Surrogate models

Background

To be able to take advantage of Robustica you must create a spreadsheet model of the system that you are investigating. For business systems, this is often fairly easy; the spreadsheet environment is where this is typically done. It is also relatively easy for many engineering, scientific and other technical system; the spreadsheet environment is familiar to many professionals in these fields. However, there are times when the system of interest can only be modeled in dedicated software (CFD, FEA and other dedicated simulations packages) or when a physical prototype must be made of the system if the operation and performance are to be investigated. In these cases, it is not immediately obvious how you can use Robustica to robustify your system's design.

Robustica can indeed be used in these cases; however, an extra step is required. You must create an approximation of the system's operation that can be used in the spreadsheet environment. Such an approximation is at times called a surrogate model and this article will cover some of the techniques that you can utilize to develop your own surrogate models.

Why surrogate models

Ideally, you would have an analytical model that is easy to enter into a spreadsheet and analyze, for your system. Even though all models are wrong, analytical models prove to be very useful when they are available. They are typically accurate, easy to enter into a spreadsheet and they can be executed very quickly. This allows you to easily investigate your system through 'what if' analysis, and to optimize it. However, some systems do not lend themselves to the easy derivation of an analytical model.

Many engineers will encounter systems that are not easily molded analytically even if the basic principles are evident. Examples of situations that have been encountered by others include:

- The priming features of a centrifugal pump. It is known that recirculation of the water within the pump housing is required until the suction line is full. However, it is not immediately evident how features within the pump casing will effect the recirculation or, for that matter, how they will affect the other performance characteristics of the pump. In such cases, CFD will be required and it will often be the case that experimentation will be the only option if any reliable data is to be acquired.
- Plough geometry for the efficient tilling of earth. In this case, the engineer will have developed an intuitive feel for the effects that changes to the cutting geometries will have on performance, but they will not be able to predict the effects or variability to optimize for a robust design. The associated problems are further complicated by the complex behavior of earth as it is cut and moves around the plough. In such a case, experimentation is often the only way data can be gained.
- The designing of an elastomeric compliant system. Compliant systems can reduce the number of parts in a system and they can also remove sliding features, which reduce cost and wear respectively. However, the nature of the compliance under operation can rarely be predicted analytically and FEA simulations can be very time consuming. This in turn makes optimization (especially a robust optimization) so difficult that many compliant systems are kept relatively simple; duck bill valves for instance.

Because of their high resource and time demands, numerical models and experimental models do not allow for the same speed of analysis and optimization that an analytical model does. Therefore, they are unsuitable for 'what if' analysis and optimization. Surrogate models are intended to bridge the gap between the numerical or experimental, and the analytical. By being based on the results from experiments and numerical simulations, surrogate models hold the pertinent information that is required for accurate prediction. In addition, by taking on an algebraic form they can be as simple and as easy to analyze as an analytical model. Thus, we use surrogate models so that we can optimize systems that do not lend themselves to the development of an analytical model.

Approximation forms

To create a surrogate model you need data from either physical or numerical experiments and an equation form that you will fit to that data. Some of the various function forms, and the data they require, will be considered here.

Polynomial

The simplest, and perhaps the most obvious, is a polynomial. The form of a first order polynomial for 3 input variables x_1, x_2 and x_3 would be as follows.

$$f(x_1, x_2, x_3) = a_0 + a_1 x_1 + a_2 x_2 + a_{1,2} x_1 x_2 + a_{1,3} x_1 x_3 + a_{2,3} x_2 x_3 + a_{1,2,3} x_1 x_2 x_3$$

Note that the unknowns are the coefficients a_i, and we have implicitly assumed that the phenomena of interest can be approximated by this first order system. Such assumptions would suggest that we might have difficulties gaining an accurate surrogate model. This is at times correct, especially for systems and phenomena that display non-linear behavior.

If the polynomial is unable to provide an accurate model, we might only use it for a local approximation. We would then generate a new model for each local region that we encounter as we search for the optimum combination of input variables. Such is the approach in Response Surface Methodology (RSM).

Another alternative is to increase the order of approximation and consider geometric and cubic functions. While this will probably increase the accuracy, the number of unknown coefficients, and the required data, increases phenomenally. This makes the consideration of high order approximations only viable for cases with few input variables.

If a polynomial model is chosen, the most efficient way to collect the data is to perform a proper design of experiments (DOE). A DOE provides a good coverage of the region being considered and makes it easier to solve for the coefficients. Any typical text on DOE will provide you with the required information to set up a proper DOE.

Dimensional analysis

The major draw back with the polynomial approximation is the workload and dimensional analysis can help with this. If we perform a dimensional analysis prior to performing any experiments, we are able to effectively reduce the number of input variables. By reducing the number of input variables, the process of developing a polynomial surrogate model takes much less time. This article will only provide a summary of dimensional analysis and its use in surrogate models. If you would like more information, take a look at *Applied dimensional analysis and modeling* by Szirtes and Rozsa and *The use of dimensional analysis to augment design of experiments for optimization and robustification* by Lacey and Steele in the Journal of Engineering Design, Volume 17 No.1.

All phenomena must obey the laws of homogeneity. By utilizing dimensional analysis, we can identify parts of the model that are essential to satisfy this requirement. After performing dimensional analysis, we are presented with a number of dimensionless groups. These groups would exist (even though they might be factored out) in the analytical model if it were possible to derive. Therefore, after performing a dimensional analysis we have essentially solved part of the model.

We can treat each of the dimensionless groups as variables in their own right when generating the polynomial. Because there will almost always be less dimensionless groups than input variables, we can perform fewer experiments to solve for a polynomial of the original order that we considered. Alternatively, we can perform the same number of experiments as was originally intended, but create a polynomial that is of higher order, and accuracy. To make the most of dimensional analysis you should also consider the reciprocal of each dimensionless group. At times, the reciprocal will better fit the polynomial approximation, and this can lead to significant gains in accuracy.

Alternative function forms

The polynomial is easy to use and understand, and it makes a good start point for the function form; however, it is unlikely that the phenomena being considered would actually be represented by a polynomial. Therefore, it is worth considering other functional forms. This of course raises the question: which forms should be considered?

Alternatives were hinted at when we considered dimensional analysis and it was suggested that the reciprocal of the dimensionless groups also be considered. One of the advantages with dimensional analysis is that it allows us to make the most of our ability to reason what alternative forms we should consider. Some examples:

- Suppose we were considering the flow through a cloth filter element that we were developing. Also, suppose that some of the input variables were weave tightness (force), yarn diameter (length), element frontal area ($length^2$) and element thickness (length). We would expect that if the element were twice as thick, it would be twice as difficult for any fluid to flow through at a given flow rate. Thus, we would expect a linear relationship between the thickness of the filter element and the pressure drop. The thickness would be part of a dimensionless group, and we would now know that the dimensionless group must be given an exponent that ensures this linear relationship between pressure and thickness.

- Consider the situation where we wish to optimize the backwashing procedure for a sand filter. The purpose of backwashing is to recover the flow coefficient of the filter (that is the ease with which water will flow through the filter), and we would like to maximize the recovery while minimizing the down (backwashing) time. The variables that we expect to affect the recovery are: the period of backwashing, the flow rate of the backwashing fluid, the mass of dispersants that have accumulated in the sand and the mass of sand in the filter. We would expect that as the period of backwashing is increased, the amount of recovery would also increase. Further, we know that the recovery cannot be greater than that which would correspond to a complete cleaning;

we can be certain that the recovery has a limit. Therefore, the function that relates the period of backwashing to recovery would have an asymptote that corresponds to a clean filter. This function would also be applied to the dimensionless number that includes the period of backwashing. A potential function that displays this behavior is the exponential decay formula below.

$$\mathrm{R}(t^*) = A_1\left(1 - e^{-A_2 t^*}\right)$$

Where t^* is the dimensionless number that includes the period of backwashing t and each A_i is a constant that must be solved for experimentally. This would not be the entire model, but if some preliminary investigation suggested that it was a good fit, it would be included in the final surrogate model. There are of course other functions that asymptote, and we could also consider those.

For cases like the second one, it would be a higher order polynomial (requiring a lot of experimental effort) that would be needed to approximate asymptotic behavior. However, a polynomial would easily approximate the linear function in the first example, but we can now say that a higher order polynomial for the thickness is probably unnecessary. Thus, once we realize what the proper form is, we can use fewer experiments to solve for the parameters in the function. This then means that we need not perform numerous experiments for higher order polynomial approximations. Therefore, we can see from the above examples the benefit in considering the likely form a function should have before simply applying a polynomial approximation.

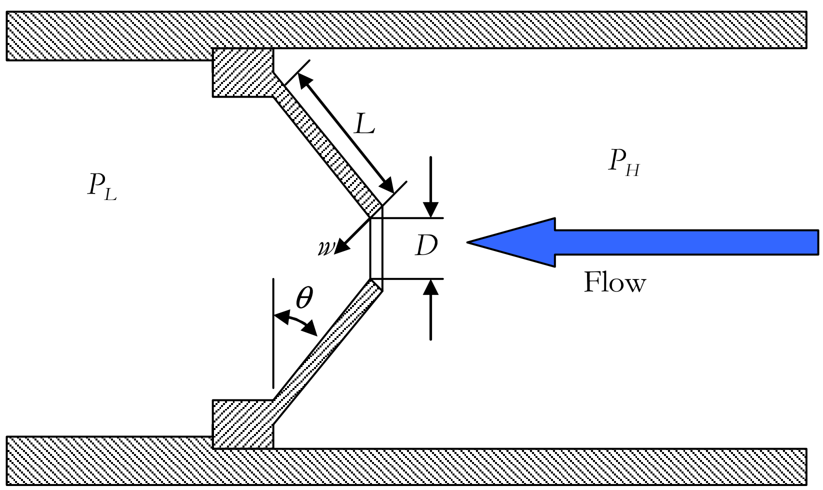
This benefit is even greater when we also utilize dimensional analysis. If we can determine the form of the relationship between the output and one input variable, then by default we have also found the relationship between the output and the other input variables within the same dimensionless group as that first input variable we considered. Thus, even fewer experiments would be required.

Generalized modeling

Perhaps one of the easiest ways to find a function form is through generalized modeling. This technique was introduced to create an adaptive controller for a machine tool subject to the effects of thermally induced distortion, see Fraser et. al.

(1998), *Modelling, identification and control of thermal deformation of machine tool structures, part 1: Concept of generalized modelling* Journal of manufacturing science and engineering Vol. 120, no3, pp. 623-631. The first step in generalized modeling is to simplify the system being considered and then develop an analytical model for the simplified system. This provides the form of the function and the exponents for each variable, making the model dimensionally correct. The coefficients within the model, which should be dimensionless, are then treated as unknowns and solved for by fitting the simplified model to data gained from experiments or numerical simulations performed on the original system.

For example, consider the development of a flow control valve, which is meant to ensure that the flow rate of a fluid remains constant despite changes in pressure. One type of flow control valve is based on a deflecting diaphragm like that shown below.



As the flow rate through the valve increases, the pressure differential across the diaphragm increases. The diaphragm then deflects under the uniform load, and the orifice diameter D reduces. This reduction increases the orifice coefficient, and the flow reduces back to its original value.

This is an idealized description of the operation and some deviation in flow rate can be expected. We would of course wish to minimize this deviation, and robustify the system so that regardless of fluctuations in the pressure (and the various characteristics of the valve) the flow remains on target. This would be much easier if we had an accurate surrogate model that we could use.

If we were to use the generalized modeling approach, we could treat the valve as a circular plate with a hole, and use an analytical model from a mechanics of materials text to model the deflection w. We could then model the reduction in D by using the w, trigonometry, and the variables L and θ. Finally, we could use an analytical equation from a fluids text to predict the flow through an orifice of a given diameter for a given pressure difference. These equations could then be combined to form a single analytical model for the change in flow versus the change in pressure difference as a function of the valve geometry and material. This analytical model would of course be incorrect and unsuitable for predictions or optimization. However, we would then treat all of the coefficients within the model as unknown and use empirical results to determine their values. This would provide an updated model that accurately predicts the performance of the valve. With this updated surrogate model, we would then perform our optimizations, including robustification, in an analytical manner.

Testing and improving model accuracy

One of the difficulties with empirical models is the determination of their accuracy, and if enough empirical data has been collected. Therefore, methods to test the accuracy of a model are of interest to us. One of the most obvious options is to perform another experiment, with each of the input variables set to values different to any tested, and compare the result with the prediction made by the surrogate model. If the result and the prediction are similar, then it would be assumed that the model is sufficiently accurate. Otherwise, the experimental results are used to improve the model's accuracy, and another verification experiment is conducted. This process can be repeated until sufficient accuracy is attained.

The above process will eventually lead to an accurate model. However, performing another experiment each time might be time consuming and the best combination of input variables for testing purposes might not be ideal for the good design of experiments. If this is the case, we might wish to compromise and choose a combination of input variable values that are more suited to the design of experiments than model evaluation. However, there is another alternative that can provide a rigorous test without more experimentation. We do this by first neglecting one of the experimental points and then reevaluating the parameters (or coefficients) of the model. We then use this lower order model to predict the output for the neglected point and compare the prediction with the experimental results. This can be done for each experimental point in turn. If each experimental point can be accurately predicted with the respective lower order surrogate model, then we can say that the surrogate model is sufficiently accurate.

Another method that can be used to evaluate the model is the trial of extreme values for each of the input variables. By analyzing the predictions, we can determine if the model makes sense, and if the behavior is indicative of an accurate model. Similar information can be found by plotting the model output against various input variables. This information can also be used to determine what extra experiments, if any, still need to be performed to improve the model; it might be found that there are certain regions of high non-linearity that require a high concentration of experimental points.

Finally, you should always perform one last experiment, after optimization, with the input variables vales set to those values found during optimization. This is to ensure that the final optimized design is indeed likely to provide the performance you predicted.

Conclusion

The above has been intended to introduce you to various techniques that you can use to develop a surrogate model. The reason for this is to allow you to utilize both Robustica and simulation software or experiments to robustify the various systems that you need to optimize. While this is only an introduction, it should now enable you to think more about how you might incorporate probabilistic techniques into your professional activities. Therefore, you should now be able to see the advantages that Robustica can offer you.