Big Data for CSG Analytics

Jaco Du Plessis and Dr Morteza Jami , School of Earth Science/Centre for Geoscience Computing

ABSTRACT

During the last two decades several thousand exploration and development wells were drilled in Surat Basin targeting the Juandah (upper) and Taroom (lower) Coal Measures of the Middle Jurassic Walloon Subgroup of the Injune Creek Group. While these wells have provided a gigantic volume of data from stratigraphy to geomechanics and coal petrology, there is a lack of a general model and data comparison in the whole Surat Basin as all the appraisal activities were conducted independently by several oil and gas companies and the government. The big data project seeks to provide a data centric platform through which this large dataset can be explored and analysed, to provide business insights; and perform some of the analysis. The programming language Python is used as it is:

- widely used for data analytics e.g. in Google's internal analytics tool (TensorFlow)
- very easy to learn and use while still providing good performance
- easy to integrate and produce data in a wide range of formats.

So far the platform has been used to analyse the distribution of seam thicknesses and net coal across the Surat basin. This recent work has confirmed that the coal seams within the Walloon Subgroup exhibit highly variable properties such as permeability and cleat density which has a particular relationship with geophysical logs, stratigraphic location of the seam and spatial location of the well in the Basin. As a result this project can reveal unique subsurface information that will help to develop a better understanding of sweet spots and trends in Surat Basin.

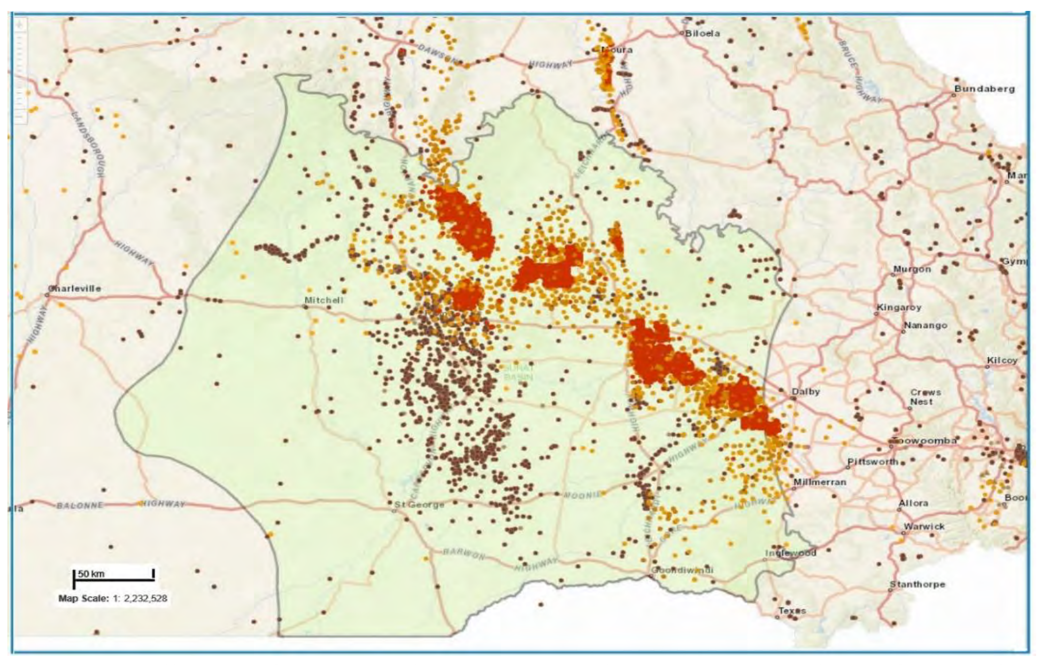

Fig 1 *Well data for Surat basin in south east Queensland from the minesonlinemap(https://www.business.qld.gov.au/industry/mining/mining-online-services/minesonlinemaps).*

SOFTWARE DEVELOPMENTS

Firstly, Big Data Analytics enable the vast amount of well data to be explored for valuable insights, such as large scale analysis to find subtle links between rate of penetration and weight on bit for various formations across the Basin. Additionally Machine Learning offers the opportunity to link well data to parameters which are expensive to test, such as permeability, with the goal of developing artificial drill stem tests via well log data to reduce costs. There are also a number of geological tasks, for example stratigraphic correlation and coal seam/package recognition, that have the potential to become at least partially automated via Data Analytics and Statistical Analysis. A general Big Data Platform will also allow for multiple geological software packages to be combined producing a more effective overall solution.

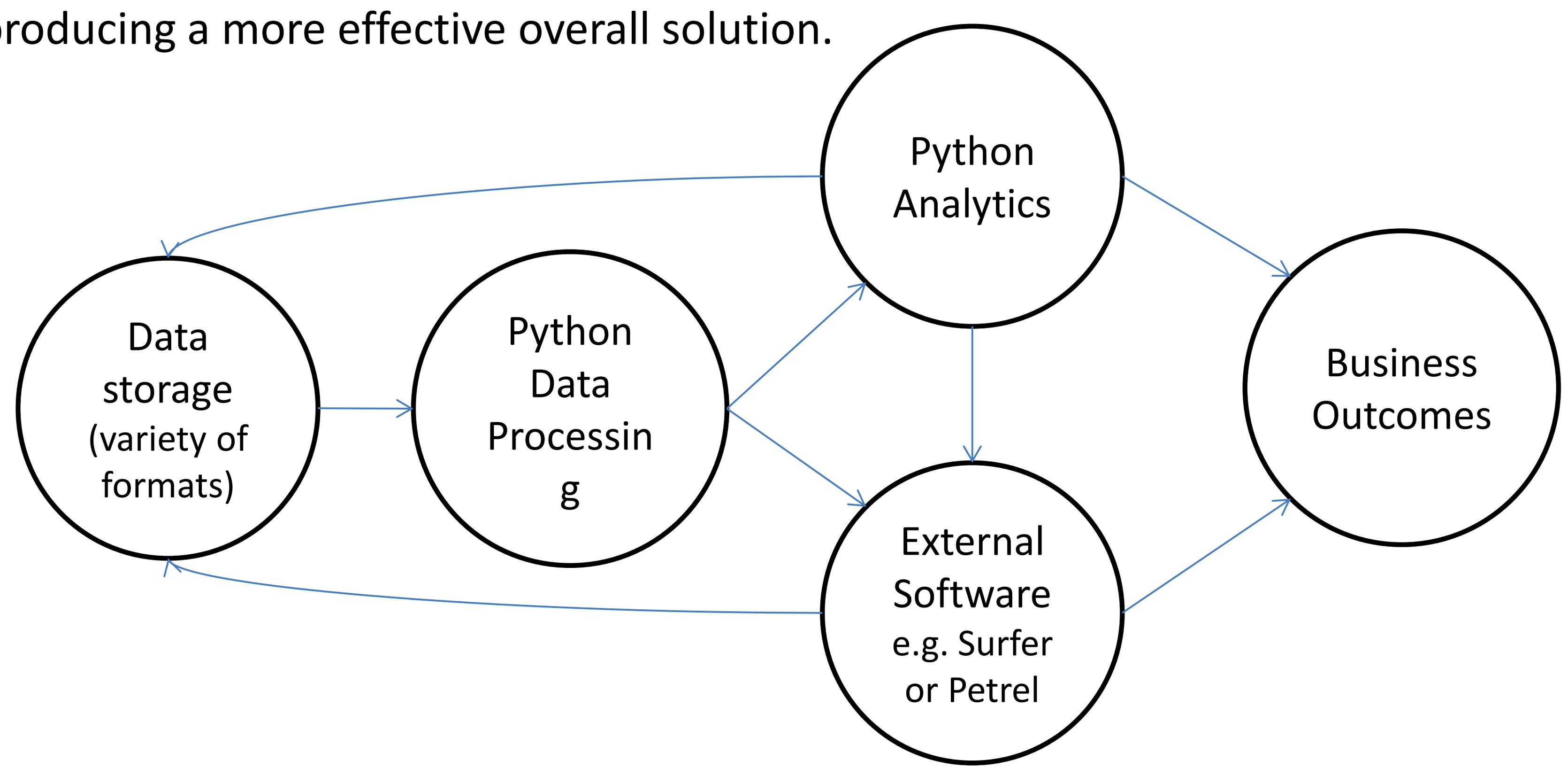

Fig 2 *Big data project workflow: Data is collected in various formats in a single location. The data may include Well header/data, Geophysical logs, Well tests, Coal/reservoir data (i.e. desorption, isotherm and production data), Drilling and Well Completion data, Production data. This data may come in any format for which a specification can be found, e.g. csv, las and sql style databases. The data is then processed in python into a form which is appropriate for analysis. From here analytics is performed in either python or external software. The results of the analytics can be passed between python and the external software, passed on as a business outcome or put back into storage for future analysis.*

The Python based data analytics platform is being developed by the Centre for Geoscience Computing at UQ. As it is a general programming language it is possible to implement a reader for any arbitrary dataset provided that a format is available. Python, as a language, has simple and convenient syntax without degrading its functionality making it easy to learn and effective to use. Adding to this, due to its integration with C, it retains quick processing speeds for time intensive tasks and allows for the use of GPUs to accelerate machine learning and integration with systems like Hadoop and Spark for huge analytics problems, all of which can be achieved without the cumbersome syntax of C. As a result, it is widely used by the scientific community for data analytics and a large selection of libraries exist, such as Pandas, Scikitlearn, Theano, Pymc3 and TensorFlow ,which are all freely available for use by the community.

EXAMPLE OF CURRENT ANALYTICS

So far the project has centred around data collection and processing in Python. Qdex well data consisting of over 16000 las files has been processed to produce an SQL database, which can be queried to quickly extract data of interest. The process of developing this database involved significant data cleanup of the las files. Functionality has been implemented to search through the database to find individual seams in wells. Figures 3a and 3b shows how this has been applied to find the thickness of the thickest seams in the Walloon Coal Measures in histogram and contour plot form.

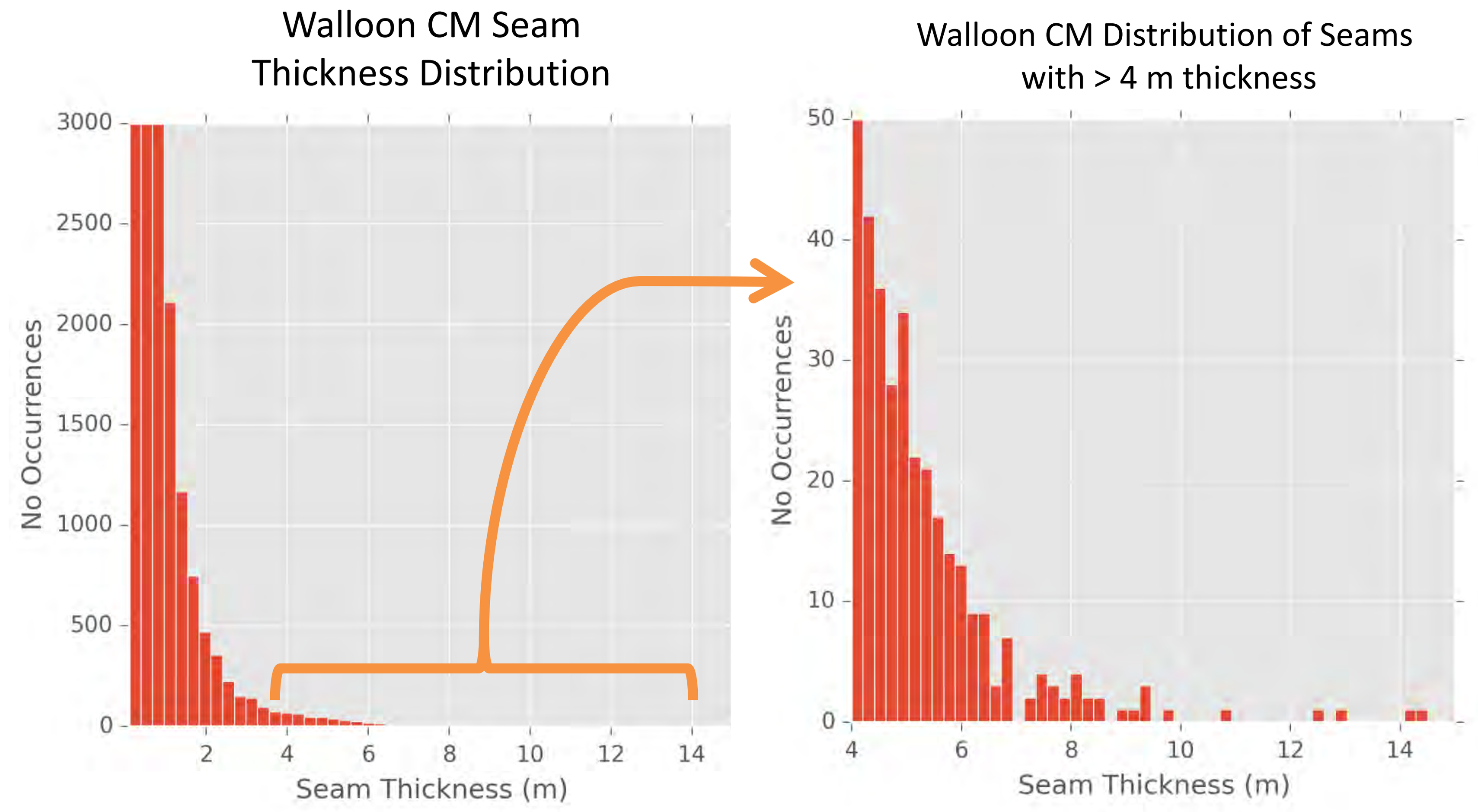

Fig 3a: *The Walloon coals are one of the key producers of gas in the Surat Basin. As a consequence, the distribution of thick seams in the Walloon sub-group is of interest as these are generally the primary producers. For the graphs above, coal cut off was < 1.80 g/cc in Density logs, coal seams with stone bands < 0.5 m were considered as one seam and the stone bands were removed from seam thickness. The Big Data system allows arbitrary analysis to be performed on coal seam data with ease, in this case finding individual seams in all the wells in Queensland. Then the stratigraphic picks are integrated to identify which seams are part of the Walloon sub-group and finally the thickest seam from each well is picked out.*

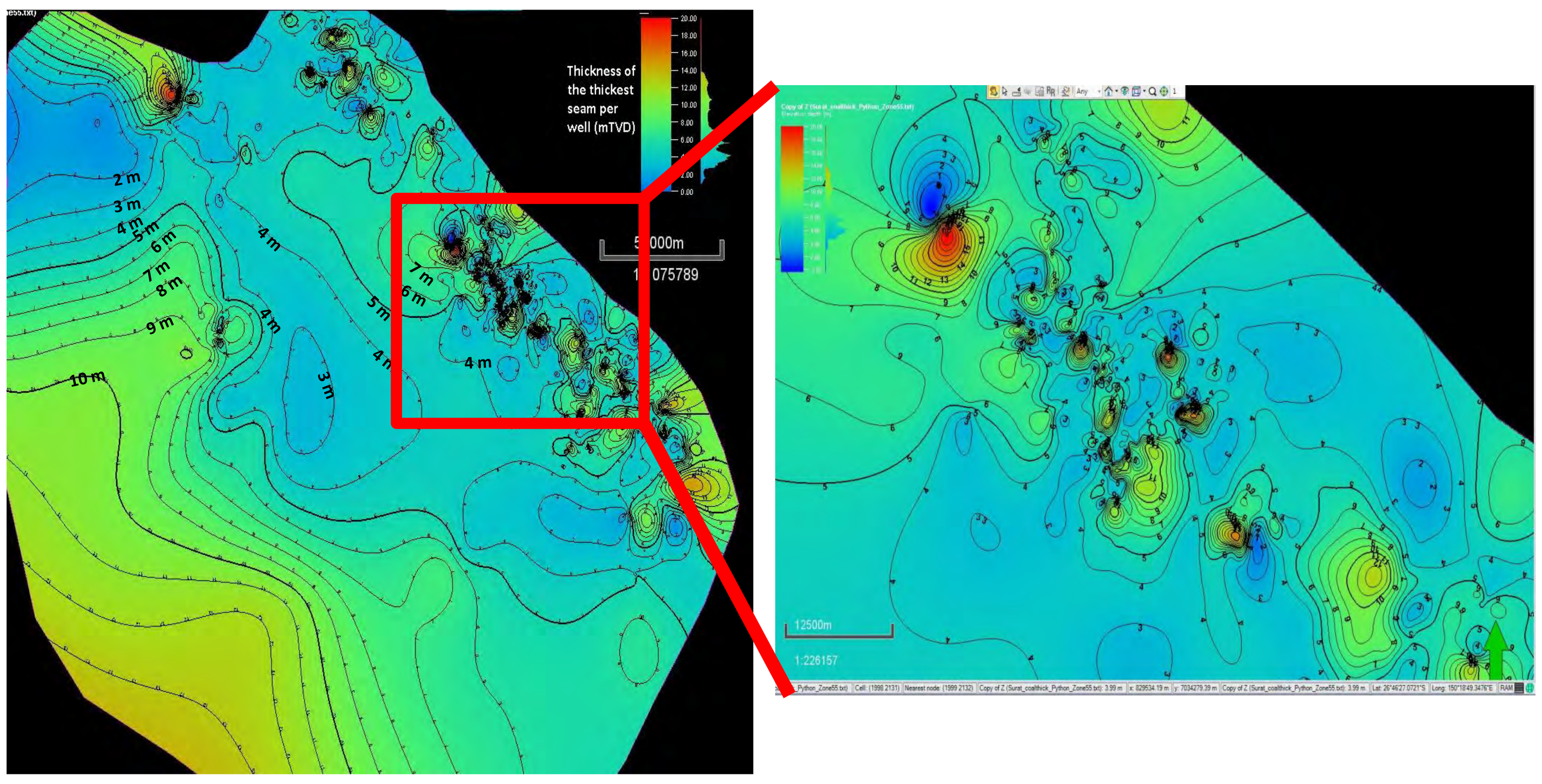

Fig 3b: *Contour map of the thickest seam thickness of more than a 1000 wells in Surat Basin. The thickest seam is picked by an automated process using wireline data. The seams with stone band < 0.5 m were considered to be one seam. These stone bands were removed from the seam thickness.*

CONCLUSIONS

Some of the key findings of this feasibility project are:

- Python is a highly suitable programming language for this platform given its ease of use, flexibility, functionality and current use in data analytics.
- The large corpus of data offers a huge opportunity to provide valuable business insights when analysed correctly.
- Significant progress has been made towards producing a useful platform which will allow for the integration of data analytics libraries in Python with well- known geological software such as Petrel.

FUTURE RESEARCH

The next stage of the project will involve investigating the link between log data and permeability with the aim of producing an artificial drill stem test. Next drilling data will be inspected to identify optimal weight on bit and other parameters in order to optimise rate of penetration through various formations. Other applications for future consideration include semi-autonomous picking of stratigraphic tops through sequence stratigraphy.