

GECCO 2015 *Industrial Challenge:* *Recovering missing information in heating system* *operating data*

Martina Friese, Christopher Schlitt, Thomas Bartz-Beielstein

¹

April 30, 2015

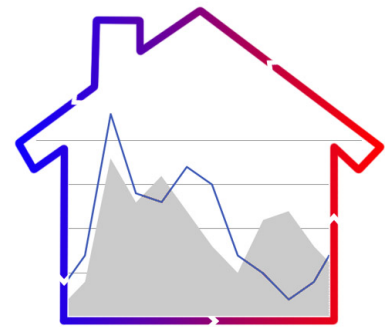
Goal of the GECCO 2015 Industrial Challenge is to develop capable procedures to recover missing information in heating system operating data. Adequate recovery of the missing data enables more accurate forecasting, which allow for intelligent control of the heating systems, and therefore contributes to a positive energy balance and reduced expenses. Thus, to be suitable for its designated use, methods must be accurate and computationally efficient. This document provides a set of rules and regulations for the GECCO 2015 Industrial Challenge, a detailed problem description, as well as contact and submission information.

¹ Cologne University of Applied Sciences, 51643 Gummersbach, Germany
martina.friese@fh-koeln.de,
christopher.schlitt@smail.fh-koeln.de,
thomas.bartz-beielstein@fh-koeln.de



1 Introduction

In times of accelerating climate change and rising energy costs, increasing energy efficiency and reducing expenses becomes a high priority goal for businesses and private households alike. An intelligent heating system, that meets the individual requirements, would contribute to the completion of this goal. Modern heating systems record detailed operating data and report this data to a central system. Here, the operating data can be correlated and analyzed to detect potential optimization opportunities and to make forecasts of future consumption behaviour. Due to various issues, like transmission or system failures, the collected data may be incomplete. This increases the difficulty of forecasting, which already is high.



THE GOAL of the GECCO 2015 Industrial Challenge is to develop capable procedures to recover missing information in heating system operating data. Adequate recovery of the missing data enables more accurate forecastings which allow for an intelligent control of the heating systems, and therefore contributes to a positive energy balance and reduced expenses.

HIGHLIGHTS of the GECCO 2015 Industrial Challenge include:

- *Interesting Problem Domain:* Recovering missing information in heating system operating data offers a challenging yet manageable test case for modern time series methods.
- *Real-world Data:* Multiple real minutely heating system operating time series are provided for training, testing, and assessing data recovery methods.
- *Fair Submission Assessment:* Prediction accuracy is determined on

test data available to the organizers, which will be made public after the competition ends.

- *Direct Link to Industry:* Bartz & Bartz GmbH ² will evaluate the winning submissions for real-world application and will be in direct contact with the winning participants, who will keep all rights to their developed methods.

The remainder of this document specifies the information needed to take part in this competition. It is organized in three parts: Section 2 introduces the problem of data recovery for heating system operating data, as well as time series data sets provided. Section 3 presents the set of rules and regulations. Finally, Section 4 gives information on how to participate in the industrial challenge.

2 Problem Description

The objective of this competition is to recover missing data in real-world heating system data. The data provided for this competition consist of various types of operating data time series. Participants of the challenge should recover any missing values in the provided data set. Hence, each submission should be the completed, equidistant time series with 1 minute intervals.

2.1 Heating System

Heating Systems are composed of a boiler which heats the water in the system, a pump to circulate the water and radiators which are wall-mounted panels through which the heated water passes in order to release heat into rooms. The circulating water systems use a closed loop, the same water is heated, pumped through the heating circuit and then reheated again. Today modern heating systems log the full details of data to enable further interpretation of the data.

2.2 Training- and Test-Datasets

The data for the GECCO 2015 Industrial Challenge contains four different time series denoting the main aspects of the heating system behaviour. Given is the heating water temperature when leaving the boiler as well as the return temperature. Additionally, the setpoint for the heating water is supplied as well as the actual required power to heat up the water (see Table 1).

The heating system that provided the data only logged the data when a value changed. Additionally the data has been revised to supply periodic, minutely data.

A first impression of these time series is given by figure 1.

For this challenge, data was removed from the original data in order to simulate the difficulties that analysis and forecasting methods usually have to cope with. To obtain a realistic problem definition the size and frequency of data gaps in other data files has been

² Bartz & Bartz GmbH gives advice to companies based on mathematical and statistical methods to enhance products and production processes and to conserve resources. Although the mathematical techniques are almost universally applicable, the Bartz & Bartz GmbH specializes in technical applications. This includes the optimization of injection molding processes in the automotive industry, manufacturing processes for sensor manufacturers and the prediction of failure probabilities of components in heating systems. Founded in 2014, Bartz & Bartz maintains contact with the world's leading research groups in the field of simulation and optimization.

Column name	Description
Timestamp	Time of measurement, given in following format: yyyy-mm-dd HH:MM:SS
Supply Temperature Setpoint	The target temperature for the supplied heating water, given in °C.
System Supply Temperature	The temperature that the water actually has, when leaving the boiler, given in °C.
Return Temperature	The temperature the water has when returning into the boiler, given in °C.
System Power	The amount of power, given in %, required by the appliance to reheat the water to the requested temperature.

Table 1: Description of the given time series data

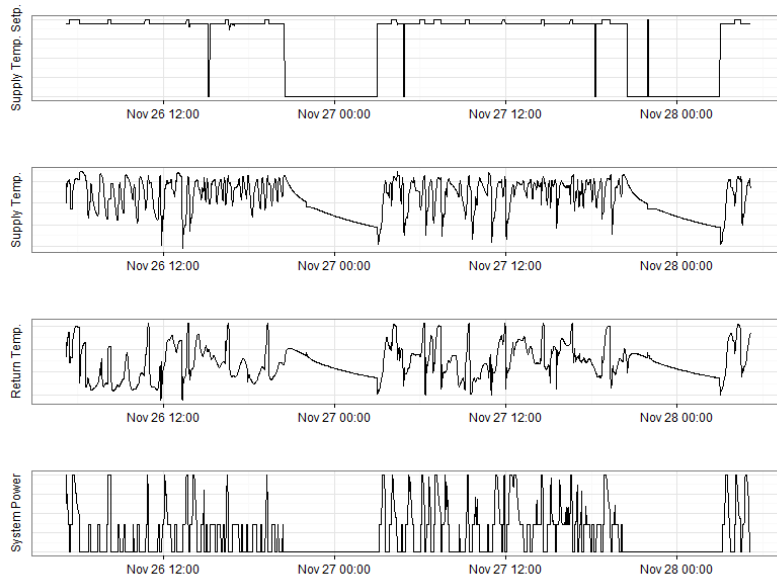


Figure 1: The plot shows an extract of about two days of the given time series data. The first row displays the target temperature for the water when leaving the boiler, while the second rows shows the temperature the water actually has when leaving the boiler. The third row displays the temperature of the water when returning into the boiler and the last row shows the amount of energy, spent that moment, to reheat the water.

analyzed and fitted by an exponential distribution function. Using this exponential distribution function, data has been removed from the provided time series. All time series contain data gaps which have to be refilled as accurately as possible. Missing values that have to be recovered are denoted with 'NA'.

The data can be roughly divided into 7 intervals, one leading and one trailing interval with complete data, and five intervals showing gaps. See Table 2 for more detailed information.

To PARTICIPATE in the competition, all missing data has to be recovered and a complete data set without missing data has to be supplied.

Data removed in	Interval
No data removed	2013-11-18 05:12:00 – 2014-01-24 11:02:00
Supply temperature setpoint	2014-01-24 11:03:00 – 2014-03-24 11:43:00
System supply temperature	2014-03-24 11:44:00 – 2014-05-22 13:24:00
Return temperature	2014-05-22 13:25:00 – 2014-07-20 14:05:00
System power	2014-07-20 14:06:00 – 2014-09-17 14:46:00
All 4 time series	2014-09-17 14:47:00 – 2014-11-15 14:27:00
No data removed	2014-11-15 14:28:00 – 2015-01-13 15:08:00

Table 2: The data can be roughly divided into 7 intervals, one leading and one trailing interval with complete data, and five intervals showing gaps.

2.3 Prediction Quality Rating

In this competition, the quality of the recovered time series is defined as the *root mean square error* (RMSE) between the recovered time series \hat{t} and the respective true time series t . Equation 1 defines the RMSE.

$$\text{RMSE}(\hat{t}, t) := \sqrt{\frac{\sum_{i=1}^n (\hat{t}_i - t_i)^2}{n}} \quad (1)$$

SUBMISSIONS are ranked by RMSE. Since four time series have to be recovered by the participants, the total RMSE used for the ranking of the submissions is defined as the arithmetic mean of the RMSE of the four provided time series. The true time series are known to the organizers and will be published after the competition ends.

3 Rules and Regulations

In order to participate in the competition, refilled time series have to be supplied. The time series delivered have to feature the same data format as the given time series. Submissions are accompanied by a two page report describing the algorithm that has been applied to recover the missing data. On request, algorithms have to be supplied as executable code, we expect the runtime of these to be reasonable. The organizers will rank the submissions by the total RMSE that is calculated as defined in Section 2.3. The winner of the GECCO 2015 Industrial Challenge will be the participant with lowest total RMSE.

Finalists selected by the jury will be invited to present their submission at the competition session, held during the GECCO conference. The winner of the competition will be announced at the SIGEVO meeting ceremony, on July 15, 2015. Therefore, each participant has to be registered at GECCO 2015.

4 *Submission*

Submissions to the GECCO 2015 Industrial Challenge should consist of:

- The recovered time series featuring the same format as the original data set, described in Section 2.2
- and a short report on the methods used, also featuring your institution and contact data (two pages maximum).

PLEASE send your submission as archive file (i.e. *.zip) via email to gecco@f10.fh-koeln.de. You can also contact the organizers via email (gecco@f10.fh-koeln.de) if you have any questions.

4.1 *Software and Data*

Example data and source code will be available for download at <http://www.spotseven.de/gecco-challenge/gecco-challenge-2015>.

4.2 *Important Dates*

- *Data Availability:* April 30, 2015
- *Industrial Challenge Submission Deadline:* June 22, 2015
- *GECCO 2015 Conference:* July 11-15, 2015

4.3 *Organizing Committee*

- Martina Friese, Cologne University of Applied Sciences
- Christopher Schlitt, Cologne University of Applied Sciences
- Thomas Bartz-Beielstein, Cologne University of Applied Sciences