EMPIRICAL MODELING OF COMPLEX SYSTEMS
IN ADVANCED LIFE SUPPORT SYSTEMS

David A. Vaccari¹, James Cavazzoni², David Fleisher² and Julie Levri¹
¹Dept. of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ; ²Dept. of Bioresource Engineering, Rutgers University, New Brunswick, NJ.

INTRODUCTION

A method is described for the empirical modeling of complex processes, called multivariate polynomial regression (MPR). By “complex systems” is meant multiple-input systems that do not have satisfactory mechanistic models to describe their behavior, and which cannot be modeled using linear methods such as multilinear regression. Behaviors that contribute to this modeling difficulty include nonlinearity and interactions at moderate to high dimensionality.

One of the few commonly available techniques for this common type of problem is the artificial neural network (ANN). However, there are a number of difficulties with ANNs. They produce models that are difficult to communicate and to analyze. They require a large number of weights in their fitting, which makes them susceptible to overfitting. There is no a priori way to determine the best model structure. Due to the large number of nodes and weights, they are capable of producing unpredictable behaviors.

The MPR technique provides the capability to describe complex nonlinear input-output relationships similar to artificial neural networks. It has advantages over neural networks including: The chance of overfitting is reduced because the resulting models tend to be much simpler. It produces models that can be written down as a polynomial equation. This makes them more tractable and easier to communicate to others. It is much simpler to optimize, perform sensitivity analysis, predict confidence intervals, and to incorporate in other models.

We will describe the MPR technique and show previous and current applications in advanced life support for long-term space travel. We propose that it may have wide-ranging potential for application to other NASA-related studies.

EMPIRICAL MODELING WITH MULTIVARIATE POLYNOMIALS

MPR models are essentially multiple regression models with added terms for nonlinearity and linear and nonlinear interactions, and for previous values of any variable if modeling a time-series. MPR models are similar to Volterra equations, and are a special case of NARMAX models [Chen & Billings, 1989]. They have similar capabilities to ANNs in ability to model complex relationships such as process transfer functions and logical relationships [Vaccari and Christodoulatos, (1992)].

MPR models use polynomials to describe the response surface of a dependent variable to any number of independent variables. For example, if Y is a function of both Q and R, an MPR model might be:

\[ Y = a_0 + a_1 Q + a_2 R + a_3 Q \cdot R + a_4 Q \cdot R^2 + a_5 Q^2 R^{-1} \]

Although a large number of possible terms of this sort could be defined, the number of candidate terms is restricted by limiting a list of possible exponents and by the maximum number of multiplicands within each term. The exponents may include –1, which enables the examination of ratios of independent variables. Non-integer values may also be used. Furthermore, models are kept simple by using a stepwise algorithm to select terms from among the candidates. Only those terms that contribute to the predictive power of the model, as measured by some criterion such as mean square error, are kept in the model. The stepwise algorithm is an iterative procedure wherein terms are added and removed until the criterion cannot be improved upon by the addition of a single term to the model or the removal of a single term from the model. This, along with other techniques such as cross-correlation, helps ensure that the model captures and generalizes behavior contained in the data without unnecessary model complexity.

PAST AND CURRENT APPLICATIONS

1. **Plant motion**: MPR was used to produce nonlinear polynomial time-series models predicting total plant projected canopy (TPCA) area versus time and temperature.

2. **Plant biodegradation**: Models were developed to relate rate and extent of nutrient recovery from inedible plant material by biological treatment. Treatment parameters included temperature (T) and a coded variable for the use of heat-pretreatment (H). One of the models developed predicted the percent total solids reduction (PTSR): PTSR = 
+80.517 - 8.8715 H - 310.89 T -1H -1. R² = 0.5532. The Figure below shows this model along with the data used to generate it. Specifically, the two curves show curvature, and the curvature depends upon a parameter, H. Linear models cannot describe this behavior.

3. **Incorporating MPR into the Energy Cascade model.** The original Energy Cascade model calculates daily crop growth rates using the following trends: a linear increase in canopy light absorption from emergence through canopy closure (occurring at time tA); a constant (maximum) light absorption after tA; a constant canopy quantum yield (CQYMAX) through the onset of senescence (occurring at time tQ), then decreasing linearly thereafter until crop maturity (occurring at time tM); a constant carbon use efficiency. The parameters tA, tQ and tM are model inputs. As tA depends on plant growth as well as development, more detailed crop models were used to produce crop specific equations for tA as dependent on light integral and [CO₂] for a given temperature regimes (using MPR). A multi-layer canopy photosynthesis model was applied to simulate the dependencies of CQYMAX on PPF and [CO₂], as such could not be determined from the available data for individual crops. Multivariable polynomial regressions (MPR) of the simulated dependencies were then incorporated into the modified Energy Cascade models.

4. **Prediction of Photosynthetic Efficiency in Crops:** The other project that is in beginning stages is a parametric study of factors affecting photosynthetic efficiency (PE). PE of a plant depends upon many external factors including relative humidity, temperature, CO₂ partial pressure, and light intensity. However, the nature of this dependency cannot be described by fundamental relationships. This is exactly the situation that calls for empirical modeling, and for which MPR modeling is likely to be most suitable.

5. **MPR use for Crop modeling:** MPR is used to simplify and increase portability of explanatory crop models for wheat, soybean, and white potato. The dependent variable was relative growth rate; the independent variables were cumulative average air temperature, light intensity, CO₂, and dry weight at previous time increment. The resulting MPR equations will be incorporated into a software program to simulate effects of off-nominal conditions on growth and development and will also be used to develop a model-based controller to compensate for these effects.

**INDEX TERMS**
Empirical models; nonlinear models; time-series models; multivariate polynomial regression; advanced life support.

**REFERENCES**