Chapter 1
Introduction to Integrated Predictive Modeling

Teresa R. S. Brandão and Cristina L.M. Silva

1.1 Introduction

When a mathematical model is properly developed, it is a potential tool for process design, assessment, and optimization. Using a mathematical expression that predicts a real observation with accuracy and precision is an efficient way to develop new products and to control systems. However, to attain a convenient model, a lot of well-guided experimental effort should be expended and the model should be validated. One should never forget that the model predicts one response in the range of experimental conditions tested and care should be taken when extrapolating to other operating conditions.

This chapter provides an introductory approach to concepts and methods involved in mathematical modeling, with particular focus on modeling quality and safety of food products.

1.2 Basic Concepts and Methods

Modeling is the use of mathematical language to describe the behavior of a system. A model is a mathematical expression that relates dependent variables(s) to independent variables(s). This relationship involves constants (i.e., parameters) that depend on intrinsic and/or extrinsic factors. In general, an observed response can be written as

\[ y_i = f(x_i, \theta_k) + \varepsilon_i, \quad (1.1) \]

where \( i = 1, 2, \ldots, n \) is the number of experimental observations/runs, \( j = 1, 2, \ldots, v \) is the number of independent variables, and \( k = 1, 2, \ldots, p \) is the number of model parameters.
parameters. $y_i$ represents a measured response at the $i^{th}$ experimental observation (e.g., dependent variable), $x_{ij}$ are fixed $j$ independent variables that define the experimental conditions at observation $i$, $\theta_j$ are unknown parameters, and $f$ is the mathematical form of the model considered. $\varepsilon_i$ represents an independent experimental error (from a normal distributed error population with mean equal to zero and constant variance).

The mathematical expressions used in process modeling can be derived on the basis of fundamental reasoning or empirical description. The first ones are called mechanistic models, since they are based on knowledge of the fundamental mechanisms involved in a process. On the other hand, empirical modeling is a black-box approach, meaning that no concern is given to the theory underlying the phenomena, and the objective is merely a convenient prediction of the observation.

Particular types of models that can be considered neither mechanistic nor empirical are ones based on probability distributions, also referred to as stochastic models. In those models, one characteristic of the system is assumed to follow a preestablished probability density function (normal, Weibull, and logistic are examples of the most used lifetime distributions).

The mathematical models chosen should predict the response variable(s) accurately and the choice depends on the adequacy of the model to describe the process, and also on the quality of the parameters. Even the most complex model cannot yield good predictions if the model parameters are not estimated with accuracy and precision.

### 1.3 Experimental Design

In mathematical modeling, the very first step is the observation of a real occurrence and experimental data gathering. The information contained in data is established at the moment they are collected and, therefore, only an assertive selection of an experimental data pattern can provide good results. This is the objective of the experimental design.

The experimental designs resulting from the application of intuitive thinking to data collection are often called heuristic designs (i.e., based on “common sense”). Heuristic designs with sampling times that are equidistant (i.e., independent variable), are certainly the preferred ones. Even so, one should be provident in the time-extension of the whole experiment. This aspect is particularly important in kinetic studies where the results can be limited if one of the extreme situations occurs: experimental points cover only a small extension of the process, or there is a deficient amount of experimental data in early phases. In processes of great variation in the early stages before stabilization, data can be collected at equidistant points on the logarithmic time scale (Boulanger and Escobar, 1994). This obviously implies that more samples are collected at short times than at longer ones. If the experimenter has a preliminary idea about the process and the model to be used, the option can be a design with sampling times chosen so that the expected change in the response between two consecutive times is constant.