Risk of Extrapolation in predictive microbiology models

Marcelo O. Masana

Instituto Tecnología de Alimentos, ITA, CA- CNIA, INTA, CC 77, 1708, Morón, Buenos Aires, ARGENTINA.

Background

Response surface and Belehrádek type models have been extensively employed in predictive microbiology to describe the effect of several environmental factors on growth parameters or on the probability of growth. Their widespread use is due to their ability to summarise the experimental results of the effect of any number of factors on the growth parameters with an acceptable degree of accuracy. In particular response surface models lead to multivariate polynomial models which can be treated by standard linear regression techniques with a straightforward fitting procedure, that can be performed with user-friendly statistical software, producing models with well characterised statistical properties. In spite of these advantages there have been also concerns about their use because of their lack of mechanistic background, their sensitivity to particular sets of data and over-parameterization difficulties (Baranyi et al., 1996).

The risk of extrapolation

The empirical nature of these models also implies that predictions must be made strictly for a combination of variables within their interpolation region defined as a Minimum Convex Polyhedron (MCP) by Baranyi et al. (1996). This MCP confines all the combinations of the environmental factors that provided the raw data for the model. The unaware misuse of these models outside their interpolation region may lead to unrealistic predictions that may hinder their confident use.

Therefore, to quantify what would be the risk of making an extrapolation when using predictive microbiology models created from a given database the Percentage Chance of Extrapolation (PCE) indicator has been introduced (Masana and Baranyi, 2000) as:

$$PCE = p_e * 100 = \left(1 - \frac{Volume(MCP)}{Volume(NVS)}\right) * 100$$

where (p_e) is the probability of extrapolation, and NVS is the Nominal Variable Space defined as the whole space confined by the minimum and maximum values of the environmental factors, where a prediction could be requested. The PCE indicator quantifies the risk that a prediction requested at random in the Nominal Variable Space could be actually outside the interpolation region or MCP.

The ratio $\left(\frac{Volume(MCP)}{Volume(NVS)}\right)$ is equal to the probability that the prediction is inside the MCP and can be approximated by the fraction of those points of a sufficiently refined grid of the Nominal Variable Space which are inside the Minimum Convex Polyhedron in comparison to all the points of the grid. It is important to realise that the determination of MCP and PCE does not depend on the particular equation describing the relationship between variables and response, but only on the combinations of factors which are used for the model creation.

The analysis of different databases used for the creation of predictive software showed that a substantial increase in the risk of extrapolation, measured as the Percentage Chance of Extrapolation, results when new factors are added i.e., during model extension. The increase in the probability of extrapolation would result from the fact that some of the new combinations including the added factor will more probably result in no growth. The risk of extrapolation could increase even more, if for practical reasons, only some combinations of the same Nominal Variable Space for the new factor are explored.

Consequently, for more complex models, the risk of inadvertent misuse of the predictive modelling software increases severely, and its consequences should be considered.

References

Baranyi, J., Ross, T., McMeekin. and Roberts, T. A. (1996). Effects of parameterization on the performance of empirical models used in 'predictive microbiology'. *Food Microbiol.* **13**, 93-91.

Masana, M. O. and Baranyi, J. (2000). Extending predictive microbiology models, implications on the risk of extrapolation. *Food Microbiol, in press.*