Climate Twins - An Attempt to Quantify Climatological Similarities

Joachim Ungar, Jan Peters-Anders and Wolfgang Loibl

AIT - Austrian Institute of Technology,
Department of Foresight and Policy Development
Tech Gate Vienna, Donau-City-Str. 1, 1220 Vienna, Austria
joachim.ungar@gmail.com

Abstract. As climate change appears, strategies and actions will be necessary to cope with its effects on environment and society in the coming decades. Current climate conditions can be observed everywhere in the world but future climate conditions can only be estimated through climate simulations which produce huge amounts of quantitative data. This data leads to statements like "temperature increase is expected to exceed 2.6°C" or similar and remain fuzzy to non-experts in climate research. The Climate Twins application is designed to communicate climate changes in an intuitive and understandable way by showing regions which have now similar climate conditions according to a given Point of Interest (POI) in the future. This paper explains how the application seeks for locations with similar climatological patterns according to the POI. To achieve this goal a method has been developed to quantify similarity between two locations' climate data.

Keywords: Climate Change, Similarity Measures, Web Mapping

1 Introduction

To allow "real world insights" about future climate impact and appropriate adaptation, one can look at model regions, where the current climate appears similar to an expected future climate of a POI. We call such region pairs with similar climate conditions (at different times) "Climate Twins". From these (remote) current Climate Twin region parts we can learn "hands on" how future climate impacts may be experienced in the POI and how to adapt there to the changing climate conditions, expected in the future.

The idea of Climate Twins is to identify regions whose current climate conditions show high similarity to the expected future climate in the POI. The Climate Twins search tool is a web-based graphical user interface (GUI) allowing to explore climate change effects based on maps of current and future climate.

To identify climatological similarity seems to be a simple exercise but the accuracy and validity of the result strongly depends on the indicators used and the similarity thresholds defined. A huge number of indicators in combination with narrow threshold ranges will reduce the number of matching regions significantly as well as few indicators combined with wide thresholds will show a big number of matching regions.

The climate indicators used here are daily mean temperatures and daily precipitation because they are seen as the most important ones and provide sufficient input for proving the concept's applicability. The most important part was to find a suitable matching method which strongly depends on the quantification of similarity between any two data vectors.

This matching method now provides

- a “unit-less” similarity value able to be combined with similarity values of other indicators,
- information of the degree of similarity to derive statements like “more similar than” or “less similar than”, and
- a consideration of many statistical properties because whole statistical distributions are being compared.

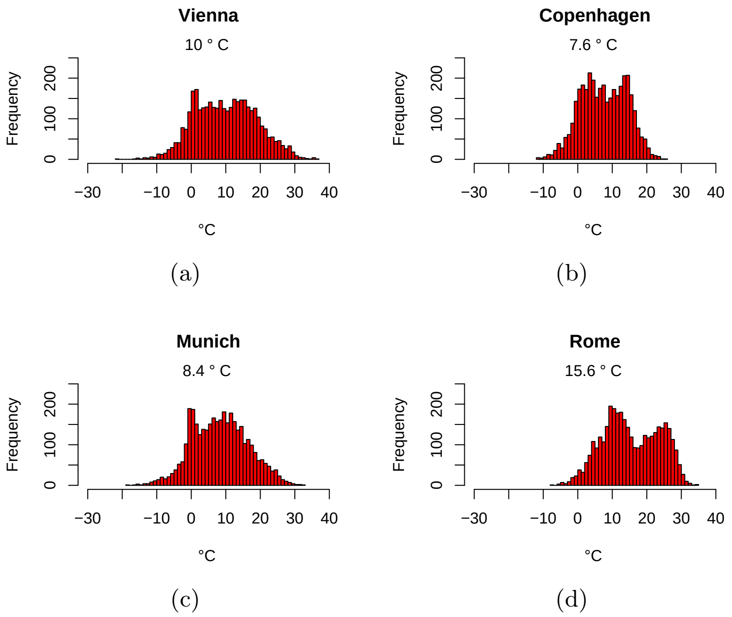
2 Theory and Methodology

2.1 Climate Data

Climate can be seen as a statistical collection of various climate variables. These variables are either measured or modeled in various time steps and therefore come as a list of values. A statistical distribution of these values can be described by three main attributes: dispersion (measure of variability), skewness (measure of asymmetry) and kurtosis (measure of peakedness) or by aggregations like mean, median or range.

There are various ways to quantify these properties but especially measuring skewness and kurtosis is challenging and the results are not always satisfactory. By using conventional methods major problems have to be faced as climate data is rather not normally distributed. Furthermore the results have to be combined to a single attribute afterwards which also leads to problems in weighting them in an appropriate manner.

Figure 1 shows frequency distributions of modelled daily temperature means. A quick visual interpretation shows that the more values are located on the right hand side, the warmer the location (e.g. Rome). The annual temperature amplitude equals the value range. As Vienna and Munich have a higher value range, the annual temperature range is wider due to their rather continental location. On the other hand Rome and Copenhagen, wich are located in maritime locations, show a narrow value range. Bipolar distributions indicate strong and distinct seasons like winter and summer with short and alternating changeovers in spring and autumn (Rome) whereas Gauss-like distributions indicate a more homogeneous climate (Vienna, Copenhagen, Munich) and so on.



(a) (b)

(c) (d)

Fig. 1. Frequency distributions of daily mean temperatures 2001 to 2010

2.2 Similarity Measures

A similarity measure in this context should define and quantify the similarity between two statistical distributions—i.e. the numerical attributes we used to describe climate. Vegelius et al. [6] did a comparative analysis on various similarity measures. His research group tested measures according to predefined criteria. The three main criteria being:

1. The result of the measurement (r) has to be a value between 0 and 1,
2. between two identical distributions r has to be 1 and
3. r has to be equal when measuring in both directions.

After the analysis the group pointed out two measures which fit to all criterias given. These two were the *Proportional Similarity* (PD, 1) and the *Hellinger Coefficient* (r_{H}, 2).

$$PD(U,V) = \sum_{i=1}^{C} min(f_{\mathrm{Ui}}, f_{\mathrm{Vi}}) \tag{1}$$

$$r_{\mathrm{H}}(U,V) = \sum_{i=1}^{C} \sqrt{f_{\mathrm{Ui}} * f_{\mathrm{Vi}}} \tag{2}$$

U, V are two distributions, r similarity value, C being "category". Both measures use relative frequencies to measure similarity. Having data on a categorical or ordinal scale these categories are already defined. Given for examples two farms with cattle, chickens and sheep it is possible to quantify the relative similarity by using the frequencies of the three species living in each farm. In this example each species is a category.

2.3 Application

Problem: Climate data though occurs on an interval (temperature) or ratio (precipitation) scale where the borders—value ranges—that enclose the categories have to be defined manually in order to use these similarity measures.

Solution: The number of categories a distribution is split into determines the resolution of the measure (and therefore the accuracy of the measurement). If there is only one category describing each distribution, this category contains 100% of the values which leads to a r value of 1. The more categories are introduced, the more convincing the similarity measurement gets but for the Climate Twins application a meaningful number of categories had to be found where on the one hand the r value provides a valid similarity indicator and computer ressources are used effectively on the other hand.

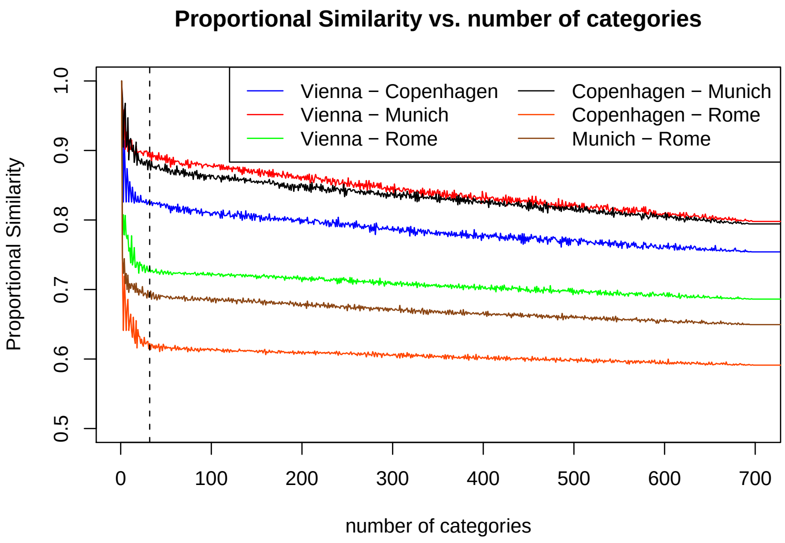


Fig. 2. Influence of category number on similarity measures (daily mean temperatures 2001-2010).

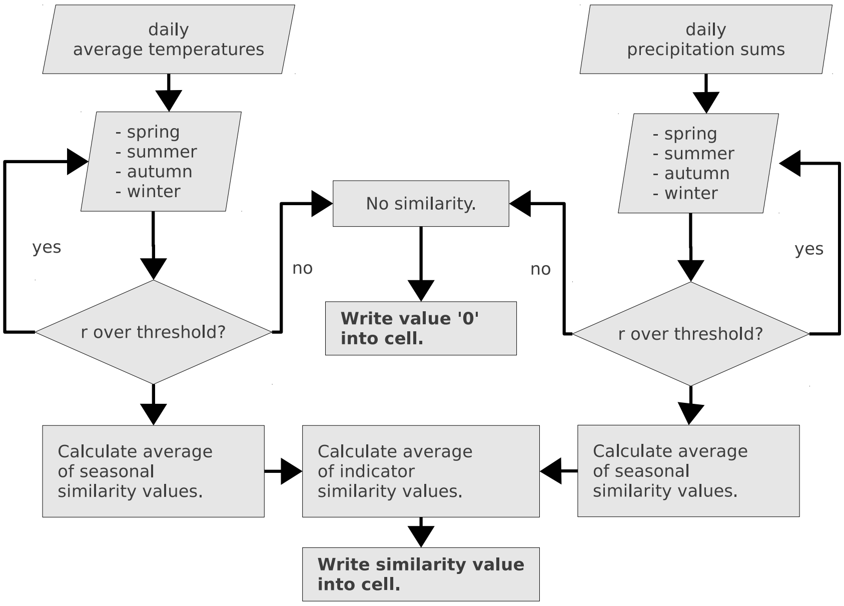

daily
average temperatures
daily
precipitation sums
- spring
- summer
- autumn
- winter
- spring
- summer
- autumn
- winter
yes
no
No similarity.
no
yes
Write value '0' into cell.
r over threshold?
r over threshold?
Calculate average of seasonal similarity values.
Calculate average of indicator similarity values.
Calculate average of seasonal similarity values.
Write similarity value into cell.

Fig. 3. Logical structure of combining r-values while define similarity between two locations. The structure is scalable, i.e. any number of indicators can be implemented as long as respective similarity thresholds are applied.

In Figure 2 the behaviour of Proportional Similarity values when increasing the category number is shown. Visually the resulting curve can be divided into three parts: (1) when having one category the r values have—as expected—a value of 1. When increasing the number the curve shows major fluctuations until it (2) stabilizes after approximately 30 categories. The r values stay (3) constant when having approximately 700 categories and more.

It is important to specify an amount of categories where valid results are possible on the one hand and the computing effort is as low as possible on the other hand. Therefore the number of categories should be a lower value of part two of the curve. Due to other reasons like the total possible value range, the categories for the Climate Twins application were set to 37 for temperature (every 2°C between -30°C and 40°C and one category each below and above that range) and 29 categories for precipitation (1 mm width each category from 0 to 10 mm, 5 mm width from 10 to 100 mm and one category for 100 mm and more).

Problem: A major problem in applying this method on climate data was losing all temporal information. A pure frequency distribution does not contain the chronological information of the distribution values anymore. Therefore two locations where the precipitation peaks occur at location A in spring and at location B in autumn will show erroniously high similarity.

Solution: The data is split up in seasonal data by aggregating to spring (MAM), summer (JJA), autumn (SON) and winter (DJF). After computing the respective seasonal r values they are recombined by averaging to an annual value.

Recombination can be done easily at least in combining seasonal r values to an annual value as the r value is "unit-less". To asses the problem in combining the similarity values of two different indicators a slider was implemented in the web application to interactively change the weighting. However, the results showed no significant change in Climate Twin result regions whilst variing the indicator weighting.

Figure 3 shows the logic behind the combination of single r values to an overall simlarity value. The most important part is that every single seasonal r value has to exeed a certain similarity threshold so that every overall similarity is mapped. If two locations match perfectly in three of four seasons but not in the fourth season, there is no point in declaring the two locations similar to each other.

Problem: As the declaration of similarity is a subjective one and up to some point an arbitrary process, so is the definition of the similarity thresholds. The thresholds should of course be tight eneough to provide a reliable result but on the other hand wide enough so that an acceptable amount of Climate Twin regions can be found. Furthermore an applicable threshold also depends on the indicator used and the category number as it can be seen in Figure 2.

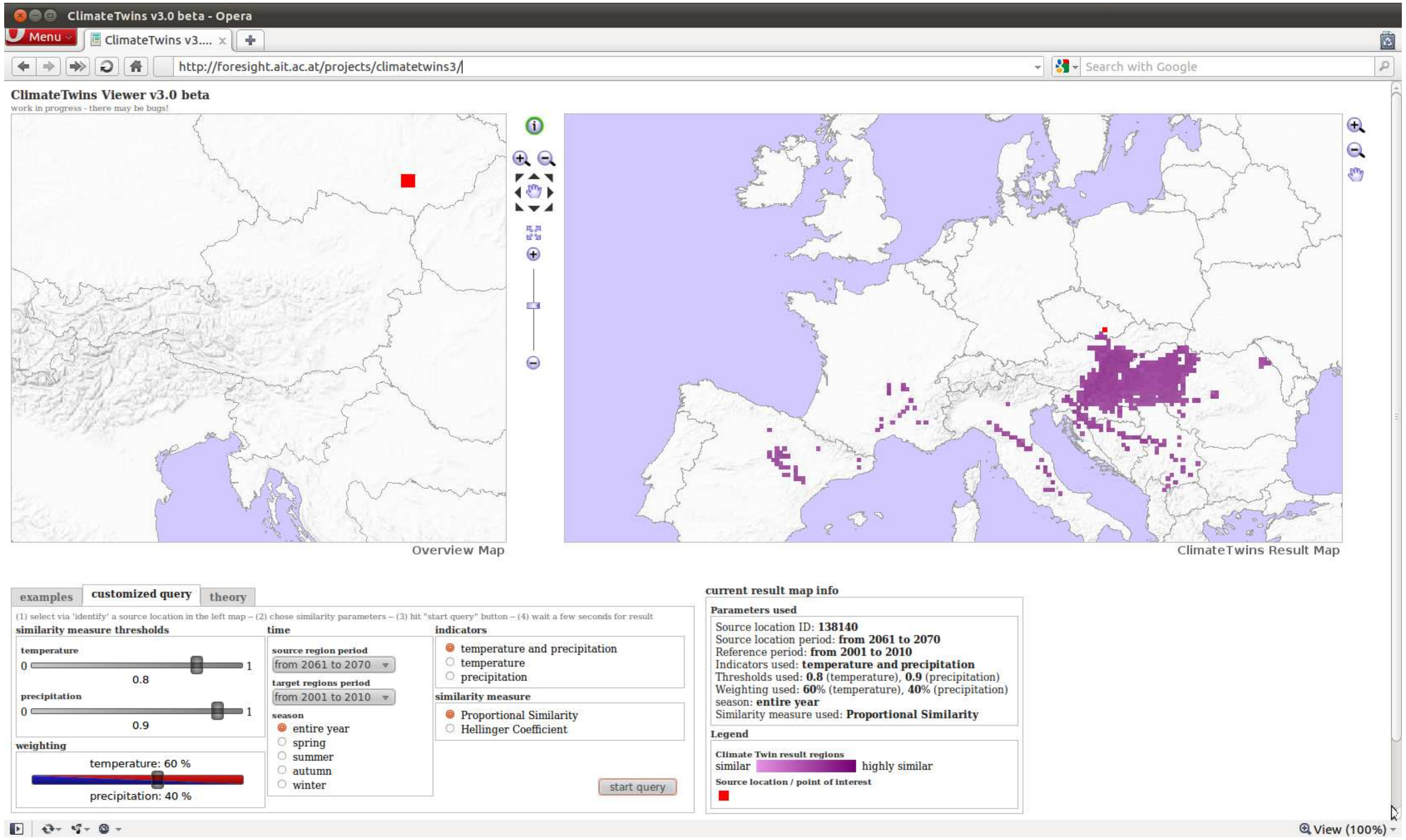
Solution: Until now no satisfying validation method or data could be found to compare the Climate Twin results with. Therefore the fictive line between "good" and "bad" results can only be drawn subjectively by visual interpretations of result maps while variing the thresholds. In the web application the user is enabled and encouraged to influence this parameter interactively through a slider. The problem of thresholds has to be faced in order to further development of this method.

3 Technical Infrastructure

3.1 Input Data

The input data is from the COSMO-CLM (COnsortium for Small-scale MOdelling - Climate Local Model) model 2.4.11 which is embedded into the ECHAM5/MPIOM global model. The model results are climate data on an hourly basis from 1960 to 2100 in a raster with a resolution of 0.165° (approx. 18 to 20 km)[2].

The input data for the Climate Twins exploration have to be prepared and "condensed" in advance for fast data retrieval and comparison. The data actually stored in the data base are the absolute frequencies of daily data aggregated seasonally and in fourteen blocks of ten years each.



ClimateTwins Viewer v3.0 beta

work in progress - there may be bugs!

examples | **customized query** | theory

(1) select via 'identify' a source location in the left map – (2) chose similarity parameters – (3) hit "start query" button – (4) wait a few seconds for result

similarity measure thresholds

temperature: 0 — 1: 0.8

precipitation: 0 — 1: 0.9

weighting

temperature: 60 %

precipitation: 40 %

time

source region period: from 2061 to 2070

target regions period: from 2001 to 2010

season

- ● entire year
- ○ spring
- ○ summer
- ○ autumn
- ○ winter

indicators

- ● temperature and precipitation
- ○ temperature
- ○ precipitation

similarity measure

- ● Proportional Similarity
- ○ Hellinger Coefficient

start query

current result map info

Parameters used

Source location ID: **138140**
Source location period: **from 2061 to 2070**
Reference period: **from 2001 to 2010**
Indicators used: **temperature and precipitation**
Thresholds used: **0.8** (temperature), **0.9** (precipitation)
Weighting used: **60%** (temperature), **40%** (precipitation)
season: **entire year**
Similarity measure used: **Proportional Similarity**

Legend

Climate Twin result regions

similar — highly similar

Source location / point of interest

🟥

Fig. 4. Screenshot of the current version Climate Twins Viewer 3 beta

3.2 Interactive Map Application

The Climate Twins map functions have been built on open source software making use of UMN map server's [11] capabilities of displaying file based geographical data and spatially enabled data through PostGIS [9] layers stored in a PostGreSQL [10] database while using JSP (Java Server Pages) technology to conduct the grid cell queries. On the client side the highly configurable Flamingo Viewer [8] is used to display the maps served by the map server and to communicate with the Climate Twins data cube.

Each click on a location in the interactive map triggers a database query and array comparison of future with current climate indicators from the selected grid cell related to the respective municipality. The comparison results of the query are stored in a 2^{nd} map view of the database, displayed in the application's second window.

4 Summary and Outlook

This paper showed the advanced matching method of Climate Twins v3 to identify locations or regions with similar climate conditions. By applying a similarity measure and configuring it for a use on interval and ratio scaled data, it was possible to build this prototype with a completely new matching method. The main advantages of this method is its scalability towards adding any number of further climate indicators and the representation of similarity's spatial fuzziness. The definition and quantification of similarity thresholds still is a challenge and not solved, yet.

The current Climate Twins application's climate data is based on the "business as usual" green house gas (GHG) increase scenario IS92b [1]. For the future, a set of climate scenarios of different GHG increase rates will be applied in order to show the "movement" of the Climate Twins areas over Europe (to be expected in south and southeast direction) with respect to larger time steps into future climate.

Semantic web technologies (developed in the EU FP7 project TaToo) may allow direct access to web sites related to Climate Twins areas in order to identify adaptation measures to cope better with further climate conditions.

References

1. IPCC. Climate Change 2007: The Physical Science Basis. Final Report Working Group 1, Intergovernmental Panel on Climate Change, Assessment Report 4., http://ipcc-wg1.ucar.edu/wg1/Report/AR4WG1_Print_SPM.pdf, Geneva, Swiss (2007)
2. Lautenschlager, M., Keuler, K., Wunram, C., Keup-Thiel, E., Schubert, M., Will, A., Rockel, B., and Boehm, U.: Climate simulation with CLM, scenario A1B run no.1, data stream 3: European region MPI-M/MaD (2009)

3. Loibl W., Beck A., Dorninger M., Formayer H., Gobiet A. and Schner W. (Ed.) (2007): reclip:more - research for climate protection: model run, evaluation, Executive summary, ARCsys. `http://systemsresearch.ac.at/projects/climate`, Vienna (2009)
4. Roeckner, E., et al.: The Atmospheric General Circulation Model ECHAM5. Part 1: Model Description, Report 349, Max Planck Institute for Meteorology (MPI), Hamburg (2003)
5. Ungar, J.: A Comparative Analysis of Region Pairs Matching Current and Future Climate Conditions. Diploma Thesis, University of Vienna, Department for Geography and Regional Research, Vienna (2011)
6. Vegelius, J., Janson, S., Johansson, F.: Measures of similarity between distributions. Quality and Quantity, 20(4):437–441 (1986)
7. Climate Twins Viewer 3, `http://foresight.ait.ac.at/projects/climatetwins3`
8. Flamingo Mapcomponents, `http://www.flamingo-mc.org`
9. PostGIS, `http://postgis.refractions.net`
10. PostGreSQL, `http://www.postgresql.org`
11. UMN Mapserver, `http://mapserver.org`