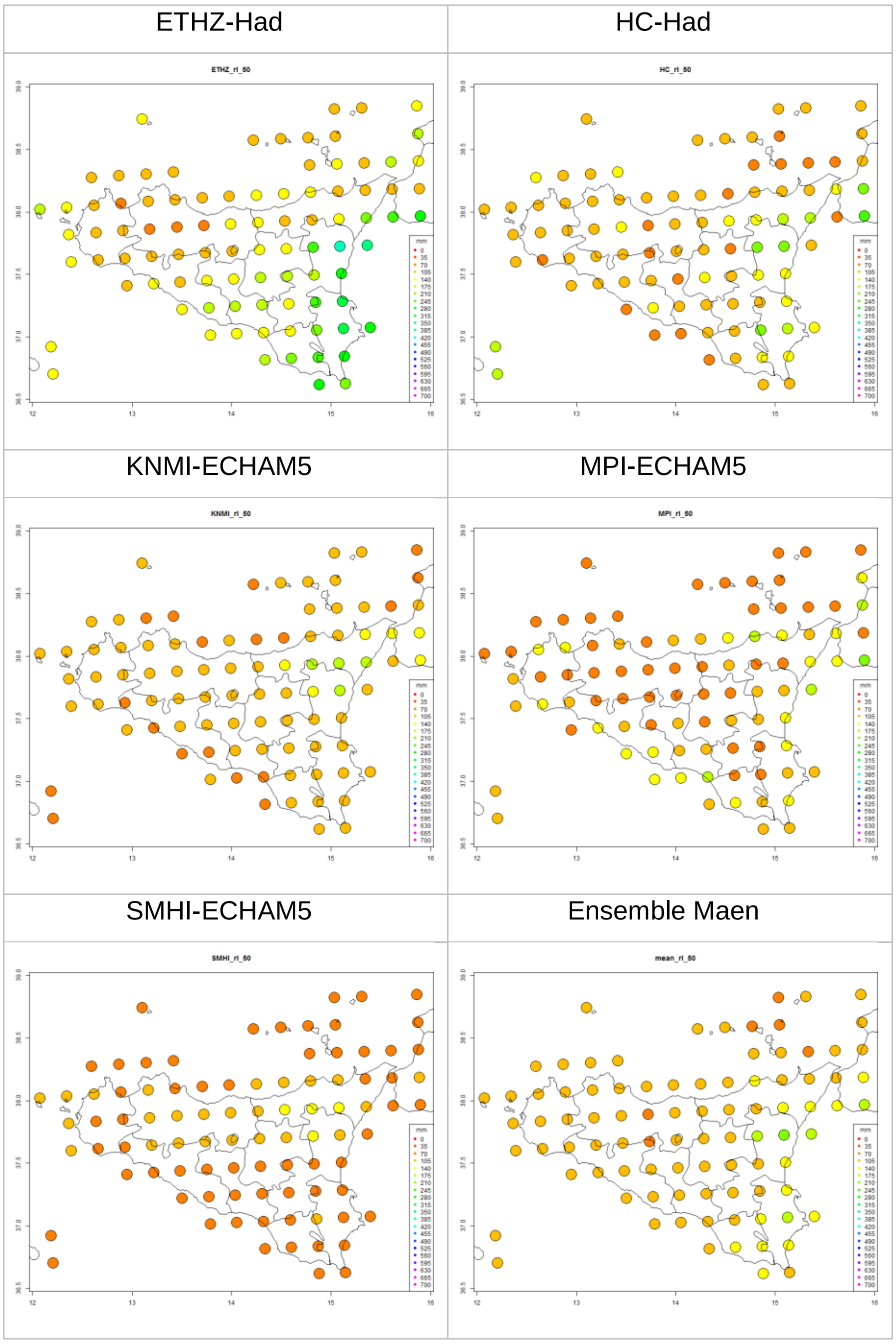

Figure 4. 50-year return levels of 1-day precipitation for the raw model data of the 5 RCM-GCM combinations and mean values of the model results. The analyses have been performed by means of GEV technique and adopting an index flood – regional frequency analysis approach.

4.1 Downscaling of the RCM records

The climate model records exhibit significantly different statistical distributions from the observed data: the disagreement between the maps in figure 4 and the correspondent one in figure 2 is a consequence of these differences. It was therefore necessary subjecting the RCM records to model output statistics.

The first task consisted in projecting the model records onto the station positions. This projection was performed with a weighted average of the model grid-points closest to the station itself, with a Gaussian weighting function that decreases to 0.5 at distance of 12.5 km.

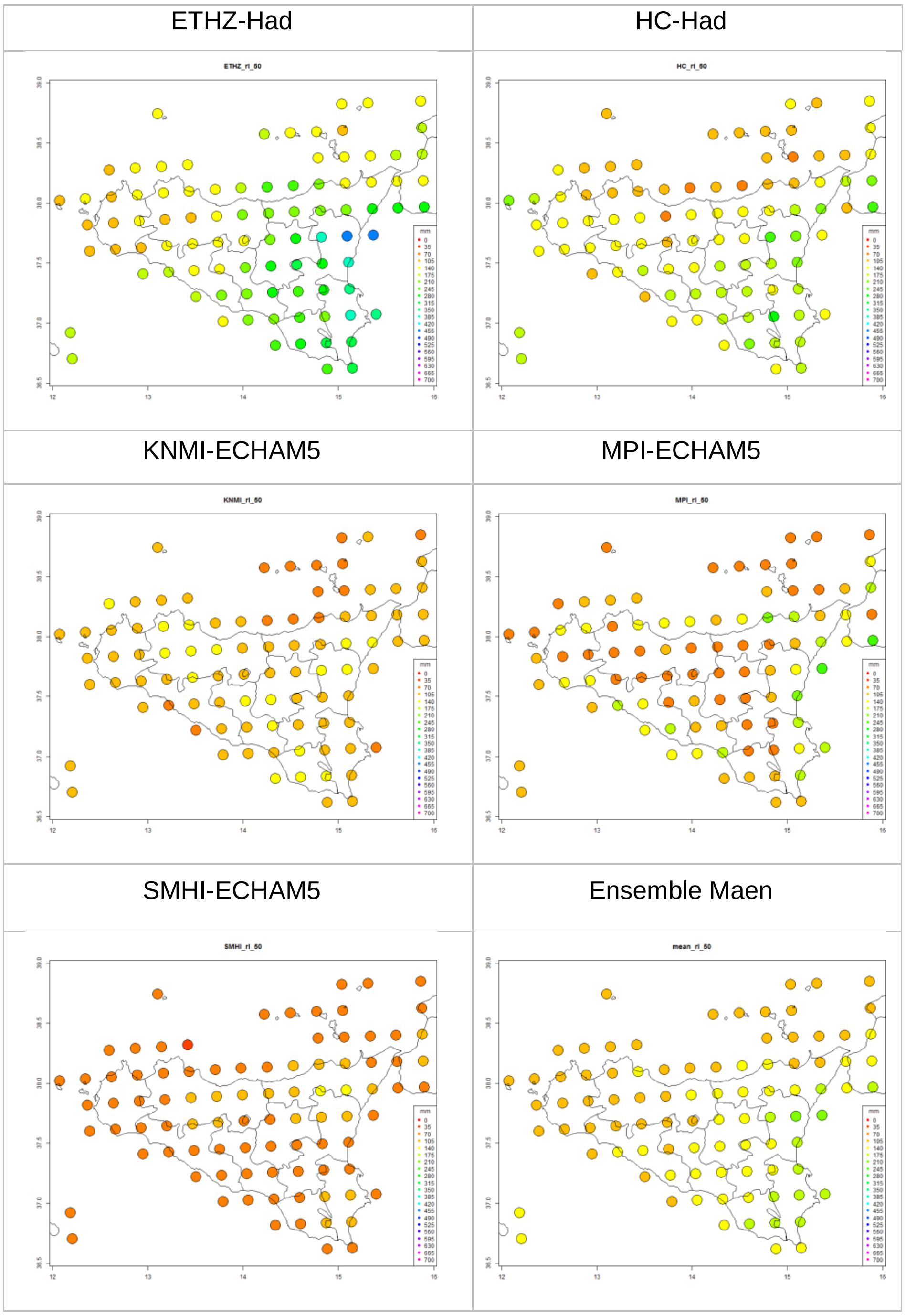

Figure 5. As in figure 4, but for the 2001-2050 period.

Once the model records were projected onto the station points, we filtered each model record with the aim of obtaining the same precipitation frequency of the corresponding observational station record in the 1961-2000 period. This filtering was performed by estimating, for each month, the observed and the modelled mean frequency of wet days and imposing the observed frequency to the modelled series by choosing a lower threshold in the daily precipitation amount so that the correct frequency is preserved (note that modelled un-filtered series have always a higher precipitation frequency).

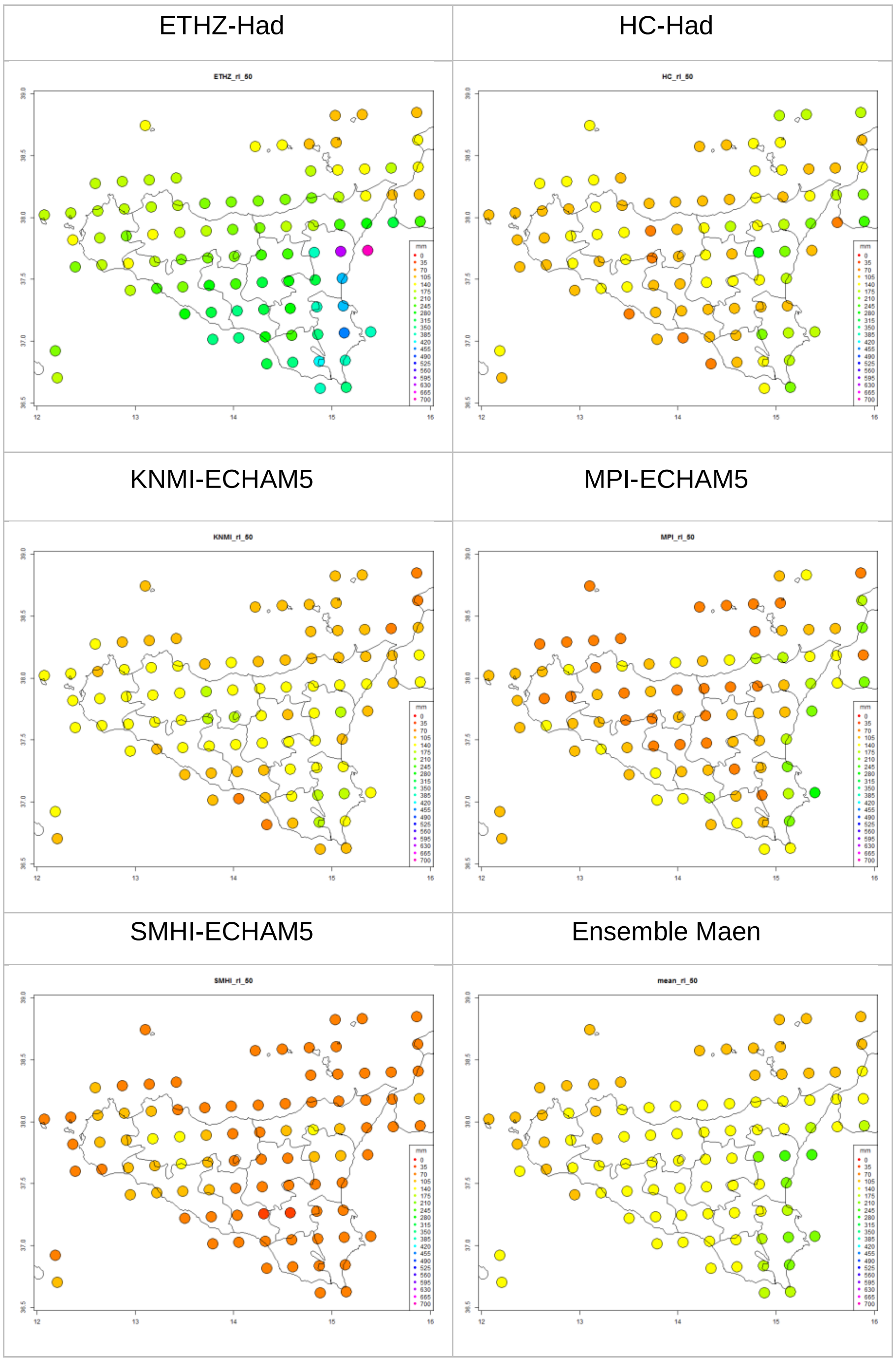

Figure 6. As in figure 4, but for the 2051-2100 period.

The filtered records were then subjected to the CDFt procedure described by Michelangeli et al. (2009). It consists in a Q-matching directly applied to CDFs. It is based on the assumption that there exists a transformation allowing to “translate” the CDF of a model variable (in our case precipitation) into the CDF representing the local-scale climate variable, i.e. at a given weather station. Let F_{Sh} and F_{Gh} stand for the CDFs of the observed local data at the weather station and the modelled data interpolated on the weather station for the past; and F_{Sf} and F_{Gf} are the CDFs equivalent to F_{Sh} and F_{Gh} but for the future. If there exists a transformation allowing to “translate” F_{Gh} into F_{Sh}, we can also apply a transformation to the future F_{Gf} to get the local F_{Sf} for the future.

To summarize, the transformation consists in estimating:

$$F_{Sf}(x) = F_{Sh}\left(F_{Gh}^{-1}\left(F_{Gf}(x)\right)\right)$$

The main limit of this procedure is that the three CDFs may cover significantly different ranges: this problem is particularly marked for some stations that have very high outliers either in the observational or in the modelled records. The consequence is that $F_{Sf}(x)$ may not completely cover the [0,1] range. In a study dealing with high precipitation values, a lower limit above 0 is not problematic. On the contrary it is necessary to manage the cases in which the upper limit is lower than 1. We managed this problem assuming that the values of x with $F_{Gf}(x)$ greater than $Max(F_{Sf}(x))$ can be downscaled in the same way (i.e. using the same multiplicative coefficient) we downscaled the highest value of future precipitation for which $F_{Gf}(x)$ is not greater than $Max(F_{Sf}(x))$. It is however necessary to underline that this solution, even though necessary, may cause strong outliers in the downscaled data.

An alternative approach to manage the disagreement between the CDFs of the modelled and observed records, could be the one introduced by Kallache et al. (2011). It consists in obtaining the GPD of the downscaled model records from the corresponding GPDs of the observational and non-downscaled model records. In our opinion, however, this method has to be studied more in detail before it can be applied to generate results that can be used within a climate service context.

In spite of all this problems, we used the CDFt downscaling approach proposed by Michelangeli et al. (2009). After downscaling, the RCM records were subjected to all the procedure outlined in section 3. It allowed obtaining return level maps for the future precipitation scenario too.

4.2 Validation of the CDFt downscaling technique

The reliability of the statistical downscaling technique has been verified considering the RCMs forced by the reanalysis over the historical 1961-2000 period, and interpolated on the station points corresponding to the 44 most complete stations. To do this, the 40-year period has been randomly divided into two 20-year sub-periods (one for calibration and one for validation). This subdivision has been applied 100 times per station by applying a bootstrap technique.

To verify the ability of the downscaling technique to correctly reproduce extreme events, the verification has been applied to the parameters of a Generalized Pareto Distribution fitted on the data and to 20-year and 50-year return levels.

Figures 7 and 8 show, for the MPI-ECHAM5 model, the scatter plots and the corresponding box-plots of the shape parameters and of the 50-year return level for winter and autumn (the rainiest seasons for Sicily).

These figures give evidence that the CDFt downscaled data give a good description of real precipitation shape parameters and, consequently, of the return levels too. Nevertheless, even if CDFt gives good results, there are some systematic overestimation of the return levels produced by this technique. These overestimations range from +9% in winter to +20% in summer, for a return level corresponding to a return period of 50 years (the overestimations are 10% and 13% in spring and autumn respectively).

4.3 Analysis of the downscaling RCM records

In figures 9-11, we present a selection of the results we obtained analysing the CDFt downscaled model records.

4.4 Uncertainties of the return level scenario

As far as the future return level scenarios are concerned, we underline that there is very large uncertainty. This very large uncertainty is evident first of all from the wide spread of the model results.

However also the downscaling procedure causes remarkable errors: they have been discussed in section 4.1. Finally also the regional frequency approach, which turns out to give excellent results for the observational records, seems to be another source of remarkable errors: they are due to a very high spatial variability of the grid-point distributions. This high variability undermines the basic hypothesis of the regional frequency approach (i.e. two near points have the same distribution after the data are normalised by means of the index flood).

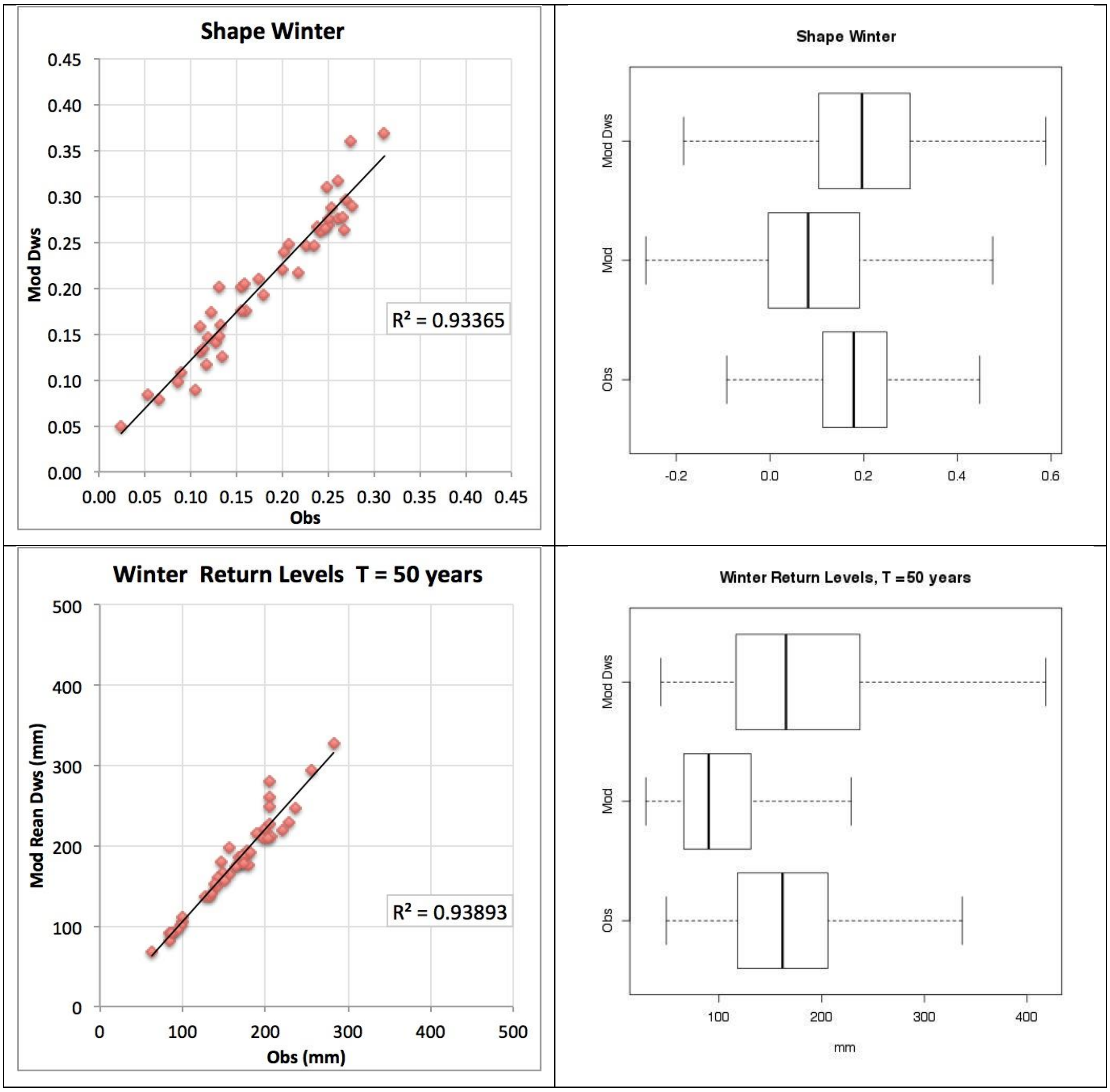

Figure 7. Left column: scatter plots of the winter shape parameters and of the corresponding 50-year return levels for the MPI-ECHAM5 model, forced by reanalysis data and the observational data. Right column: box plots of the same data and of the corresponding values obtained for the not downscaled model data.

The consequence of these large uncertainties is that the return levels from the RCM scenarios cannot yet be used by an organisation as Sicily Regional administration to address planning issues. The most robust result of our analyses is, in our opinion, the general tendency of the model data to present an increase in the highest percentiles of the distributions of precipitation data. These results highlights the risk of a future increase of the present-day return level, but assessing with an adequate confidence a quantitative estimation of this increase is probably beyond the state of the art of the present day research on this issue.

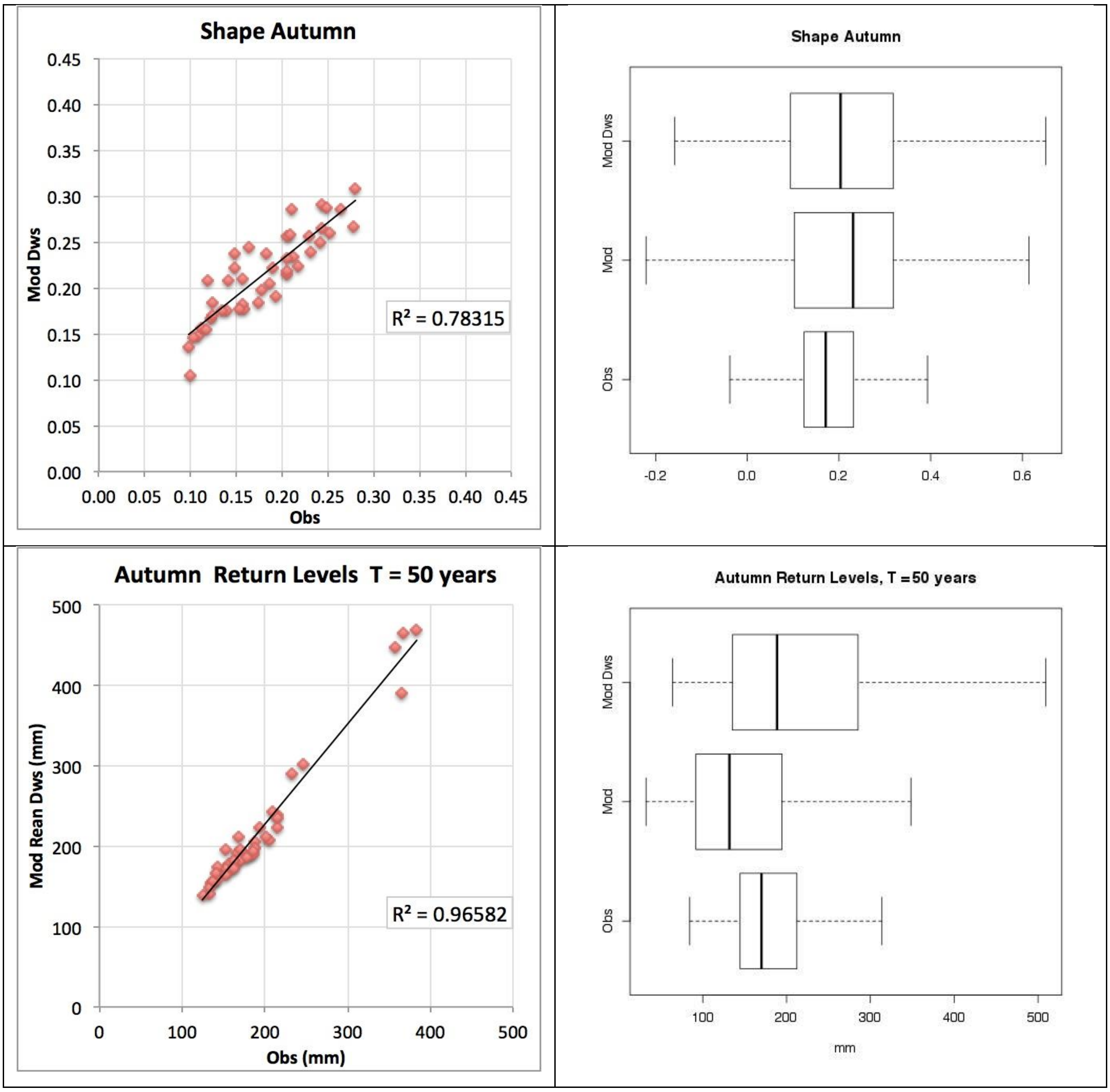

Figure 8. As in figure 7, but for autumn.

5. Conclusions

The activities performed within task 4.4 of the ECLISE Project allowed getting a better knowledge of the present-day spatial distribution of the probabilities of occurrence of heavy precipitation events over Sicily. The results have reasonable uncertainties and they can already be uses to provide return levels within a climate services context.

On the contrary, the return levels estimated from the RCM scenario data cannot yet be used within a climate services context. This limitation is due both to very large uncertainties of the model results and to the fact that the available downscaling techniques have to be studied more in detail.

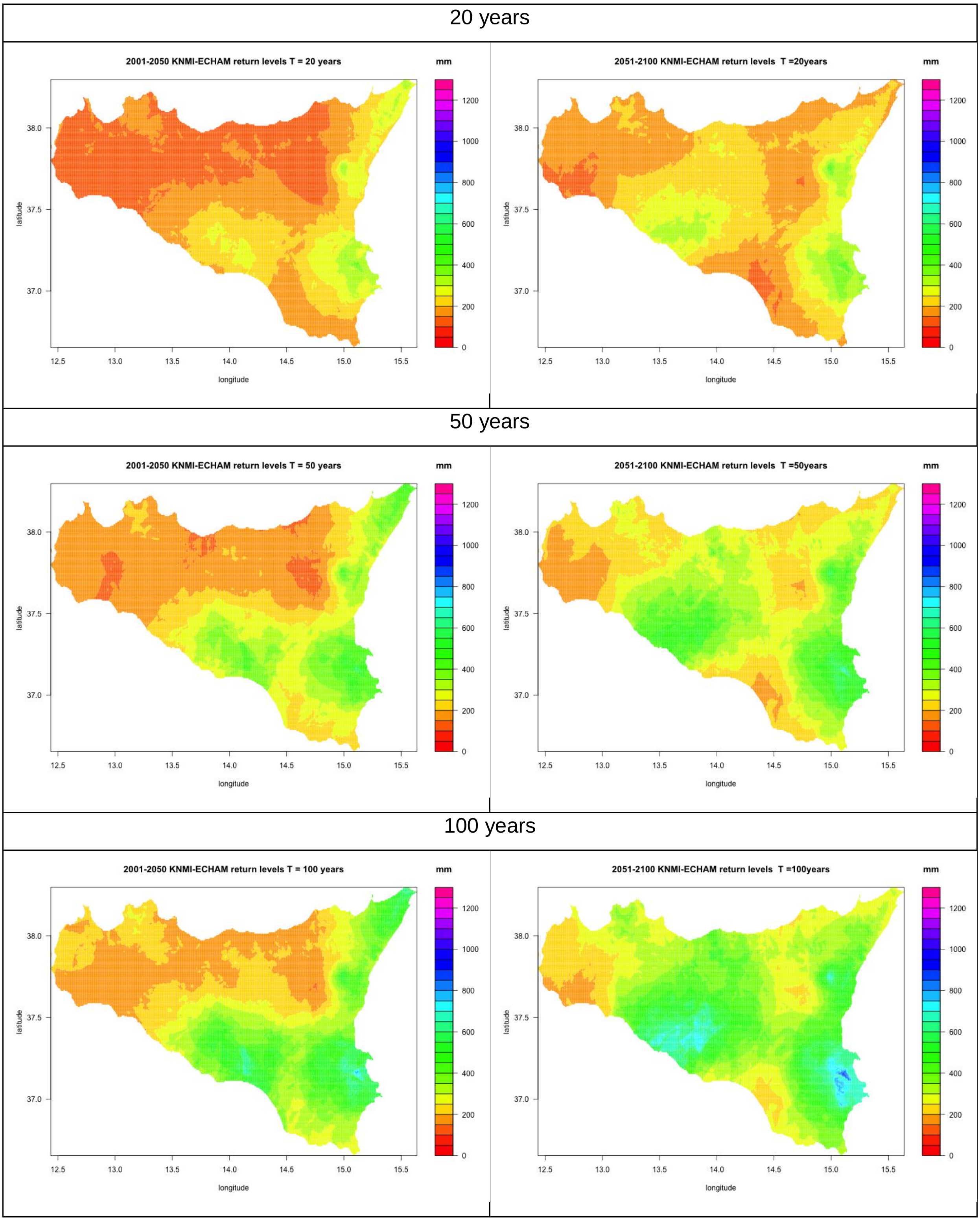

Figure 9. KNMI-ECHAM5 RCM return levels for 20, 50 and 100-year return periods. The analyses have been performed as for the maps in figure 2.

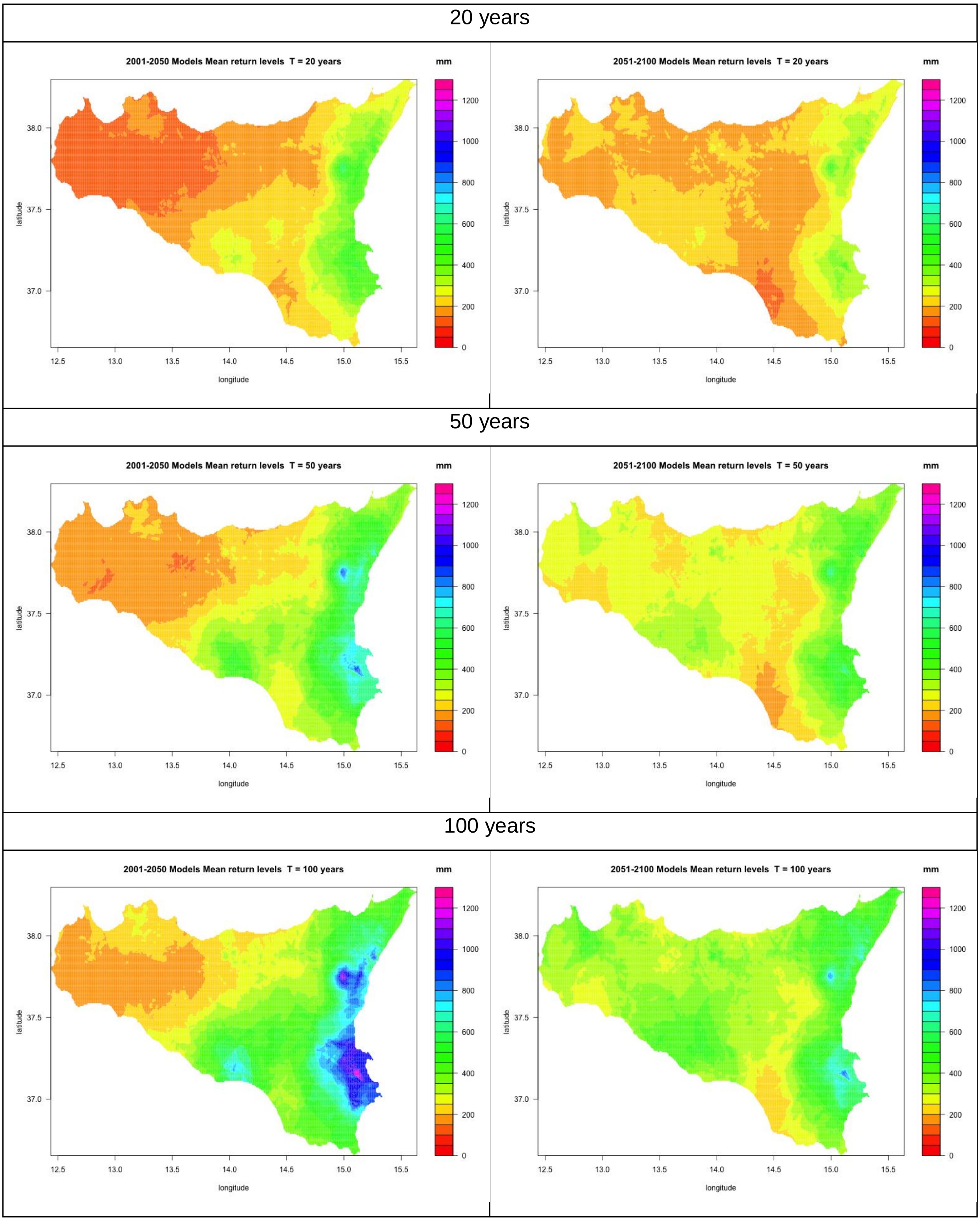

Figure 10. RCM Ensemble mean return levels for 20, 50 and 100-year return periods. The analyses have been performed as for the maps in figure 2.

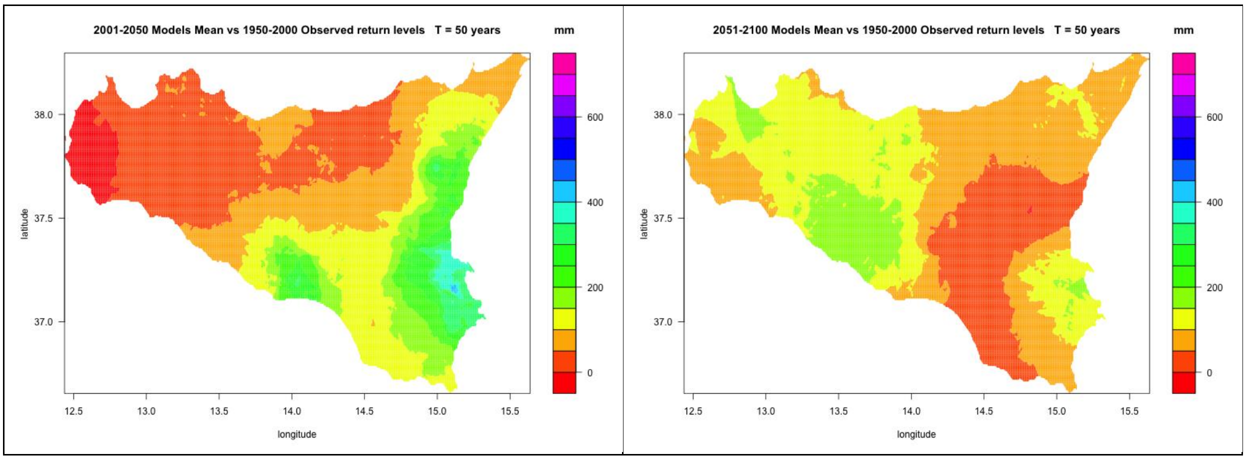

Figure 11. Differences between the ensemble mean 50-year return levels for the two future sub-periods (2001-2050 and 2051-2100) and the corresponding values for the 1951-2000 observational data.

6. References

Fowler H. J. and Kilsby C. G., A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000, International Journal of Climatology, 23: 1313 - 1334, DOI: 10.1002/joc.943, 2003.

Garzoglio M., Distribuzione spaziale delle precipitazioni estreme in Sicilia. Degree Thesis in Physics. Milan University (Università degli Studi di Milano), 2014.

Hosking J. R. M. and Wallis J. R., Regional Frequency Analysis: An approach based on L-Moments, Cambridge University Press, Cambridge, 1997.

Kallache M., Vrac M., Naveau P. and Michelangeli A., Nonstationary probabilistic downscaling of extreme precipitation, Journal of Geophysical Research , 116, DO5113, DOI: 10.1029/2010jd014829, 2011.

Michelangeli, P. A., M. Vrac, and H. Loukos, Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophys. Res. Lett., 36, L11708, doi: 10.1029/2009GL038401, 2009.

Reed D.W., Faulkner D.S. and Stewart E.J., The FORGEX method of rainfall growth estimation. I: Review of requirement, Hydrology and Earth System Sciences 3(2): 187 - 195, 1999.

USGS (United States Geological Survey). GTOPO30 Documentation. Available at https://lta.cr.usgs.gov/GTOPO30, 1996.

Links to concrete results:

http://www.eclise-project.eu/

http://www.isac.cnr.it/climstor/ECLISE-project.html

http://www.sias.regione.sicilia.it/ECLISE

References to activity meetings: The objectives of task 4.4 have been presented at the ECLISE Kick-off meeting (De Bilt - 09 March 2011); the methods and results have been presented at the First and Second ECLISE meetings (Norrkhoping - 6-7 March 2012; Chania – 23-26 April 2013).