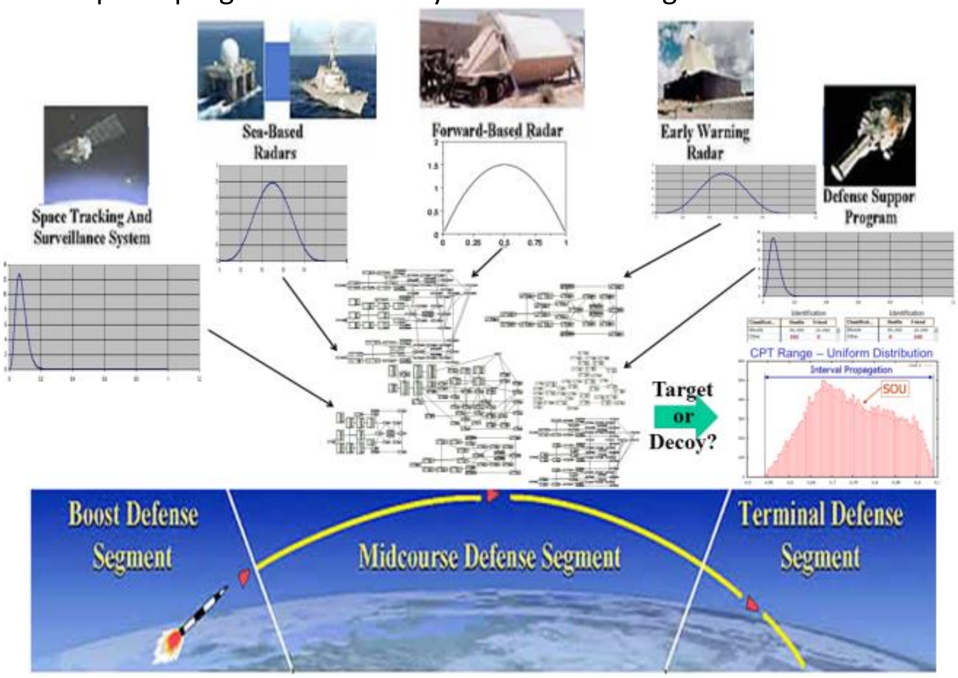
SOU: Second Order Uncertainty

SOU is a statistical acronym for "Second Order Uncertainty", which is the uncertainty of the uncertainty. SOU represents the precision of a statistical estimation. In simulation and modeling of future events, predictions such as weather forecasts use probabilistic models that are fed information from a variety of sources such as satellite data from a number of sensors. All the data is dynamic meaning it changes with time. The forecasting models first convert the sensor and other data into probability distributions, expressing the accuracy, that used by the prediction engines to estimate the likelihood of a future event, such as will it rain tomorrow.

The most common statistical prediction engine is the Bayesian Network (BN). A BN is typically defined in an esoteric mathematical statement such as "*A Bayesian network is a probabilistic directed acyclic graphical model (a type of statistical model) that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).*" In fact, a BN can be defined as a probabilistic expert system. Perhaps a more meaningful description however is that a BN is graphical programming language that allows the user to write executable computer code expressing a Bayesian inference using graphical objects in a flowchart format. When the BN flowchart is compiled, the result is executable computer code that can be dropped into a larger system computer program such as a system's control logic.



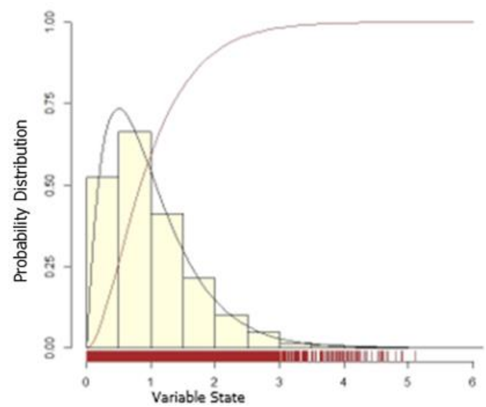
Standard BNs suffer from a lack of precision in two areas:

1. Uncertainty in the evidence provided to the network, and
2. Uncertainty in the processing engine which is embodied in the CPTs found at each network child node.

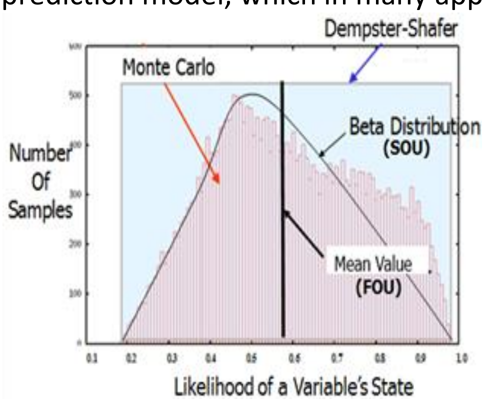
Expressed using the filter analogy these two sources of imprecision are the system's measurement and processing noise respectively.

The evidence provided to the BN from the outside can come from multiple sources, including surveys, measurements and heuristics such as rules-of-thumbs and subject matter expert (SME) opinion. If the evidence is a heuristic such as from an SME opinion, the SME seldom is comfortable in providing a precise value and would prefer to specify the variables state with a range, typically interpreted as a three-sigma extreme, as well as the expected mean.

The second source of uncertainty associated with processing is from the estimates found in the BN's conditional probability tables (CPTs) which encode Bayes inference. BNs have the inherent assumption that the conditional probabilities embedded in the network all have approximately similar uncertainty, which is seldom true. The uncertainty found in the CPT values is a result of the same issues described above for the evidences. The CPT cells are either populated by human experts or automatically from learning algorithms. In both cases, the individual cells in the CPT are presented as discrete values which represent mean estimates. Furthermore, if the source of the CPT information is a human expert it is even more convoluted by the indecision of SME. That is, it may prove difficult to get the knowledge out of the experts in a form that can be converted into probability distributions. This is compounded if the SME does not understand or have confidence in the developed BN.



Even if the input probability distributions are continuous, they quantized into a number of digital ranges called "states" to approximate the actual probability distribution. Each state is typically assigned a single mean value to represent the uncertainty value of that range. In many cases the variation across the state range can be significant which introduces uncertainty in the estimated state uncertainty, i.e., the precision. The more the states the smaller the range and the higher the precision. But lots of states results in a very slow prediction model, which in many applications is unacceptable and a fast prediction is needed.



The GCAS SOU software provides the best of both worlds for predictive analytics: namely, it provides both high precision and high speed using its higher order propagation engine and a graphical editor for constructing models including second order effects. The Missile Defense Agency sponsored the development of the software for their computer forecasting models used to recognize and distinguish objects that are flying in a cluster. The predicted results are comparable in accuracy to those obtained using Monte Carlo simulation that take 1,000's of times longer to execute.

There are many applications where SOU is needed in the commercial sector. Basically, any situation where a quick and accurate decision is needed in a time-constrained environment will benefit from SOU.