The BIG DATA phenomenology of (wheat) grain hardness measurements and utilization.

R. S. Anderssen[(1)] and R. Haraszi[(2)],

(1) CSIRO Data61; ANU Mathematical Science Institute; La Trobe Dpartment of Mathematics and Statistics; (2) Campden BRI, Chipping Campden, UK.

The recovery of information about the hardness of cereal grains, to be utilized both scientifically and technologically, is an excellent example of the multifaceted challenges which arise in dealing with biological Big Data investigations.

The three basic Big Data stages that are involved are:

(i) the growing and harvesting of the wheat and the collection and cataloging of the seeds,

(ii) the measurement of the individual crush response profiles (iCRPs) of the collected seeds, the averaging of a suitably large number of iCRPs to obtain the average crush response profiles (aCRPs) and the deconvolution of the aCRPs to obtain their rheological phenotype phases (RPPs) for the various wheat varieties under investigation [1], and

(iii) the development of an electronic database for all the iCRPs recorded, for all the various situations examined, so that they can be subsequently utilized to track how wheat responds to changing environmental conditions such as global warming.

The situation (i) represents an example of the type of big data problem that arises in plant breeding, where, for example, in a wheat breeding experiment, a large system of plots are planted using an appropriate experimental design [3] which involves considerable resources in terms of the associated infrastructure and the management of the multitude of activities involved.

Here, the main focus is (ii), where the extent is examined to which the information contained in the RPPs can be used to monitor the effect of environment locality and environmental changes on wheat. The recovery of the RPPs from the iCRPs can be viewed as a bagging-and-boosting process where weak images are averaged to give strong images.

The situation in (iii) highlights that modern technology allows information to be utilized for subsequent investigation and partially reduces the need to keep large collections of seeds, as their rheological properties, though reasonable stable if stored at -20^{o}C, do change with time. The iCRPs recorded for the same wheat over a number of years and at different locations allows new grain hardness questions, which arise in subsequent scientific and technological deliberations, to be tested on historic wheats.

What is already known [5, 6] is that the RPPs change when the circumstances with respect to the grain (e.g. wheat) change such as for different varieties in the same environment (the genotype effect) and for different environments for the same wheat (the environmental effect).

In part, the focus for the talk is an assessment of the extent to which grain hardness can be used to monitor environmental change.

The underlying rationale on which this assessment is based includes:

(a) Various publications [1, 5, 6] have established that the RPP's are quite sensitive to genotype, with the protein and starch compositions defining indirectly the genotype.

(b) The environment, in which the grain is grown, affects the composition and relative proportions of the various protein groups (e.g. gliadin to glutenin ratio) and starch granule types within the grains [2, 4, 10].

(c) The RPPs correspond to indirect measurements of the protein and starch composition of the botanical layers that form wheat grains as well as other properties such as milling performance [1, 5, 6].

(d) Consequently, the RPPs can be used to characterize how different environments affect grain structure, and, thereby, how observed changes in the RPPs of different grains have been affected by the environment.

0.1 How should grain hardness be determined?

How grain hardness is measured is important. For example, Surma *et al.* [9] conclude that grain hardness is independent of environment which contradicts the comments in various papers [2, 4, 10].

This is a consequence of Surma *et al.* using the SKCS hardnes index (HI) to measure grain hardness which is known to be a highly problematic procedure, as explained in Anderssen and Haraszi [1], which can, due to the way that HI is calculated, even generate negative meaningless values.

In addition to the results presented in [5], which support the need for an improved grain hardness measure, such as their rheological hardness index (RHI), the problematic nature of HI also represents motivation for having a more clearly defined measure of grain hardness which relates to the rheological response of grains to their compression on an SKCS4100 device. The end-use properties of wheat flours are strongly influenced by grain hardness, such as biscuit texture [8] and bread volume.

0.2 Measuring grain hardness using the SKCS4100 device.

Using a Perten Instrument's SKCS 4100 device, one can measure for individual cereal grains (such as wheat) the force with which the grains resist their compressive crushing as a function of time. The resulting profiles are called individual crush response profiles (iCRPs). The huge variability in the iCRPs is illustrated in Figure 1.

As explained by Osborne and Anderssen [7], Anderssen and Haraszi [1] and others, if a suitably large number of iCRPs are averaged, the resulting averaged CRP (aCRP) yields a rheological encapsulation of the strength and fragmentaion of the internal botanical layers in the grains being crushed, as shown in

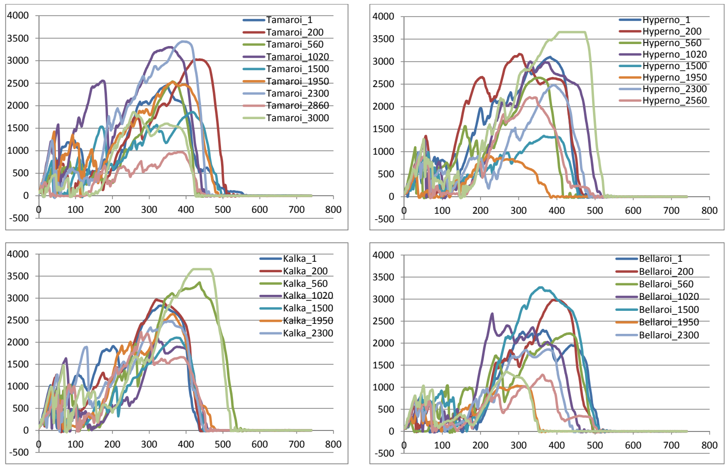

Figure 1: iCRPs for some representative durum wheats

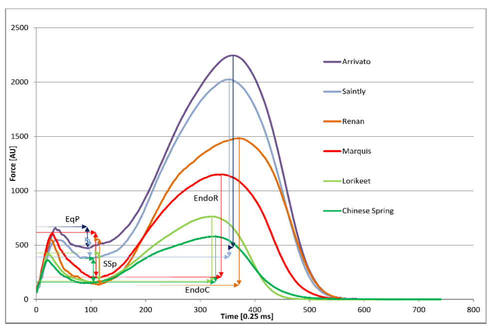

Figure 2: aCRPs for representative soft, hard and durum wheats, where these aCRPs have been obtained by averaging more than 1000 iCRPs.

Figure 2. In particular, from a cereal science perspective, the aCRPs represent independent rheological validation that the internal botanical layers in different wheat varieties behave in quite different ways and therefore can be viewed as indirect measurements of the differences in terms of their internal molecular structures.

The key features in the aCRPs can be summarized as the rheological phenotype phases (RPPs) as explained and illustrated in Haraszi *et al.* [5, 6]. Their importance relates to the fact that they yield a simple and quantitative basis for comparing the response to their crushing of different wheat varieties which in turn can be related to milling performance [6].

The successive steps starting with the measurement of the iCRPs, followed by the averaging of a suitably large number to generate the aCRPs and finishing with the cataloging of the RPPs can be viewed as a hierarchical *"bagging and* *boosting"* process where the weak images of the current stage generate the strong images of the next stage.

References

[1] R. S. Anderssen and R. Haraszi. Characterizing and exploiting the rheology of wheat hardness. *Euro. Food Res. Tech.*, 229:159–174, 2009.

[2] C. S. Blumenthal, E. W. R. Barlow, and C. W. Wrigley. Growth environment and wheat quality: the effect of heat stress on dough properties and gluten proteins. *Journal of Cereal Science*, 18:3–21, 1993.

[3] M. D. Casler. Fundamentals of experimental design: Guidelines for designing successful experiments. *Agronomy Journal*, 107:692–705, 2015.

[4] F. M. Dupont and S. B. Altenbach. Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of cereal science*, 38:133–146, 2003.

[5] R. Haraszi, A. Juhasz, M. Sissons, M. Rakszegi, L. Tamas, and R. S. Anderssen. Rheological Hardness Index for Assessing Hardness of Hexaploids and Durums. *Cereal Chem.*, 90:430–438, 2013.

[6] R. Haraszi, M. Sissons, A. Juhasz, G. Kadkol, L. Tamas, and R. S. Anderssen. Using rheological phenotype phases to predict rheological features of wheat hardness and milling potential of durum wheat. *Cereal Chemistry*, 93:369–376, 2016.

[7] B. G. Osborne and R. S. Anderssen. Single-kernel characterization principles and applications. *Cereal Chemistry*, 80:613–622, 2003.

[8] A. Pauly, B. Pareyt, M. A. Lambrecht, E. Fierens, and J. A. Delcour. Flour from wheat cultivars of varying hardness produces semi-sweet biscuits with varying textural and structural properties. *LWT-Food Science and Technology*, 53:452–457, 2013.

[9] M. Surma, T. Adamski, Z. Banaszak, Z. Kaczmarek, A. Kuczynska, M. Majcher, B. Lugowska, W. Obuchowski, B. Salmanowicz, and K. Krystkowiak. Effect of genotype, environment and their interaction on quality parameters of wheat breeding lines of diverse grain hardness. *Plant Production Science*, 15:192–203, 2012.

[10] C. Zhao, T. Ning, N. Jiao, B. Han, and Z. Li. Effects of genotype and environment on protein and starch quality of wheat grain. *Ying yong sheng tai xue bao (Journal of Applied Ecology*, 16:1257–1260, 2005.