Report:
Velibor Želi

Guide on how to Obtain and Process Climate Model Data

Nowadays, there exists a number of atmospheric modeling groups, which focus on making, developing and running atmospheric models. Many of these models are *climate models*, special models that simulate processes in the atmosphere that occur over a large time scale. The main idea for existence of climate models is to understand in which way the climate is going to change, depending on the antropogenic effects (since these are the effects we can control). Models that run over the entire domain of Earth are know as *global climate models*, which are composed of coupled atmospheric and oceanographic models. In these coupled simulations there exists a feedback between ocean and atmosphere, which is rather important since the ocean acts as a large heat container for the atmosphere. Other models are known as *regional climate models* which run over a specified area, and usually their usage is in case study of climate change around lakes, rivers, mountains, etc.

This paper emphasizes on the problems related to data processing for anyone who would like to evaluate the climate change and has never done it before. Target audience are students of undergraduate and graduate level that would like to grow there skills and learn where to obtain the climate model data, and how to work with it. Requirements are that the user has a GNU/Linux OS and is familiar to programming.

Collecting the Output Data

Since there is a considerable research in the area of climate change, many modeling groups are dedicating time to improving and developing of climate models, and large research centers have their own climate models. These models use input information and based on the principles of physics they work out the equations and produces an output. Sound fairly simple, but these are complex programs that are ran by a group of people and need a lot of computational resources and time to run. After the process of computation is finished, each model produces the output values for different variables, sorted by time, which are called *output data from the model.*

It is possible for a regular user to obtain climate model output data at the portal of Earth System Grid Federation (ESGF). This portal contains links to the output data for many different projects and different climate models. The website address for the ESGF portal is http://pcmdi9.llnl.gov/esgf-web-fe/, though to download data one has to register as a user. After the registration is complete, user can filter through the data by specifying the attributes, that way making it easier to find specific output data. For example, it is possible to filter the data by the name of the model which produces it, or by project for which the data was made. There are many filters at the left side of the browser window that can be used for specifying the attributes of the output data: *project*, *model*-name of the model, *institute*-name of the institute that produced the data, *time frequency*-6 hourly, 12 hourly, daily data output, *variable*-if there exists an interest in particular variable[1], *ensenble* and many other. In this way when the user enters more information about required data the filter will be better and at the end there should be only one result for downloading, one that user was searching for. Process of downloading can go directly over the browser or it can be done through the use of Wget program and Wget scripts, the second option is recommended. After filtering through the data and selecting it by "Add To Cart", user should go to "Data chart" bellow the search engine and check the "Show all" option (it is *mandatory* to do so if following this procedure). There is a "WGET All Selected" button which makes the Wget scripts with all the locations at the web server from which Wget program will download the selected data. It is also suggested that Wget scripts are made containing less than 1000 files (the best way would be, if possible, to make a unique Wget script for every set of data), because if this is not the case, there will be problems due to the fact that Wget scripts are limited to 1000 files. It is possible to split one large Wget script into parts and download them, but that is not going to be covered in this paper. Once the Wget scripts are made internet browser will download them and they will be ready for usage. To make the script executable we use change file mode command with the name of Wget_script as an argument from the terminal: `chmod` u+x Wget_script, and than simply run the script from the terminal. This will start the downloading process, and the data will be downloaded at the directory where the Wget_script is located. All the data downloaded from the ESGF portal is packed in form of netCDF (network common data form) files, and these files contain climate model results that will be used for processing.

The data that was downloaded and processed for the purpose of this guide belongs to the set of output from the Canadian Earth System Model

[1] List of variable names at: http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.png

(CanESM2). It important to download set of data that is for the experiment called *historical run*. This is considered to be a “control simulation”, and it is ran for the period in between 1850 and 2005 because in that period there are measurements that can be compared with the model output, and the model can be evaluated. Also one needs to download the data that belongs to the *future simulation* of climate for the same model. In our case it is data for the period of 2006–2100. These are two different simulations for the same model that are going to be compared. Also the frequency of the output was one value for each month (monthly data).

Data Processing

While working on this problem I used Matlab as a tool for data processing, but any other programming language would suffice. For the users that will work with Matlab it is vital to install the *nctoolbox* which allows importing data from netCDF files into the variables.

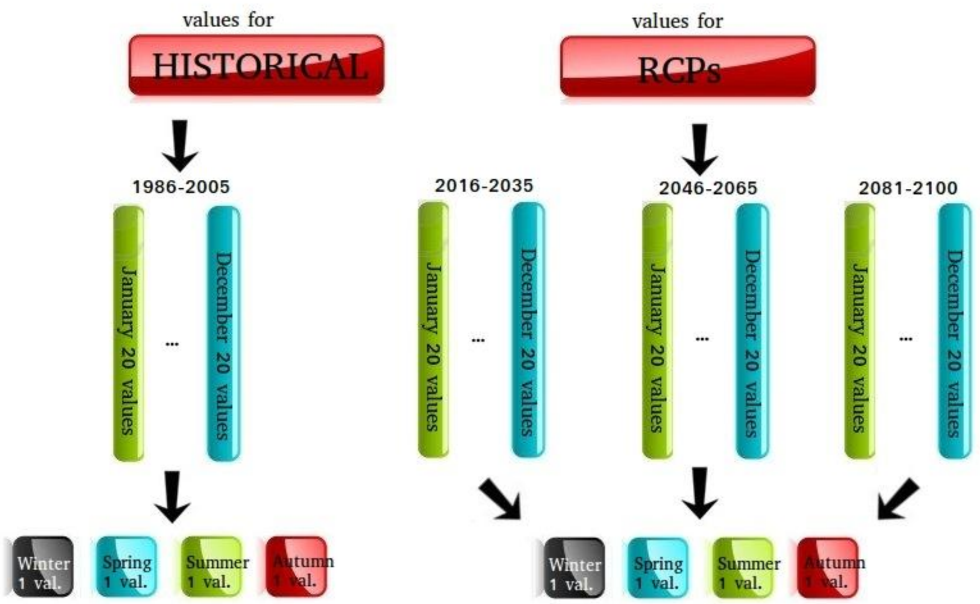

Figure 1: Graphical representation of the procedure used in data processing.

It is common to use a period of 30 continues year for describing climate, this is recommended by World Meteorological Organization (WMO), but that is not an objective number and in case of some variables smaller period would suffice. In this paper 20 year period will be used for describing the climate. First of all it is necessary to import the values from 20 continues years, for four periods. For the period of 1986–2005 data is obtained from the historical simulation output file, and for three periods 2016–2035, 2046–2065 and 2081–2100 data is gathered from the future simulation output. For each time period, values of each month should be imported into a different variable. That way one can make a mean value for every month in the year. This value represents the mean climate value for each month individually in the period of 20 continues years. To make the problem more simple one can merge values of three months into a single value by making a mean value for three months. Following this procedure (graphically represented in Figure 1), the data is now mean valued for each seasons in the period of 20 years, and each value represents seasonal climate value. Way in which seasons are split is the following:

- December, January and February are winter season,
- March, April and May are spring season,
- Jun, July and August are summer season,
- September, October and November are winter season.

When data is prepared in this way there is no problem to subtract the value that represents seasonal climate in the future simulations with the value of the historical simulation. That way obtained results, in a very intuitive way, represent *absolute climate change* between two climates.

Though the graphics made out of these results are comprehensive, they can still be improved. By applying statistical analysis on the data it is possible to locate which areas have experienced "significant" climate change. Method used for this purpose is called *Student's t-test*. For preforming a t-test one should not use previously mentioned seasonal values, but instead make some data manipulation. When importing the values for each month into variables, these values should not be mean valued because it is important to have population variance when preforming the t-test. That way each month has 20 values, and after merging three months into a season, each season will be represented with 60 values. Now t-test (in Matlab there is a function called ttest2) can be preformed for the same seasons but between future simulations and historical simulation. Results of preforming this test give *student's distribution* and from it one can obtain p-value. The p-value is the probability of obtaining results as extreme or more extreme than the ones observed given that the null hypothesis is true[1]. Thus, the smaller the p-value, the more evidence you have to reject the null hypothesis. Usually the rejection criterion is set to $\alpha = 0.05$. In this case, the null hypothesis is that the climate of two different periods are the same, so if the p-value is less than α, null hypothesis is rejected and as a result two climates are significantly different from each other. Processing the data in this manner gives a *significance of climate change relative to each another*.

Plotting the results over an area of interest is vital, because the main goal is to produce comprehensive plots with as much information as possible. This is done by making a contour of the absolute climate change, and on top of that contour, mark the locations on the grid with p-values lower than α, which means that the climate change for that specific grid point is significant. Figures 2, 3, 4, 5 are made by following this procedure in Matlab. They represent results for seasonal climate change of CanESM2 model, in Europe, for temperature and precipitation in the period of 2016–2035, 2046–2065 and 2081–2100 compared with the climate of 1986-2005.

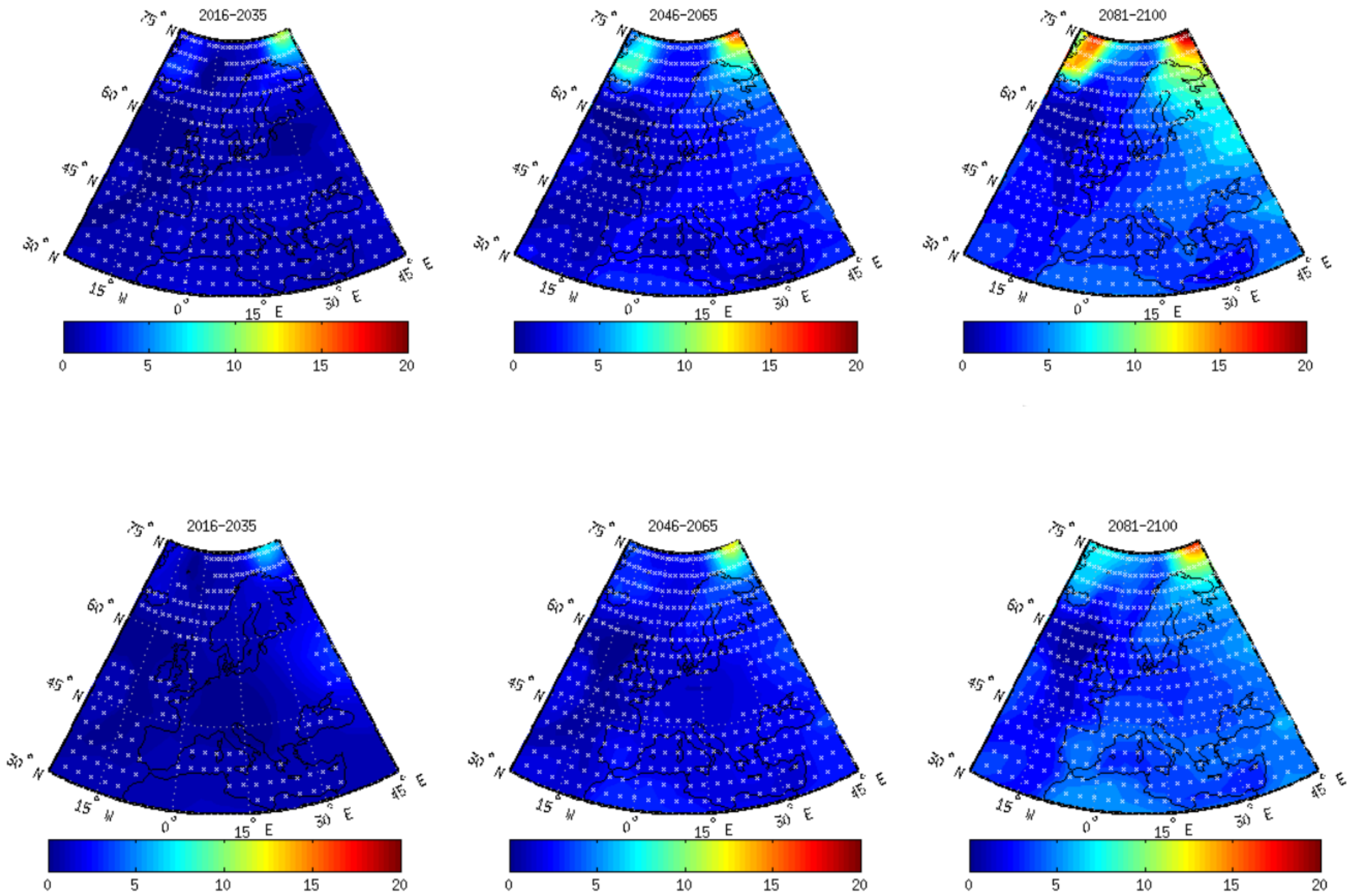

Figure 2: Air Temperature difference and climate change significance for winter (upper row) and spring season (lower row).

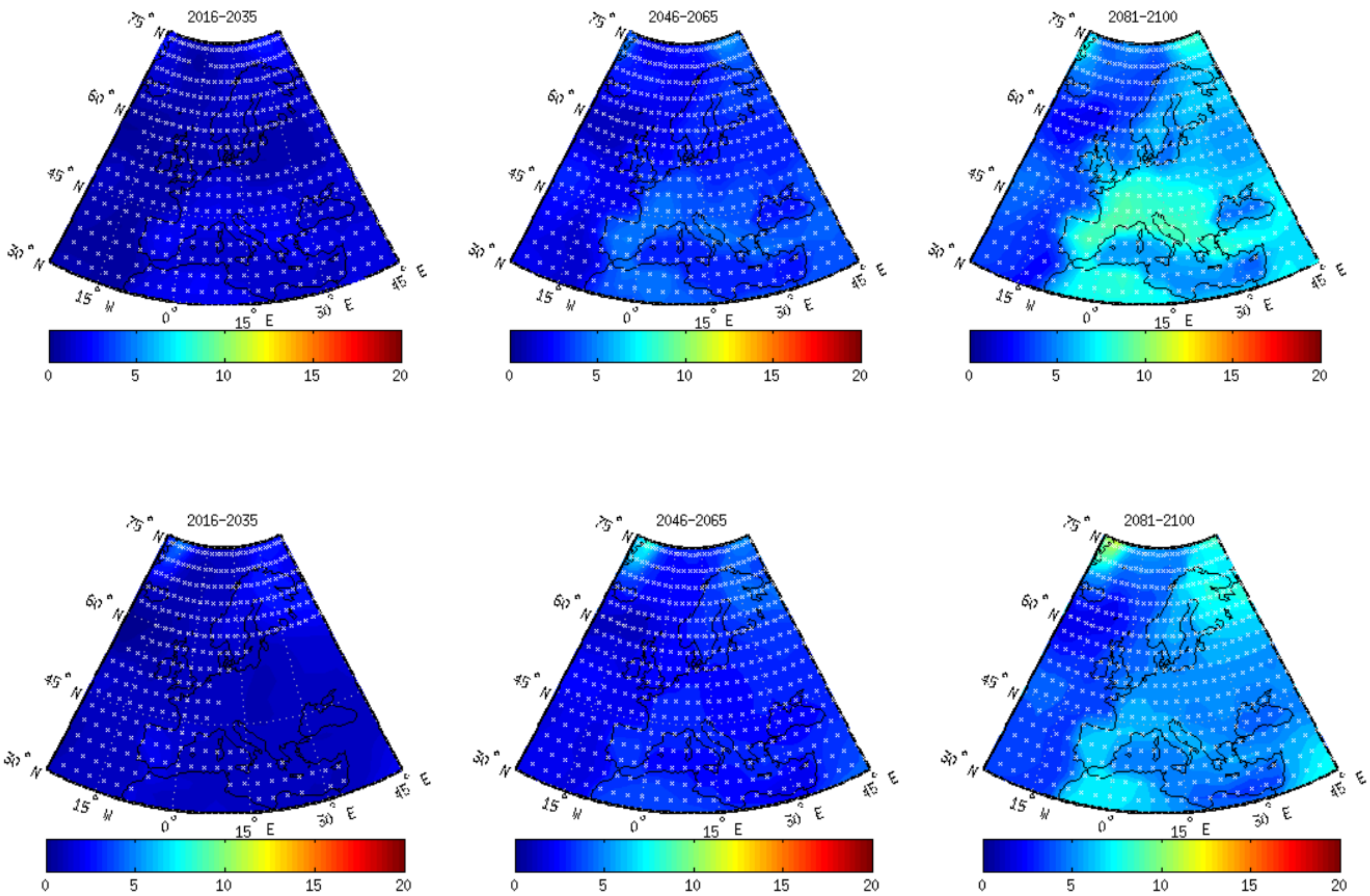

Figure 3: Air Temperature difference and climate change significance for summer (upper row) and autumn season (lower row).

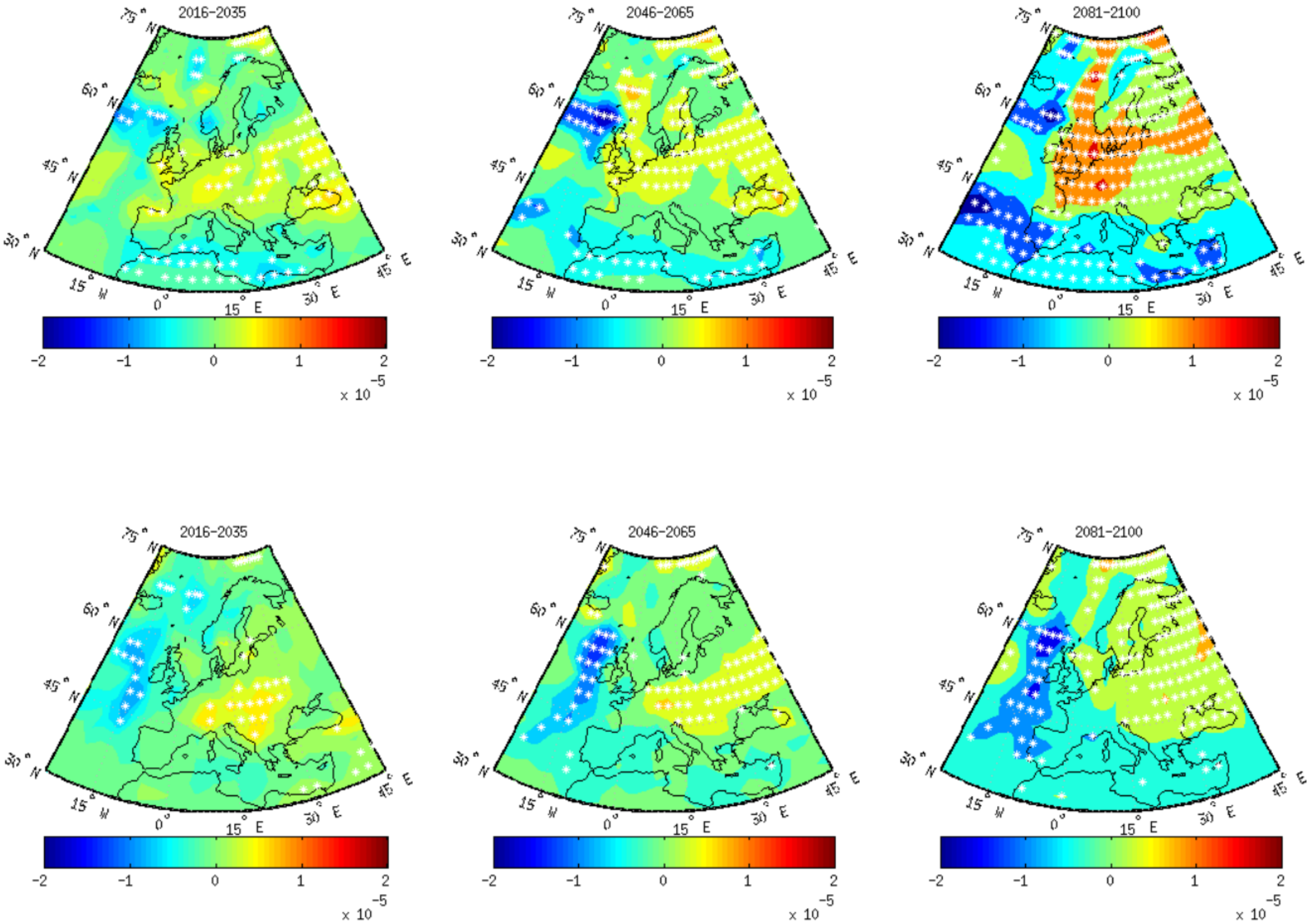

Figure 4: Difference in amount of precipitation and climate change significance for winter (upper row) and spring season (lower row).

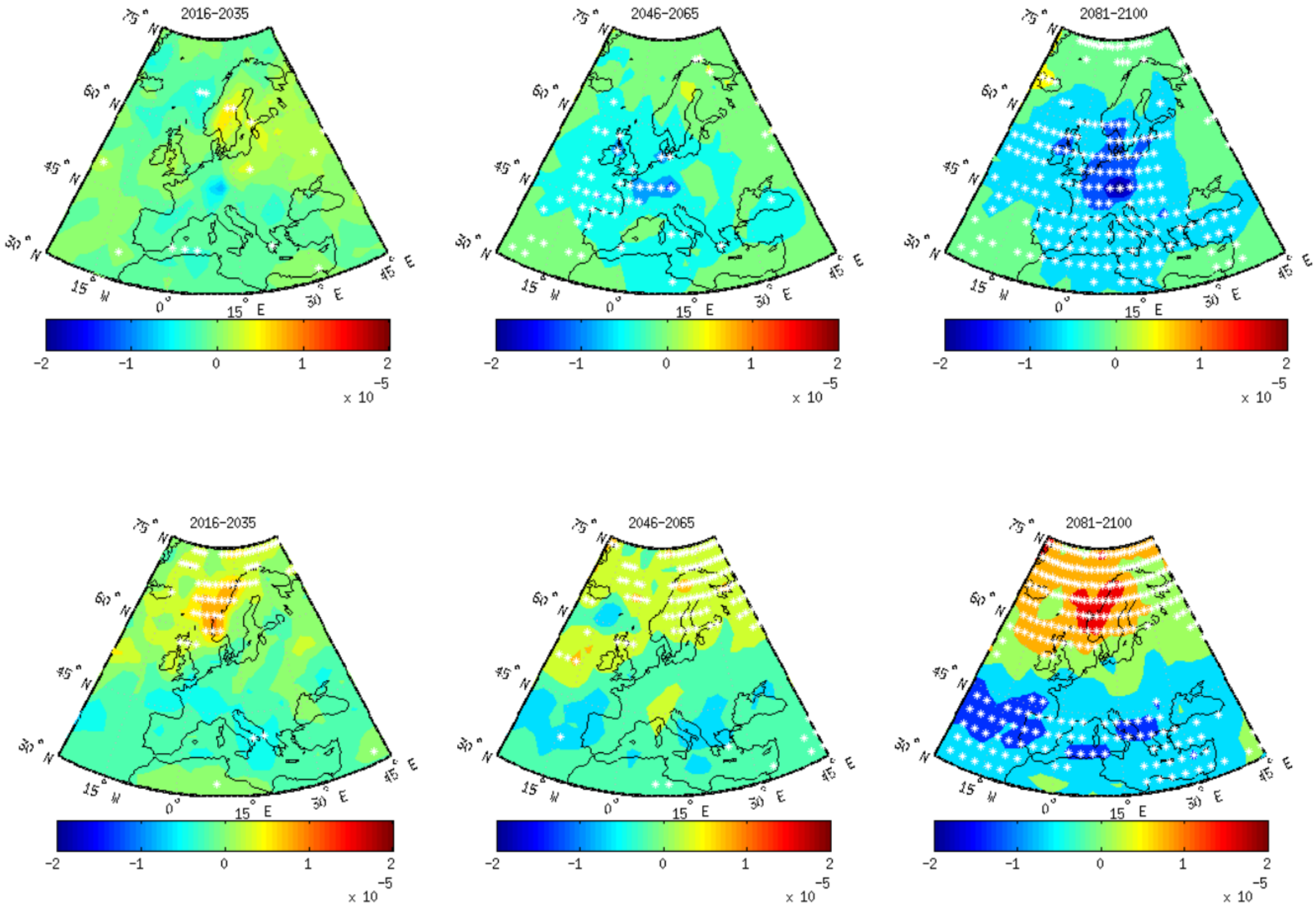

Figure 5: Difference in amount of precipitation and climate change significance for summer (upper row) and autumn season (lower row).

Acknowledgments

I wish to express my gratitude to Prof. Alfredo Rocha for his patience and guidance, and also to Tiago Luna for helpful discussions.

---

[1] Woodward W. A. Elliot A. C. *Statistical Analysis, Quick Referance Guidebook, SAGE.* 2007.