Big Data: An Open Source Solution

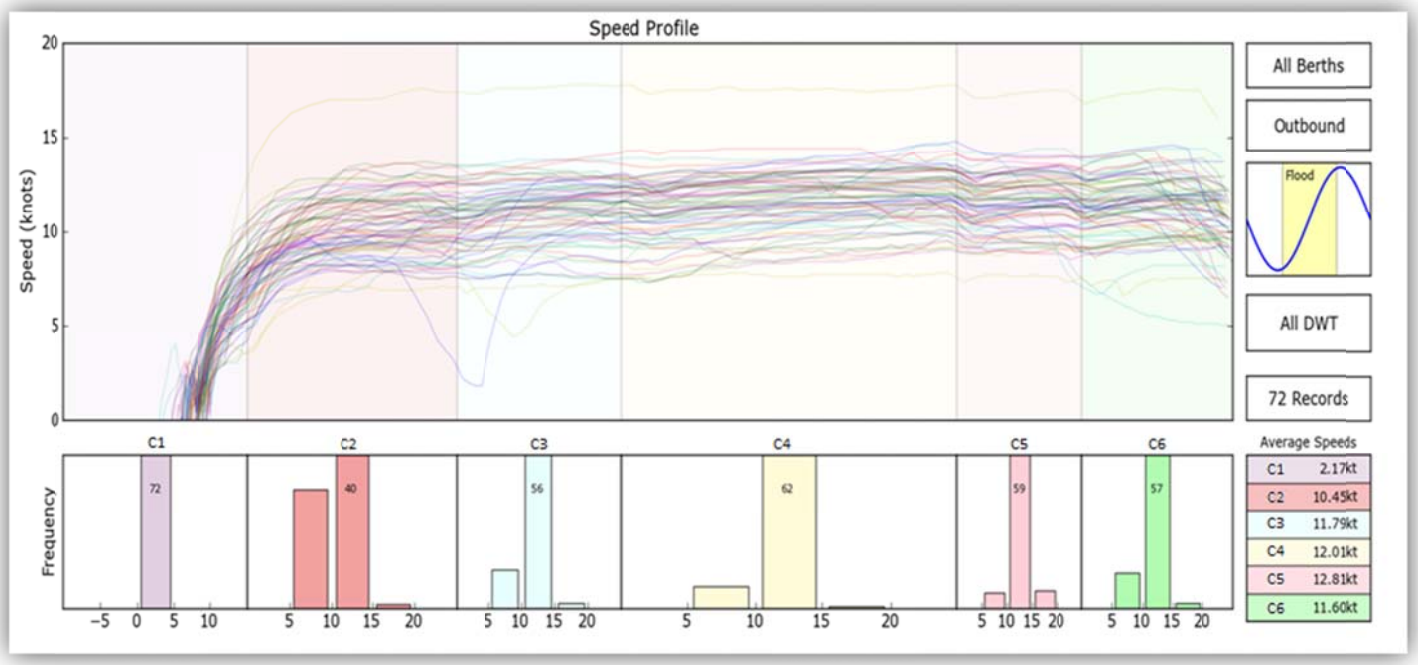

Figure 1: Python-generated chart summarising ship movements though a series of commercial port channels.

The AECOM Analytics and Simulation team thrives on finding smarter and more efficient solutions to some of the world's most challenging projects. We routinely simulate transport and resource related operations research problems. Commercial seaports, container terminals, mining equipment and rail networks are represented as dynamic simulation models in order to determine the impact of varying infrastructure options or operational rules.

Analysis and visualisation of datasets in a way that can be understood by non-technical people is crucial to any simulation study. This includes analysis and transformation of historical operational data for generating model inputs and gathering an understanding of the current situation. Summarising simulation outputs– the position of each element in the system at every time step, and all associated elemental and system properties- is also important to be able to draw conclusions and analyse outcomes quickly to inform business decisions.

In the past, the team at AECOM have relied heavily on Microsoft Excel and its programmable VBA (Visual Basic for Applications) macros to assist with data analysis. However, Excel has severe limitations on processing large amounts of data (often of gigabyte magnitude) and is unable to open data files over a certain size. Thus, **a more powerful way to analyse data prompted a move from Excel to Python**.

Python is an open source, cross platform programming language. With an emphasis on code readability, it is an excellent first choice for those who wish to learn programming and is no more difficult than VBA coding.

Thousands of freely available, user-created libraries exist online to assist with every aspect of data analysis and visualsation. Implementing charts from the Matplotlib package, for example, greatly expands on Excel's graphing capabilities. Furthermore, Python is able to interface with other open source technologies. Simulation output data files that are gigabytes in size can be handled and summarised by data filing programs such as HDF5, a flexible and efficient data library designed to manage high volume and complex datasets. Large datasets are extracted to a HDF5 database where they can be analysed and visualised by Python packages.

Customised charting is just one of the many ways Python can enhance data analysis. For example, AECOM conducted analysis of raw vessel traffic services (VTS) data for historical vessel movements in a commercial port. As depicted in Figure 1, Python charting allowed vessel speed profiles in the various channel segments to be visualised in various formats. This information was used to visually demonstrate and quantify the variability in vessel speeds due to factors such as destination (berth) and tide condition. Furthermore, the information is used for decision support particularly around operational issues.

Not only have open source technologies been used by AECOM to deliver more powerful and customisable solutions than off-the-shelf commercial packages, they're free; a win-win for both service providers and clients alike.

AECOM is a global provider of technical and management support services to a broad range of markets, including transportation, facilities, environmental, energy, water and government. Our teams deliver solutions that create, enhance and sustain the world's built, natural and social environments. Advisors, engineers, designers and planners work together as an integrated team that is committed to client success.

The Analytics and Simulation Team in particular provides decision support through identification of capacity bottlenecks and system inefficiencies – ultimately improving profitability, reducing CAPEX/OPEX and reducing client risk. Our work includes some of our region's most iconic projects such as the Port of Gladstone Expansion, the Cross River Rail Project and the award winning Oakajee Supply Chain in Western Australia.

For further information, please contact Kari Stuart
Senior Simulation Analyst | AECOM Analytics & Simulation | Kari.Stuart@aecom.com