Manual: 6.2. Assessing Model Quality

Navigate to the show page for your soft sensor. You will be taken there automatically after you have pushed the model and apply button on the edit page or you can select the show link on the list of soft sensors accessible through the ISS menu.

On the top of the page, you can select a time period for the time series display. This will select data for the original tag and the soft sensor tag over this time period and display it in two ways. The top diagram will be a simple time-series plot and the second diagram will plot the measured value against the computed value. Ideally the two tags will have equal values. Realistically, they will differ at least by some inherent random variation that is a result of the measurement process. On the plot of both tags versus each other, you will find a straight black line. This is the ideal line you would get if both values were always perfectly equal. Based on this line, you can judge the degree to which they agree.

The third plot is the probability distribution of the differences between the modeled and measured values. Ideally, this is a bell-shaped curve with its center at zero deviation. The curve should be symmetrical about its center and be sharply peaked. Look at where the bell shaped curve drops off on either side. There is a point at which the sharp descent from the center turns into a gentle approach to the axis; this is the "elbow" of the distribution. If this elbow on either side occurs at a difference value comparable to the measurement uncertainty of the original tag, then all is well. Should this plot have more than one significant and pronounced peak in it, then the model has a systematic problem that is almost certainly the result of a sub-optimal selection of independent variables, i.e. the model is lacking an important source of information.

Further down the page, you can see some statistical quantities to make this assessment more precise. If you are not convinced of the model's quality, please edit either the training time periods, the exclusion rules, or, most importantly, the selection of independent variables. As a last resort, you could increase the number of free parameter by increasing the number of hidden neurons. Please note that it is possible for an independent variable to be a source of disinformation that could harm model accuracy. Consider making a soft sensor for the NOX emission of a power plant and using your heart rate as an independent variable. At best, this will have no effect on the model but it is more likely that the model will suffer because your heart rate adds structure to the data without any causation.

In the exceptional case that you have carefully checked the selection of independent variables as well as all other settings and still get a poor model, then you may conclude that there is a essential piece of information in your process that you are not currently measuring or have excluded from the model.