Deliverable D3.6

Description of the

final version of all gridded data sets of the climatology of the Carpathian Region

Contract number: OJEU 2010/S 110-166082
Deliverable: D3.6
Author: Igor Antolovic, Dragan Mihic
Date: 12.07.2013
Version: Final

Contents

1. Introduction

The aim of this deliverable is to describe the final version of the gridded datasets of the climatology of the Carpathian region and to extend the concepts that have been previously presented in deliverable D3.4. The final gridded data are stored in a PostgreSQL database, therefore this deliverable will compare the initial data format (ASCII files) with the final binary representation of the gridded data and finally present the two download options.

2. Raw gridded datasets

The term "raw" refers to the fact that all gridded data were initially produced in simple ASCII format. Though this format is easy to read from a standard text editor and to parse from a third-party application, this format is not efficient enough for fast querying.

2.1 Time resolution

Daily gridded data series for the period 1961-2010 (18262 days).

2.2 Area of interest and spatial resolution

Area: between latitudes 44°N and 50°N, and longitudes 17°E and 27°E, except Bosnia and Herzegovina.
Spatial resolution: $0.1^{\circ} \times 0.1^{\circ}$ grid
Number of grids: 5895 = 6161 (whole area) – 266 (Bosnia and Herzegovina)

2.3 The countries

Data homogenization, quality control, completion and gridding were implemented per country using the software Multiple Analysis of Series for Homogenization (MASH) and Meteorological Interpolation based on Surface Homogenized Data Basis (MISH).
The countries (partially) covered by the dataset are (in anti-clockwise direction): Hungary and Croatia (1), Serbia (2), Romania (3), Ukraine (4), Slovakia (5), Poland (6), Czech Republic (7). The grids per countries are shown in Figure 1.

Figure 1. The area of interest with country names (except Bosnia and Herzegovina)

2.4 The gridded daily data series compiled in one file

The gridded series from the individual countries have been compiled in one file per variable, with the names **{Variable}DailyGrid.ser**, see Table 1.
Such a file includes 5895 gridded daily data series for the period of 1961-2010. Each column corresponds to one grid series, and the file format is:

row 1: index of the series (j=1,…,5895)
column 1: year (i4)
column 2: month (i3)
column 3: day (i3)
column 3+j: series j (f8.2), j=1,…,5895

The spherical coordinates λ, ϕ of the grids with the country index (1-7) are listed in the file **PredtandfilaGrid.dat** for the 5895 grids of the area.

2.5 The variables

The variables and their compiled gridded daily data series files are shown in the Table 1.

Table 1. File names for compiled daily variables

Variable name	File name
Minimum air temperature	**TminDailyGrid.ser**
Maximum air temperature	**TmaxDailyGrid.ser**
Mean air temperature	**TaDailyGrid.ser**
Precipitation	**PrecDailyGrid.ser**
Sunshine duration	**SunshineDailyGrid.ser**
Cloud cover	**CcDailyGrid.ser**
Global radiation	**RgDailyGrid.ser**
Relative humidity	**RhDailyGrid.ser**
Vapor pressure	**PvapDailyGrid.ser**
Surface air pressure	**PairDailyGrid.ser**
Snow depth	**SnowdDailyGrid.ser**
Snow water equivalent	**SnowweDailyGrid.ser**
10m wind speed	**WindVVDailyGrid.ser**
2m wind speed	**WindVV2mDailyGrid.ser**
10m wind direction	**WindDDDailyGrid.ser**
Maximum 10m horizontal wind speed	**WindMaxDailyGrid.ser**

3. Final gridded database

In order to get a more efficient data storage, the data needed to be converted from ASCII to binary format. What is more important is that most of the data is decimal type with a maximum of two decimal digits.

3.1 Data Storage

Data storage will be illustrated on the example of the Mean Air Temperature grid.

- The **TaDailyGrid.ser** ASCII file takes about 800MB
- All temperature values inside **TaDailyGrid.ser** are in decimal format with two decimal digits (i.e. Ta= 14.45 °C)

If all values in the **TaDailyGrid.ser** are examined it can be concluded that they roughly fall in the range -40.00 to +50.00 °C. In order to convert the data to integer (i.e. eliminating the decimal

point) these values can be multiplied with 100, thus resulting in a new range from -4000 to +5000. By checking Table 2 it is clear that this range falls into the "Signed Short" category meaning that every value can be stored by using 2 bytes.

Table 2. Integer types and ranges

Integer type	Bytes	Range
Unsigned Short	2	0 to 65535
Signed Short	2	–32768 to 32767
Unsigned Integer	4	0 to 4294967295
Signed Integer	4	–2147483648 to 2147483647

The total amount of bytes needed to store the gridded data in binary format can be calculated as:

Size =50 (Years) * 365 (Days) * 5895 (Num. of grid points) * 2 (Bytes per value) ≈ 205 MB

This result shows that gridded data in binary format takes 4 times less space than the original ASCII format. The other advantage of binary data is that it can be accessed faster.

3.2 Storing gridded datasets into PostgreSQL database

The process of storing gridded datasets into PostgreSQL database has several steps:

- Read data for each day from the original ASCII files
- Convert data from ASCII to binary
- Create a new row in the appropriate table

Every table row contains the following columns:

period_id – This column contains a period code (1- Daily, 2 – Monthly, 3 – Yearly)
date_start – This column contains the start date of the period
data - This column contains the gridded data in binary format

Every variable/indicator is stored in a separate table in the database. The advantage of this approach is that every individual variable/indicator can be backed up or restored without affecting other variables/indicators in the database.

For the purpose of the CARPATCLIM project all gridded data has been converted to binary data and imported into a PosgreSQL database. The process of installation of the PostgreSQL database and the required data is described in the installation manual included in the deliverable D3.13.

3.3 Data download

All gridded data sets are stored in the PostgreSQL database while the users can obtain a specific grid (gridded data for specific variable/period/country) by using the CARPATCLIM Web portal interface. It is important to emphasize that the users have two main options for data download:

- **Single grid download** – Gridded data for a specific set of selected options (variable, period, country, altitudes). The downloaded file is a .zip archive containing the following files:
  - Gridded data in ESRI GRID format (suitable for additional analysis within popular GIS applications like ArcGIS, Quantum GIS, GRASS GIS, SAGA GIS)
  - Interpretation sheet containing all necessary information for interpretation of the downloaded data
  - Rendered image of the selected gridded data (suitable for presentations or Web site upload)
  - Computational methodology containing mathematical equations used in the process of computing the selected variable/indicator.

- **Bulk download** – Complete gridded data sets for the period 1961-2010. This type of download is suitable in the case that user wants to process the original data in a third-party application without using the CARPATCLIM Web portal. The downloaded file is a .zip archive containing the following files:
  - Gridded data in ASCII format
  - Interpretation sheets containing all necessary information for interpretation of the downloaded data sets
  - Disclaimer text
  - File with spherical coordinates λ, ϕ of the grids with the country index (1-8) - **PredtandfilaGrid.dat**
  - Computational methodology containing mathematical equations used in the process of computing the selected variable/indicator

Detailed information regarding data download with step by step descriptions can be found in the deliverable D3.12.

3.4 Bulk download file names and format

When a user wants to download data over the bulk download, after filling the simple form with some basic personal information, receives the .zip file with gridded data file accompanied by several files containing disclaimer text and all necessary information about the downloaded data sets. Interpretation sheets and computational methodologies are in pdf format, while files containing disclaimer text, gridded data and coordinates of the grids are in ASCII format. File format described in the interpretation sheets is basically presented in the chapter 2.4 of this document. There are some differences regarding temporal resolution of the gridded data, the country index and names of the files. Depending on the temporal resolution of data columns two and three (month, day) are missing. Files with monthly data do not have column 3, while files with yearly data do not have columns 2 and 3. Hungary and Croatia have separate country indexes (1 and 8, respectively).

The complete data sets give total coverage of the CARPATCLIM interest area in the period 1961-2010. Each grid consists of 61 rows and 101 columns. The data start at the NW upper corner of the files, position 50° lat N and 17° long E and are arranged in latitudinal bands. The data ends at the SE lower corner, position 44° lat N and 27° long E. The ASCII (plain text, space delimited) file has a data format as follows:

One column is one series, and the format is:
row 1: indices of the series (j=1,…,5895)
row i: rows containing grid data (i>1), each row containing one day, month or year

- column 1: year
- column 2: month (missing for yearly data)
- column 3: day (missing for monthly and yearly data)
- column 3+j: series j (j=1,…,5895)

Longitude and Latitude of the grid points with the country indexes (1-8) and their altitudes are listed in the file PredtandfilaGrid.dat (ASCII, plain text, space delimited, with header row describing the columns) for the 5895 grid points of the CARPATCLIM area. The country names corresponding to the country indexes are as follows:

1. Hungary
2. Serbia
3. Romania
4. Ukraine
5. Slovakia

6. Poland
7. Czech Republic
8. Croatia

Grid cell with no value: -9999

Data files have generic names: CARPATGRID_{variable/indicator acronym}_{temporal resolution}.ser List of variables/indicators with their acronyms and temporal resolutions (frequency) is presented in the Table 3. More detailed description about variables/indicators and computational methods can be found in the deliverable D3.7.

Table 3. List of variables/indicators

No.	Indicator Acronym	Description	Frequency
1.	TA	Average mean air temperature	Daily/Monthly/Yearly
2.	TMIN	Average minimum air temperature	Daily/Monthly/Yearly
3.	TMAX	Average maximum air temperature	Daily/Monthly/ Yearly
4.	PREC	Accumulated total precipitation	Daily/Monthly/ Yearly
5.	WS10	Average 10m horizontal wind speed	Daily/Monthly
6.	WS2	Average 2m horizontal wind speed	Daily/Monthly
7.	WD10	10m wind direction	Daily
8.	WMAX10	Maximum 10m horizontal wind speed	Daily
9.	SUN	Sunshine duration	Daily/Monthly/Yearly
10.	CC	Average cloud cover	Daily/Monthly
11.	RG	Global radiation	Daily/Monthly
12.	RH	Average relative humidity	Daily/Monthly
13.	PV	Mean vapour pressure	Daily/Monthly
14.	PA	Mean surface air pressure	Daily/Monthly
15.	SNOW	Snow depth	Daily/Monthly
16.	SWE	Snow water equvivalent	Daily/Monthly
17.	FD	Number of frost days (Tmin < 0^oC)	Monthly/ Yearly
18.	PFD	Percentage of frost days (Tmin < 0^oC)	Monthly/ Yearly
19.	SD	Number of summer days (Tmax > 25^oC)	Monthly/ Yearly
20.	PSD	Percentage of summer days (Tmax > 25^oC)	Monthly/ Yearly
21.	HD	Number of hot days (Tmax > 30^oC)	Monthly/ Yearly
22.	PHD	Percentage of hot days (Tmax > 30^oC)	Monthly/ Yearly
23.	PAI	Palfai Drought Index	Yearly
24.	SPI-3	Standardized Precipitation Index – 3-months	Monthly
25.	SPI-6	Standardized Precipitation Index – 6-months	Monthly
26.	SPI-12	Standardized Precipitation Index – 12-months	Monthly
27.	SPEI-3	Stand. Prec. Evapotranspiration Index – 3-months	Monthly

28.	SPEI-6	Stand. Prec. Evapotranspiration Index – 6-months	Monthly
29.	SPEI-12	Stand. Prec. Evapotranspiration Index – 12-months	Monthly
30.	RDI-3	Reconnaissance Drought Index (3-months)	Monthly
31.	RDI-6	Reconnaissance Drought Index (6-months)	Monthly
32.	RDI-12	Reconnaissance Drought Index (12-months)	Monthly
33.	PDSI	Palmer Drought Severity Index	Monthly
34.	ID	Number of ice days (Tmax < 0^{o}C)	Monthly/ Yearly
35.	PID	Percentage of ice days (Tmax < 0^{o}C)	Monthly/ Yearly
36.	EHD	Number of extremely hot days (Tmax ≥ 35^{o}C)	Monthly/ Yearly
37.	PEHD	Percentage of extremely hot days (Tmax ≥ 35^{o}C)	Monthly/ Yearly
38.	ECD	Number of severe cold days (Tmin < -10^{o}C)	Monthly/ Yearly
39.	PECD	Percentage of severe cold days (Tmin < -10^{o}C)	Monthly/ Yearly
40.	GSL	Growing season length	Yearly
41.	WD	Number of wet days (RR ≥ 1 mm/day)	Monthly/ Yearly
42.	PWD	Percentage of wet days (RR ≥ 1 mm/day)	Monthly/ Yearly
43.	EWD	Number of wet days with (RR > 20 mm/day)	Monthly/ Yearly
44.	PEWD	Percentage of wet days with (RR > 20 mm/day)	Monthly/ Yearly
45.	M1DTOT	Maximum 1-day total rainfall	Monthly/ Yearly
46.	M5DTOT	Maximum 5-day total rainfall	Monthly/ Yearly
47.	ARI	Aridity index	Monthly
48.	MI	Moisture index	Monthly
49.	EI	Ellenberg index	Yearly
50.	CDD6	Cooling degree days (summer)	Yearly
51.	HDD6	Heating degree days (winter)	Yearly
52.	GDD8	Growing degree days (extended summer)	Yearly
53.	PET	Potential evapotranspiration	Monthly

4. REFERENCES

D 3.4 Preliminary version of gridded datasets of the climatology of the Carpathian Region

D 3.7 Final report on the production of the climatologies of the Carpathian Region

D 3.12 Final and public version of the implemented web site with full functionality that hosts all relevant information on the Climate Atlas of the Carpathian Region, including a public download functionality with the web site for all gridded datasets of the climatology

D 3.13 Complete software of the website including all web pages, data and source code of the web site, including a description of system requirements and instructions for installations

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-No. 41, pp. 27-46.

Szentimrey, T., 2008: Development of MASH homogenization procedure for daily data, Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary, 2006; WCDMP-No. 68, WMO-TD NO. 1434, 2008, pp. 116-125.

Szentimrey, T., 2011: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, pp. 64. Szentimrey, T., Bihari, Z., 2006: MISH (Meteorological Interpolation based on Surface Homogenized Data Basis), COST Action 719 Final Report, The use of GIS in climatology and meteorology, Edited by Ole Einar Tveito, Martin Wegehenkel, Frans van der Wel and Hartwig Dobesch, 2006, pp. 54-56

Szentimrey, T., Bihari, Z., 2007a: Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis), Proceedings from the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2004, COST Action 719, COST Office, 2007, pp. 17-27

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=About_the_ESRI_grid_format