

**”CDRec”**  
Centroid Decomposition  
Recovery & Imputation Tool

# User’s Guide

**Eszter Börzsönyi**

University of Zürich  
Department of Informatics  
Database Technology Group

*Part of Master Project “Recovery of Missing Values based on Centroid Decomposition”  
supervised by M. Khayati and Prof. Dr. M. Böhlen in 2013.*

Ver.1, 29.01.2013

# Table of Contents

Overview .....	3
Installation .....	3
Running the tool .....	3
Windows .....	3
Mac, UNIX or Linux .....	3
Data source .....	3
Navigating the user interface.....	4
Display function .....	4
Recovery function.....	5
Decomposition function .....	7
Input file format.....	7
Troubleshooting.....	8

## Overview

This document provides an in-depth guide to understand the usage of the software tool CDRec. It gives help with the installation and launching, then guides through the full functionality of the tool by introducing the elements of the graphical user interface, and the way of user interaction.

## Installation

The software tool is packed in a file called CDRec.jar. You need to copy this file to a folder on your machine.

### Running the tool

You must have Java SE 6 or later installed to launch CDRec. See Troubleshooting section in case of any errors during launch.

#### Windows

Start the command line. (E.g. press the Windows key, type “cmd” then press Enter.)

Change the directory to which you have copied CDRec.jar. (E.g. type “cd C:\My\Path\CDRec”, then Enter.)

Copy and paste the following command into your command line, then press Enter.

```
java -jar CDRec.jar
```

#### Mac, UNIX or Linux

Note, that you need a window manager on UNIX or Linux systems to launch CDRec.

Start the Terminal and change the directory to which you have copied CDRec.jar. (E.g. type “cd ~/my/path/CDRec”, then Enter.)

Copy and paste the following command into your command line, then press Enter.

```
java -jar CDRec.jar
```

## Data source

CDRec connects to a dedicated Oracle database located at a fixed address of the university lab. The semantics and schema of the database is fixed. CDRec reads and exports data of time series from and into this database. If the database is not reachable from the launching machine, CDRec cannot be used.

## Navigating the user interface

The top bar of the main windows of CDRec is a navigation bar (tab bar) to select the functionality you wish to use.

The default function is “Display”, and you can select “Recover” or “Decomposition” (fig. 1).

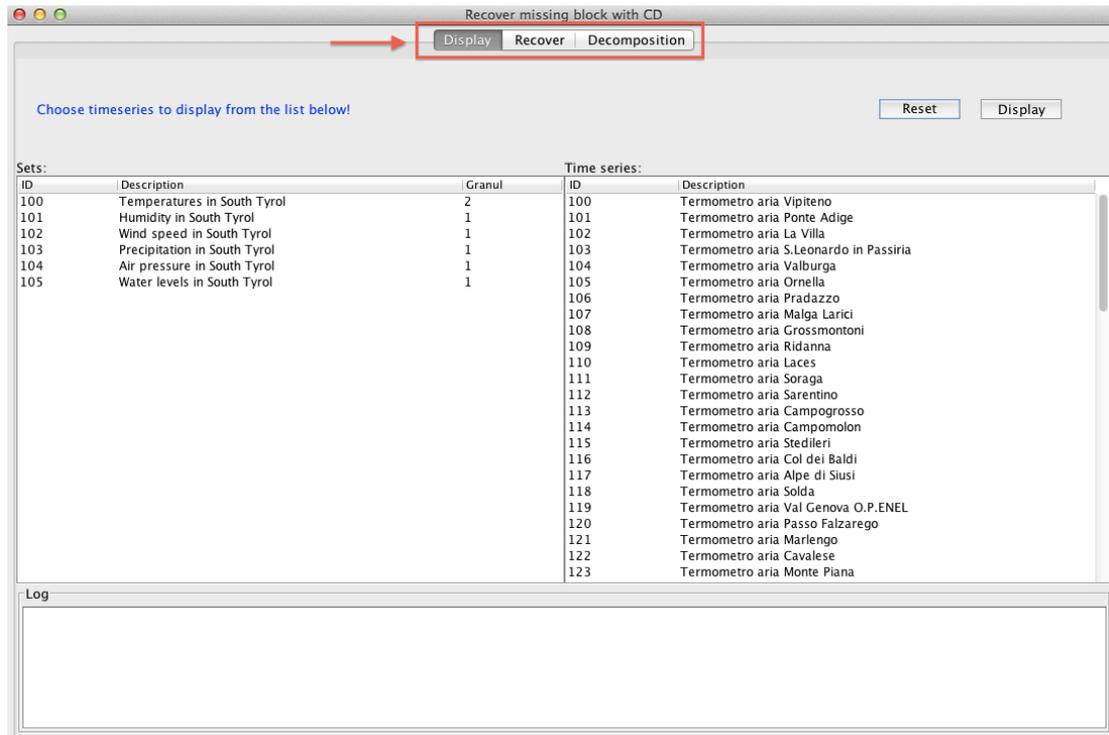


Figure 1 - The main screen

## Display function

The display function gives you the ability to explore and visualize time series data stored in the database.

First select a **set of time series** from the left hand list (marked as 1 on fig. 2). This updates the list on the right hand side (marked as 2 on fig. 2), so now you can choose any **time series** you would like to display.

If you click on any of the elements of the right hand side list, its ID number will be added to a list of **selected time series** (marked with 3 on fig. 2). You can reset this list anytime by pressing the “Reset” button on the pane (mark 4 on fig. 2).

Once you have selected your time series to display, (at least one), press “Display” (mark 4 on fig. 2) to let CDRec start reading the database and creating the graphs for you.

Please note, that while CDRec prepares the charts for you, you cannot perform any other operations to prevent unwanted modification of the data. You will see information on the progress inside the **log box** (marked as 5 on fig. 2).

Once the data is read, and the charts are produced, they appear in a new window (fig. 3). If you selected multiple time series, they will get color-coded to help you distinguish.

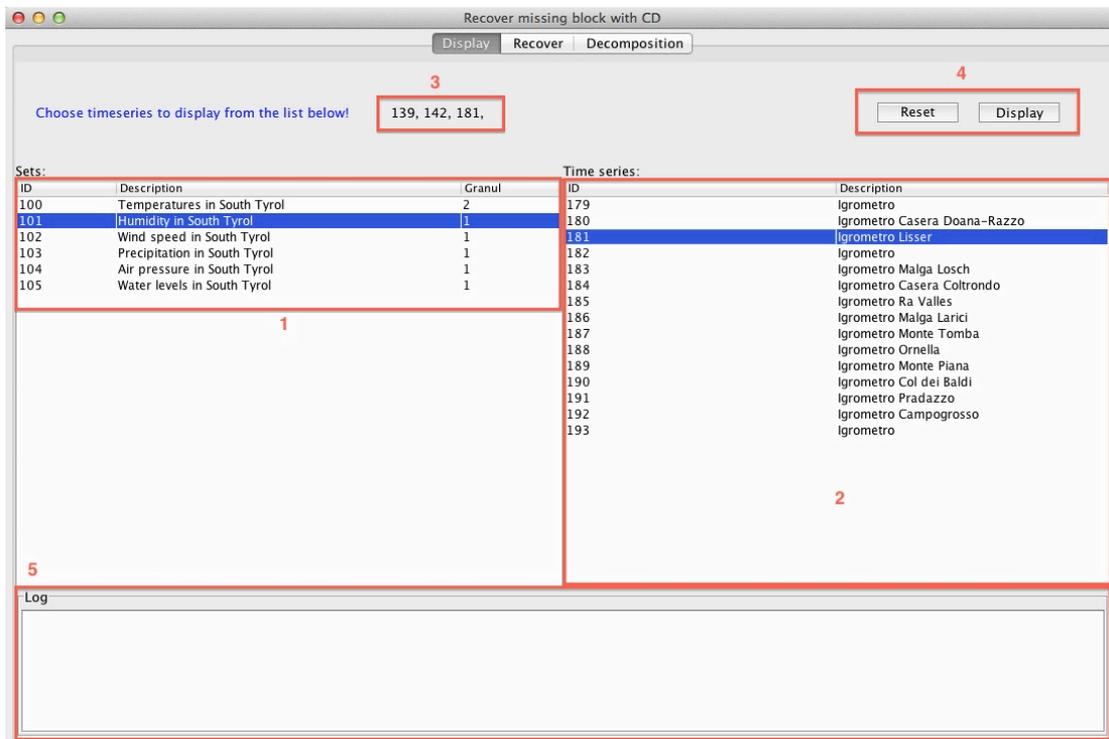


Figure 2 - Display pane

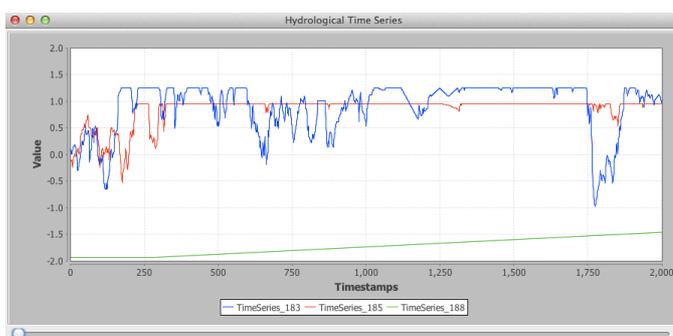
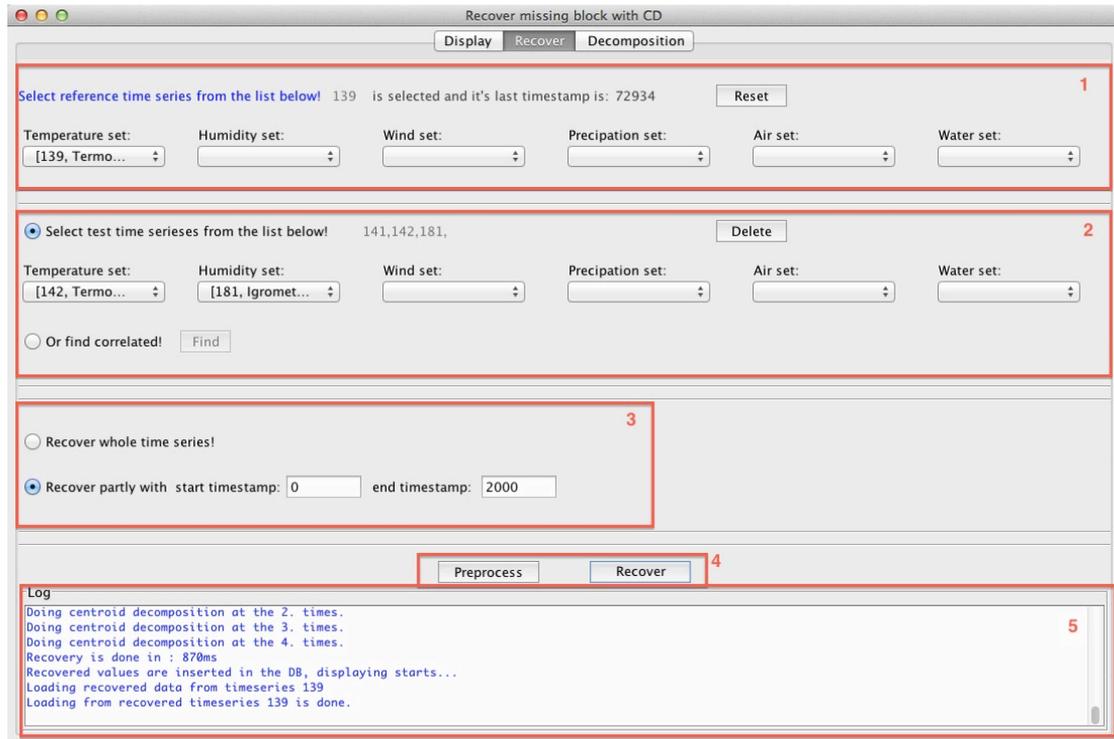


Figure 3 - Visualized time series

## Recovery function

First let's choose the time series of which you wish to recover missing values. This is called the **reference time series**.



You can do this in the section marked as **1** on fig. 4. Use any of the list boxes to see the set of available time series by their subject (temperature, humidity, wind, etc.). Note, that you can only select one reference time series. The selected time series will be the *last one* you clicked in any lists of the section. You can check the ID of your selection next to the blue label. Click the “**Reset**” button to clear the selection.

Once you have selected the reference time series, you need to choose a set of other time series to help the recovery (in the section marked as **2** on fig. 4). These are called **test time series**.

You have two choices to select test time series:

- **Select time series manually** from the lists. For this, click the first radio button called “Select test time series from the list below”. Then below you can browse all the available time series by subject again. Each one you click in any of the lists, will be added to your list of selections. You can see the list of IDs of your selected time series next to the radio button. Click “**Delete**” to clear the list of selections.
- **Find correlated time series**. For this, select the second radio button called “**Find correlated!**”. To start the automatic correlation-based selection process, click the button “**Find!**”. Please wait until CDRec finds the best time series for you, which can take a while.

After you have selected the test time series as well, you can specify the **time range for the recovery** in the section marked as **3** on fig. 4:

- **Recover the whole time series:** take all measurements, and run recovery on every missing block between the first and the last timestamp found in the database.
- **Recover segmented time series:** specify a starting and end timestamp. The recovery will only run on the missing blocks within the two specified time stamps.

Once you have selected the time range as well, you can start the preprocessing. This needs to be done before the recovery itself. This contains the step of reading the selected time series data from the database, identifying missing blocks, and preparing the data values for running the recovery. You can start the preprocessing by clicking the **“Preprocess”** button (in the section marked as 4 on fig. 4). If the button is disabled, make sure you took all the steps above successfully.

As the preprocessing finishes, the **“Recover”** button turns enabled. Click it, and let CDRec do the recovery steps. Once it finished, the visualization of reference and test time series will appear in a new window, where you can check the recovery.

During every processing, the pane will stop responding to prevent unwanted actions. The **log box** (section 5 on fig. 4) will inform you about all the actions.

## Decomposition function

This function offers the ability to use the core algorithm (“Centroid Decomposition”) on arbitrary input data, i.e. matrix.

The user must specify an input file that contains the elements of the matrix in a specified format described below, and in a text box on the UI pane (see section 1 on fig. 5).

Click **“Open a file...”** button to browse for the input matrix (inside section 1 on fig. 5). Choose the input, so that the browse window closes.

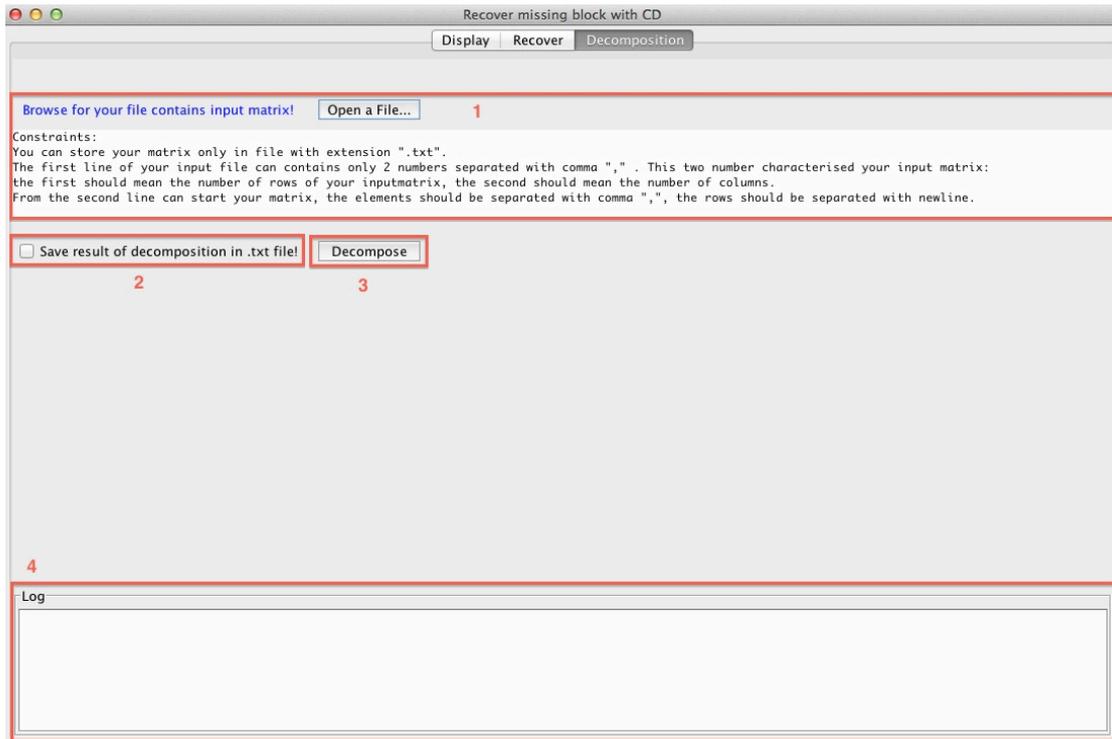
In the checkbox marked as 2 on fig. 5, you can choose the option of saving the output of the decomposition into a .txt file.

Once the input file has been specified, you can click the button **“Decompose”** (marked as 3 on fig. 5), which triggers the start of the process. You can find the output matrix (if chose to export) in the working directory.

You can inform about the process steps in the log box shown as 4 on fig. 5.

### Input file format

The input file should be a text file in which the first row must contain an integer followed by a comma, then another integer, (and a new line character). These numbers should specify the row- and column count of the matrix respectively. Then the file should contain the elements of the matrix row by row. I.e., column count times a float value, separated by commas, and a new line character, for every row of the matrix. (The float value can be any string which the Java method `Float.parseFloat(String)` accepts.)



## Troubleshooting

Errors when trying to launch CDRec:

- If the first line of the error message says “java.sql.SQLException: Io exception: The Network Adapter could not establish the connection”, the database could not be reached. Check the network connections, and make sure the server is online, and the hardcoded user credentials still work.