Authors: Mark-John Bruwer and John MacGregor, December 2007

Introduction

This brief communication is focused on the data requirements for rapid product development projects.

Background

ProSensus' technology for the rapid development of new products (or reformulation of existing products), in its most general form, operates simultaneously in the three "domains" (or "degrees of freedom") associated with product development. These are:
a) selecting the raw materials,
b) determining the blend ratios for these raw materials,
c) determining the process operating conditions (batch or continuous)

Therefore, data are needed that cover (in an informative manner) all three domains. The researcher may expect that one or more of the domains are unimportant to the particular product development problem. However, it is best to still acquire suitable data for all three domains and then determine from the data what effect variations within each domain have on the key final product properties.

We will now address the data requirements for each of these domains.

Data requirements for raw materials domain

In this domain we need to thoroughly characterize each potential raw material type. Examples of different measurement types include:

   1. Basic chemical characteristics: melting point/ boiling point, glass transition temperature, molecular weight, chain length, viscosity, heat capacity, hydrophilicity, pH, latent heat, Gibbs

free energy,   *etc*.

    2. More complex measurements: FTIR, NIR, NMR, GC and/or mass spectrometry spectra

    3. Other relevant measurements to characterize physical and chemical characteristics such as particle size distribution,   *etc*.

    4. NOTE: different raw materials may have a different set of available measurements. The measurements will usually not be common to all raw materials. In this case, we still keep all measurements available for each raw material even if the union of these variables is not common to all raw materials. For a raw material that is missing a particular measurement, we set that measurement to a status of "missing" (a blank element in spreadsheet). These missing measurements are easily handled by the latent variable modeling methods.

Data requirements for blend ratios (mass fractions) domain

The data required to optimize the blend ratios typically come from blending designed experiments where a representative selection of raw materials have been blended in differing mass ratios so as to span the space of possible raw materials combinations in varying relative amounts.

Data requirements for process operating conditions domain

The key here is that, given a particular raw materials selection and certain blend ratios, we need to optimize the manufacturing procedure, much like a chef tries to perfect a recipe. Given the high level of automation in modern manufacturing facilities, we do not adjust the key process variables directly. Instead, we adjust the setpoints for these key variables (continuous processes) or their set point trajectories (batch processes).

An implicit assumption (that is outside the scope of the product development problem) is that the control systems (e.g., "low-level" PID control loops, or "supervisory-level" MPC controllers) are well-designed and well-tuned, so that the process value closely tracks its set point. To this end, for continuous processes, we also do not want process data during transient conditions (e.g., while the control system is responding to large process deviations due to significant disturbances or grade transitions). Rather, for continuous processes we want representative "steady state" data. This may mean collecting data over a significant time period and averaging it to filter out the (less major) transients.

For batch processes we need all the information pertaining to the "recipe", for example:

1. Total mass charged,
2. order in which the raw materials are charged (if relevant),
3. key set point trajectories (e.g., temperature),
4. key batch maturity variable(s) (e.g., percent conversion) vs. time,
5. key event tags (e.g., agitator turned on/ off, initiator charged, *etc*)
6. any other information that may have an affect on the product produced.

Continuous processes are more of a "steady state" problem and we need set point values to be applied to the key process manipulated values or their average values over a steady state period and the corresponding mass ratios and Y variable averages during those periods.

An important point in terms of process operating conditions: In addition to the technology for product development (across the three domains described above), ProSensus has complimentary technology that enables process operating data across multiple manufacturing operations to be used together in the modeling exercise. For example, process operational data may be available at bench scale, pilot scale and commercial scale. Furthermore, the process measurements may be different in each case, or the process itself may be significantly different (e.g., batch versus continuous, or two different process designs for a continuous process, *etc*.). Our technology can handle all of these scenarios and will extract the common, underlying, "latent structure" that enables the relevant information from all these different data sources to be combined. The advantage of using data from multiple processes is that we get a richer dataset for modeling. For example, significant designed experiments may have been run on the pilot plant during product development, but experimentation on the commercial plant may be much more restricted.

Figure 1 shows the data structure described above. The various data matrices are:

X_{DB} Matrix with raw material properties for all possible

$X_i T$ The raw material properties

Z_i Process operating conditions used on the ith manuf

R_i Blend mass fractions for the subset of raw materials

Y ᵢ The final product properties for each product (or gra

$$X_1^T$$

$$X_{DB}$$

$$Z_1 \quad R_1 \quad Y_1$$
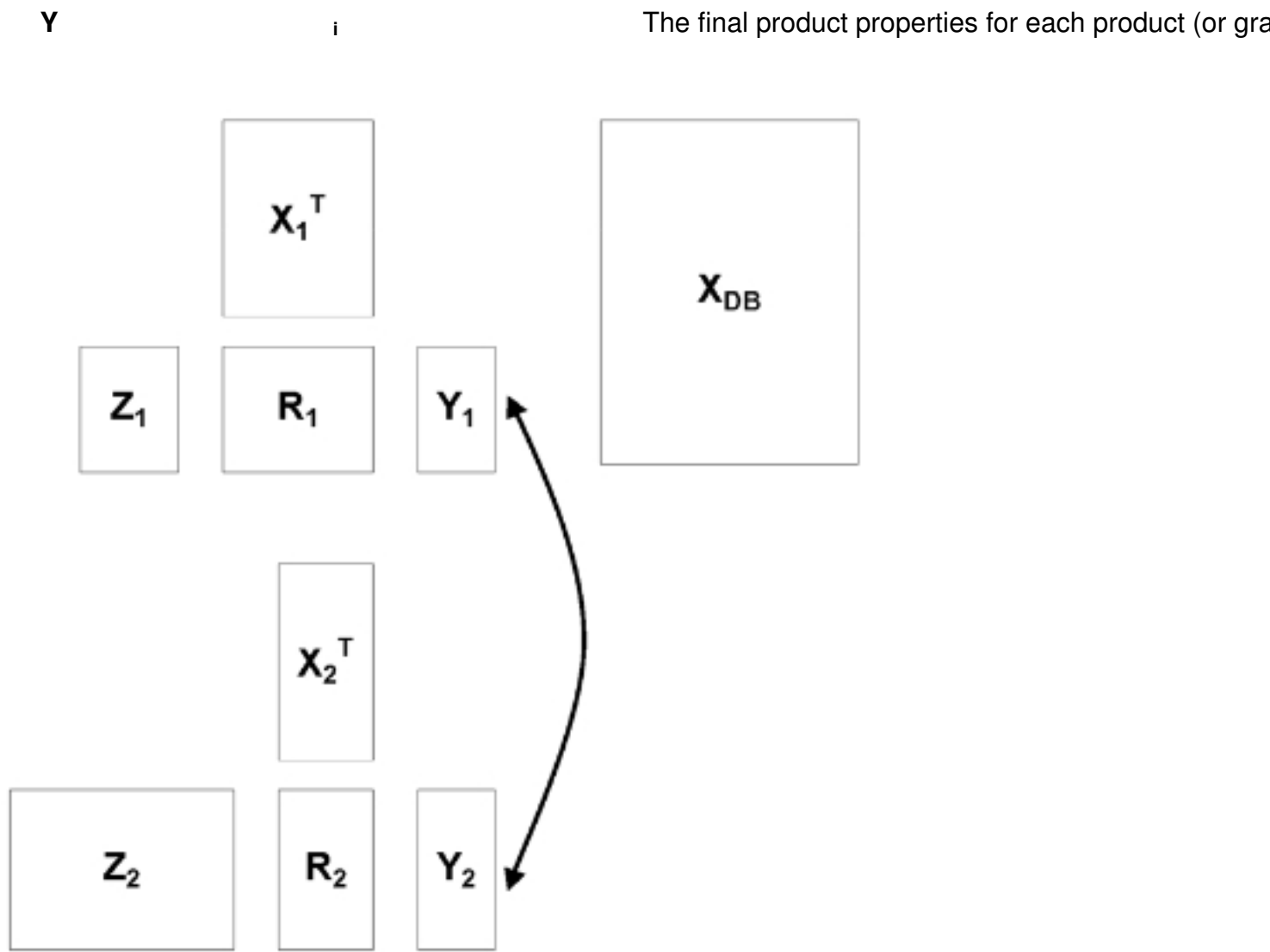
$$X_2^T$$

$$Z_2 \quad R_2 \quad Y_2$$

Figure 1: Data structure for rapid product development; multiple plants

Another important factor is that the process may be run differently under different scenarios, such as summer versus winter. This issue is only important if it affects the set points of the manipulated variables (i.e., the "recipe"). For example, say we have a reactor with a cooling jacket. One of the "recipe variables" might be the temperature of the material in the reactor. In winter a lower flow rate of cooling water to the jacket is needed since the water is much colder. In this scenario the "recipe" would not actually change between summer and winter because the set point for the temperature in the reactor doesn't change. Instead the flow control valve position for the cooling water flow changes to maintain set point, but this is hidden from the

"recipe layer" of the control system. So, in this example the "recipe" wouldn't change between summer and winter. In other examples, however, the recipe itself (i.e., the set points or set point trajectories of the key manipulated variables) may be changed in response to major, known shifts in environmental conditions, raw material properties (e.g., new lot of raw materials with quite different measured properties). In this case, we would treat each such scenario as a unique recipe to be included in the dataset.

If actual data for each batch or for each steady state period of a continuous process is not readily available, then one could use the computer recipe set-points (for process conditions, mass ratios) and product property targets for each grade produced by the company. This is not as good as actual data because it assumes that these are actually being followed, and it cannot use the natural batch to batch variations in the process variables, mass ratios and product quality to provide additional information. In this case one only has as many rows as one has nominal products.

Design of experiments

Traditionally, the design of experiments is done separately in each domain (raw materials selection, blend ratios and process operating conditions). If such experiments have been performed in the past then these would be used as the initial dataset to launch a product development project.

However, there is often a lot of synergy/ interaction between the three domains. Therefore, ProSensus has also developed technology for performing additional designed experiments simultaneously across all three domains. It has been shown that this can enable much greater efficiency in terms of the number of experiments needed to achieve a sufficiently accurate model for a new product to be designed.

Final product quality measurements

The final key information needed in the dataset is a set of measurements on the final product that capture the characteristics relevant to determining its "fitness" for the end-use of the product. The important issue here is that the underlying characteristics that determine fitness should be observable from the set of measurements.

Each entry in the database (e.g., each batch for a batch process, or each "steady state period" for a continuous process) should have an associated set of final product property measurements. The product sample(s) tested should be representative of all product produced during this period (batch, or "steady state period").