

RCUK Statistics and Applied Probability Landscape Event

Report of the RCUK Statistics and Applied Probability Landscape Event held at
the DoubleTree Hilton, Bristol on the 26th September 2016

Written by Hannah Pearson & Anke Davis (EPSRC)



Executive Summary

Aims

The aim of the day was to bring together representatives from the Statistics and Applied Probability community to investigate the connection between fundamental Statistics research and its applications.

Outcomes

There is significant world-leading Statistics and Applied Probability (SAP) research in the UK and strong industry, governmental and end user demand for skilled SAP researchers and research. In addition the UK supports several large institutions to undertake research in topics related very directly to Statistics and Applied Probability. There has also been increased support for training and fellowships for researchers in Statistics and Applied Probability. Although there has been considerable growth in the portfolio there are still significant concerns that there remains a growing un-met need for skilled statisticians and applied probabilists in academia and business, industry and government.

There are significant opportunities for Statistics and Applied Probability research to contribute to and lead in current research opportunities particularly building on connections between uncertainty, big data and modelling and inference with application areas.

Next steps

It has been recommended that the Statistics and Applied Probability community work together to promote the benefits of multi-disciplinary working particularly to link fundamental and application aspects of Statistics and Applied Probability, exploiting opportunities arising from recent investments.

The research councils have been working together in monitoring and highlighting the opportunities for Statistics and Applied Probability researchers in standard research funding modes as well as through additional funding opportunities. It will be investigated how more collaborative research with application areas and industry can be encouraged. The developments of the people pipeline and the links to the community of The Alan Turing Institute will be monitored.

Introduction

Statistics and Applied Probability is an area of strength for the UK and one which is connected across all the research council portfolios and the wider research landscape.

The UK has an international reputation for expertise in Statistics and applied Probability Statistics and their applications. There are strong connections between Statistics and Applied Probability to an array of applications in science, industry, business and government providing economic, industrial and societal impact in a range of applications and sectors, for example healthcare, finance and the environment. As such Statistics and Applied Probability is an important research area which is core to and supports a number of other research areas, including The Alan Turing Institute for data science. This importance is reflected in the 2012 Deloitte report (Measuring the Economic Benefits of Mathematical Science Research in the UK), 2014 EPSRC Statistics and Applied Probability theme day, and the 2014 Statistics Strategy from the Office of National Statistics.

Alongside the importance of the research area, previous reviews have highlighted that although the research area has grown there are still significant concerns over capacity and the need for continued growth. This is particularly important in light of the relevance of this research field to significant real-life challenges including but not limited to the growth in data science (big data), the need for uncertainty quantification, healthcare and health diagnostics, finance and business, economics and the environment. These areas are now being characterised by the rapidly changing nature of the data streams being generated and from which inferences and knowledge must be extracted.

Background

In 2014 EPSRC held a review day as part of the delivery plan with specific focus on changes since the international Review and to consider the current state of Statistics and Applied Probability and opportunities and issues still to be addressed. Following the review day EPSRC produced an action plan for the research area and under the last delivery plan, the EPSRC's Statistics and Applied Probability portfolio grew and is currently approximately 2% of the whole portfolio. Alongside the growth of Statistics and Applied Probability within EPSRC, all research councils have supported and contributed to the landscape with application relevant research.

Biotechnology and Biological Sciences Research Council (BBSRC)

Statistics comes under the Exploiting New Ways of working strategy sector at BBSRC and statistical approaches broadly underpin many of the analyses of research data within BBSRC funded awards. Cutting-edge bioscience is critically dependent on the availability of modern research infrastructure and the adoption of new ways of working. BBSRC supports these through research grants, and training students and researchers in the latest skills and techniques. This helps to ensure that the UK bioscience base has the capability to continue to drive forward global scientific research. BBSRC's vulnerable skills survey identified maths, Statistics and computation as among the skills areas where the UK bioscience base was vulnerable, and these areas are now being supported through in vivo Skills Awards and Strategic Training Awards for Research Skills.

Engineering and Physical Science Research Council (EPSRC)

Statistics and Applied Probability is one of 115 research areas within EPSRC. Although it sits within the Mathematical Sciences theme it has cross cutting relevance across the whole of EPSRC's portfolio with strong links to the ICT, Digital Economies and Healthcare Technologies themes. During the last delivery plan the Statistics and Applied Probability research area was considered a priority with the aim to grow the relative size of the portfolio. As such a number of investments have been made including The Alan Turing Institute and a number of programme grants. EPSRC supports research at a range of career stages with fellowships encouraged at all career stages, Centres for Doctoral Training as well as standard research grants within the area of Statistics and Applied Probability. The current portfolio consists of 87 grants totalling approximately 2% of the total EPSRC portfolio.

The portfolio in Statistics and Applied Probability covers research in statistical methodology and the development of new probabilistic techniques inspired by applications. The portfolio covers a wide spectrum of research in stochastic and probabilistic modelling and inference in stochastic systems.

Economic and Social Research Council (ESRC)

Statistical literacy is vital to the social sciences as it allows researchers to understand and interpret numerical data and undertake analysis to answer questions relevant to the social environment. The use of Statistics in the social sciences is increasingly widespread and is relevant to many areas of ESRC's remit. Quantitative research methods are particularly linked to statistical analysis of data and ESRC has invested heavily in this area in recent years. We aim to both build capability as well as develop innovative methodologies needed to address new research challenges.

Medical Research Council (MRC)

The advance and implementation of statistical techniques is a thread which runs through the entire MRC portfolio and strategy. All biomedical and health research is underpinned by the appropriate design and application of data collection and analysis.

MRC has invested £100m over the past 5 years in informatics research (computational information processing) through the Farr Institute of Health Informatics (analysing health records/system data), a series of Medical Bioinformatics Awards (analysing basic biomedical science big data) and investment in Genomics England's 100,000 Genomes project (analysing the genome), all of which are built around advances in Statistics. MRC has now committed a further £37m to establishing a new national informatics institute.

MRC also supports advances in Statistics and Applied Probability and their application in biomedical and health sciences through many of our research Units, such as the Clinical Trials Unit, the Epidemiology Unit and the Integrative Epidemiology Unit. MRC has a long history of supporting statistical science; MRC's £3m PA BioStatistics Unit is over 100 years old.

MRC continues to explore new avenues for advancing statistical science in medical research – the £4.5m PA Methodology Research Programme panel (co-funded by NIHR, <http://www.mrc.ac.uk/funding/browse/methodology-research-programme/>) supports open, response-mode projects in this field, and we look to further grow this field through the next generation - capacity building in quantitative skills has been identified as a key priority, through schemes such as the Skills Development Fellowship (<http://www.mrc.ac.uk/skills-careers/fellowships/skills-development-fellowships/>). These areas will remain a priority as biomedical research evolves and the need to handle large complex data sets and to integrate data from different sources becomes ever greater.

Natural Environment Research Council (NERC)

Statistics, mathematical modelling, computer science, environmental informatics and related disciplines underpin many areas of environmental science, and much of the research we support as part of our Discovery Science and Strategic Research portfolios. These quantitative skills are recognised as skills gaps in environmental researchers and NERC has targeted these areas in previous calls for Advanced Training Short Courses for PhDs and postdoctoral researchers. NERC has recently funded a Centre for Doctoral Training (CDT) in Risk and Mitigation using Big Data and the current CDT call is for Modelling and Quantitative Skills in Ecology and Evolution. NERC is also supporting six fellowships in Environmental Omics.

Science and Technology Facilities Council (STFC)

STFC is particularly interested in the development of new techniques for handling and analysing very large and complex datasets. Facilities such as the Planck and Gaia satellites, the Large Hadron Collider at CERN, and the forthcoming Large Synoptic Survey Telescope and Square Kilometre Array radio telescope are producing, or will produce, huge amounts of data, driving the need for better and faster ways of dealing with such high volumes and high rates of data flow. Similarly, ever increasing data volume and complexity from the experiments at the national and international facilities are driving new techniques and capabilities.

As well as funding the UK groups involved in developing the data processing techniques associated with these facilities, STFC also funds the associated scientific analysis of the data and provides advanced computing infrastructure, such as the DiRAC supercomputing facility. STFC also hosts the JASMIN computing system for NERC science and the STFC Hartree Centre provides supercomputing and data analytical services in support of a wide range of scientific and industrial requirements.

To help develop data analytical skills STFC is in the process of setting up a Centre for Doctoral Training in data intensive science. The centre, scheduled to start in autumn 2017, will encourage interdisciplinary approaches and close collaboration between academic and industrial partners.

Aims and Objectives of the Theme Day

The aim of this Statistics and Applied Probability Landscape Event was to build upon the International Review and the 2014 Review Day developing connections between the fundamental and applied aspects of the portfolio and looking forward to the future of the portfolio at the start of a new delivery plan.

The aim of the day was to bring together representatives from the Statistics and Applied Probability community to investigate the connection between fundamental Statistics research and its applications.

The specific objectives may be summarised as:

- To gather a wider picture of the Statistics and Applied Probability landscape
- To understand the connections between fundamental researchers and researchers into the applications of statistical techniques. Investigating whether statistical advances developed in one field are informing and building on techniques developed in other fields.
- To investigate how we currently linking across different communities and funding streams
- To understand the challenges and opportunities of inter-disciplinary research
- To look at future opportunities for the Statistics and Applied Probability portfolio

Attendance at the event

There were 48 delegates in attendance at the event from a range of disciplines including mathematical sciences, environmental sciences, ICT, economics, social sciences, medical and biological sciences. The full attendance list is at Annex 1.

STAPLE Outcomes

The event had four key sessions (for the full agenda, see Annex 2) building on an overview of the landscape from the Royal Statistical Society. The first session developed a SWOT analysis of UK Statistics and Applied Probability. These next two sessions investigated the detail of the key challenges and opportunities to enable a range of stakeholders to consider actions required and the impact of these actions on maximising the benefit to the future of UK Statistics and Applied Probability. The final session was to scope out some real-world challenges that require multi-disciplinary research with a strong need for Statistics and Applied Probability researchers.

As well as the discussion sessions there were some short application-focused presentations. These provided an overview and examples of the diversity and importance of the research area to other disciplines as well as connections between different disciplines.

2.1 Strengths, Weaknesses, Opportunities and Threats

The aim of the session was to understand the Statistics and Applied Probability landscape in terms of strengths, weaknesses, opportunities and threats.

Activity

In order to understand the Statistics and Applied Probability landscape, an initial analysis was undertaken in groups. To begin with, participants were asked to focus on what is strong and weak in the UK, including areas of research, university departments, industry strength, funding related issues, capacity related issues and the people pipeline. With opportunities and threats it was considered to be important to think more broadly than just the UK.

Outcomes

Participants highlighted the application of Statistics in many areas as well as the strong international reputation it has. The importance of SAP to government, industry and end users was acknowledged although it was also recognised that engagement in some areas could be strengthened. Furthermore, the flow of people is of concern with expertise often being concentrated into a few strong groups. Opportunities identified included the Global Challenges Research Fund, improved engagement with industry and increasing needs for multi-disciplinary working. Threats identified include shortages in the people pipeline and endemic misconceptions and often misuse of Statistics.

2.2 Interdisciplinary connections

The aim of the session was to think about current and potential connections across SAP portfolio as well as connections to other disciplines. Participants in groups identified real-world challenges and then explored the questions within these challenges and the research areas and researchers required to address such questions.

Activity

In order to start to investigate the current connections and potential scope for new connections, the groups created a short list of real-life challenge areas which require Statistics and Applied Probability research.

Outcome

Participants identified eighteen different challenges in the areas of health and the environment, energy and transport, and more generic issues on data and modelling that are listed in Annex 3. Six challenges were prioritised by the participants and chosen for further elaboration.

How are 'models' and the world related?

This challenge encompasses model design, including dealing with uncertainty, robustness, model verification and assumptions and how a model is a reflection of the world. Specific examples addressing social, environmental, health and political questions were identified.

Ageing population and modern living

This challenge includes the development of assisted living technologies, looking at changing work conditions and patterns, including consideration of health, demography, architecture/design of buildings, cities and transport. The complexity of the challenge lies in the interaction of several sub-challenges occurring at the same time, as well as considering associated ethical issues. Personalised medicine and care are some of the research challenges in the area, where SAP has a key role to play. Design of housing and transport systems, urbanisation, and sustainability which are increasingly making use of new data streams from new sensors, including real time activity recording were identified.

Energy: Statistical modelling of supply, demand & behaviour

This challenge considered the need for the optimisation of consumer energy behaviour and the modelling of energy supply networks. Underlying questions involve the understanding and modelling effects of climate change, the need to reduce pollution and changing consumer behaviour on energy demand and supply. Statistical modelling of individual level data and individual decisions are some of the specific research challenges associated with this theme.

Trees and Networks: Modelling with data

Processing and interpreting random data for example in telecommunications and biology constitute to this challenge. With increasing quantity and quality of data, the challenges of this area lie in developing systems to make best use of such data in all application areas.

Combining different sources of information

New technologies are delivering many different data streams, which need to be linked and connected within an application area, examples might include in the health care area (for example individual sensors) , combined with more conventional medical exams, to biochemical and genetic data to deliver improved individual health care.

Challenges of this nature will require multi-disciplinary teams. Of importance is the careful consideration of to the sampling frames, the data collection processes, and whether we are dealing with designed or observational data.

Scalable Decision Making under Uncertainty

A generic challenge that was identified was scalable decision making under uncertainty, of very wide applicability, for example in disaster response and in medical decision making. One aspect is the need for real-time applications decision making, based on incomplete information, which may need the use of complex models. Underlying questions involve data asymmetry, sharing, privacy, transparency and ethics. Specifically, model robustness and refinement are important.

2.3 Challenges within the Statistics and Applied Probability portfolio

The aim of the session was to look forward to the future of Statistics and Applied Probability research and what challenges and actions need to be undertaken to ensure a healthy SAP community within the UK.

Activity

In order to investigate the impact of the challenges prioritised from the first session a more detailed analysis was considered in groups. Participants were asked to detail the challenge and provide potential approaches that would be needed to address the challenge. Detail about the benefits of tackling this challenge as well as considering all potential stakeholders that may be involved in the challenge were also captured.

Outcomes

The participants identified the importance of multi-disciplinary working in addressing the challenges. They also highlighted the need for media and communication training and for increasing awareness of SAP at school level to increase the profile of the area. Despite the high level of demand for trained statisticians, expertise is missing at all levels of the career spectrum and a need for more postgraduate training was highlighted. The growth of Data Science was identified as an opportunity but that the SAP community needed to assert more ownership in this area where SAP is a fundamental component. Further challenges include supporting joint working with industry where different terminology, intellectual property rights and established practices can be barriers. It was proposed to highlight more success stories to show the successful partnerships between industry and academy.

2.4 Future of the Statistics and Applied Probability portfolio

The aim of the session was to consider the opportunities and future for the Statistics and Applied Probability portfolio and the role of all the research councils in securing that future.

Activity

In order to investigate the potential and impact of the opportunities prioritised from the first session a more detailed analysis was considered in groups. Participants were asked to detail the opportunity, consider the impact within other research domains, industry and users. The delegates

were also asked to consider actions that could be taken to maximise the opportunity and who needs to be involved in pursuing the actions to increase the impact.

Outcomes

The Global Challenges Research Fund was highlighted as an opportunity to address research questions including uncertainty with an emphasis on the development of multi-disciplinary teams. For engagement with industry, gradual development and expansion of projects started at for example the graduate level was suggested, followed by, for example, seed funding to continue the collaborations. The increasing demand for statistical skills requires an increasing need to teach Statistics at the undergraduate level and beyond. Exposure of students to companies during their final year projects would provide an additional way of highlighting the demand for statisticians. The growth of Data Science provides opportunities for statisticians to be involved in multi-disciplinary teams to address the associated research questions.

Overall conclusions

There is significant world-leading Statistics and Applied Probability research in the UK and strong industry, governmental and end user demand for SAP researchers and research. In addition the UK supports some large institutions for research in topics related very directly to Statistics and Applied Probability. There has also been increased support for training and fellowships for researchers in Statistics and Applied Probability research.

Although there has been considerable growth for the Statistics and Applied Probability research area there are still considerable concerns over capacity in particularly at all career stages. This concern has been voiced by the 2010 International Review and confirmed by the community at the 2014 Review Day. EPSRC recently published its new research area rationales and continues to be committed to grow the SAP portfolio as part of its Balancing Capability strategy.

There are a number of interfaces between Statistics and Applied Probability with other research and application areas across all research councils. Some of these are well developed but there is still work to be done to improve the multi-disciplinary working which is critical for meeting challenges and current emerging research areas. There is concern within the community about how multi-disciplinary proposals are supported and funded across all the research councils. It would be beneficial to have a steer from the research councils about how to approach multi-disciplinary research and proposals. In conjunction with this it was highlighted that the community could be critical of each other particularly for multi-disciplinary proposals where reviewers did not feel they add expertise in all areas. There is a need for the community to support itself within peer review particularly for proposals which crossed interfaces within and between the research councils.

Statistics and Applied Probability is underpinning a range of research topics and subjects but the importance is not always well communicated, reflected in the perception of Statistics and Applied Probability not always being recognised outside of the community. It is important that the community and other stakeholders work to demonstrate the importance of Statistics and Applied Probability researchers for example through increased public engagement.

Statisticians are important team members in cross-disciplinary research teams and are used to collaborating broadly in a host of disciplines. It is important to maximise on opportunities presented from current topics including big data allowing for statisticians to define their role and definition within these big topics.

Statisticians are typically located within mathematical departments but increasingly scientists with statistical expertise are members of other academic teams and departments highlighting the importance of Statistics in a range of application areas. It is important that future statisticians are trained to have application skills as well as fundamental statistical methodologies. As a result it is important that statisticians involved in training are able to demonstrate the benefits of fundamental Statistics as well as the application benefits and potential impacts.

It is important that the community take advantage of the current opportunities for research council funding, including the Global Challenges Research Fund which focusses on research for overseas development (ODA). There are significant opportunities for Statistics and Applied Probability research to contribute to this stream of research particularly building on connections with application areas and real-world topics such as uncertainty and modelling.

Next steps and Recommendations

The research councils will work together to collate the outputs from this meeting and put together an action plan for this research area.

In addition this input will feed into our thoughts on the area of Statistics and Applied Probability and has been considered as part of EPSRC's Balancing Capability strategy as published in mid-February 2017.

Below is a list of suggested actions that the research councils and the community may look at over the next few years:

- Investigate the processes of cross council and multi-disciplinary working and proposals.
- Statistics and Applied Probability community to work together to promote the benefits of multi-disciplinary working particularly linking between fundamental and application aspects of Statistics and Applied Probability. This is especially important with regards to the peer review system.
- Consider where Statistics and Applied Probability fit into the industrial strategy
- Investigate the current and expected balance between the Statistics and Applied Probability aspects of the portfolio.
- Explore how the research councils can encourage more collaborative research in Statistics and Applied Probability and application areas.
- Explore how the research councils can encourage more collaborative research in Statistics and Applied Probability with industry.
- Showcase success stories of partnerships between academia and industry.
- Monitor the effects of The Alan Turing Institute and develop links between the SAP community and the institute and other large investments.
- Look into current and other potential mechanisms for developing and supporting training and the people pipeline.
- Work together with the community (in Statistics and other areas of mathematical sciences) to exploit big data opportunities, building on the excitement and potential that researchers are experiencing.

Annex 1: Attendance list

FIRST NAME	SURNAME	AFFILIATION
Niall	Adams	Imperial College London
Deborah	Ashby	Imperial College London
Ron	Bates	Rolls-Royce
Nigel	Birch	EPSRC
Veronica	Bowman	DSTL
Peter	Challenor	University of Exeter
Jonathon	Chambers	University of Newcastle
Andrew	Chesher	University College London and CeMMAP
Paul	Clarke	ESRC MiSoC Research Centre, University of Essex
David	Crosby	MRC
Idris	Eckley	Lancaster University
Alison	Etheridge	University of Oxford
Glenn	Everett	Office of National Statistics
Jonathan	Gair	BioSS
Steven	Gilmour	Kings College London
Mark	Girolami	University College London
Michael	Goldstein	University of Durham
Philippa	Hemmings	EPSRC
Daniela	Hensen	BBSRC
Chris	Holmes	University of Oxford
Adrian	Hilton	University of Surrey
Christopher	Jennison	University of Bath
Andreas	Kyprianou	University of Bath
Joanna	Lake	ESRC
Duncan	Lee	University of Glasgow
Malwina	Luczak	Queen Mary, University of London
Adele	Marshall	Belfast University
Andrew	Mead	Rothamsted University
Alex	Mijatovic	Kings College London
Katharine	Moore	EPSRC
John	Moriarty	Queen Mary, London
Guy	Nason	University of Bristol
Hannah	Pearson	EPSRC
Chris	Quince	University of Warwick
Matthew	Roberts	University of Bath
Gareth	Roberts	University of Warwick
Marian	Scott	University of Glasgow
Mark	Strong	Sheffield University
Jonathan	Tawn	Lancaster University
Charles	Taylor	University of Leeds
John	Terry	Exeter University
Almut	Veraart	Imperial College London
David	Westhead	University of Leeds
Darren	Wilkinson	Newcastle University
Daniel	Williamson	University of Exeter
Simon	Wood	University of Bristol
David	Woods	University of Southampton
Nikolaos	Zygouras	University of Warwick

Annex 2: Agenda

RCUK Statistics and Applied Probability Landscape Event (STAPLE)

September 26th 2016, Bristol – Double Tree Hilton

AGENDA

09.45	<i>Arrival, registration and Coffee</i>
10.15	Welcome and Introduction
10.20	Overview of Statistics landscape
10.45	What is Statistics?
11.00	Discussion session - Research area analysis
12.00	Opportunities and threats prioritisation
12.15	<i>Lunch</i>
13.00	Application Statistics overview presentations
13.15	Discussion session - Interdisciplinary connections within the research area and between researchers
14.00	Discussion session - Challenges within the Statistics portfolio in particularly the challenges within cross disciplinary working
14.40	<i>Tea/Coffee</i>
15.00	Discussion session - Opportunities and the future of the research area
16.00	Summary and Final discussion
16.30	<i>Close</i>

Annex 3: Outputs from sessions 1 – 4

Session 1 – Research area (SWOT) Analysis

Common themes for strengths and weaknesses for the research area are summarised below. In addition, following from the SWOT analysis the key opportunities and threats were chosen and clustered. They were agreed as follows Opportunities:

STRENGTHS

- Underpins a lot of areas
- Statistics has application in many areas including medical/economics/environmental
- There is a strong international reputation
- There is an acknowledgement of the importance of Statistics within government, industry and end users
- Statistics is rooted within mathematics in the UK
- Statistics has impact including the ability to model complex systems providing skills and techniques for real world problems.

WEAKNESSES

- Some interfaces such as data science/machine learning/bio-informatics
- People pipeline
- Concentration into a few strong groups
- Perception of Statistics and it's definition (particularly in the hype and involvement in data science)
- Lack of engagement with industry in some areas

OPPORTUNITIES

- Global Challenges Research Fund
- Engaging with Industry
- Meeting hugely growing demand
- Multi-disciplinary working
- PhD training
- Data Science

THREATS:

- Raising the profile of Statistics/ dealing with the misconceptions of Statistics
- People pipeline
- Data science
- Statistics community gets left behind
- Multi-disciplinary working
- Working with industry

Session 2 – Interdisciplinary connections within the research area

GRAND CHALLENGES

- Ageing population
- Cyber security
- Spread of disease
- Obesity
- Dementia
- Climate change
- Food Security
- Sustainable intensification
- Driverless cars
- Energy supply
- How are models and the world related
- Scalable Decision making under uncertainty
- Energy statistical modelling of supply, demand and behaviour
- Trees and Networks – Modelling with data
- Combining different sources of data
- Sampling emerging populations
- Handling large-dimensional heterogeneous data
- Small and expensive data

The challenges that were developed further are shown below:

Name of Grand Challenge:	How are 'models' and the world related?	Ageing population	Energy: Statistical modelling of supply, demand & behaviour
What does the grand challenge involve?	Two ways: <ul style="list-style-type: none"> - Examples during the question - Methodology - with very many examples How to design models, approximation and evaluation, uncertainty quantifications, scales of operation, robustness, verification, predictability, stochastically data and best models (sampling) assumptions, model rejection	<ul style="list-style-type: none"> - Health, demography, insurance/pensions, building design/architecture/sustainable cities, social environment, transport - Medics, architects, applied maths, statisticians, economics, imaging scientists, engineers/manufacturers - Assisted living technologies/devices - Changing work conditions/patterns - Intergenerational issues - Leisure/entertainment 	<ul style="list-style-type: none"> - Optimising consumer energy behaviour: model, predict, modify - Big data modelling, network models, econometric models - Industries: Electricity, gas, transport, energy storage - Statistical computing
What are the other underlying questions involved?	<ul style="list-style-type: none"> - Social and political – communication - Safety (introduce constraints) - Timeliness 	<ul style="list-style-type: none"> - Social care → how to look after people - Health in diseases of aging e.g. dementia/arthritis - ODA → how the effects are different in in different countries - Climate change – impact of extreme weather on aging populations - Ethical issues 	<ul style="list-style-type: none"> - Controlling climate change - Reducing other air pollution - How to shift behaviour - Health benefits of 'active transport'
What are the specific research challenges in this area?	<ul style="list-style-type: none"> - Big models, scaling from small models, representability - Sparse data - Poorly understood processes - Uncertainty - Inter-linked models - Systems of systems 	<ul style="list-style-type: none"> - Predicting trend in life expectancy - Personalised care/medicine - Drug developments – chronic/managing rather than curing (complexity of studies, patients prone to multiple conditions) - Learning from data streams – non-intrusive data sources - Manufacturing/engineering of devices – metrics to produce & measure them - Robotic careers - Experimental design of regimes to see what works best 	<ul style="list-style-type: none"> - Statistical modelling of individual level data (e.g. smart meter data) - Modelling of individual decisions - Algorithms for working with these models - Move to real time modelling
Examples of research and connections already relevant to this Grand Challenge?	Food security, natural hazards, energy, climate, water security, health/disease	<ul style="list-style-type: none"> - Personalised medicine - Housing & sustainable architecture - Internet of Things - Sport science: Fitbit & higher end devices that show activity - Privacy awareness of analytical methods: security, what is stored, who knows what impact this can have? - Analysing data streams 	<ul style="list-style-type: none"> - Energy theme - Electricity local forecasting community - Demand modelling

Name of Grand Challenge:	Trees and Networks: Modelling with Data	Combining different sources of information	Scalable Decision Making under Uncertainty
What does the grand challenge involve?	<p>Application areas:</p> <ul style="list-style-type: none"> - Telecommunication, cloud media, social networks, consumer data - Genetics, ecology, Interpreting 'omics data - Combinatorics - Terrorism - Transport - Finance <p>Research Areas:</p> <ul style="list-style-type: none"> - Random processes or Random media - Network modelling - Random networks - Random geometrics - Graphical models <p>Techniques:</p> <ul style="list-style-type: none"> - Processes on random media - Combinatorics: inference on network - Graph theory - Bayesian Networks 	<ul style="list-style-type: none"> - Chemical engineering - Mining engineering - Cyber security - Network battle space - Users defined by examples <p>Multi-disciplinary team is essential: industrial involvement as appropriate</p>	<p>Applications:</p> <ul style="list-style-type: none"> - Disaster response, biomedical – integrating heterogeneous/big data, web scraping, real-time applications <p>Communication challenge</p> <p>Techniques:</p> <ul style="list-style-type: none"> - modelling, optimisation, machine learning, inference + extrapolation, hierarchical statistical modelling, model uncertainty + stability <p>Users:</p> <p>government agencies, companies, individuals, supermarkets, internet companies, finance sector, infrastructure</p>
What are the other underlying questions involved?	<ul style="list-style-type: none"> - As the quantity and depth/variety of data increase, how can we get more out of it, access all application areas? - Making network data more useful - Identify trends 	<ul style="list-style-type: none"> - Informed by specific example above 	<ul style="list-style-type: none"> - How to value heterogeneous outcomes and how to quantify utility of outcomes - Data asymmetry, data sharing/privacy, data ethics, transparency - Having 'common language' across disciplines - Bias caused by 'digital divide' - Changing existing practice - What should be optimised (needs of many/few)
What are the specific research challenges in this area?	<ul style="list-style-type: none"> - Inference – given partial knowledge of the system, plus noise, how can we get insight from data in specific applications/scenarios - Mathematical structure vs statistical applicability - Computation traceability 	<ul style="list-style-type: none"> - How do you collect data & design the experiment - Computation & algorithms - Dimensionality - Intractable likelihood problems 	<ul style="list-style-type: none"> - Computation and scalable computation (parallelisation, distribution) - Dimensionality and heterogeneity - Model refinement - Model robustness under misspecification - Interface between e.g. machine learning & Applied Probability - Repurposing of mechanistic models & integration into e.g. stochastics
Examples of research and connections already relevant to this Grand Challenge?	<ul style="list-style-type: none"> - Complexity Scientists' approach to networks - Industrial handling of network data e.g. marketing agencies - Tree approximations of networks - Stochastic geometry - Processing random media 	<ul style="list-style-type: none"> - Programme Grant EQUIP 	<ul style="list-style-type: none"> - Formulating multi-objective utility functions - Communication about risk in cancer screening - Met office joint modelling from heterogeneous data - Current NERC call on risk modelling - Work in systemic risk in finance - Stratified medicine – integrates genetics

Session 3 - Challenges within the Statistics portfolio in particularly the challenges within cross disciplinary working

Multi-disciplinary working

- Challenge detail
 - Funding processes are not configured to support multi-disciplinary working
 - More aspects to a proposal are not supported at peer review
- Potential solutions
 - Calibrate raw scores in panels
 - Have a statistical panel
 - Train statisticians to be more positive in their reviews
 - Allow and value dual council submissions (as equal partners)
 - Have multi-disciplinary funding streams

Raising the profile of Statistics/ dealing with the misconceptions of Statistics

- Challenge detail
 - Hot topic areas have become popular (Uncertainty, big data, machine learning) but it is not evident about the importance of Statistics within these subjects
 - What is data science and what is the need for statisticians within it?
 - Misconception of what Statistics is and what it can do
 - Recognition of the benefits of Statistics.
- Potential solutions
 - Training statisticians in communication
 - Better training of non-statisticians
 - Better teaching in schools
 - Work successful with industry and then make it known.
 - Broader training of statisticians

People pipeline

- Challenge detail
 - There are not enough people at all levels in all places. This has got worse since the last review day despite the increase in CDTs and the increase of funding within Statistics and Applied Probability portfolio
 - Need for more PhD studentships
 - Loosing people to industry
 - Need to have statisticians in academia teaching connections and applications of Statistics
 - Diversity – how to retain women?
- Potential solutions
 - Longer term guaranteed funding post PhD
 - More MSc funding
 - More competitive salaries to compete with industry
 - Deeper collaboration with industry
 - Better teaching in schools
 - More relevant/applied sciences
 - Examples that capture students imagination
 - Sustained relationship with mathematicians

Data science

- Challenge detail
 - Communicating results
 - Pressure from industry on salaries and opportunities

- Data Science and machine learning is moving into the Statistics space. Where are the statisticians on these projects?
- Potential solutions
 - Take ownership
 - Use data science as a useful (but limited) tool within Statistics

Stats community gets left behind

- Challenge detail
 - Demand led – responding to a need
 - More people doing Statistics outside the community mean a fragmented community
 - Slow publication timelines
- Potential solutions
 - Communication – develop relationships and collaborations
 - Training – the curriculum needs to be updated
 - Strategic thinking and planning including a recommendation to ‘grow’ Statistics within the research councils.

Working with industry

- Challenge detail
 - Changing established practises
 - There is a need for a common language
 - Find common basis for partnerships development
 - Resolve IP issues
 - Joint funding does not have an easy method to be supported
- Potential solutions
 - Success stories are needed – build a track record of success
 - Develop new joint funding models to help underpin increased academic research motivated by industry
 - Allow time for funding to be secured from industry as it often needs a business case.

Session 4 - Opportunities and the future of the research area

Global Challenges Research Fund

- Detail of the opportunity
 - All research aimed at improving people’s lives has a large uncertainty component
 - These needs to be multi-disciplinary but the involvement of statisticians early will enable better decisions to be made at all levels.
- Action to maximise impact
 - More engagement is needed at institutional and disciplinary levels
 - Applications to fund should be explicit as to how uncertainty will be treated

Engaging with Industry

- Detail of the opportunity
 - Further involvement with business and industry
 - Interfaces with government/industry/applications areas which needs to include teaching as well as research
 - Engaging with industry creates problems and data as well as the two-way people pipeline.
 - Joint PhD and postdoc funding.
 - Leverage teaching skills in academia and take them into industry
- Action to maximise impact
 - Graduated projects, seed funding then increase the scale. Detail of the opportunity

- Action to maximise impact
- More support for multi-disciplinary call programmes

Meeting hugely growing demand

- Detail of the opportunity
 - Demand for Statistics and statisticians is growing
 - There is a massive need for Statistics methods and UK are seen as currently leading in this area
 - Statistics is important and government and companies are aware of this.
 - Managing need verses demand
 - There is a wide need but people do not know it
 - There are not enough people and the demand is only getting greater
- Action to maximise impact
 - Need to increase the number of graduates equipped with skills
 - Convert post-grad training to Statistics
 - Improve teaching of Statistics to undergrads
 - Final year/MSc dissertation with industry

Multi-disciplinary working

- Detail of the opportunity
 - Interaction between machine learning/computer science/engineering/applied mathematics
 - Possibilities to exploit new computing resources
 - Multi-disciplinary RCUK by increasing funding opportunities and the perception of stats in other disciplines
 - Training – introducing Statistics in other disciplines earlier
 - Looking for new ways to support Statistics to avoid Silos.
- Action to maximise impact
 - Change the process to make it easier to bid for cross-council funds
 - Allow time in the process for communication across disciplines
 - Involvement in GCRF activities – helping to formulate problems
 - Supporting statisticians working in other departments
 - Funding activities around departments interacting

PhD training

- Detail of the opportunity
 - Statistics needs to be supported widely across the country including for PhD students
- Action to maximise impact
 - PhD studentships to incentivise strong undergrads into Statistics
 - PhD students to be allowed in grant applications
 - Cross council funding and joint funding with industry of PhD students/CDTs
 - Show importance of statisticians to students including the focus of mathematical and real world challenges
 - Training opportunities need to provide a rounded experience
 - Possible national CDT for training specific schemes for stats PhDs.

Data Science

- Detail of the opportunity
 - To work with other disciplines to take ownership as users of data science have a better understanding of what they are doing.
 - Everybody is excited by it leading to new funding opportunities, new collaborations and new methodological problems in Statistics
 - Multi-disciplinary – gain different perspectives on challenges

- Better access to and processing of data sets
 - Broadening and developing links with The Alan Turing Institute
- Action to maximise impact
 - Joint training e.g. MSc
 - Developing a common language
 - Statisticians need to proactively be involved in proposals at the early stage
 - Expanding The Alan Turing Institute partners and involving more statisticians