Chemometrics and Intelligent Laboratory Systems 157 (2016) 96–103

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Setting simultaneous specifications on multiple raw materials to ensure product quality and minimize risk

John F. MacGregor [a],*, Zheng Liu [a], Mark-John Bruwer [a], Boris Polsky [b], Glenn Visscher [b]

[a] ProSensus Inc., Ancaster, ON, Canada
[b] Mondelez International, Whippany, NJ, USA

A R T I C L E I N F O

Article history:
Received 6 April 2016
Received in revised form 22 June 2016
Accepted 27 June 2016
Available online 28 June 2016

Keywords:
Specifications
Design spaces
Product quality
PLS
Monte Carlo simulation
Food products
Pharmaceuticals
Chemicals

A B S T R A C T

Establishing meaningful multivariate specification regions on the multiple properties of a single raw material has been presented and illustrated by Duchesne & MacGregor Duchesne and MacGregor (2004) . However, the manufacture of most final products usually involves the use of many raw materials each with multiple measured properties and from different suppliers. Setting specifications separately on each of these materials is unreasonable since it is the simultaneous combination of the properties of all the materials that will affect final quality. This paper presents an approach to determining the acceptability of new lots of raw materials from multiple suppliers and of assessing the suitability of combining specific lots of materials currently in inventory that will minimize the risk of manufacturing a poor quality product. Multivariate statistical models based on PLS are used to determine the importance of all the properties of each of the materials and to develop the specification methodology. Use of the models for achieving improved control over the product quality is also discussed.

1. Introduction

Establishing specifications on incoming raw material properties that will assure a high probability of manufacturing an acceptable final product is an important problem. DeSmet [3] and Duchesne & MacGregor [4] emphasized that quality of a material is a multivariate property in which the simultaneous combination of all the material's measured properties is what is important rather than each of its properties individually. They proposed procedures for defining multivariate specification regions on all the measured properties of a single material. This involved building multivariate latent variable models using PLS (Partial Least Squares or Projection to Latent Structures), Geladi [5], Kourti and MacGregor [6] relating the final quality of the manufactured product to all the measured properties of the incoming raw material. The dimension of the latent variable space of these models is usually much lower than the number of raw material properties due to correlation among the raw material properties and their correlation with the product quality. Therefore the specification regions were established in this reduced dimension latent variable space. Raw material lots meeting specifications would have scores that fall within the specification region in the latent variable space and also have residual squared prediction errors (SPE) that are acceptably small (e.g. within 3 SD limits).

To illustrate this methodology, we present an application to setting specifications on an API (Active Pharmaceutical Ingredient) used in the manufacture of a drug product. Seven properties were measured on each lot of API and since this API was used in more than one drug product manufactured at different locations, the company wanted to set specifications on the API to determine which drugs could satisfactorily use this API. In the pharmaceutical industry this is often referred to as a design space on the raw materials.

Fig. 1 is a plot of a preliminary specification region based on a PLS model built on a small initial number of API lots for one of the drug products. The specification region is defined in the space of the two statistically significant latent variables (summary 1 and 2 in Fig.1). For each lot of API two latent variables scores are computed from the measured seven properties and plotted as shown in Fig. 1. The scores for the API's that led to good product and poor product are shown and a crude specification region (rectangle) defined for the API that included the scores for the good and excluded the scores for the poor API's.

The scores for a new lot of API (+ in the plot) fall outside the specification region and hence this API lot should be rejected for use in this product. Note that one of the seven measured properties was missing. However, the missing data handling capabilities of PLS still allow for

* Corresponding author.
E-mail address: john.macgregor@prosensus.ca (J.F. MacGregor).

http://dx.doi.org/10.1016/j.chemolab.2016.06.021
0169-7439/

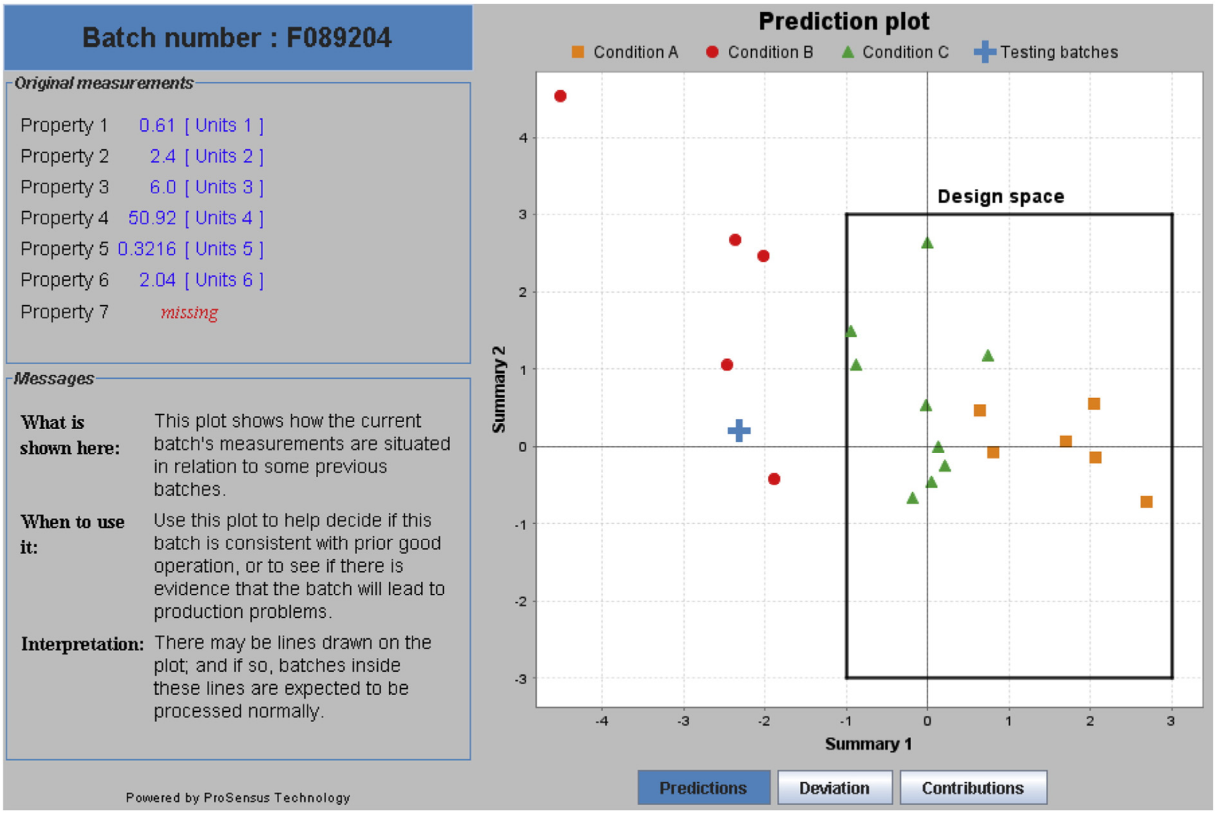

Fig. 1. Specification region for an API in latent variable space. API's yielding good quality are denoted by Δ and □ (those with □ were just more difficult to process) and those yielding poor quality are denoted by filled O. The scores for a new API lot are denoted by a (+).

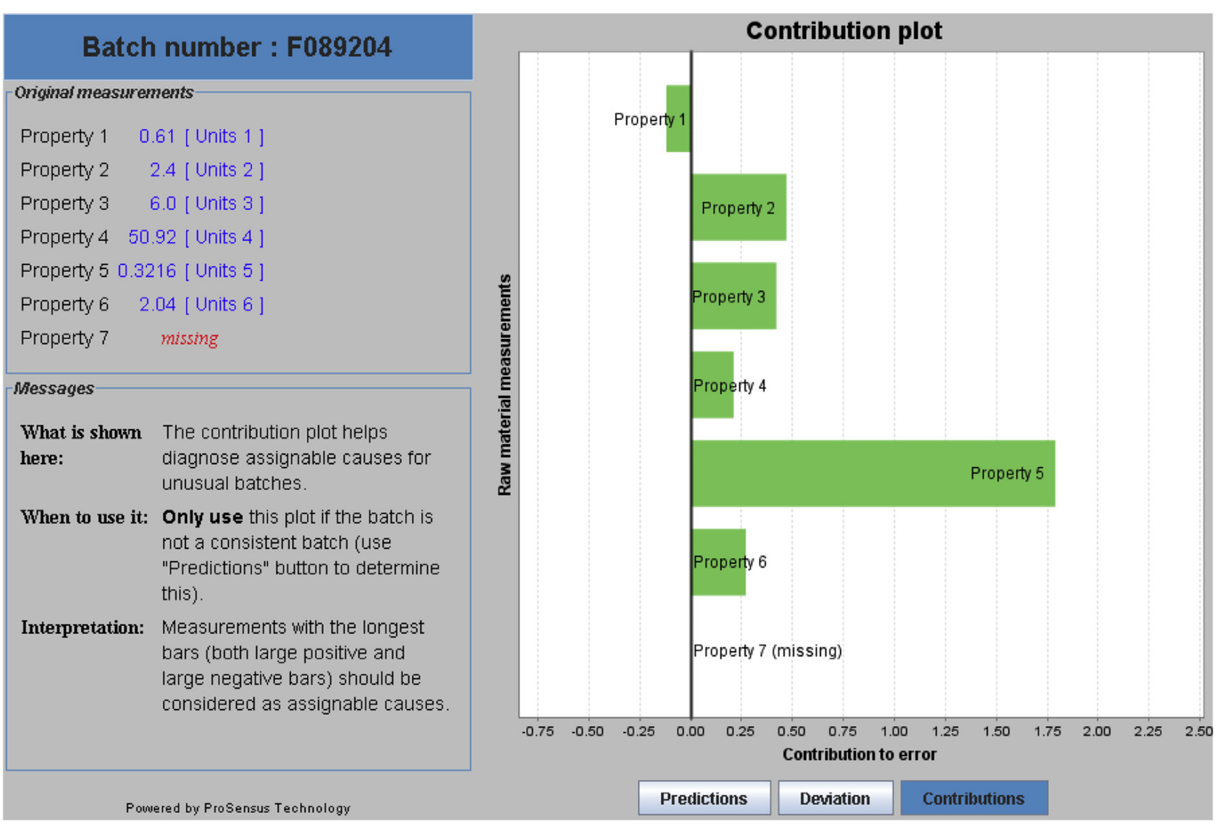

Fig. 2. Contribution plot for the out of specification batch (+) in Fig. 1.