Institution: Lancaster University

Unit of Assessment: 10 Mathematical Sciences

Title of case study: Adoption of Lancaster's changepoint methodology across a broad range of sectors results in significant economic, societal and environmental impact.

Period when the underpinning research was undertaken: 2010 – present

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Idris Eckley	Professor	2007-present
Paul Fearnhead	Distinguished Professor	2001-present
Rebecca Killick	Senior Lecturer	2013-present

Period when the claimed impact occurred: 2015 – present

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact

The prevalence of sensors and data collection systems in our everyday lives means that data sequences are being generated at unparalleled rates in both commercial and public sector settings. Lancaster's fundamental research on identifying the times at which changes occur in data sequences, together with the development of accompanying open-source software, has led to impact realised across a diverse range of public, private and charitable organisations both nationally and internationally.

Illustrative impact examples from the census period include:

- **Digital communications**: Helping to monitor and assure the nationwide operational performance of BT's critical digital infrastructure.
- **Online grocery retail:** Improving Tesco's operational forecasts, resulting in substantial directly attributable savings.
- **Environmental monitoring and improvement:** Helping to assure high-quality data from long-term monitoring programmes, identifying tipping points within long-term lake water quality records in the English Lake District, and monitoring the effect of resulting remediation measures.
- **Supporting safety of space travel:** via inclusion within a NASA Standard Testing Procedure used to certify that manufacturer submitted survival suit designs maintain safe levels of carbon dioxide during suited operations.

End-user demand for these tools has led to Lancaster's methods being incorporated by NAG and MathWorks, two internationally recognised, commercial numerical computing environment providers as well as open-source environments. As a consequence, our research is now accessible to millions of data analysts globally, with users in Government agencies (national and international); tech start-ups; FTSE100 and NASDAQ corporations.

2. Underpinning research (indicative maximum 500 words)

The research underpinning this case study consists of work carried out at Lancaster University from 2010 onwards by Idris Eckley, Paul Fearnhead and Rebecca Killick (formerly PhD student and faculty member since 2013), culminating with the 2012 publication of a landmark paper [3.1] in the Journal of the American Statistical Association. This work introduced the PELT (Pruned Exact Linear Time) changepoint search method for univariate signals. PELT is a computationally efficient and exact method that detects multiple changepoints within a data sequence in linear time, under conditions that are suitable for a broad range of changepoint problems. This makes PELT highly suitable for big data settings.

PELT's development, coupled with the release of associated open-source R software and documentation [3.2], has enabled a step-change for the user community who are now readily able to analyse longer series in a matter of seconds. Combined, these two papers [3.1,3.2] have been cited over 1,800 times and have been widely adopted across a range of disciplines, notably environmental sciences and finance.

Subsequently, the group have developed several further methods, including those highlighted in [3.3-3.6]. These have typically been in response to methodological challenges emerging via our collaborative partnerships with key industrial research sponsors including BT, DSTL, Howz, Shell and Tesco. Innovations include generalisations of the PELT approach to (i) a nonparametric setting, where fewer assumptions about the precise data structure are required [3.4], and (ii) a setting that allows for the detection of anomalies [3.5].

Engaged partnership: A key feature of our underpinning research has been to ensure that our research is significantly influenced by research insights derived from our substantive and long-standing collaborations with a range of partners in different sectors. To date, this approach has resulted in 14 funded and co-supervised PhD students, 3 PhD internships and project partners' support on various EPSRC proposals. As a number of our corroborating letters indicate, this philosophy has also enabled us to:

- Produce novel statistical methodologies that integrate knowledge of specific, key industrial challenges;
- Provide cutting edge, statistical thought-leadership to a variety of sectors so that they may capitalise on latest changepoint research developments;
- Develop environments for best practice sharing and knowledge exchange.

For example, the impact we describe with Tesco arose from proof of concept work initiated as part of a short-term research project by group members and a PhD student in 2016. Similarly, the impact that our methods have realised in BT has been catalysed through our substantial collaboration during the REF census period, including 5 co-funded PhD projects and core partner activity within the EPSRC-funded StatScale Programme Grant and EPSRC/BT Prosperity Partnership.

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

3.1. Killick, R., Fearnhead, P., Eckley, I.A. (2012) Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association. 107, 1590-1598. 1087 citations*
3.2. Killick, R., Eckley, I.A. (2014) changepoint: An R Package for Changepoint Analysis Journal of Statistical Software. 58, 3. 831 citations*
3.3. Haynes, K., Eckley, I.A. and Fearnhead, P. (2017) Computationally efficient changepoint detection for a range of penalties. Journal of Computational and Graphical Statistics. 26, 1, 134-143. 72 citations*
3.4. Haynes, K., Fearnhead, P. & Eckley, I.A. (2017) A computationally efficient nonparametric approach for changepoint detection. Statistics and Computing, 27, 1293–1305. 46 citations*
3.5. Fisch, A.T.M, Eckley, I.A. and Fearnhead, P. (2018) A linear time method for the detection of point and collective anomalies. arxiv.org/abs/1806.01947. 16 citations*
3.6. Jewell, S.W., Hocking, T.D., Fearnhead, P. and Witten, D.M. (2020) Fast nonconvex deconvolution of calcium imaging data. Biostatistics. 21, 709-726. 25 citations*

* Source: Google Scholar

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

The paragraphs below provide a summary of a range of non-academic impact exemplars of Lancaster's changepoint research, all established since 2015. These are corroborated by personal communications and, in some cases, also by published documents citing [3.1-3.6].

4.1 Applications in Industry

a) **Improving national telecom provider's operational performance:** BT's long-term collaboration with Lancaster has led to the uptake of our research to improve the UK's leading telecom provider's operational performance [5.1]. BT describe Lancaster's work as *"critical in the operation of BT's networks in order to manage the performance of a network that is driven by the behaviour of millions of users and applications. The scale of the network … means that it is impossible to use human eyes, except at the macro-scale where many smaller problems may be missed."* Methods developed within [3.1, 3.4 and 3.5] are embedded in a number of different operational areas, including BT's Internet Peering

Platform. This key platform connects all BT's internet traffic, i.e. millions of people daily, with other network and content providers. Since 2019, Lancaster University's change and anomaly detection tools have been used to monitor the Internet Peering Platform in real time, triggering anomalies to BT's network operations teams to help them monitor and assure the performance of this critical digital infrastructure. During the COVID pandemic, these methods have also been used to rapidly identify material changes in financial risk, enabling alerts to be generated quickly to support operations [5.1].

b) **Improving retail forecasting and efficiency:** Multinational groceries and general merchandise retailer Tesco reports that use of Lancaster's changepoint research has been crucial in enabling the Data Science team to improve their commercial forecasts of stocks held at online shopping distribution centres nationwide. When compared against their previous system, incorporation of [3.1] within their forecasting suite has directly resulted in substantial cost savings [5.2].

c) **Underpinning growth of healthcare tech startup:** Howz is a tech startup developing healthcare technology to support independent living for the elderly and those with dementia. Core to their business is the collection and analysis of thousands of signals each day from door, motion, plug-based and other activity sensors around the home. These are monitored to provide a warning of changes in activity pattern that help trigger follow-up healthcare actions. The automated partitioning of these signals using [3.1-3.3] has proved transformational for Howz. Examples include (i) Howz's partnership with EDF, established in 2018, selling Howz kits in partnership with EDF Energy; and (ii) in response to the COVID pandemic, working in partnership with the Surrey and Borders Partnership NHS Foundation Trust, to roll out the TIHM (Technology Integrated Health Management) service to 1000 homes for people with dementia and their carers. Such rapid developments have *"resulted in a step-change increase in data volumes being analysed by Howz".* Notably *"Lancaster's highly efficient, and trusted changepoint methods proved crucial in our ability to robustly up-scale our analytics and monitoring in response to this opportunity"* [5.3].

d) **Revenue growth through improved targeting:** Peak Ltd is a rapidly growing UK-based AI consultancy who have a variety of connections with Lancaster's changepoint group. In recent work with a FTSE-listed tool and equipment hire company, with over 300 depots nationwide and thousands of products, Peak were able to achieve significant cost reductions by use of [3.1] within an improved inventory management system. *"The change-point-enhanced forecasts play a crucial role"* in the approach developed, and are now used throughout the clients' UK operations. This has led to a *"fourfold increase in their Return on Capital Employed (ROCE), a key business metric"* [5.4].

4.2 Applications in Public Sector and Non-governmental organisations

a) **Supporting NASA Safety protocols:** Lancaster's work [3.1] forms part of the recently published NASA test protocol for quantifying inspired carbon dioxide (CO2) within extravehicular and launch, entry, survival pressure suits [5.5a]. Specifically [3.1] is used to identify the start and end of the respiratory CO2 waveform. As the test protocol describes, [3.1] was chosen *"given the number of change points, an optimization algorithm based on dynamic programming with early abandonment was used"*. This test procedure now underpins the development of current and future space suits used by NASA to ensure that they are safe for use [5.5b].

b) **Environmental data quality monitoring and improvement:** The UK Centre for Ecology and Hydrology (UKCEH) is an independent not-for-profit research institute responsible for highly regarded, long-term environmental monitoring programmes. UKCEH has been using [3.1] to ensure that the "*underpinning long term data remains fit for purpose, free from errors and any artificially induced effects"* [5.6]. Previous attempts to apply automated quality assurance have been hampered due to the volume of data and computational costs. However, PELT [3.1] *"has been a game changer",* to the extent that the approach has now been embedded within a standardised quality assurance workflow. The resulting approach is realising savings estimated to be in excess of GBP100,000 per annum, and helping to provide the UK with rigorously assured environmental monitoring data sets.

[3.1] has also been used by UKCEH to explore potential tipping points and regime shifts across its whole portfolio of long term monitoring data sets. For example, it has been used to support the long-term monitoring of lake water quality within the English Lake District, a UNESCO World Heritage site. *"[I]dentifying regime shifts is pivotal both for ecosystem and economic value as the lakes are a focal point for tourism in the area, attracting around 15 million visitors every year"* [5.6]. Here, [3.1] *"was used to assess regime shifts in water quality and hypothesised drivers in the Cumbrian lakes data having accounted for factors such as seasonality and changes in protocols."* Analyses using [3.1] identified key changes in phosphorus levels across a number of lakes, demonstrating increases in concentrations in the mid to late 1980s, consistent with previous analyses. Subsequently, a number of remediation measures including grants to reduce runoff from farms and improved water treatment facilities have been introduced [5.6].

c) **Quantitative monitoring of international fish stocks:** Oceana is a nonprofit ocean conservation organization focused on influencing specific policy decisions on the national level to preserve and restore the world's oceans. Contract research undertaken by Lancaster University in 2016 on Oceana's behalf used [3.1, 3.3] to study historic international fish stock abundance levels to identify those stocks in decline or recovery, and the times at which certain changepoints occurred. This work has been used to explore the effects of different interventions and policy measures on a regional and species basis. In particular, dates of known interventions have been compared against estimates of when changepoints occurred in various species stock levels, with a view to informing future policy directions [5.7].

4.3 Implementations of Changepoint Research in Software

a) **Open-source software:** the work in [3.1, 3.2] has been incorporated within a very popular open-source R package, changepoint, which has been in continuous development since 2010. The package has been downloaded over 250,000 times during the REF assessment period [5.8]. Quantifying the value of the impact generated by open-source implementations of our algorithms is not possible. However, a sense of scale can perhaps be obtained from the diversity and reach of the known user community for this package, which includes those working in Government agencies (national and international); tech start-ups; FTSE100 and NASDAQ corporations.

b) **MATLAB:** is a long-established commercial computing environment and programming language developed by MathWorks, with a community of over 4,000,000 users in 185 countries and installations at over 100,000 business, government and university sites [5.9a]. The direct impact of [3.1] for MathWorks and its user community began with Release 2016a, when MathWorks added the function findchangepts to its Signal Processing Toolbox, including algorithms from [3.1], "*which made it possible to implement change point detection in a way that's not prohibitively slow"* [5.9b]. The importance of the work was further underlined in Release 2017b when MathWorks implemented [3.1] within a new routine called ischange in core MATLAB – *"a reflection of its importance and wide applicability to our user community*" [5.9b].

c) **The Numerical Algorithms Group Ltd (NAG):** Established in 1970, NAG is a not-for-profit company providing expertise in numerical engineering via computational software, consulting and high-performance computing services. An example of [3.1]'s direct impact for NAG can be seen by the PELT algorithm's inclusion within the NAG toolbox since Mark 25's release in 2015 [5.10a]. Several sectors are known to have engaged with NAG's changepoint work including those working in investment banking and consumer marketing [5.10b].

d) **Allen Institute Application Program Interface.** The Allen Institute for Brain Science, based in Seattle, was established to accelerate understanding of how the human brain works, with the potential for ground-breaking health impacts. They use the algorithm FastLZeroSpikeInference [3.6] as a key part of the downstream analysis to deconvolve calcium imaging traces from the brain, with results from our algorithm on approximately 60,000 neurons being made publicly available via the web and their application program interface [5.11a]. Impact from this resource includes work by a pharmaceutical company on

how a weight loss drug acts on the brain and helping to find new ways to treat drug addiction [5.11b].

5. Sources to corroborate the impact

5.1 Letter of evidence from Senior Research Manager, BT in support of Impact 4.1a.
5.2 Letter of evidence from Data Science Manager – R&D, Tesco in support of Impact 4.1b.
5.3 Letter of evidence from Chief Operating Officer, Howz in support of Impact 4.1c.
5.4 Letter of evidence from Data Science Team Leader, Peak Ltd. in support of Impact 4.1d.
5.5 (a) NASA Technical Report in Support of Impact 4.2a: (Link to resources); **(b)** Corroboration of impacts claimed in 4.2a provided by NASA (contact details provided).
5.6 Letter of evidence from Head of Statistics, UK Centre for Ecology and Hydrology in support of Impact 4.2b.
5.7 Corroboration from Oceania confirming impacts claimed in 4.2c (contact details provided).
5.8 Download data obtained using the cranlogs R package (Csárdi, 2019).
5.9 (a) MathWorks company overview document: (Link to resources); **(b)** Letter of evidence from VP Marketing, MathWorks in support of Impact 4.3b.
5.10 (a) NAG website evidencing Impact 4.3c; **(b)** Corroboration of impacts claimed in 4.3c provided by NAG (contact details provided).
5.11 (a) Release Notes for data in the Allen Brain Atlas in support of Impact 4.3d: (Link to resources); **(b)** Allen Institute website describing impact narratives associated with 4.3d: (Link to resources).