

<b>Institution: Newcastle University</b>		
<b>Unit of Assessment: 10</b>		
<b>Title of case study: Statistics-based Data Analytics for Industry - A Focus on Small and Medium Enterprises</b>		
<b>Period when the underpinning research was undertaken: 2000-present</b>		
<b>Details of staff conducting the underpinning research from the submitting unit:</b>		
<b>Name(s):</b>	<b>Role(s) (e.g. job title):</b>	<b>Period(s) employed by submitting HEI:</b>
Dr Shirley Coleman	Technical Director and Principal Research Associate	1990-present
Dr Jian Shi	Reader in Statistics	2002-2020
<b>Period when the claimed impact occurred: 2014-present</b>		
<b>Is this case study continued from a case study submitted in 2014? N</b>		
<b>1. Summary of the impact</b> (indicative maximum 100 words)		
<p>A School of Mathematics, Statistics and Physics research team has delivered significant impacts within Small and Medium-sized Enterprises (SMEs) through the innovative design and application of statistical methods and data analytics.</p> <p>Economic and commercial, as well as societal and environmental impacts, have occurred within the marine, utilities, automotive, social housing, and service sectors, corroborated by evidence. Total benefits include:</p> <ul style="list-style-type: none"> <li>i) increases in annual sales of greater than £4.5M;</li> <li>ii) increases in annual exports of greater than £4M;</li> <li>iii) increases in annual profits of greater than £1M;</li> <li>iv) the creation of intellectual property;</li> <li>v) the creation of at least 13 new jobs.</li> </ul>		
<b>2. Underpinning research</b> (indicative maximum 500 words)		
<b><u>The SME challenge</u></b>		
<p>Developments in statistical and data analytics technologies are relatively easily disseminated through academic communities and large, well-resourced organisations. It is more difficult for a SME to become aware of and subsequently exploit the best available techniques. SMEs contribute 50% to the UK economy and employ 82% of the UK workforce. Within Europe, SMEs employ around 65 million people.</p> <p>A mathematical sciences research team at Newcastle University (NU), led by Dr Shirley Coleman and including Dr Jian Shi, has developed and transferred statistics-based data analytics tools into a wide range of industrial and business SMEs [P1-P4]. Initially supported by €750,000 from the European Regional Development Fund [G1] and €500,000 from the European Social Fund [G2], and subsequently through five Knowledge Transfer Partnerships within the REF period [G3-G7], the group has, inter alia, embedded generic (e.g. [P5]) and targeted (e.g. [P6, P7, P8]) Newcastle research within more than 100 SMEs. The following case studies provide examples.</p>		

**Marine sector**

In collaboration with Royston Diesel Power, using research reported in [P1] and [P3], location data from company systems was combined with open access meteorological data on tide and wind to analyse fuel consumption and emissions for active marine engines. Summed rank cumulative sum techniques [P6] and other original statistical process control methods [P7] were applied to identify process step changes that led to negative environmental impacts and to build a predictive model to allow efficient scheduling of services.

**Utilities sector**

In collaboration with Advanced Engineering Solutions (AES), Gaussian process regression methodology [P5] was extended for detection and sizing of pipeline defects. AES detection equipment records changes in magnetic flux as eight external sensors traverse regions of interest. The flexible Gaussian process method automatically maps the signals into measures of remaining wall thickness, allowing defects to be identified without strong parametric assumptions. Empirical Bayesian methods are used for hyperparameter estimation and principal components for multivariate summaries.

**Automotive sector**

In collaboration with Rain Data, a portfolio of data science reliability methods for the automotive industry was developed and applied in [P8], based on big data such as 14 million part return records and 6 million replacement part records. The portfolio includes a novel funnel plot control chart and normal mixture survival analysis. The methods allow the SME to provide the sector with near real time reliability information for parts and vehicles, for use in product matching, stock control, pricing and planning.

**Social housing sector**

Social housing accounts for 50% of rental properties in the UK. In collaboration with Orchard Information Systems, Coleman and colleagues [P3] developed methods to exploit the large reservoir of data on rent balances, property repairs and empty properties to identify tenants in danger of falling into arrears. The methods blend machine learning and data visualization techniques, specifically random forests, T-distributed stochastic neighbor embedding and SHAP plots.

**Service sector**

In collaboration with Enzen Global, official statistics and other publicly available open data were integrated with company data [P1] to quantify the effect of improved localized weather measurements for predicting gas consumption. Structural equation modelling and Bayesian network analysis evaluated the relationships between socio-economic factors and gas consumption.

**3. References to the research (indicative maximum of six references)**

[P1] Coleman, S (2016). Data mining opportunities for small to medium enterprises from official statistics. *Journal of Official Statistics* 32, 849-866. doi.org/10.1515/jos-2016-0044

[P2] Coleman, S.Y., Gob, R., Manco, G., Pievatolo, A., Tort-Martorell, X. and Reis, M. (2016). How can SMEs benefit from big data? Challenges and a path forward. *Journal of Quality and Reliability Engineering International* 32, 2151–2164. doi.org/10.1002/qre.2008

## Impact case study (REF3)

[P3] Vicario, G. and Coleman, S.Y. (2020). A review of data science in business and industry and a future view. *Applied Stochastic Models in Business and Industry* 36, 6-18. doi.org/10.1002/asmb.2488

[P4] Coleman S.Y. (2019) Data science in Industry 4.0. *Mathematics in Industry* 30, 559-566. doi.org/10.1007/978-3-030-27550-1\_71

[P5] Shi JQ, Murray-Smith R, Titterton DM. (2005) Hierarchical Gaussian process mixtures for regression. *Statistics and Computing* 15(1), 31-41. doi.org/10.1007/s11222-005-4787-7

[P6] Stewardson, D.J. and Coleman, S.Y. (2001). Using the summed rank cusum for monitoring environmental data from industrial processes. *Journal of Applied Statistics* 28, 469-484. doi.org/10.1080/02664760120034180

[P7] Coleman, S. Y., Arunakumar, G., Foldvary, F. and Feltham, R. (2001). SPC as a tool for creating a successful business measurement framework. *Journal of Applied Statistics* 28, 325-334. doi.org/10.1080/02664760120034063

[P8] Smith, W.S., Coleman, S.Y., Bacardit, J. and Coxon, S. (2019). Insight from data analytics with an automotive aftermarket SME. *Journal of Quality and Reliability Engineering International* 35, 1396– 1407. doi.org/10.1002/qre.2529

### Grants

[G1] European Regional Development Fund to engage and transfer research into SMEs (€750k).

[G2] European Social Fund to engage and embed statistical skills within SMEs (€500k).

[G3] KTP programme with Royston (£134k). 2014.

[G4] KTP programme with Enzen Global (£134k). 2014.

[G5] KTP programme with Orchard (£140k). 2016.

[G6] KTP programme with Rain Data (£133k). 2016.

[G7] KTP programme with Royston (£137k). 2017.

### 4. Details of the impact (indicative maximum 750 words)

Economic, commercial, societal and environmental impacts have occurred within the marine, utilities, automotive, social housing, and service sectors.

#### Impact within the Marine sector

Royston Diesel Power are specialists in diesel power, sales and repairs, with a significant customer-base within the marine and offshore industry. The research work has led to the following benefits since 2016 [E1]:

- an increase in **annual sales turnover of £4M;**
- an increase in **annual exports of £4M;**
- an increase in **annual profit of approximately £1M;**
- the creation of **5 new jobs** within the business;
- an **investment of greater than £1M** in new software and hardware implementation and support.

Statistical data analytics research was used to develop modules for Royston's fuel monitoring product Enginei. *"The research work enabled us to develop new opportunities via our existing customer base and global shipping databases. The data analytical capability has added significant value to customer data through the creation of high-value operational intelligence for the customer."* [E1]. As a result of the research collaboration with Newcastle University, Royston began a 3- year collaborative R&D programme with Innovate UK with a focus on "whole vessel energy monitoring". The **value of the programme was £1.2m**.

### **Impact within the Utilities sector**

Advanced Engineering Solutions (AES) are specialist pipeline and pipeline equipment engineers. The research has generated **annual turnover of £275,000** [E2] for AES through the capturing of new service contracts. Modern statistical methodology has become an integral part of AES's analysis processes.

The well-recognised AES pipeline condition assessment process is used on a global basis. *The statistical analysis method for pipeline defect identification, detection and sizing has been implemented and utilised within our processes, used by water companies across the UK, and is becoming a key component within technological services provided to our collaborative organisations in France (Suez) and Australasia (Asset Integrity Australasia).*" [E2].

### **Impact within the Automotive sector**

Rain Data specialise in data and cataloguing services for the automotive sales after-market. The research work achieved the following benefits [E3a, E3b]:

- an increase in **annual sales turnover of £60k** (estimated to be £180k within 3 years);
- an increase in **annual profit of £35k** (estimated to be £95k within 3 years);
- the **creation of 2 new jobs**.

Newcastle University research has improved Rain Data services to their customers through establishing data analytics capacity within the business. The implementation of data analytics tools has increased levels of data automation and accuracy; avoided the outsourcing of data cleansing and analytics activities; enhanced the profiling of stock; enabled better prediction of car trends and likely mechanical failures; reduced customer returns thus increasing client satisfaction; and improved stock level efficiency [E3a].

### **Impact within the Social Housing sector**

Orchard Information Systems provide property management solutions and services to housing organisations and councils throughout the UK. Social housing research impact benefits include [E4a, E4b]:

- the **development of Intellectual Property (IP)** related to predictive models for rent arrears and void properties. This IP has been embedded into Orchard products and services. The **prediction of voids is a sector first**;
- the **creation of 2 new jobs**;
- supporting the **development of Income Analytics**, Orchard's flagship product.

Data analytics has become a key component of new solutions and business insight for Orchard clients. The generation of new products has strengthened Orchard's position in the sector and provided competitive advantage.

**Impact within the Service sector**

Enzen Global provide consulting, technology, engineering and innovation services to the energy and water industries. Core benefits, as a result of the creation of an improved data analytics capability through the research, are [E5, E6a, E6b]:

- **an increase in annual sales turnover of £292,000**
- **an increase in profit of £70,000**
- **the creation of 4 new jobs**

The subsequent year-on-year increase in sales turnover and profit could be up to 10 times higher, although Enzen have not undertaken a full cost-benefit analysis.

The research has been embedded within Enzen's Knowledge Centre and enabled the company to **expand their range of business products and services**. The Newcastle University data analytics research has enabled applications to diverse challenges within the utilities service sectors across the UK [E5]. Work with **Northern Gas Networks** established improved gas demand estimations by using local rather than regional weather measurements. A collaboration with **Scotia Gas Networks** identified key relationships between socio-economic factors and residential demand. A project with **Wales and West Utilities** resolved public reported escapes (PREs) of gas yielding financial optimisation benefits.

*"...this work with Newcastle University opened up a new revenue stream for Enzen and a new way of thinking about data and the opportunities that it provides". [E5].*

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

[E1] Testimonial from the Chief Executive Officer of Royston Limited. *Provides evidence of the financial impact.*

[E2] Testimonial from the Technical Director of Advanced Engineering Solutions (AES). *Provides evidence of the financial impact and global reach of the research.*

[E3a] Rain Data KTP final report. *Provides evidence of the main achievements of the research (pages 2-3), financial benefits (pages 5-6) and new jobs created (page 7).* [E3b] Declaration by Partners. *Provides evidence of completion of project deliverables.*

[E4a] Orchard KTP final report. *Provides evidence of developing Income Analytics (page 7), new jobs created (page 9) and developing IP (page 12).* [E4b] Declaration by Partners. *Provides evidence of completion of project deliverables.*

[E5] Testimonial from the Client Partner at Enzen Global. *Provides evidence of the main benefits of the research.*

[E6a] Enzen Global KTP final report. *Provides evidence of the financial impact (pages 7-8) and new jobs created (page 10).* [E6b] Declaration by Partners. *Provides evidence of completion of project deliverables.*