Intelligent Health Monitor (IHM)
Intelligent Soft Sensor (ISS)
Advanced Process Optimizer (APO)
Advanced Process Controller (APC)

Input Data Specifications

Version January 2018

1. Introduction

The software products of algorithmica construct a machine learning model of the dynamics of an industrial plant from its historical data. If the underlying physics and chemistry of the industrial process remains the same over the long-term, then the modeling process will be able to deliver an accurate, precise and reliable model for the measurement concerned.

Two data files must be provided: A list of tags and a data file. The text below specifies the format and contents for each of these files in detail.

2. List of Tags

Each time-series in the data has certain descriptive characteristics that are provided in this file. This file will be a text file containing 15 TAB-delimited columns and it will be saved in the UTF8 encoding.

A sample file (both as a text and as a Microsoft Excel file) is available from algorithmica.

Here follows a detailed explanation of each column

Column Name	**Requirement**	**Description**
Tag	required	The unique identifier for a time-series. This is often an alphanumeric string of characters used in the DCS or historian to label a tag.
OPC Tag	required	Often the same as the column "tag," this is the unique identifier used by the DCS. IHM uses this column in order to query the current value of the tag from the OPC data source.
Sensor name	recommended	A short description of what this tag is.
Description	optional	A longer description of what this tag is.
Units	required	The physical units of the tag, e.g. '°C'
Minimum	required	The smallest value allowed. Any values lower than this will be ignored as not physically possible.
Maximum	required	The largest value allowed. Any values larger than this will be ignored as not physically possible.
Low green	optional	Within the range of allowed values [minimum, maximum] we may define three sets of fixed alarm levels from the inside out: green, yellow and orange.
High green	optional	
Low yellow	optional	
High yellow	optional	
Low orange	optional	
High orange	optional	
Delta	required	The measurement uncertainty of this tag in the same units as the value itself. Please note that the full uncertainty of a measurement is far larger than the uncertainty of its sensor. In fact, the full

		uncertainty is usually dominated by the compression factor set by the data historian[1].
Limit	required	This column is either 'TRUE' or 'FALSE' depending on whether this tag is to receive a dynamic limit or not.
Control	required	The tag is to be put into one of three classes. A tag is **controllable** if an operator can proactively and directly make changes to its value, i.e. it is a set-point. A tag is **uncontrollable** if the operator has no influence of its value whatsoever, e.g. the weather. All other tags are **semi-controllable** in that the operator has some influence over them but cannot change them directly.
Type	required	The tag is obtained either automatically by reading it from an OPC data source or by manual input. This column is either OPC or Manual depending on this choice. A tag should almost certainly be read by OPC and only in special circumstances should we incorporate manual inputs.

Financial prices are treated on the same footing as other tags in APO. They will also be values in the data table and they should be semi-controllable. The operator cannot actually influence the price of course but the prices are not boundary conditions for the physical operation of the plant.

Indeed, every tag that appears in the goal function should be semi-controllable. If it were controllable, the optimization would be trivial, i.e. set it to its maximum. If it were uncontrollable, the optimization would be impossible since no change can be effected. Thus, we should only use controllable or uncontrollable tags in the goal function if it is necessary to get the right value out but it should principally consist of semi-controllable tags.

The meaning of control and delta are interrelated. If a tag is uncontrollable, its value is a boundary condition. For example, if the external temperature is e.g. -20°C we cannot suggest changing to a condition that was present during summertime at +30°C. The delta decides the comparability of conditions. If Delta is 5°C, then all conditions in the range from -15°C to -25°C are comparable to a condition with -20°C. If Delta is lowered, the number of comparable points decreases and vice-versa.

The number of comparable points is important because it represents the basis for the mathematical model of this constellation of uncontrollable tags. Please note, that all

[1] A data historian usually only records a value in its database if it differs from the last recorded value by more than some fixed amount, the compression factor.

uncontrollable tags must be comparable in a similar way as they are all boundary conditions.

If we increase the number of uncontrollable tags and/or lower their Deltas, then we reduce the number of comparable points in the past and the quality of model suffers as a result alongside the potential to optimize.

If an improvement is calculated, a suggestion will be output for those controllable tags that need to be changed by an amount greater than its Delta. If the change of a controllable tag is smaller than its Delta, then it is not part of the suggestion as the current and optimal values are considered comparable.

The Deltas for controllable and uncontrollable tags must be selected with care as this has consequences for model behavior.

Columns that are not required, may be left blank.

This file may have a header row with the names of the columns in it.

3. Data File

The data file is an ASCII text file with its first column being a timestamp and then N more columns of floating point numbers where N is the number of tags in the tag file. This file is semicolon-delimited. This file may also have a header row with the names of the columns in it.

The timestamp may be in one of two formats:

- The standard international time format: dd.mm.yyyy hh:mm:ss.xxx
- The ISO 8601 format: yyyy-mm-dd hh:MM:ss.xxx

In both cases, the use of milliseconds '.xxx' is optional.

Please note that the time stamp format popular in the USA (mm.dd.yyyy) is not supported as this leads to format confusions with the international format.

It is essential that the order of the columns in the data file is the same as the order of the tags in the list of tags.

4. OSI PI Users

Users of the very popular data historian PI by the company OSI Soft, have another option for supplying their data.

The command-line tool 'piconfig' may be used to extract the data of any tag over a historic period. Usually one extracts the data of a single tag into a single file meaning that the data of N tags would yield N files. In this case, the data file is simply a list of the individual file names containing the individual data for a single tag.

The individual files are ASCII text files looking either like

```
*> 121F6050.OP, 17-Feb-13 00:00:00, 17-Feb-13 23:59:59, *
61.84849,GOOD,17-Feb-13 00:00:05.68701
61.76492,GOOD,17-Feb-13 00:00:20.65601
62.00223,GOOD,17-Feb-13 00:00:23.67201
61.62254,GOOD,17-Feb-13 00:00:24.67201
```

or like

```
;121F6050.OP;
04.05.2013 12:32:40;-0,1645781;
04.05.2013 13:02:20;0,1094844;
04.05.2013 13:02:40;No Data;
04.05.2013 13:03:00;0,215625;
04.05.2013 14:59:00;0,309375;
04.05.2013 14:59:20;0,05243749;
```

depending on the settings chosen for 'piconfig'.

In both cases, the string '121F6050.OP' is the name of the tag that we expect to find in the first column of the tag file.

Example Data File

Column Separator: Semi-colon

Time	CHA01CE011	LAE11CP010	LAE11CT010	LBA10CF001
01.01.2015 00:00:00	1.1	2.1	3.1	4.1
01.01.2015 01:00:00	1.2	2.2	3.2	4.2
01.01.2015 02:00:00	1.3	2.3	3.3	4.3
01.01.2015 03:00:00	1.4	2.4	3.4	4.4

Example List of Tags

Column separator: TAB

Tag	PLS Tag	Sensor Name	Description	Units	Minimum	Maximum	Low Green	High Green	Low Yellow	High Yellow	Low Orange	High Orange	Delta	Limited	Control	Type
CHA01CE011	CHA01CE011	Power		MW	-1	60				54			0.31	TRUE	controllable	OPC
LAE11CP010	LAE11CP010	Pressure		Barg	0	150				105			0.75	FALSE	uncontrollable	OPC
LAE11CT010	LAE11CT010	Temp-erature		°C	0	300				180			1.5	FALSE	semi-controllab	OPC
LBA10CF001	LBA10CF001	Flow		t/h	15	150				108			0.68	FALSE	semi-controllab	OPC