Optimizing Extraction from Seeds with INForm

The squash Momordica Cochinchinensus is a perennial plant indigenous to tropical regions. Its' dry roots are used as an expectorant and anti-inflammatory in Chinese medicine. The seeds are also used for treating various conditions by topical use. Saponins have been identified as the active components.

The extraction of the biologically active constituents is usually affected by factors like plant materials, solvents, and equipment and processes used.

In this note, we report on the use of neural networks for modelling the cause-and-effect relationships between process and formulation conditions, and the amount of materials obtained from the extraction. Genetic algorithms have been used for the optimization; both of these technologies are used in the **INForm** formulation optimization system.

Extraction from Seeds

The present study uses data from Professor Dang Van Giap of the School of Medicine in Ho Chi Minh City, Vietnam. Squash seeds were collected from ripened fruits and they were baked either in a heating device, or over live charcoals. Dang and his colleagues carried out 16 experiments, with 4 input variables:

- Heating method (heating device or live charcoal)
- Ethanol concentration
- Extraction temperature (30^o or 80^o)
- Ratio of seeds to ethanol

16 different experiments were carried out varying these inputs, and the total dried extracts, the fat-soluble residue, and the saponins were measured.

Modelling the Process

Dang and his coworkers used an early version of **INForm** to develop neural network models, and their work has been repeated here with **INForm** v3.7. 2 of the data records were withheld for testing, and the remaining 14 used to develop a model.

The high quality of their data means that good models could be developed even with such a small number of experiments, with R^2 values of 0.95, 0.99 and 0.98 for the training data sets respectively. Good values for R^2 for the test data sets were also found – these were in excess of 0.98 for the total residue and the fat-soluble residue, and 0.75 for the saponins. These ANOVA values were found for a network with 2 nodes in the hidden layer, trained using the RPROP backpropagation algorithm. Separate neural network models were developed for each of the properties

Figure 1 shows the actual and predicted values for the fat-soluble residue, for the training data set.

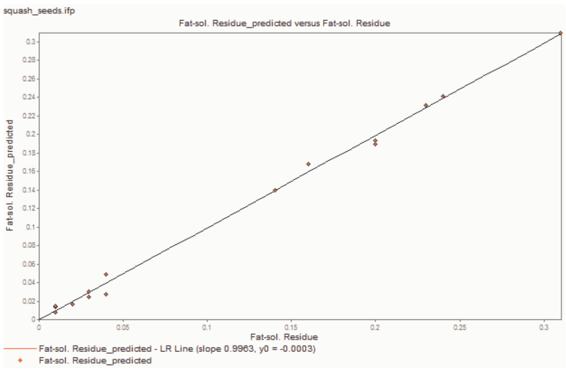

Figure 1 Actual vs. predicted Fat-soluble Residue values

Figure 2 shows similar results for the saponins.

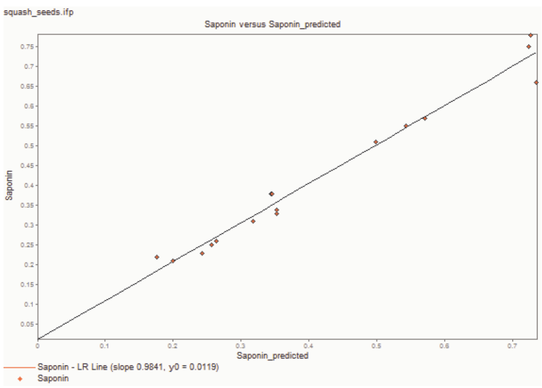

Figure 2. Actual vs. predicted Saponin values

These indicate that the models should be sufficiently reliable to allow predictions and to permit optimizations to be made.

Response surfaces can be generated, showing how the inputs affect the properties. These showed that the fat-soluble residue was largely unaffected when the heating device was used instead of heating over live charcoals. The same was true for the saponins, and for the total dried extracts.

Intelligensys Ltd, Springboard Business Centre, Stokesley Business Park, Stokesley, N. Yorks TS9 5JZ, UK
Tel:(+44) 1642 715380 Fax:(+44) 1642 715301 E-mail: sales@intelligensys.co.uk

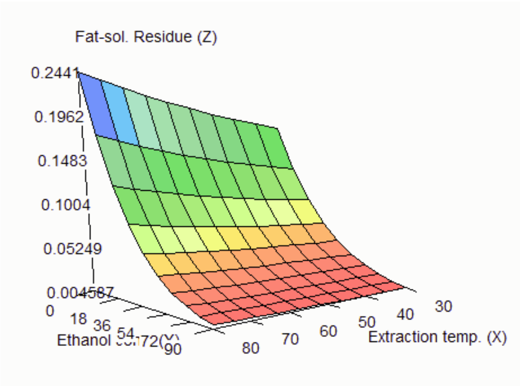

Figure 3. Dependence of fat-soluble residue on extraction temperature and ethanol concentration. Seed/ethanol ratio is 0.14 in this plot, heating method is over charcoal.

The amount of saponins was greatly affected by the extraction temperature, as Figure 4 shows. In the experimental results, the extraction temperature took one of two values, 30° or 80°. Therefore, the intermediate values in the plot are interpolated from the model. At low temperatures, the seed/ethanol ratio had virtually no effect on the amount of saponins, while at high temperature there was some effect with the amount of saponins decreasing as the seed/ethanol ratio increased.

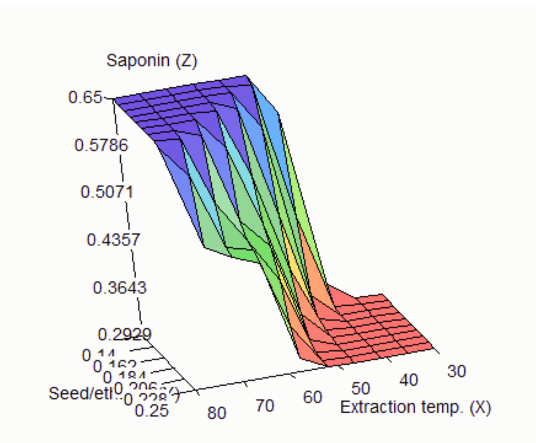

Figure 4. Saponins as a function of seed/ethanol ratio and extraction temperature

Optimizing the Process

Dang *et al* used **INForm** to optimize their process, setting the conditions as follows:

Total dried extracts – as high as possible
Fat-soluble residue – as low as possible
Saponins – as high as possible

These were specified using **INForm's** fuzzy logic capability, which is used in determining the fitness function for the genetic algorithms.

They did not specify the relative importance of the properties, and in the present study we assumed that the amount of saponins was the most important (with a weighting of 10, on a 1 to 10 scale) with the other properties weighted at 8. Figure 5 shows the Optimizer Configuration screenshot, with the desirability function for the saponins superposed upon it.

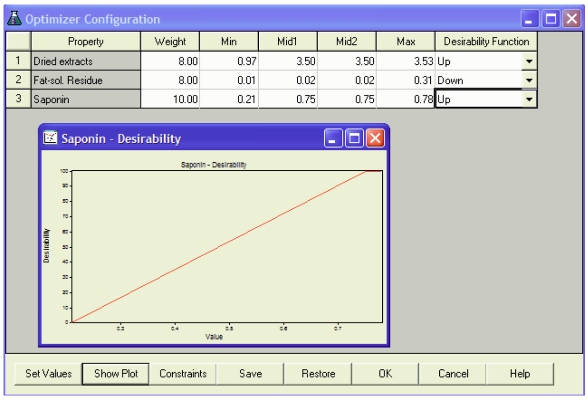

Figure 5. Optimizer configuration specifying desired values and relative importance (weights)

There was one constraint on the system, that the ethanol concentration was in excess of 60%.

With these requirements, the following optimum was found:

Dried extracts	3.18
Fat-soluble residue	0.01
Saponins	0.72

under the following conditions:

Heating method:	over live charcoals
Ethanol concentration:	90%
Extraction temp.:	80°
Seed/ethanol ratio:	0.20

Ethanol concentration and extraction temperature both lay at the top of the ranges used in collecting the experimental data. A sensitivity analysis showed that ethanol concentration was not crucial in achieving these objectives.

A further option was investigated, by fixing the heating method to be the heating device. The same conditions were found, indicating that the exact nature of the heating is not important.

Conclusions

Good models (assessed by ANOVA statistics) could be developed from a relatively small number of experiments, and these were used successfully in optimization.

Intelligensys Ltd, Springboard Business Centre, Stokesley Business Park, Stokesley, N. Yorks TS9 5JZ, UK
Tel:(+44) 1642 715380 Fax:(+44) 1642 715301 E-mail: sales@intelligensys.co.uk