Appendix:

Important instrumentation boundary conditions for best prediction or optimization service

The quality of data is fundamentally important for mathematical modeling. We therefore describe the requirements in detail regarding the metadata of all measurement variables.

We need, for every measurement variable, the following five metadata:

1. The theoretical minimum: If a measurement lies below this value, the measurement must be in error. Such errors happen due to a variety of instrumentation problems. An example is a percentage below 0, a temperature below 0 degrees Kelvin, etc.
2. The theoretical maximum: Just as the minimum in reverse.
3. The measurement accuracy: What is the uncertainty in the measurement? For example, if we measure a temperature to be 100 +/- 5 degrees Fahrenheit, then the uncertainty is 5.
4. The unit: What physical unit is the value measured in?
5. Objective metadata
   a. For NEMO™, we also need the boundaries between the good (green), critical (yellow) and failure (red) regions of the predictive chart. These are typically within the above theoretical minima and maxima.
   b. For OMEN ™, all sensors must be classified into one of three categories. The *uncontrollable* sensors are those that cannot under any circumstances be controlled by the operators in the plant control room, e.g. weather, stock market prices, supplier quality, etc. The *controllable* sensors are those that can be controlled directly by the operators in the plant, e.g. temperatures, flow rates, etc. The *semi-controllable* sensors are those that can be controlled not directly but only via being correlated with a controllable sensor, e.g. vibrations on a pump are caused by a higher throughput setting that is directly controllable. The OMEN ™ model will only give suggestions regarding controllable sensors.

These metadata are necessary to establish a correct model that can then serve to give either predictions or suggestions for optimization. The reason is that in practice, we observe two problems. First, every dataset contains outliers that may not be used for modeling. Second, every measurement is uncertain and this uncertainty is increased via a (necessarily integrative) analysis.

The first problem is solved by knowing minimal and maximal values. This allows us to immediately identify and erase the most radical of outliers – the actually illegal values. If this is not done at the pre-processing stage, the model with probably represent the outliers and not the system rendering it useless.

The second problem is solved by knowing the measurement accuracy. When we know this, we can propagate the uncertainty through the model and arrive at an uncertainty of the result. This allows an interpretation as to the practical value of the result. For example, if a temperature of 50 degrees is measured with an accuracy of 5 degrees, this has a different meaning as if the accuracy were 30 degrees. The decision to do something on the basis of this information is much sounder in the first instance than in the second. Thus it is essential to know one's accuracy.

The numerical value of the uncertainty is a complex topic that we briefly introduce here. (Should you not know your accuracies, we are happy to help you determine them.) The accuracy originates from a variety of sources:

1. The accuracy of the instrumentation. This is very easy to determine as the manufacturer of the instrumentation equipment will usually specify it. We must however keep in mind that this accuracy is not usually a constant but almost always depends upon environmental properties such as temperature, humidity and such. These are often unknown and so we must set this uncertainty equal to some estimated upper limit.
2. The variability of the medium. If we wish to measure a temperature in a vat, we ask ourselves what "temperature" may mean. We know that the temperature depends upon the location in the vat and thus, that the temperature on the walls of the vat, where is the sensor is located, cannot be equal to the average temperature. As the difference between the "temperature" and the measured quantity is not always constant, but depends upon environmental conditions, this is not a systematic error but a true uncertainty. A systematic error occurs, when the measurement always differs from the true value by the constant (e.g. because of using an instrument inappropriately.)
3. Invisible changes in the system. If a sensor degrades over time, it will change its measurement properties. A dirty sensor, for example, becomes insulated and now measures with greater time delay and uncertainty. Via a slow change in the environment of the sensor, which is usually not measured or measurable, the measurement values are numerically modified even though the real values of the observed medium have not changed. This effect adds a time component to the uncertainty of a sensor.

The accuracy that we are after is an upper limit resulting from the sum of these three effects. For example: A temperature sensor has an accuracy of 0.1 degrees, is located in a distillation column with temperature differences of up to 2 degrees and drifts because of dirt accumulation in between cleaning intervals up to 1 degree. Thus the accuracy is 3.1 degrees.

Modeling cannot be more accurate than the accuracy of the measurement. Indeed, modeling will produce results that are, in general, more uncertain.

It is not necessary to determine the various influences on the accuracy with scientific rigor. Statements may be blanketed, such as "all temperatures are accurate to 4 degrees." The essential point is that uncertainties are not unrealistically small and thus raise hopes for modeling that cannot be fulfilled. Especially the instrumentation tolerance given by manufacturers of sensors is usually very small and often given as the accuracy – this would create unrealistic expectations.