

Impact case study (REF3b)

Institution: WestCHEM
Unit of Assessment: Sub-panel 8 – Chemistry
Title of case study: PolySNAP Computer Software for enhanced processing and classifying of crystallographic and spectroscopic data
1. Summary of the impact

PolySNAP is an extensive commercial computer program developed at WestCHEM to process and classify large volumes of crystallographic and spectroscopic data. It is a market-leading product sold and supported by Bruker Corporation (a manufacturer of scientific instruments for molecular and materials research selling products world-wide) and is used in laboratories throughout the world supporting business in the pharmaceutical, materials, mining, geology, and polymer science sectors. The PolySNAP software was and continues to be sold in combination with all Bruker x-ray powder diffractometers.

2. Underpinning research

Context

The research undertaken at WestCHEM addresses the problem of analysing large quantities data from x-ray powder diffraction and one-dimensional spectroscopic data, in particular, from Raman and IR spectroscopy [1,5]. In commercial research settings, it is now routine to collect thousands of x-ray powder-diffraction patterns in a day using crystallisation robots and a fast powder diffractometer. Many pharmaceutical and materials laboratories carry out such work routinely, one example being in the pharmaceutical industry where companies need to find all the possible polymorphs of a drug candidate both to protect their intellectual property and to ensure that more efficacious forms are not excluded. Using crystallisation robots, thousands of potential polymorphs are crystallised in small quantities and subjected to x-ray powder diffraction and, if possible, to Raman spectroscopy.

Key Research Outcomes

The motivation for the research carried out at WestCHEM was to explore ways to classify the data automatically and to provide associated data visualisation tools. The PolySNAP software classifies the patterns automatically using cluster analysis and multivariate statistics showing groups of patterns that are similar and thus belong to the same polymorph as well as highlighting new patterns that do not match the library of known forms [2-4]. From its first release, PolySNAP's approach has been revolutionary in that it uses every measured data point, whereas previously software used only the heights and positions of the top 5-10 peaks [6]. This makes it far more robust in its statistics especially with regard to poor quality data, which are common in high-throughput studies.

The WestCHEM research has led to a new integrated approach to full powder-diffraction pattern analysis. It incorporated wavelet-based data pre-processing, parametric and non-parametric statistical tests for full-pattern matching, and singular-value decomposition to extract quantitative phase information from mixtures. Every measured data point is used in both qualitative and quantitative analyses. The success of this new integrated approach was demonstrated through examples using several test data sets [6].

Key researchers

The initial research was carried out at WestCHEM under a grant from the Ford Motor Company (USA and Germany) starting in 2000. Major developments were commissioned by Pfizer in Kent and involved Professor Chris Gilmore (Prof at WestCHEM 1973-2010) along with Dr Gordon Barr (PDRA at WestCHEM 2002-10), and Dr Wei Dong (PDRA at WestCHEM 2003-07). Grants and other income continued from various sources until 2010 including an EPSRC Adventurous Research in Chemistry grant (2006).

Impact case study (REF3b)**3. References to the research**

References 2, 3 and 5 best illustrate the quality of the research:

- [1] 'High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles', C.J. Gilmore, G. Barr, and J. Paisley, *J. Appl. Cryst.* (2004), 37, 231-242. (doi: [10.1107/S002188980400038X](https://doi.org/10.1107/S002188980400038X))
- [2] 'High Throughput Powder Diffraction: II Applications of Clustering Methods and Multivariate Data Analysis', G. Barr, W. Dong and C.J. Gilmore, *J. Appl. Cryst.* (2004), 37, 243-252. (doi: [10.1107/S0021889804000391](https://doi.org/10.1107/S0021889804000391))
- [3] 'High-throughput powder diffraction. III. The application of full-profile pattern matching and multivariate statistical analysis to round-robin-type data sets', Gordon Barr, Wei Dong, Christopher Gilmore and John Faber *J. Appl. Cryst.* (2004), 37, 243-252. (doi: [10.1107/S0021889804013743](https://doi.org/10.1107/S0021889804013743))
- [4] 'High Throughput Powder Diffraction: IV Cluster Validation using Silhouettes and Fuzzy Clustering', G. Barr, W. Dong, and C.J. Gilmore, *J. Appl. Cryst.* (2004), 37, 874-882. (doi: [10.1107/S0021889804020990](https://doi.org/10.1107/S0021889804020990))
- [5] 'High Throughput Powder Diffraction V: The Use of Raman Spectroscopy with and without X-ray Powder Diffraction data', G. Barr, G. Cunningham, W. Dong, C.J. Gilmore and T. Kojima, *J. Appl. Cryst.* (2009), 42, 706-714. (doi: [10.1107/S0021889809022924](https://doi.org/10.1107/S0021889809022924))
- [6] 'PolySNAP3: a computer program for analysing and visualizing high-throughput data from diffraction and spectroscopic sources', G. Barr, W. Dong and C.J. Gilmore, *J. Appl. Cryst.* (2009), 42, 965-974. (doi: [10.1107/S0021889809025746](https://doi.org/10.1107/S0021889809025746))

PolySNAP was developed using grants to Prof Gilmore from Ford Motor Company (2000), Pfizer (2003, 2004, 2005), Bruker, the Cambridge CDC (2006), and EPSRC ('Adventurous Research in Chemistry', 2006) for a total value of £712k.

4. Details of the impact**WestCHEM develops software to support high-throughput systems**

At the same time that the precursor to PolySNAP, the SNAP-1D software package, was being developed at WestCHEM to facilitate an integrated approach to full x-ray powder-diffraction pattern analysis, the pharmaceutical industry was developing automated crystallisation systems using robotics, and was beginning to generate large volumes of one-dimensional diffraction patterns. However, there was no software available to process, classify, and visualise such data. PolySNAP was written under contract from Pfizer specifically to address this problem. It was the first software to do this and used novel statistics and visualisation methods. Data visualisation is a vital component of handling large amounts of data. With PolySNAP, time expenditure was reduced from days to seconds for >100 data sets. It is an almost impossible task to classify large data sets without these techniques and so any investment in the new robotic hardware would have been pointless. The software was described by one Pfizer employee as "like having a new spectroscopic method made available". A confidential list of PolySNAP users up to 2010 is available, which contains 72 users with commercial licences (Source 2). The software was the property of Glasgow University until 2010 when Professor Gilmore retired, at which time it was sold to Bruker.

PolySNAP supplied worldwide with Bruker diffractometers

Bruker purchased the rights to the program from Pfizer in 2002-2003 for ca. £125k (the same as the original cost to Pfizer) and started to market PolySNAP commercially under licence from Glasgow University (Source 1). Until 2010, license fees to the University totalled £84,164 (from 2008 £37,894) (Source 3). Bruker continued to develop this software and Gilmore set up a limited company partly for this purpose (Allander Science Ltd.) and continues to act as a consultant to Bruker. Dr Dong was employed full-time by Bruker during the assessment period. The development of PolySNAP also led to a free computer program called dSNAP that uses the same

Impact case study (REF3b)

ideas to classify and visualise the results of searches on the Cambridge Structural Database (see below).

New versions of the software were released: PolySNAP-2, PolySNAP-M, and the latest PolySNAP-3. PolySNAP-3 is unique in that it allows data from multiple sources, techniques or data collection strategies to be incorporated into the analysis. For example, data from powder diffraction can be combined with differential scanning calorimetry and Raman data to give a combined analysis and more accurate classifications of samples [5, 6]. This method is proving so important that Bruker are currently extending it to include x-ray fluorescence data in the new releases of PolySNAP. It saves money because it releases the time of highly paid scientific staff and it is also objective, avoiding the risk that exists with this sort of data that interpretation can become subjective.

The relationship between Bruker and the University continued until 2010 when Prof Gilmore retired. At that point, Bruker paid the equivalent of two years' license fees (€30k) to purchase the software outright. PolySNAP continues to be a part of the company's development plan (Source 4). Current list price when PolySNAP is purchased separately is €2,700, and approximately 120 have been sold since 2004, while Bruker also offers another version, m-SNAP, at a list price of €5,000. PolySNAP continues to be sold in combination with all Bruker x-ray powder diffractometers.

PolySNAP utilised across industry and education

Although the impact beyond the sale of the software is very difficult to quantify, the beneficiaries of this software are many:

- The pharmaceutical industry: Most large pharmaceutical companies use PolySNAP: Novartis, AstraZeneca, Pfizer, Roche, Syngenta, Sanofi Aventis, and many smaller companies. It dramatically reduces both the timescale and the reliability of sample screening allowing even very large datasets (>100,000 samples) to be pre-screened in minutes to find the subset of patterns most relevant to a given target. Prior to PolySNAP, such searches would have used reduced patterns in which the powder data are presented as a small set of intensities and positions for the top 5-10 peaks. PolySNAP was the first program to use the full measured data in all its analysis steps. High-throughput data are of poor quality and using traditional reduced patterns is ineffective at best and often worthless.
- Mining industry: e.g., companies such as Anglo Platinum and Mintek in southern Africa, where PolySNAP software is used to assess the quality of ore samples prior to processing into pure metals, and to group huge quantities of materials of different origin. The software allows the companies to carry out rapid comprehensive assessments of ore variability in terms of bulk mineralogical composition and facilitating decision-making.
- Materials Science: identifying new materials, studying phase changes with temperature and pressure, simple quantitative analysis of mixtures.
- Forensic science: e.g. South African Police Force have used PolySNAP to match samples from crime scenes with a database of known samples. It is especially valuable when used with mineral theft.

A Principal Scientist, Materials Science, at AstraZeneca explains the value of PolySNAP (Source 5):

"Identification of a suitable solid form for development of a crystalline material still remains a great challenge for the pharmaceutical industry. Due to the different physical properties (e.g., dissolution, stability) of different polymorphs, extensive screening is performed during the early stages of development to ensure the solid form architecture of the proposed molecule is fully understood and there have been a number of costly issues where compounds have had to be reformulated due to changes of form after launch.

In screening for different solid forms, both manual and automated technologies have been utilised and a number of high throughput platforms have been developed that utilise common analytical techniques, e.g., PXRD for analysis of solid form. These techniques yield large quantities of analytical data which is quite difficult to comprehend and can rapidly overwhelm operators with excessive data. [...] PolySNAP utilises a wide range of mathematical

Impact case study (REF3b)

techniques to help manage large data sets and successfully guides the scientists into not only the presence of the different forms but, more importantly, how these forms have been manufactured and hence how they can be controlled. It provides a range of pattern-matching methods and allows the user to have full control over understanding the data avoiding 'black box' type results leaving the operator in control and reducing large datasets to a much smaller number of samples which require interpretation by scientists. The results from PolySNAP allow rapid identification of 'safe' areas of working (e.g., solvents, temperature, concentrations, etc.) and where further understanding is required. Most recently, the extra dimension of adding the capability of other analytical data (e.g., spectroscopic) has provided further confidence and opportunities in the pattern matching."

A one-month trial version of the PolySNAP software is available from WestCHEM, accompanied by a tutorial and trial data (Source 6).

Visualising the Cambridge Structural Database with dSNAP

The WestCHEM research that led to PolySNAP also underpinned a free computer program called dSNAP (Source 7) that classifies and visualises the results of searches on the Cambridge Structural Database (CSD), a repository of over 500,000 crystal structures derived from x-ray and neutron diffraction. The CSD is by far the most important source of structural data for small molecules, but while it is easy to search this database, it can be a problem to interpret the results. dSNAP is an important tool for this. dSNAP simplifies the problem using the same methodology as PolySNAP: it clusters the search results into groups that are very similar and thus allows the user to treat the group as a single entity so reducing the number of structures that need to be examined. Additional tools are provided to assist in this. A confidential list of 166 licensed dSNAP users up to 2010 is available (Source 2).

5. Sources to corroborate the impact

- [1] Evidence of use of software by Bruker: **Bruker website** <http://www.bruker.com/en/products/x-ray-diffraction-and-elemental-analysis/x-ray-diffraction/xrd-software/applications/xrd-software-applications/polysnap.html>
- [2] A **confidential** list of PolySNAP users up to 2010 is available. [To be treated as commercially sensitive] The list also contains 166 licenced dSNAP users up to 2010. This software is free, and is now downloadable without a licence, so users after 2010 are not registered on the spreadsheet.
- [3] There is a **confidential** spreadsheet of royalty income from Bruker arising from sales of the PolySNAP software up to 2010. [To be treated as commercially sensitive]
- [4] Product Manager XRD at Bruker AXS GmbH in Karlsruhe Germany can be contacted to confirm the value of PolySNAP to Bruker products.
- [5] Statement from Principal Scientist, Materials Sciences, AstraZeneca R&D provides evidence of the value of PolySNAP in materials analysis
- [6] A one-month trial version of the PolySNAP software can be downloaded by sending an e-mail to snap@chem.gla.ac.uk. The program comes with a tutorial and trial data.
- [7] The dSNAP software can be downloaded from:
http://www.chem.gla.ac.uk/snap/PolySNAP_index.html