Manual: 4.5. Assess Model Quality

There are three steps to assessing the quality of a trained APO model:

1. Checking plausibility
2. Reading the model report
3. Fine-tuning the model

We will describe the first one here and reserve the next two sections for the report and fine-tuning.

The model plausibility looks at the historical data in relation to the specifications made for each tag and tries to assess if these specifications make sense. The first thing checked is the measurement range of each tag. From the historical data, it is computed how many points lie inside and outside of the interval from minimum to maximum. This interval is intended to mark any and all normal and reasonable values. Therefore, only a very small minority of historical observations should lie outside this interval. However, we frequently find that some tags have a great many points outside this interval.

The minimum and maximum values are important because any point outside this allowed range is excluded from modeling. Generally, these values need adjustment before enough valid points are available to train a sensible process model. In the plausibility form, you will see not only the percentage of historical points that lie outside this interval but also the smallest and largest values ever measured for this tag. This information should help to guide you in making the appropriate corrections.

The second check concerns the measurement uncertainties of the uncontrollable tags. The modeler clusters the available historical data on the basis of the measurement values and uncertainties of the uncontrollable tags. For each such cluster, a separate optimization model is built. This is done in order to avoid suggesting a operational condition that cannot be reached because one or more of the uncontrollable tags are different. As they are uncontrollable, they mark boundary conditions to the process. The more tags are marked as uncontrollable and the smaller their uncertainties are, the more restrictions are put on the model. A more restrictive model will be able to optimize less.

In this check, please examine the selection of uncontrollabel tags. Perhaps some of them can be removed from the model if they do not really represent a boundary condition. Please also have a look at the uncertainty of these tags. The plausibility form will tell you how many clusters are formed on the basis of this tag. If in doubt, please start with larger uncertainty values to obtain a reasonable model. This can always be lowered at a later time to make the model more restrictive.

The third plausiblity check concerns the validity of historical data. Only valid points are used for training. A point is valid if all controllable tags are within their minimum to maximum ranges and if all validity rules are satisfied. You may also have defined a number of custom conditions that must be met. A common such custom condition is to require the plant to achieve a steady-state for validity. This effectively excludes all transient states from learning. If we have very few valid points in the history, then there may be too little data to learn from. As a plant's operations should generally be normal, having few valid points indicates that some settings need correcting.

After these three checks have been passed, the machine learning should be able to train a reasonable model for your process. Please train the model and have a look at the model report next.