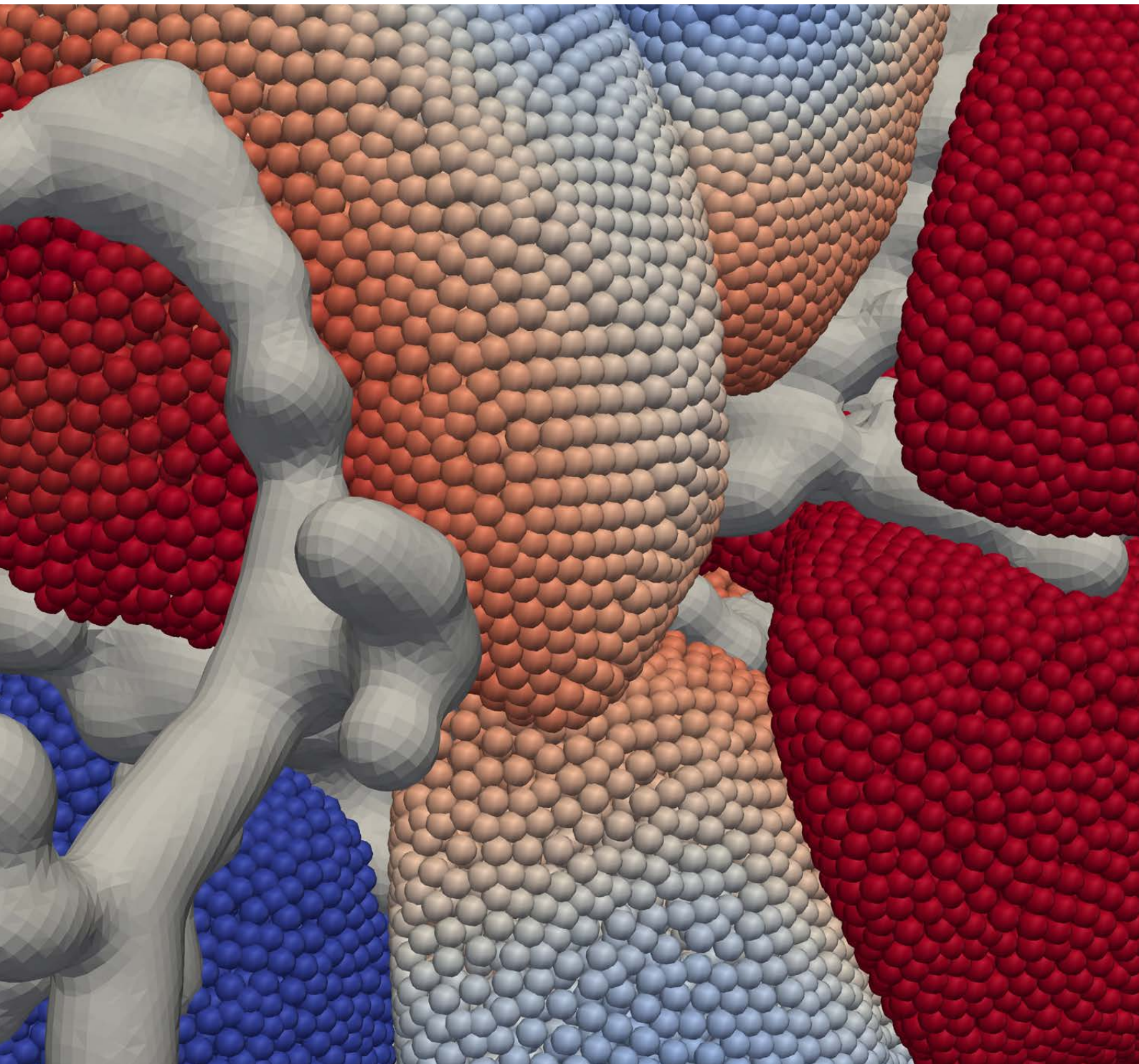
THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE

2014

The University of Auckland

Centre for eResearch

Enabling Computationally-Intensive Research



Contents

New Zealand's path to prosperity, as a small nation, is paved with pragmatic and collaborative approaches to translating advanced practices and technologies into the research system – NeSI is a key enabler in making this happen.

NeSI is a leading example of a national collaboration across the research system. NeSI brings together high-tech skills and infrastructure from across its investors, connecting with a broader range of researchers around the country to enhance their research.

NeSI does this from within the sector as the specialist capabilities being harnessed aren't easy to build and sustain, and are strongly defined by the research they support. An unincorporated body, NeSI receives investment from New Zealand universities, Crown Research Institutes and the Crown. The investors are the University of Auckland, University of Canterbury, NIWA (National Institute of Water & Atmospheric Research Ltd), Landcare Research, the University of Otago and the Ministry of Business, Innovation and Employment.

NeSI has grown the pre-existing capabilities at these investing institutions, with the University of Auckland as Host and the Centre for eResearch as one of five key collaborators. As we look ahead to 2015, NeSI's investors have renewed their commitments for a further four years, and the team are looking forward to continuing in their support role working with researchers nationwide.

NICK JONES
Director, New Zealand eScience Infrastructure

NeSI
New Zealand eScience Infrastructure

– ***a national collaboration supporting researcher needs for computation and analysis***

Acknowledgement

We thank all the authors from the University of Auckland in contributing these case studies for their research aspirations. We'd also like to thank the New Zealand eScience Infrastructure (NeSI) for providing High Performance Computing platforms, data services and the computational support team.

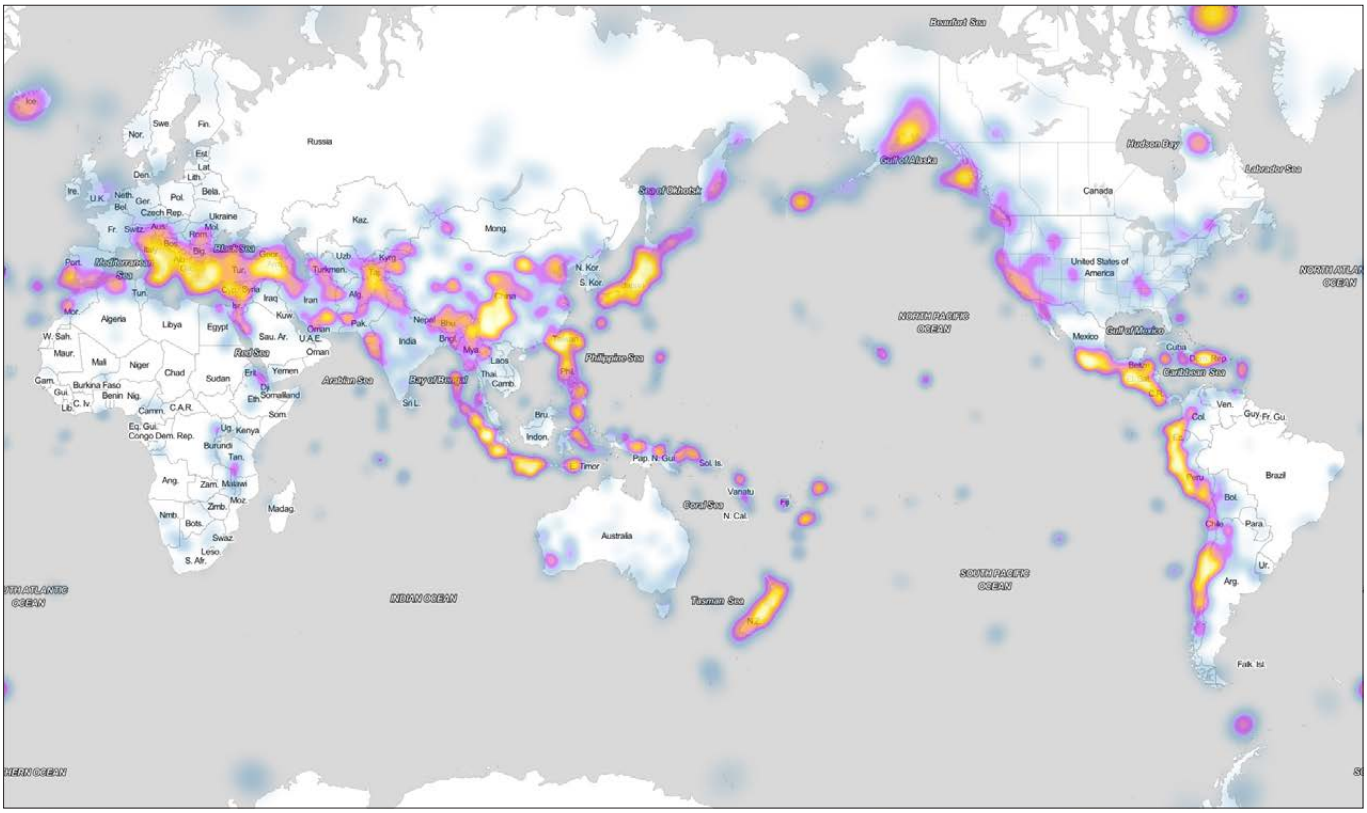
Message from the Director

2014 has been an extremely busy year for the Centre for eResearch. In partnership with NeSI, we have added some more hardware into our High Performance Computing facility (the Pan cluster), and a lot more researchers, projects and applications into our portfolio. We passed the landmark of 50 million CPU hours delivered and used by researchers since our record keeping began in January 2012. Over 50 scholars have now passed the mark of using over 500,000 CPU hours in support of their research. In many branches of research, access to massive amounts of computing power is now a prerequisite for those who aspire to publish in quality journals, and this trend is set to continue as more fields adapt to the opportunities afforded by large-scale analysis and simulation. These computing resources have been used to enable much diverse research, from archaeology to zoology; from cosmology to nano-materials. The Centre's offerings have grown too. We now provide a Research Virtual Machine (RVM) service that supports interactive research applications, both on Linux and Windows, we also run a large-scale visualisation facility (using 20 displays working in synchrony) and we offer a variety of workshops to up-skill researchers in eResearch and high performance computing.

The report presented here details some of these achievements, showcasing a small cross-section of the projects we have worked on, providing metrics to show the extent that we have raised the computational capability of researchers across the University, and lists the research outcomes that have resulted from these efforts.

In 2015, we look forward to beginning new work to better support research data, to the expansion of our RVM and visualisation capabilities and to a continued partnership with the refunded NeSI. Get in touch with us if you face computational challenges with your research.

PROFESSOR MARK GAHEGAN
Director, Centre for eResearch, The University of Auckland

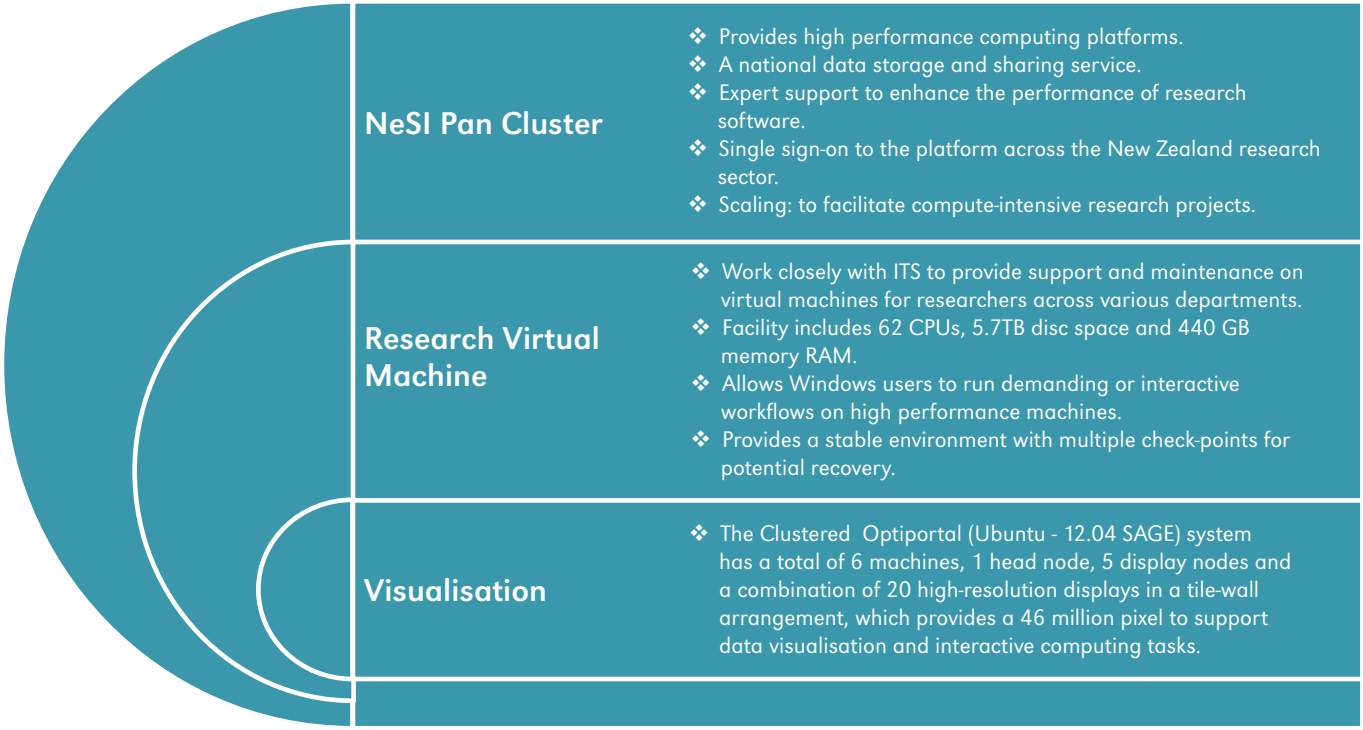


Mapping references to earthquake in over 2 million Wikipedia articles and travel blog entries.

Enabling computationally-intensive research

Research facilities and support

The Centre serves as the gateway for researchers within the University to access the NeSI High Performance Computing facilities. We actively engage with researchers to assess their needs, provide training, support, advice and consulting services in order to help them to achieve better computing performance and research outcomes. The facilities and support we offer are outlined as below:



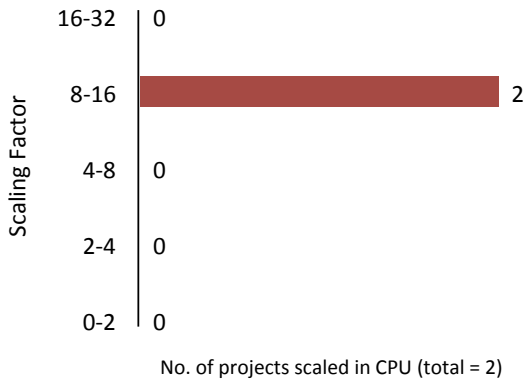
Scaling

Definition: In many ways, the NeSI Pan cluster offers significant advantages over desktop computers for compute-intensive projects. This is because the cluster helps overcome common limiting factors such as the availability of memory and the number of processors available. This section illustrates how the cluster helps to scale different research projects under following limiting factors:

Throughput	The cluster enables researchers to run more jobs at a time than they would be able to run without the cluster. A typical scenario is where several similar simulations are performed to test the importance of key parameters in an underlying model. Being able to test various conditions simultaneously can greatly improve a researcher's productivity by decreasing the time it takes to compute such simulations.
CPUs	Many scientific software applications can utilise multiple CPU cores concurrently across one or more computers. The cluster enables researchers to run such applications with a much higher CPU core count than they would otherwise have access to. This reduces the amount of time required to produce results and allows researchers to expand the scope of their investigations.
Memory	Each cluster node has more memory than a standard desktop computer. This category covers scenarios where researchers can work with larger data sets or more detailed models because of the amount of memory available to them, either on a single cluster node, or by leveraging memory across multiple cluster nodes.
Optimisation	This category covers scenarios in which research codes and workflows have been optimised to reduce the amount of time required to generate results. This includes general improvements made to program efficiency as well as extensions to enable parallel processing. Improved software configuration and installation are also included in this category.

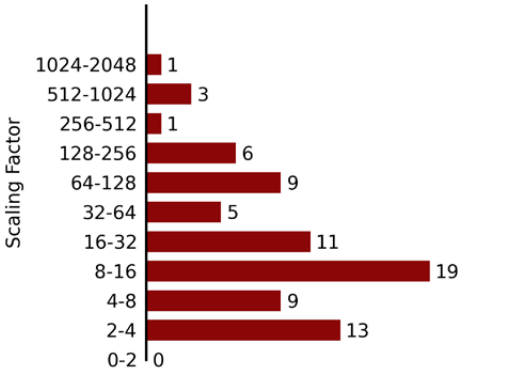
Each of the graphs that follow shows the total number of research projects that have been scaled (horizontal axis) across the various factors described above (vertical axis). The scaling factor value shows the improvement in productivity from using the cluster.

For example, the graph below shows that 2 projects have reported an improvement of between 8-16 times due to the higher number of CPU cores available to them.

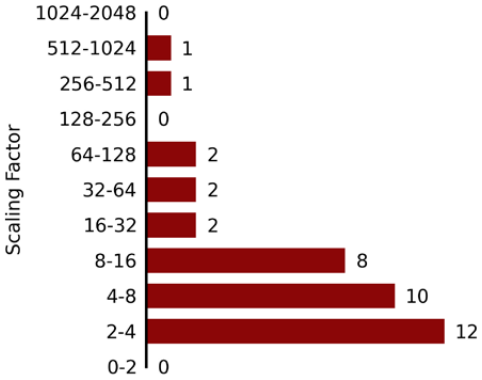


No. of projects scaled in CPU (total = 2)

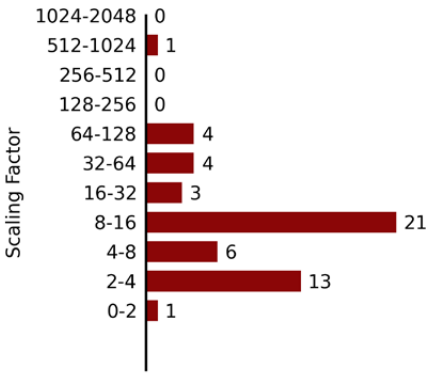
A breakdown of improvements achieved for the University of Auckland researchers using the NeSI Pan cluster against the four factors of throughput, CPUs used concurrently, memory and application optimisation. As the graphs show, the majority of researchers achieved scaling from 2-32 times across these various factors, with a few achieving 500 or even 1000 times increase in their capabilities. Exactly how much scaling is achieved depends on many factors, including the exact nature of the algorithms used, the volume of data produced and consumed, and the time spent in improving the various aspects of the computational workflow.



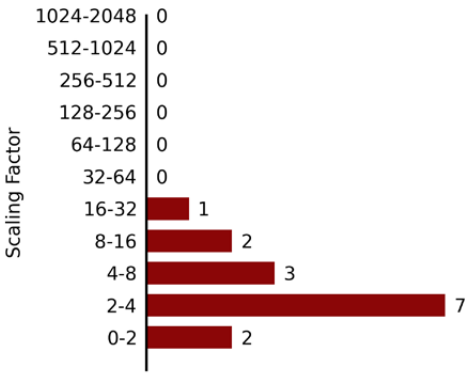
No. projects scaled in Throughput (total=77)



No. projects scaled in CPUs (total=38)



No. projects scaled in Memory (total=53)



No. projects scaled in Optimisation (total=15)

Training for researchers

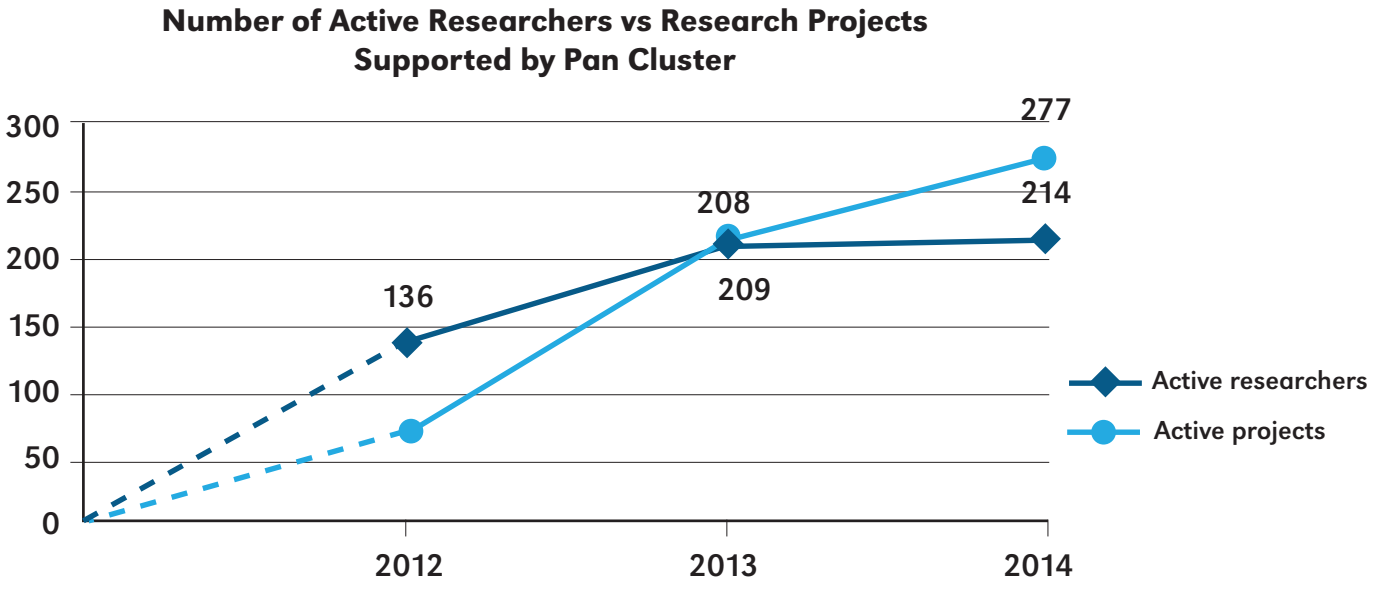
The Centre provides a series of training workshops available for researchers to learn some of the skills required to gain advantage from our advanced computing facilities. These workshops range from courses for inexperienced compute users to more sophisticated software modelling assistance. The courses run throughout the year, as needed, with introductory training offered every week.

Key aspects of the courses:

- What is high performance computing?
- Concept of parallel programming.
- Shared and distributed memory programs.
- Hardware and software resources.
- Access, job submission and data management.
- Best practices and security measures.
- SLURM installation, extension and support.
- How to migrate workflows between job schedulers.
- Using new job scheduling command line tools.
- Advanced job submission features.
- Hands on training and guidance.
- Expert consultation for in-depth support needed can be arranged through appointments with our consultants.
- Development of high performance workflows.
- Support in data visualisation.
- Software and systems development for interdisciplinary projects.

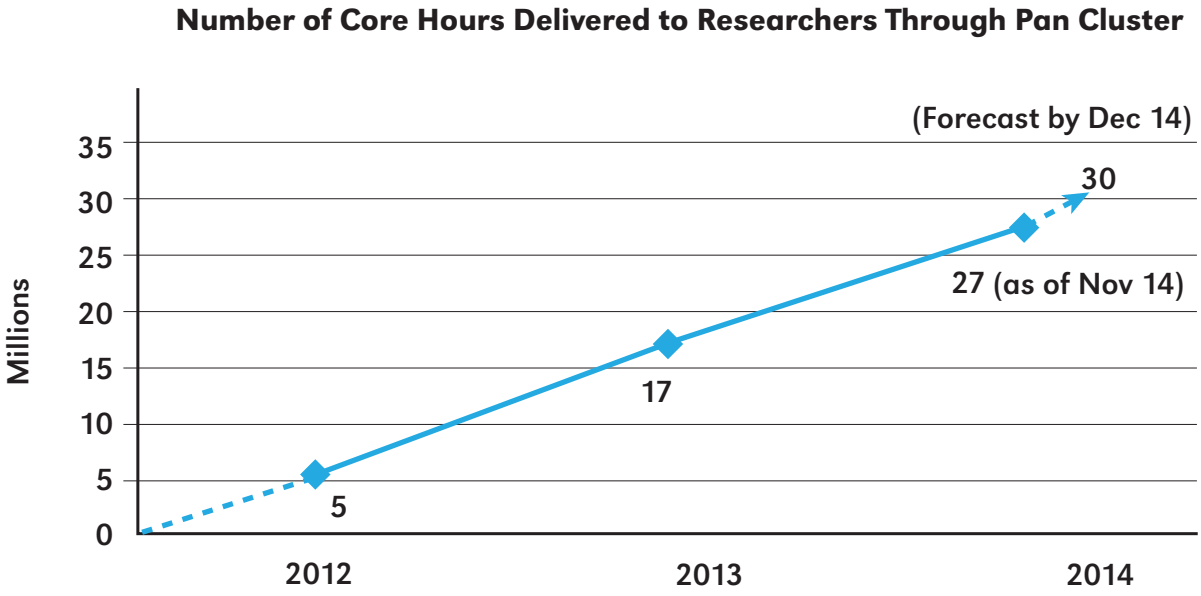
Research outcomes enabled

The Centre and the NeSI staff currently support hundreds of researchers and their research projects across the University. The use of the High Performance Computing facility continues to grow at a fast pace. The graph below shows the increase in our activities over the last 3 years. The number of active researchers appears stable but in fact there is a lot of turnover each year, particularly among the postgraduate community. We add several new users each week.



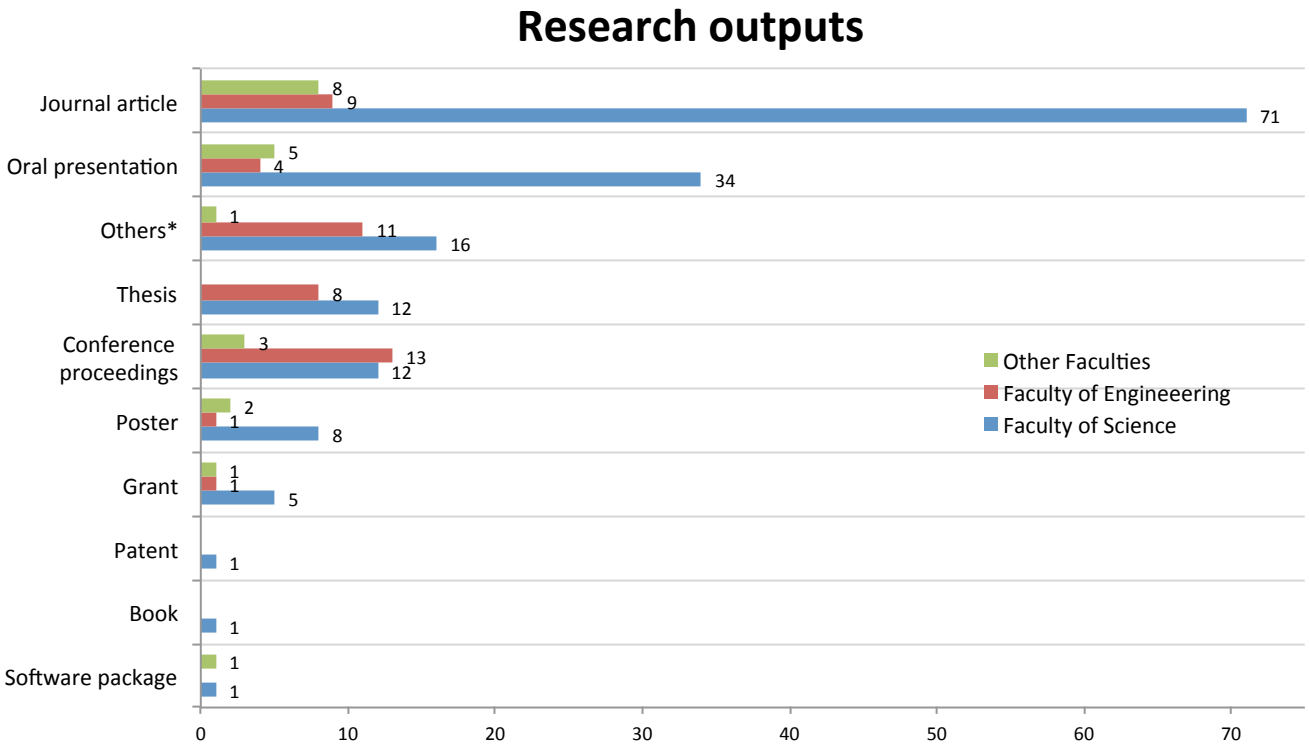
The above figures are the numbers of researchers who have submitted at least one job during the year. Projects closed are not re-counted in the subsequent year.

The increased number of core hours delivered to researchers during each investment year is consistent with the growth of the HPC capacity through multiple investments in hardware. By 2012, the incremental investments made to the Pan cluster were close to 2.97 million. A further 1.9 million was invested in 2013 which brought the total investment in Pan cluster to 4.87 million by the end of 2013.



The above figures illustrate the millions of core hours used each year by the University of Auckland researchers through the NeSI Pan cluster. The dotted line prior to 2012 is an estimate due to lack of data, while the dotted arrow in the end forecasts core hours that should be achieved by the end of 2014.

The chart below categorises research outcomes collected between 2013 and 2014 by type and by faculty. The information was gathered by survey and sent to all of our users. About an 89% response rate was achieved—so some outcomes will be unknown to us. The outputs under *others are either in the form of international collaboration success, awards, PhD student internal reports or papers in progress.
A detailed list of all of these research outcomes is provided in the Appendix.



Other Faculties include Faculty of Medical Health Sciences, Faculty of Arts and Liggins Institute.

Case Study 1

The formation of surface archaeological deposits in arid Australia

Ben Davies, School of Social Sciences

How does the archaeological record form? The deposits that archaeologists study are the product of both behavioural and geophysical processes, both of which operate at varying tempos over time. However, archaeological studies often emphasise the role of human behaviour, looking for signals of human activity within arrangements of artefacts and features. Geophysical and post-depositional processes are typically thought of as disturbances to signals of human behaviour, generating an incomplete record that archaeologists are tasked with reconstructing. Different deposits, depending on their contexts, are considered to be more or less informative based on impressions of degrees of preservation.

Changes which occur to archaeological deposits throughout their existence cannot always be neatly separated from the residues of an original (or any) occupation event. Interpretations which attempt to fit synchronic behavioural reconstructions on to a record perceived as incomplete presume a subtractive rather than generative nature for archaeological deposits, ignoring potentially informative patterning based on perceptions of preservation. An alternative is to view the archaeological record not as partial or fragmentary, but as a complete outcome from the long-term accumulation of processes occurring at a given location. From this perspective, it is the job of the archaeologist to associate recorded patterning within the record with a combinative set of formational processes rather than discrete sociocultural events.

This study is aimed at understanding the formation dynamics of surface archaeological deposits in order to identify what kinds of patterns might be associated with different formational processes, both behavioural and geophysical.

Figure 1: *The remains of one of many heat-retainer hearths from surface archaeological deposits at Rutherford's Creek, western New South Wales*

Surface archaeology in Australia's arid zone

Australia's arid zone is dominated by surface scatters of stone artefacts and cooking hearths. These are notoriously difficult for archaeologists to interpret due to stratigraphic mixing caused by long term erosion, and are often incorporated into settlement models based on an exclusively behavioural model But what these deposits lack in terms of vertical integrity is made up for by the high degree of visibility over large spatial extents. This kind of record can offer archaeologists an opportunity to understand prehistoric use of space at the landscape scale, something typically hampered by the expense and logistics of excavation.

Previous research at Rutherford's Creek in western New South Wales has produced a record of radiocarbon dates which have been obtained from ancient cooking hearths. When taken together, these show a general trend of greater numbers of dates in the more recent past, as well as noticeable gaps in the occupation history for the area. The overall increasing trend may be consistent with an increasing population, while the gaps may be explainable as human absence or the effects of dispersal and congregation. Both of these patterns, however, may also be consistent with differential preservation rather than any particular activity set. Furthermore, there is some evidence to suggest that people in the past were regularly carrying away stone artefacts in fairly large numbers. This is seen in the low amount of cortex (the outer weathered surface of stone nodules) found in archaeological assemblages, indicating that the mobility of the people who manufactured the artefacts was high as cortical pieces are taken away. The spatial areas covered by the archaeological assessments, however, are small in comparison with the probable space being utilised in the past, and it is not known how different land use schemes might produce different patterning at the local scale.

To assess these issues, we constructed a set of agent-based computer simulations to evaluate the distribution of archaeological remains in space, and how general patterns of erosion and deposition through time might affect the character of surface deposits.

Simulating the formation of landscape-scale archaeological deposits on the NeSI Pan cluster

Spatial simulations were developed using the NetLogo software package. The first set simulates the changing visibility of surface deposits by modelling the erosion and deposition of surface sediments alongside the construction of cooking hearths. Hearths are constructed by programmed agents at a given rate, and these become hidden, visible, or destroyed by sequences of geomorphic events. At the end of the simulation, the resultant distribution of surface hearths are sampled and compared with field data. The second set of simulations resamples a database of stone artefacts made in a laboratory, and simulates their dispersal using different mechanisms (random walks, correlated walks, etc.). A GIS extension allows the simulations to be run in a spatial environment analogous to Rutherford's Creek, using different landscape features (lakes, creeks, etc.) as attractors for human occupation.

The data and code were then uploaded to the NeSI cluster, where batches of controlled experiments were automated using a customised shell script developed with the help of the NeSI staff. The simulations disperse millions of artefacts in less than an hour, and generate thousands of simulated populations of radiocarbon dates in rapid succession according to different geomorphic models.

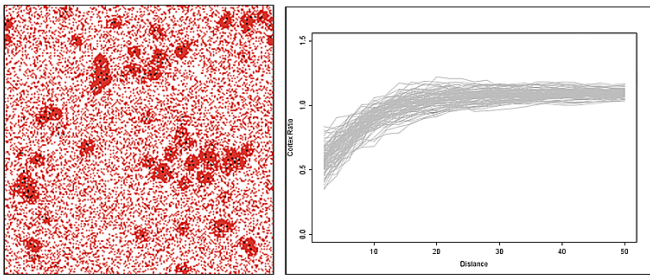


Figure 2: A simulated distribution of stone artefacts (left) and the average accumulation of lithic cortex over distance from 100 similar simulations. Preliminary output from the simulations suggest that low levels of cortex within field assemblages are consistent with high levels of artefact transport.

Once these reach densities comparable with those recorded in the field, the results are returned from the cluster, sampled appropriately, and compared to those obtained from Rutherford's Creek. Without the speed of the NeSI cluster and the experiment automation devised by the NeSI staff, this would be a slow and gruelling procedure.

What's next

Preliminary results from this study suggest that the patterns in stone artefact assemblages and hearths are more consistent with explanations that include high levels of mobility among human groups. Changes in the levels of cortex among recorded stone tool assemblages are parsimonious with repeated, short-term occupation with little redundancy in place use. Additionally, modelling the formation of the radiocarbon record indicates that both the increasing trend in radiocarbon dates and gaps seen in the record may be consistent with large scale erosion and deposition events. These simulations can be used to develop field tests which may help to differentiate between erosion-driven gaps and those associated with human absence. This kind of insight is important for the development of informed heritage management policy as well as for developing a better understanding of the human past.

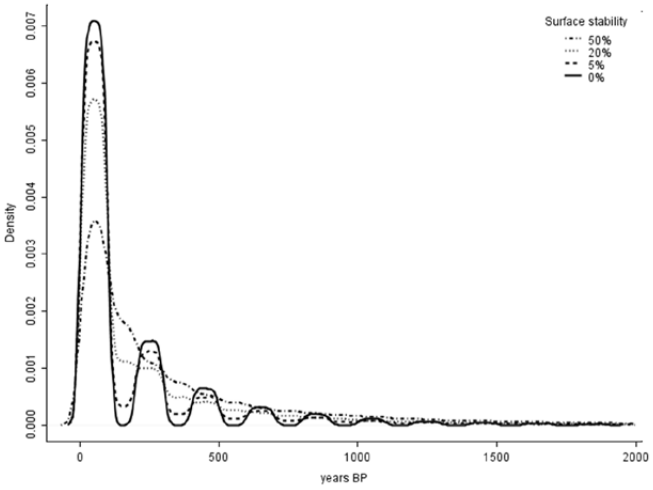


Figure 3: A comparison of age probability densities for simulated surface hearths under different degrees of surface stability, with simulations using a 2:1 depositional/erosional scheme and geomorphic events spaced at 100 year intervals. Preliminary results indicate that widespread surface instability can produce the appearance of targeted gaps in radiocarbon chronology

Case study 2

BEAST, Bayesian evolutionary analysis sampling trees

Alexei Drummond, Bioinformatics, Department of Computer Science

BEAST is a cross-platform program for Bayesian MCMC analysis of molecular sequences. It is entirely oriented towards rooted, time-measured phylogenies inferred using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST uses MCMC to explore tree space in such a way that each tree is weighted proportional to its posterior probability. We include a simple to use user-interface program for setting up standard analyses and a suit of programs for analysing the results.

BEAST 2 (www.beast2.org) develops new features to allow resuming an MCMC chain, and real time tracking of ESSs while running a chain.

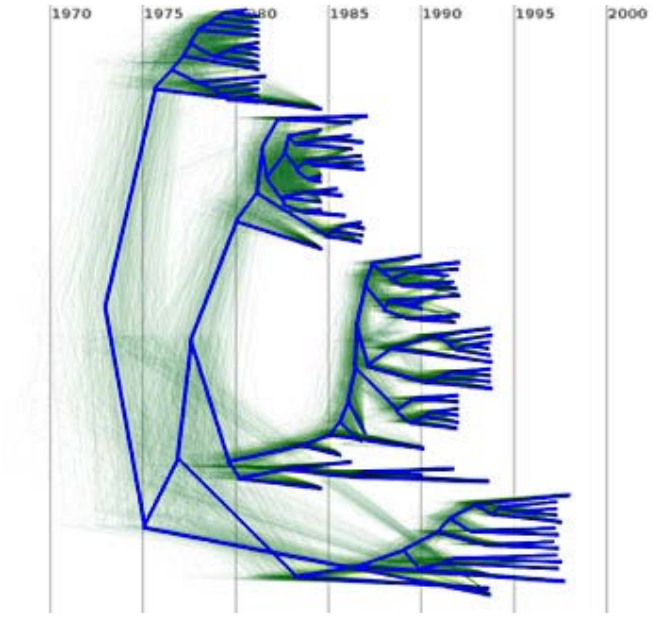
It is easily extendible, for example, the BEAST 2 packages supports multi-chain MCMC, some experimental likelihood calculations that are potentially faster than the base implementation, and a spread sheet GUI for manipulating models.

BEAST 2 has a plug-in facility. BEAST 2 packages provide more models, such as

- Birth Death Serial Skyline Model
- Reversible jump substitution model
- SNAPP Species trees from SNP and AFLP analysis
- Substitution Bayesian Model Averaging
- Sampled Ancestors

They can be easily installed/un-installed from BEAUti (BEAST's GUI). More packages are being developed (http://www.beast2.org/wiki/index.php/Main_Page#BEAST_2_packages).

BEAST developers use the NeSI Pan cluster for running simulation studies to validate new models, and run large analyses with questions of considerable scientific interest. For example, BEAST was used for dating the split of Cichlid species between Africa and South America, determining the origin of Indo-European languages (see http://language.cs.auckland.ac.nz), and analysing epidemiological histories of diseases such as HIV and influenza.



Output of a BEAST analysis of viral samples taken over different years – samples from the same year tend to cluster together. Blue line represents a summary tree of the posterior sample. Green lines indicate uncertainty in the tree distribution. It illustrates how both topology and node height estimates vary in the posterior sample.

Case study 3

Finding genetic variants responsible for human disease hiding in universe of benign variants

Klaus Lehnert and Russell Snell, Jessie Jacobsen and Brendan Swan, School of Biological Sciences

Introduction and research

Our programs aim to unravel the genetic basis of human diseases, using new approaches enabled by recent step-changes in genetic sequencing technologies (aka the "$1000 genome"). The human genome comprises 3 billion loci, and individuals typically differ from this 'reference' at millions of sites. These differences are the result of a complex interplay between ancient mutations, selection for survival fitness, mating between populations, events of near-extinction, and strong population expansion over the last ~150 years. A constant supply of new mutations creates new variants that are extremely rare. Some of these variants are directly responsible for disease, and others cause genetic diseases in unknown combinations.

We combine classical genetics approaches with genome sequencing to identify potentially disease-causing variants for experimental validation. One of our focus areas are neurological diseases, and we started a large project to understand the genetic nature of autism-spectrum disorders (www.mindsforminds.org.nz/), an often debilitating neurodevelopmental condition with increasing prevalence in all human populations. We expect that identification of genetic mutations will help us to better understand the disease process and identify new targets for therapeutic intervention. We take similar approaches to support mutation discovery in clinical research in collaboration with ADHB.

This is 'big data' research – we typically process 100 billion 'data points' for each family, and the data analysis and storage requirements have pose significant challenges to traditionally data-poor biomedical analysis.

What was done on the Pan cluster and how it helped

Using the parallel processing options available on the Pan cluster we were able to derive optimal combinations for multiple inter-dependent parameters to align billions of short sequence reads to the human genome. These sequence read are strings of 100 nucleotide 'characters' (one of the four DNA 'bases' plus 'not known'), including a confidence score for each base call, and contain a small number of differences to the reference string – the patient's individual and population-specific variants. The non-uniform nature of the human genome reference further complicates the similarity search. The goal is to assign a unique position for each read in the genome, using mixed algorithms employing string matching to 'seed' the alignment and a combination of string matching and similarity scoring to extend the alignment through gaps and differences, taking into account the confidence score for each base call. Through parameter optimisation and multi-threading we successfully reduced run times from several hours to seven minutes per patient.

The second step of each patient's genome analysis aims to derive 'genotypes' for each of the millions of loci that differ from the reference in each patient. A genotype is a 'best call' for the two characters that can be observed at a single position (humans have two non-identical copies of each gene, one inherited from each parent).

Genome sequencing creates 20-500 individual 'observations' for each of the 3 billion genome positions. The observations may be inconsistent, and/or may be different to the reference. We obtain a 'consensus call' for each position through a process that first proposes a *de novo* solution for the variant locus (i.e. not influenced by the reference), and then applies complex Bayesian framework to compute the most probable genotype at each locus.

This process requires approximately 2000 hours of computation for a single family. However, on the Pan cluster we can apply a classic scatter-gather approach: we split the genome into dozens of segments, compute genotypes and probabilities for all segments in parallel, and then combine the results from the individual computations to generate a list of all variants in each individual. Total compute time remains unchanged, but the process completes within a few hours!

We use standard bioinformatics software programs such as Burrows-Wheeler aligner, the Genome Analysis Toolkit (GATK), samtools, and others; these work very well but require multi-dimensional parameter optimisation.

Being able to obtain results in hours instead of months allows us to analyse more patients than otherwise possible, and the time gained for variant interpretation increases the number of cases we can analyse.

The NeSI approach has allowed us to perform this research in collaboration with colleagues in Christchurch and Dunedin. The Pan cluster is the perfect home for this collaboration, and the use of a shared computation platform leverages the development effort in the three research teams.

What's next?

NeSI's high-bandwidth connection and compute facilities will allow us to extend our collaboration across the Tasman, and we look forward to working with new collaborators in Australia. And of course there are disease-causing variants to be discovered!

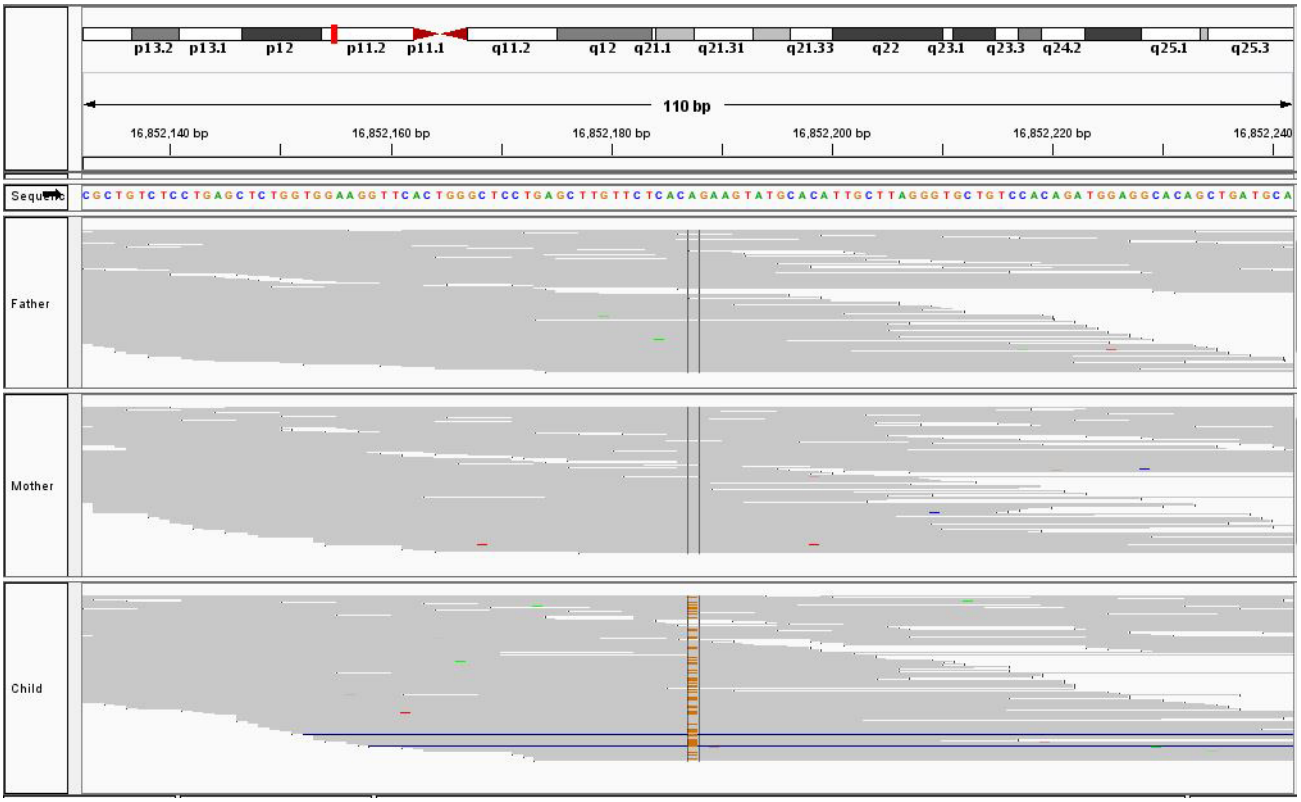


Figure 1: A new mutation – is it real, and does it cause disease?
A visualisation of the alignment of 360 million 100-nucleotide reads for a parent-child trio against an unrelated genome reference. The window is zoomed on 110 of the 3 billion bases in the human genome (horizontal axis), the vertical axis shows features in the reference genome (top), the reference sequence (colour-coded), followed by three panels displaying several hundred reads (grey lines) obtained from the genomic DNA from father, mother, and child. Differences to the reference sequence are indicated in colour, positions identical to the reference are in grey (majority). Our analysis on Pan has clearly identified a new mutation in the child (orange squares, bottom panel), and it appears to affect only one of the child's two gene copies.

Case study 4

Multigene environmental DNA data analysis for New Zealand genomic observatory

Alexei Drummond, Dong Xie, Department of Computer Science, Andrew Dopheide, School of Biological Sciences, NZGO (Genomicobservatory) team

The project aims to produce a comprehensive phylogenetic and environmental characterization of the terrestrial species in a well-defined New Zealand model ecosystem using modern sequencing, informatics, niche modelling and field ecology approaches. The project is a collaboration between several universities and research centres in New Zealand and Australia, the Department of Conservation, Auckland Council, and iwi, and internationally as part of the Network of Genomic Observatories. It also aims to provide a long-term research program structure for collaborative, interdisciplinary research projects at the intersection of ecology, evolutionary biology and genomics.

The percentage of OTUs at the 97% clustering threshold assigned to phyla. Unclassified OTUs, OTUs containing low-complexity sequences, and OTUs from phyla that are represented by less than 0.1% of the OTUs are grouped into the "Others" category.

In the project, we are able to measure broad diversity of eukaryotes from soil using an environmental DNA approach. Environmental DNA (eDNA) approaches typically focus on microbial communities within the soil and tend to use single gene marker regions. Here we evaluate a suite of DNA markers coupled with Next Generation Sequencing (NGS) that span across the tree of life. Sequences analysis, such as Operational taxonomic units (OTUs) identification by molecular markers, taxonomic assignment, and biodiversities estimation, is a main part of this evaluation.

The raw reads file in a FASTQ format was then passed into a UPARSE pipeline (Edgar, 2013) to identify OTUs, which includes quality filtering, length truncation (300 bp), dereplication, abundance sorting, OTU clustering, chimera filtering and mapping OTUs. The outputs of the pipeline were a FASTA file containing OTU sequences and a mapping file between OTUs and reads for each given OTU clustering threshold. The community matrix was created from the mapping file by retrieving the site information added in the sequence label previously, and the matrix described species abundance (OTU counts) according to sampling sites.

Jost's biodiversities (Jost 2006) are respectively calculated regarding community matrices of six eDNA methods using R package *vegetarian* (Charney and Record 2012). Rarefraction curves for diversities are further estimated using a 97% threshold for OTU identification by subsampling the minimum number of OTUs of sampling sites (subplots) using R ecology package *vegan* (Oksanen et al 2013). BLAST+ was used to classify the taxonomy of OTUs, and the classification result was interpreted to taxonomic assignment by phyla.

To learn more about the project, please refer to the project webpage www.genomicobservatory.cs.auckland.ac.nz and the database link https://data.genomicobservatory.cs.auckland.ac.nz.

Case study 5

Testing what cosmic inflation really predicts

Layne Price, Department of Physics

Overview

A period of accelerated expansion in the early universe has recently received outstanding support from precision measurements of the cosmic microwave background. This "inflationary" epoch occurs at high energy scales where the exact behaviour of physics is uncertain. We focus on the numerical exploration of the dynamics of early universe models that have many degrees of freedom, which generically arise in high energy theories such as supersymmetry or string theory.

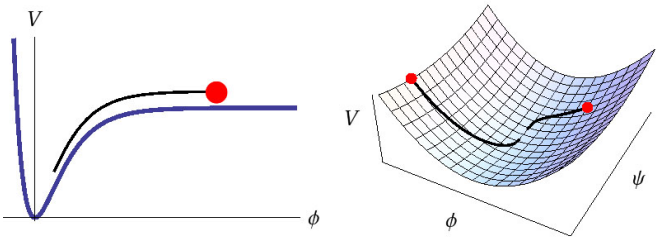
Inflation in the era of precision cosmology

Inflation is an early period in the universe's history when the cosmological expansion accelerated. Paradoxically, this acceleration helps to explain both the general smoothness of the cosmic microwave background (CMB) and the statistics for the tiny hot and cold spots that cause the CMB to deviate from perfect homogeneity. Inflation has recently received significant observational support from measurements of the temperature and polarization of the CMB, which were taken by the space-based telescopes WMAP and Planck, as well as the ground-based telescope BICEP2.

Inflation operates at energy scales that are vastly higher than any other known physical process, with typical energies that are 14 orders of magnitude higher than what can be obtained at the Large Hadron Collider at CERN. Before these recent developments in precision cosmology, the exact physics at these scales was exclusively the playground of speculative ideas on quantum gravity, string theory, and supersymmetry. While simple inflationary models seem to be able to reproduce the CMB data, only recently have people begun to explore the consequences of inflationary models that incorporate features generally expected to be present in high energy particle physics.

The role of high performance computing

In our work we have focused on exploring the generic predictions of more complicated particle physics models of inflation that incorporate many active degrees of freedom. We have developed a series of numerical tools that are able to perform efficient simulations of the relevant physical processes at work in the early universe. Deriving a complete picture of what one of these models predicts requires extensive sampling of a high dimensional parameter space, the numerical evaluation of a coupled system with hundreds of nonlinear differential equations, the simulation of the plasma dynamics that generate the CMB, and the construction and evaluation of observationally relevant statistics.

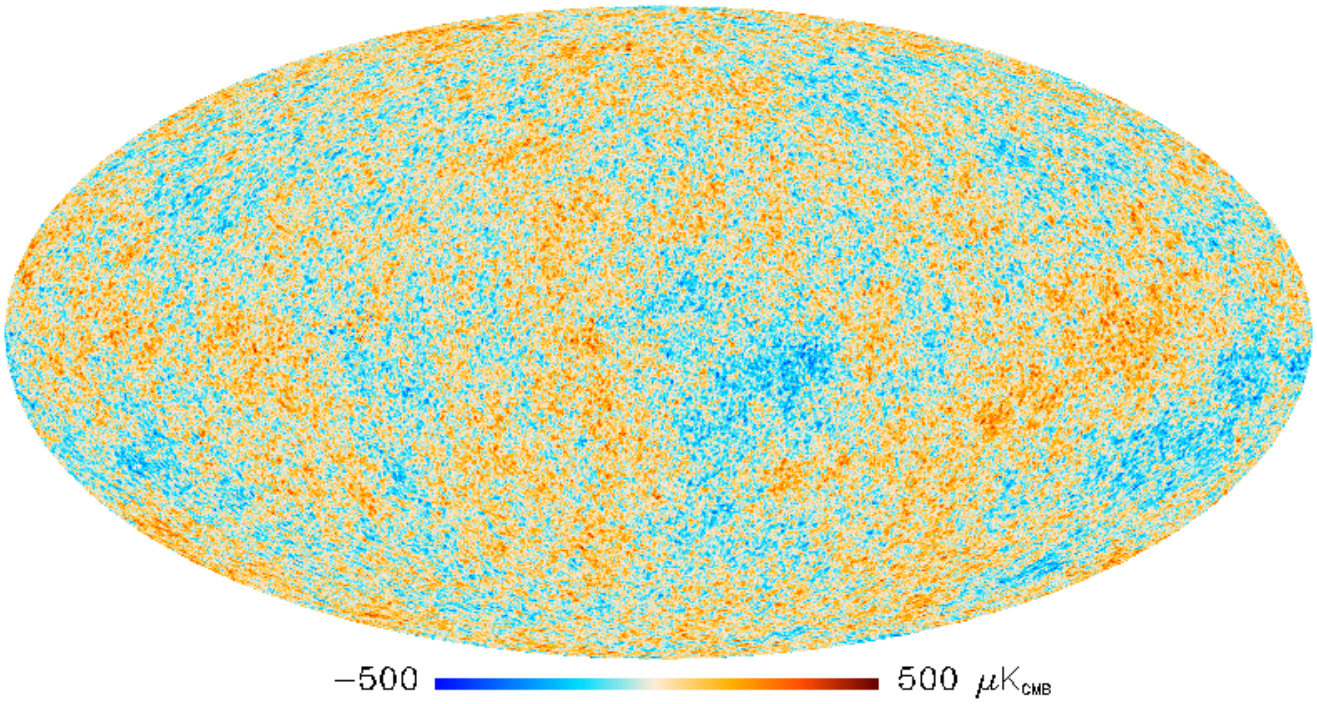


This shows the scalar field potential energy density V as a function of one scalar quantum field (left) or two fields (right). The red dots indicate some possible initial positions for the fields, while the black lines show the paths the fields would take during inflation. With one field you can only go one way down the potential; all initial conditions give the same outcome. With more than one field the paths are different for different initial conditions and each of these predict slightly different statistics for the CMB.

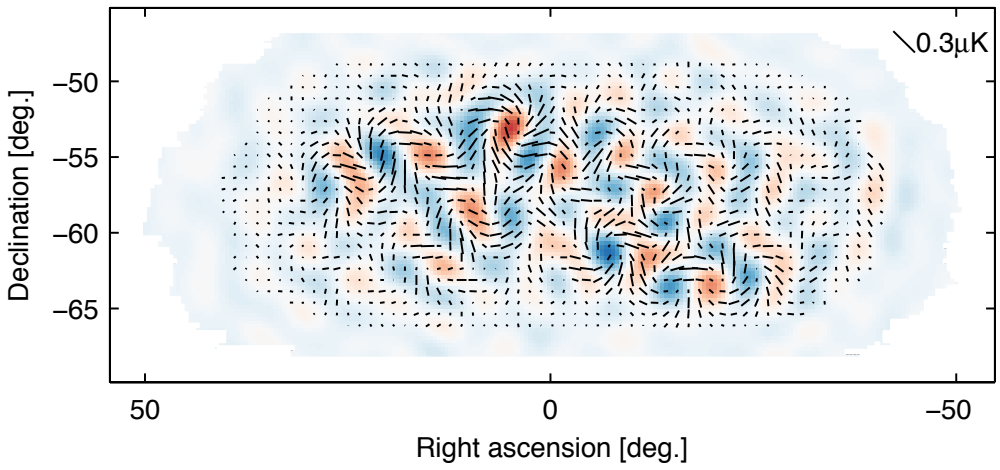
We have leveraged the computational resources of the Pan cluster to perform MPI-parallelized simulations of millions of possible early universe scenarios. We have focused extensively on the role that the universe's initial conditions play in changing the end-state of a given universe. This allows us to gauge how special our particular universe is amongst the range of possibilities. We have shown that due to non-linear interactions in the early universe that the initial conditions do not have to be fine-tuned (JCAP 1307 (2013) 027; arXiv:1304.4244) and that they have only a very minor impact on the CMB statistics (Phys.Rev.Lett. 112 (2014) 161302; arXiv:1312.4035), greatly simplifying the physics. These results lend support to the fact that inflation does not necessarily require a fine-tuned initial state in order to produce a CMB that is statistically similar to ours. While we have many international collaborators, the New Zealand computing assets and the support available from the Centre for eResearch have enabled us to complete our research quickly and with more resolution than we expected.

Future work

The next step in this project is to use lattice simulations of partial differential equations to evaluate the role of spatial gradients in the initial conditions. We will also perform a complete statistical analysis of the Planck and BICEP2 data for these more complicated inflation models by calculating the Bayesian evidence and posterior probabilities. This will require extensive numerical work where we have to explore the entire allowed parameter space and weight each possible set of parameters against how likely it is that it could reproduce our CMB. The work that we have done to-date provides a series of significant first steps toward this more ambitious goal.



The hot and cold spots in the cosmic microwave background (CMB) as measured by the space-based telescope Planck. The CMB is at an average temperature of 2.7K and these fluctuations occur only at the level of micro-K, making a high resolution picture like this a true experimental achievement. [arXiv:1303.5062]



The detection of the B-mode polarization of the CMB by the BICEP2 telescope in Antarctica. The simplest explanation for this polarization spectrum is due to gravitational waves in the early universe, with an amplitude consistent with what is predicted by the simplest inflationary models. [arXiv:1403.3985]

Case study 6

Phylogeny and phylogeography of the family kyphosidae (Perciformes: teleostei)

Steen Wilhelm Knudsen, School of Biological Sciences

Introduction

Comparing DNA sequences of multiple genes in various species is commonly applied today for inferring relationships among species and finding the best phylogenetic tree that explains how species have evolved and how they are related.

In this study we focused on the sea chubs (family *Kyphosidae*) that are abundant consumers of macroalgae on both temperate and tropical reef systems. Several species have extremely broad distributions. Juveniles of *Kyphosus* species are often found well offshore among drift algae, suggesting a high potential for oceanic dispersal. The relationships and taxonomic status of the 16 valid species of sea chubs (genera *Hermosilla, Kyphosus, Neoscorpis* and *Sectator*) known worldwide have long been problematical due to perceived lack of character differentiation, complicating ecological assessment.

We inferred the relationship and distribution of all known species of sea chubs using a combined analysis of partial DNA fragments from mitochondrial markers (12s, 16s, cytb, tRNA -Pro, -Phe, -Thr and -Val) and three nuclear markers (rag1, rag2, tmo4c4), in total comprising 5960 bp from 118 individuals. Synonyms among Atlantic and Indo-Pacific taxa show that several sea chub species are more widespread than previously thought.

Using the Pan cluster

The main work carried out on the Pan cluster comprised parallel analysis using the software MrBayes and BEAST, and was run across multiple CPUs where each result afterwards was combined and the best tree inferred through the lengthy analysis that occasionally can take more than a day to complete. Furthermore, by applying fossil calibrations on the groups of species in the tree while the analysis runs, the time for speciation of different species and groups can be inferred and provide a chronogram – an evolutionary tree that shows points in time of speciation for various groups.

The chronogram shows that *Kyphosus* originated relatively recently, in the early Miocene. *Neoscorpis lithophilus, Kyphosus cornelii* and *Hermosilla azurea* are basal in the topology, implying a subtropical, Indo-Pacific origin for the family. The Southern Indo-Pacific Ocean hosts the greatest diversity in sea chubs, and Western Australia and South-Eastern Africa appear to be refuge habitats for ancestral lineages. The tropical species of *Kyphosus* and *Sectator* represent very recent divergences, indicating that herbivory in the group originated at higher latitudes, and that carnivory in the autapomorphic *Sectator ocyurus* is derived. Several clades of perciform herbivorous reef fishes had established themselves on tropical reefs in the Eocene. Kyphosids colonised low latitude reefs much more recently, and speciated rapidly.

Running these analyses on the Pan cluster made it possible to infer evolutionary trees that are based on more comprehensive data analysis than would be possible – within a feasible timeframe – to prepare from analysis run on a single desktop computer. It also made it possible to test multiple settings and cross influence of parameters to ensure the resulting trees were the most optimal reflection of the evolution among the sea chubs that could be inferred from the inherent DNA data.

Being able to infer these evolutionary trees through the use of parallel computing on Pan cluster was fundamental to this study, and ensured that the evolutionary scenario inferred was based on rigorous testing of the underlying data.

Our research has resulted in two scientific publications in the journal Zootaxa, and we are currently working on two additional scientific papers, and two chapters for a guidebook on New Zealand fishes.

Figure 1: A photo of Kyphosus gladius (above) and Kyphosus sydneyanus (below). We discovered that Kyphosus gladius had gone unnoticed in Western Australia, as it has been confused K. sydneyanus in the past. But careful examination of radiograph photographs and detailed analysis of the DNA sequence variation in the different species of Kyphosus lead us to conclude that K. gladius should be described as a new species for Western Australia. These findings were published last year in Zootaxa (vol. 3599). Photo by K. Clements.

Phylogeny and phylogeography of the family Kyphosidae (Perciformes: Teleostei)

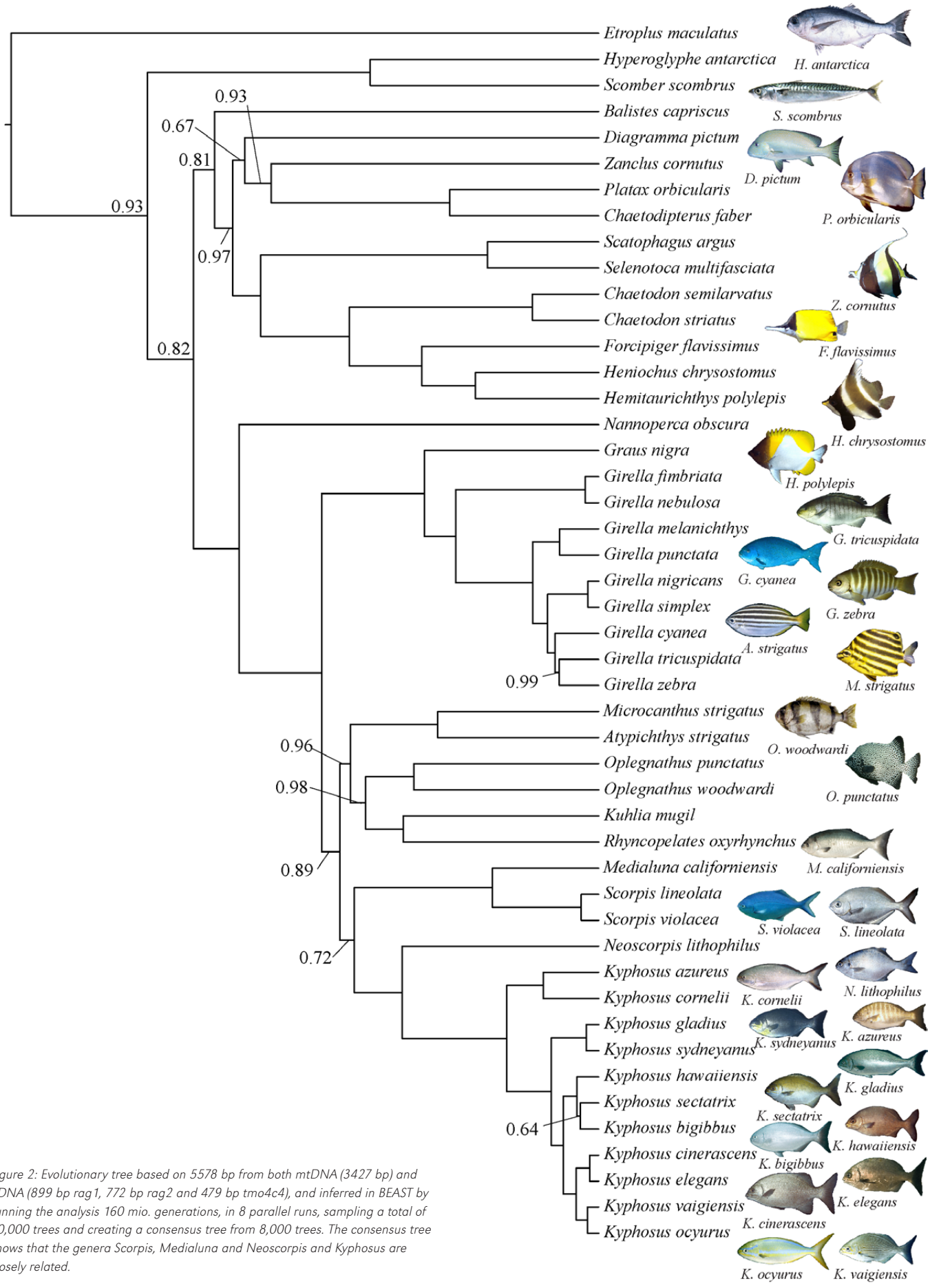


Figure 2: Evolutionary tree based on 5578 bp from both mtDNA (3427 bp) and nDNA (899 bp rag1, 772 bp rag2 and 479 bp tmo4c4), and inferred in BEAST by running the analysis 160 mio. generations, in 8 parallel runs, sampling a total of 10,000 trees and creating a consensus tree from 8,000 trees. The consensus tree shows that the genera Scorpis, Medialuna and Neoscorpis and Kyphosus are closely related.

Case study 7

Molecular phylogenetics uses genetic data to reconstruct the evolutionary history of individuals, populations or species

Stephane Guindon, Department of Statistics

Understanding biological processes generally involves the comparison of various species. Deciphering the origins of a complex organ, such as the human eye for instance, very much relies on the analysis of same organ in other animals. It is through such comparison that the working of this organ can be inferred. This mechanistic approach to understanding biological systems has been, and still is,very successful. Phylogenetic trees depict the evolutionary relationships between species and thus provide a relevant framework to the comparative approach in biology. Phylogenetic trees are therefore much more than a tool to classify species.

Genetic sequences convey a wealth of information about evolution. Differences between contemporary sequences are the consequence of the accumulation of mutations during the course of evolution. Using a relevant statistical approach, one can then reconstruct the tree that best explains the observed differences between sequences. More specifically, mutations are modeled as a Markov process that "runs along'' the phylogeny. The inference then relies on the likelihood function, i.e., the probability of observing the sequences at the tips of the tree given the whole phylogenetic model, i.e. the Markov model of mutations as well as the phylogenetic tree itself.

Estimating the parameters of phylogenetic models is challenging. First, the number of possible binary trees that connect the species (or populations) grows exponentially with the number of tips. In practice, when reconstructing phylogenies, it is therefore impossible to examine all possible trees. Heuristic algorithms have thus to be designed in order to explore the space of possible solutions efficiently. Second, as the size of available data sets increases, Markov models of increasing complexity can be fitted to genetic sequences. Estimating the parameters of these models then relies on sophisticated optimization or sampling techniques.

The PhyML software

In collaboration with colleagues from the CNRS in France, I have developed the software PhyML that implements fast algorithms for estimating trees that maximize the likelihood function. The original article that describes these algorithms [1] has been widely cited and PhyML is now one of the most popular software in evolutionary biology. The maintenance and development of this software requires constant testing. New methods and algorithms are regularly incorporated in PhyML. Assessing the stability and the performance of these new methods before making them available to the scientific community is therefore paramount.

Given the increasing size of data sets available, the NeSI Pan cluster allows me to run thorough tests by analyzing hundreds, sometimes thousands of sequence alignments in parallel. Being able to monitor the average speed and accuracy with which phylogenetic models are inferred with PhyML is essential in order to make sure this software performs well compared to its multiple competitors.

Modelling competition between species

PhyML also serves as a platform to implement research projects. Together with Louis Ranjard (SBS) and David Welch (Department of Computer Science), we have recently designed a statistical model that uses phylogenetics in order to explain the current spatial distribution of species [2]. In particular, our model relies on the idea that a given species may be found at a particular location simply because it happened to be the first to colonize this territory during the course of evolution. In other words, competition between species may prevent secondary colonization. Using phylogenetics helps us decipher the ordering in which different species colonized their habitat, therefore providing precious indication about whether or not competition could indeed have played a role in this process.

An important aspect in statistical modeling of stochastic processes is to assess the accuracy and precision with which the parameters of these models can be estimated. Large scale in silico experiments are here very helpful. One first generates data according to the proposed model, using specific (or randomly chosen) values for its parameters. The second stage of the experiment consists in using these simulated data and infer the model parameters. Ideally, the estimated values and the 'true' ones (i.e., the values that were used in the first stage of the experiment) should be highly correlated.

The NeSI Pan cluster provided us with adequate computing resource to perform these experiments for our competition model. We were thus able to show that the parameters of our model can indeed be recovered from the available data (see Figure 1), hence

demonstrating the validity of our approach. Each data point in these scatterplots was obtained from the Bayesian estimation of the corresponding parameters using Markov Chain Monte Carlo. These inference techniques enjoy good statistical properties but are very demanding from a computational perspective. We would simply not have able been able to produce these results without the computing resources provided by the NeSI Pan cluster.

References

[1] A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood S Guindon, O Gascuel Systematic Biology (2003) 52 (5): 696-704.

[2] Modelling competition and dispersal in a statistical phylogeographic framework L Ranjard, D Welch, M Paturel, G Stephane Systematic Biology (2014) 63 (5): 743-752.

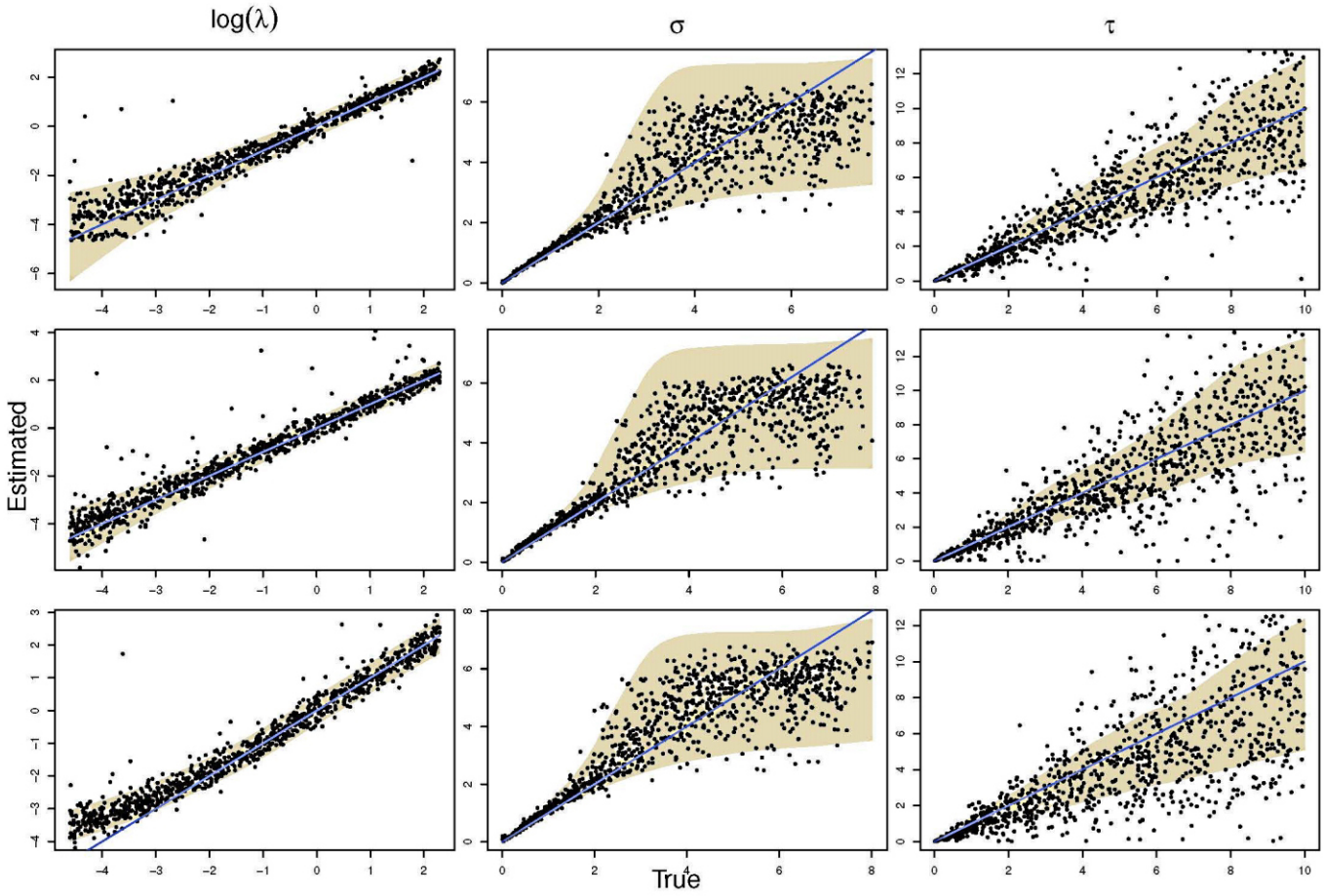


Figure 1: Posterior estimates of competition (λάμδα) *and dispersal* (σίγμα και ταυ) *parameters — 100 taxa, 150 locations (top row), 100 locations (middle) and 50 locations (bottom).*

Case study 8

Number theoretic algorithms in cryptography

Shi Bai, Steven D. Galbraith, Department of Mathematics

Modern public-key cryptography is about communication in the presence of adversaries, allowing users to communicate confidentially without requiring a secret key to be distributed by a trusted party in advance [1]. The security of public-key cryptosystems is usually based on the presumed hardness of certain number-theoretical problems such as integer factorization and lattice problems. Therefore, it is important to understand the difficulties (both theoretical and practical) of the underlying computational problems. The project aims to understand the security of modern cryptosystems by investigating the computational efforts that are required to attack them.

Our main goal therefore is to investigate new and improved algorithms for solving certain important computational problems. The NeSI Pan high performance cluster provides significant computational power and parallel scalability for us to investigate these algorithms. Our research has practical importance for the security of modern cryptosystems: suggesting safe parameters for them as well as constructing new schemes and protocols.

Schemes

There are two scenarios in the scope of our research: pre-quantum and post-quantum cryptosystems. At present, many public-key cryptosystems rely on the presumed hardness of integer factorization and computing discrete logarithms. Such cryptosystems are often referred-to as the pre-quantum schemes as there exist efficient algorithms to attack them on quantum computers. In contrast, new mathematical foundations for cryptography such as lattices are becoming interesting as they can be used to construct post-quantum cryptosystems (since they are believed to be secure even against attackers with quantum computers).

Integer factorization

The first scenario is concerning the factorization of large integers. The RSA cryptosystem relies on the assumed difficulty of the large integer factorization problem. Figure 1 shows the RSA-1024 challenge, a 309-decimal digit number, which has not been factored so far. The number field sieve (NFS) is asymptotically the most efficient algorithm known for factoring large integers. Its time complexity is sub-exponential. It consists of several stages, the first one being polynomial selection. The running time of NFS depends on the quality of the polynomials. The project aims to investigate efficient algorithms for polynomial selection. The NeSI Pan cluster provides computational resources for the project, as the computational requirements of such a computation are substantial. For large integer factorization, a few thousand jobs are usually running on the cluster at a time. Such tasks would not be possible to accomplish without the usage of the NeSI cluster.

135066410865995223349603216278805969
93888147560566702752448514385152651 0
60485953383394028715057190944179820 7
28216447155137368041970396419174304 6
49658927425623934102086438320211037 2
95872576235850964311056407350150818 7
51067659462920556368552947521350085 2
87941637732853390610975054433499981 1
150056977236890927563

Figure 1 RSA-1024

Lattice-based cryptosystems

The second scenario is to study the computational hardness of certain lattice problems. A lattice can be visualized as a periodic arrangement of points in space (Figure 2). The difficulty is that there are many different ways to write down a "basis" for a lattice. Lattice- based cryptosystems often rely on the assumed difficulty of finding certain vectors that satisfy nice properties (e.g. shortness) in a lattice. At present, lattice-based cryptosystems have relatively large keys and/or ciphertext size compared to traditional cryptosystems such as RSA. Hence, an important research problem in lattice-based cryptography is to obtain practical public key schemes that are still provably secure under lattice assumptions. The NeSI cluster helps us to investigate the security of lattice-based systems through large-scale computation.

Real-world application

The research has practical implications for the real world, as cryptography is critical for electronic transactions everywhere, such as e-mail, e-commerce and mobile communications. It is important to make sure the cryptographic algorithms are secure and efficient. One of our research outputs helps to build more efficient and provably secure digital signatures on lattices (e.g. [3]). Our research also helps us identify potential vulnerabilities and weaknesses of cryptographic algorithms and hence suggest safe parameters and constructions for crypto-protocols [2]. For example, researcher Zachary Harris used our software cado-nfs to find vulnerability in Google [5] (using a too small RSA key for email authentication). Recently, two researchers Fabien Perigaud and Cedric Pernet from the Cassidian Cybersecurity used our software cado-nfs to reverse-engineer a ransomware[6].

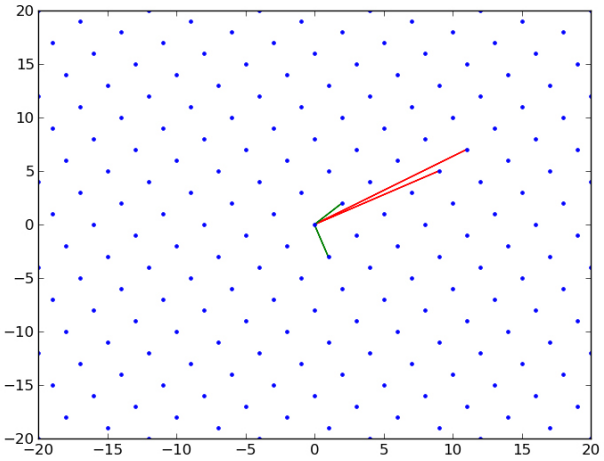


Figure 2: Two bases in a 2-dimensional lattice

Network and collaboration

We are participating in international collaborative research projects that bring together researchers from France, Germany, USA and other countries. We have been involved in co-developing the open-source software cado-nfs (http://cado-nfs.gforge.inria.fr), which has been used in the computation on the Pan cluster. We hope to attack some large-scale computational problems in feasible time through our collaboration. Locally, our research, computation and experiments would not have been possible without access to the NeSI computing facilities and help from the Centre of eResearch staff.

References

1. Ronald L. Rivest. Cryptography. Handbook of Theoretical Computer Science. J. Van Leeuwen (ed.), 1990.

2. Shi Bai and Steven D. Galbraith. Lattice decoding attacks on binary LWE. ACISP, Lecture Notes in Computer Science, Vol. 8544, 2014.

3. Shi Bai and Steven D. Galbraith. An improved compression technique for signatures based on learning with errors. In J. Benaloh (Ed.), CT-RSA 2014, LNCS 8366 (2014) 28–47.

4. Shi Bai, Richard Brent and Emmanuel Thome. Root optimization of polynomials in the number field sieve. Mathematics of Computation, 2014.

5. Zachary Harris, How a Google Headhunter's E-Mail Unraveled a Massive Net Security Hole, http://www.wired.com/2012/10/dkim-vulnerability-widespread.

6. Fabien Perigaud, Cedric Pernet, Bitcrypt broken, http://blog.cassidiancybersecurity.com/post/2014/02/Bitcrypt-broken.

Case study 9

Estimating migration rates in the budding yeast Saccharomyces cerevisiae

Sarah Knight, School of Biological Sciences

Background

Eukaryotic microbes are key ecosystem drivers, but we have little theory and few data elucidating the processes influencing their observed population patterns. Because of their large population sizes, and ease of transfer, one might expect microbial populations to be well mixed, but there is increasing evidence showing that many are not homogeneous but structured. Understanding the processes that drive population structure and connectivity has implications for understanding the evolutionary trajectories of these organisms.

The budding yeast *Saccharomyces cerevisiae* is of significant commercial importance due to its role in the production of bread, wine, beer and other alcoholic beverages but is also widely used by the scientific community as a model research organism. While we have a vast knowledge of its cell biology, genetics and increasingly its ecology and evolution, studies of its population patterns are often confounded by geography.

Since microbes are inherently difficult to observe directly, we rely on genetic methods to infer their movements. As part of a larger investigation into the population dynamics of *S. cerevisiae* in New Zealand, we sampled *S. cerevisiae* from vineyards and surrounding native bush in major Sauvignon Blanc growing regions and genotyped 850 individual isolates at eight microsatellite loci.

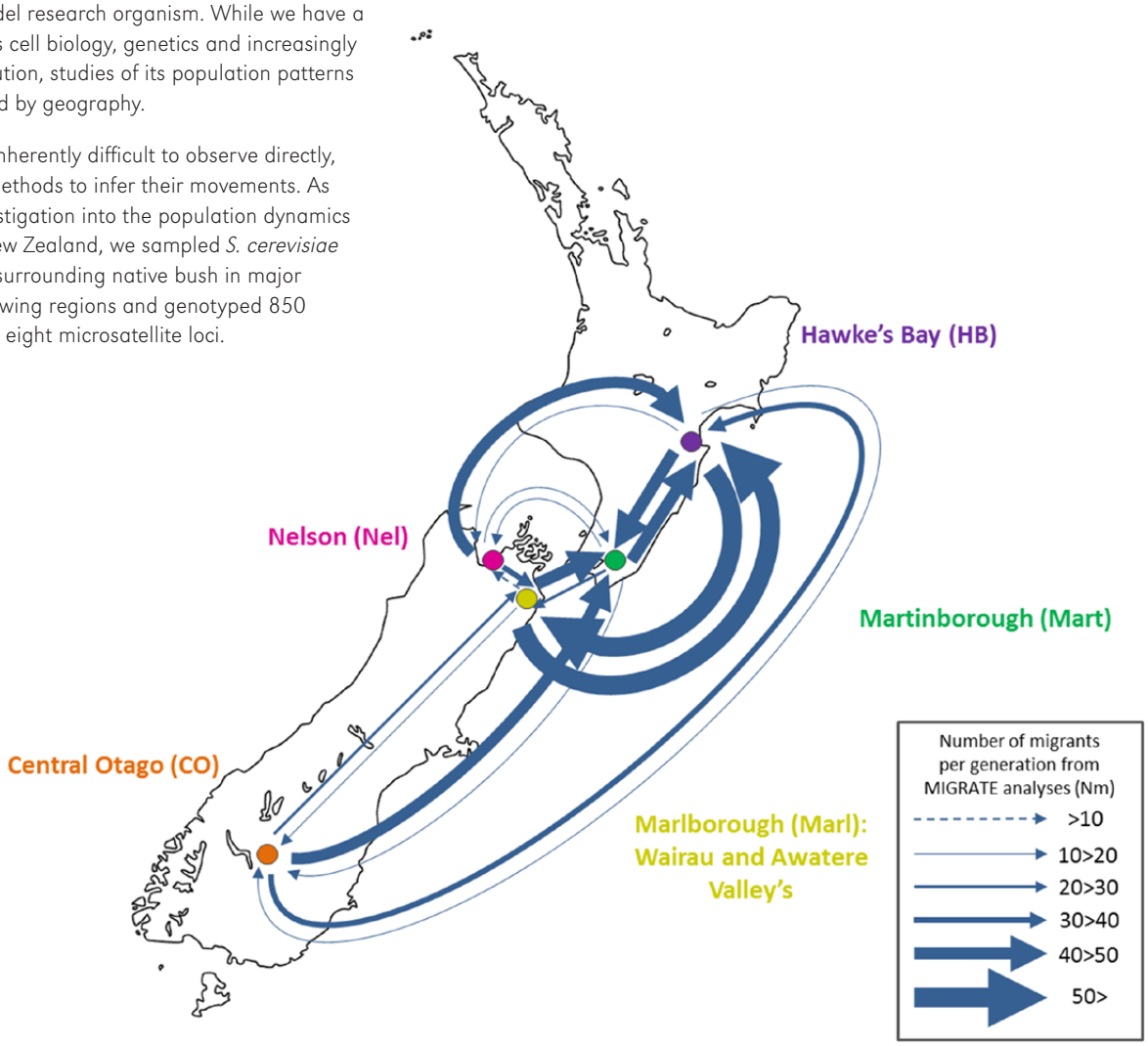


*Figure 1: Adapted from *Knight and Goddard (In press). Migration rates between regions as calculated by MIGRATE.*
**This research has recently been accepted for publication by The ISME Journal.*

After removing identical genotypes from the same samples due to the clonal nature of growth of S. cerevisiae during the culturing process, the final dataset comprised 369 microsatellite profiles. This dataset was used to infer patterns of population structure and connectivity between these regions and between managed and unmanaged ecosystems within each region (*Knight and Goddard, *In Press*).

Estimating migrations rates using the Pan cluster

Analyses of genetic data to infer migration rates are typically very computationally expensive. We used a Bayesian coalescent approach implemented in MIGRATE (Beerli, 2006; Beerli, 2009) that ran ten replicate MCMC chains of one million steps in length across all eight loci. With the help of staff at NeSI, optimal settings were determined and we managed to run each chain for each locus in parallel, essentially reducing the time it took to run the analysis by 80 times! Even then the analysis still took weeks to run and without the resources provided by NeSI this analysis would not have been possible.

What's next?

Our results suggest that migration of *S. cerevisiae* in New Zealand mimics the movements of fruit by the wine industry, with larger migration rates into the larger wine producing regions. While this work is now largely completed, it would be interesting to examine what differences exist in the population dynamics of microbial eukaryotes that do not have such a close association with human activity. Can we still detect human influenced dispersal? Or are these populations more highly structured?

References

Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22(3): 341-345.

Beerli, P. (2009). How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? Population Genetics for Animal Conservation. G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli and C. Vernesi. New York, USA, Cambridge University Press: 42-79.

Case study 10

Improving the treatment of heart disease

Susann Beier, Faculty of Medical and Health Sciences

Coronary artery disease is the most common killer in the Western world. One in four people will die from a blockage in one of the coronary arteries, preventing supply of blood and oxygen to the heart.

A common treatment for narrowed arteries is 'Percutaneous Coronary Intervention' or PCI, where a wire mesh tube, or 'stent', is inserted into the narrowed blood vessel and expanded to hold it open. The expansion of the stent compresses the abnormal build-up of atheroma narrowing the artery. This is a highly effective and relatively low risk treatment, it avoids surgery, and enables rapid improvement in the patient's condition. Unfortunately, stents fail in 25% of patients, when the previous narrowing reoccurs within the stent.

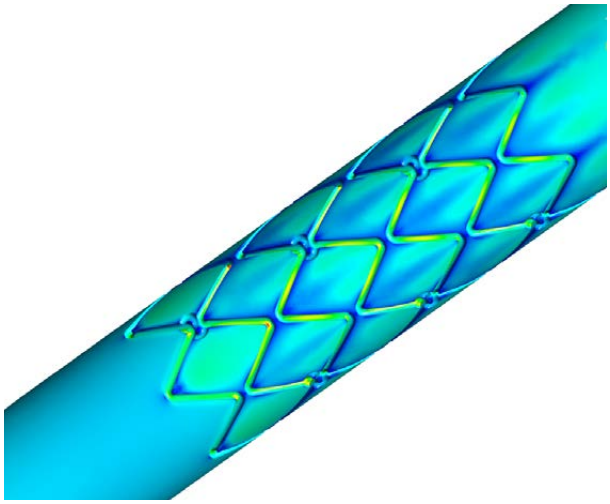


Figure 1: Stress in the wall of a coronary artery after stent implanted

New stents which release medication (known as 'drug-eluting' stents) promised a solution to this problem, with the slow local release of drugs supressing the regrowth of the tissue responsible for re-narrowing. A short-term improvement is achieved, but late failure rates still carry the risk of causing sudden death.

It is known that the presence of the stent changes blood flow, which in turn alters the growth of tissue, and thereby affects treatment outcome. Currently, there are more than 250 different stent designs available, each of which has a different effect on the 3D pattern of blood flow. The stent *design* is therefore critically important in determining treatment success or failure. Clinical trials with groups of patients have been unable to define the stent design features which determine success or failure due to the complexity of the problem. More detailed quantitative analysis is needed to answer the question of how different stent designs affect the complex blood flow through the coronary arteries.

My research aims to study the link between stent design features, the changes they cause in blood flow (haemodynamics), and the impact on patient outcome. Computational Fluid Dynamic (CFD) modelling was used to quantify the haemodynamics for various stent designs. This allowed the calculation of blood velocity, the stress the blood flow induces on the vessel wall, and to detect regions of rapid changes by simulating the vessel shape, circulation and blood rheology. Two commonly used stent designs were modelled to determine their respective strengths and weaknesses.

In addition to the computational simulations, a number of blood vessels were scaled up, manufactured in plastic using 3D printing and connected to a pump to simulate the heart. They were then placed in an MRI scanner to measure the blood flow with PC-MR (Phase-Contrast Magnetic Resonance). A comparison was then performed between the computational CFD and experimental MRI data with good agreement found.

Simulating coronary blood flow on the Pan cluster

Many simulation time steps were required to accurately represent the rapidly changing flow conditions in the coronaries with more than 2,500 time steps computed for a single heartbeat. To ensure full development of the flow, four heartbeats were modelled, and only the last was analysed. Each simulation contained approximately 20 million elements, and used multi-threading over six cores with 40CPUs and with 20GB of memory each, and took around 48 hours to solve.

This task would have not been possible without the Pan cluster, as the memory requirements would have exceeded local desktop capabilities well before a possibly month long solution was obtained. We were able to run many simulations on the cluster, allowing us to systematically investigate more than 20 different stent design aspects, and their impact on blood flow in the coronary arteries.

Next steps

Due to the Pan cluster resource we are now extending our research to more complex vessel geometries such as vessel branching. Having gained a greater understanding of the features that are critical to idealised stent design, the next milestone is to model a stented coronary artery from an actual patient. This is a highly complex undertaking but the knowledge that can be gained will have an important impact in the treatment of patients with heart disease.

Case study 11

Modelling dual reflux pressure swing adsorption (DR-PSA) units for gas separation in natural gas processing

George Zhang, Department of Chemical and Materials Engineering

Pressure swing adsorption is an alternative gas separation method using adsorption to separate gas mixtures. DR-PSA is an advanced configuration of pressure swing adsorption with two product refluxes and an internal loop. DR-PSA can achieve perfect separation in theory.

Dynamic modelling of adsorption units is considerably hard. Material balance, energy balance, heat transfer, mass transfer has to be considered in dynamics, thus rigorous numerical models are often built using a dedicated simulator: Aspen Adsorption. Aspen Adsorption is a module of the Aspen Suite and has built-in partial differential equations for mass balance, energy balance as well as adsorption isotherms. Besides these, Aspen Adsorption can model pressure-flow networks in a rigorous way. Conventional PSA usually contain two columns working in parallel - one in adsorption mode and the other in regeneration mode. To simplify the model in Aspen Adsorption, it introduced an "interaction unit" to decrease computational intensity. However DR-PSA has two refluxes and four adsorption parts thus four adsorption columns need to be independently and simultaneously simulated while the pressure-flow relationships of the system also has to be guaranteed. Besides these, the stiff manner of the system requires the integrator step size to be very small (as low as 1E-5 seconds), otherwise the integrator would either crash or result in a wrong predicted value.

As the dynamic model of DR-PSA has never been built before, a very large number of runs are required to match experimental data and to study the effect of each operation parameter. All these contributed to extra computational resource required. On a normal desktop (3.3Ghz Quad Core Intel i5 processor with 8GB ram), one simulation scenario would take more than 10 hours to finish and the system only allows 3 scenarios to be run at once. During last phase, around 70 different scenarios were run and the whole work took more than 4 months to finish. Now using the high performance computer, 6 simulations can be run concurrently, and instead of 2 weeks for 18 simulations it takes only 3 days.

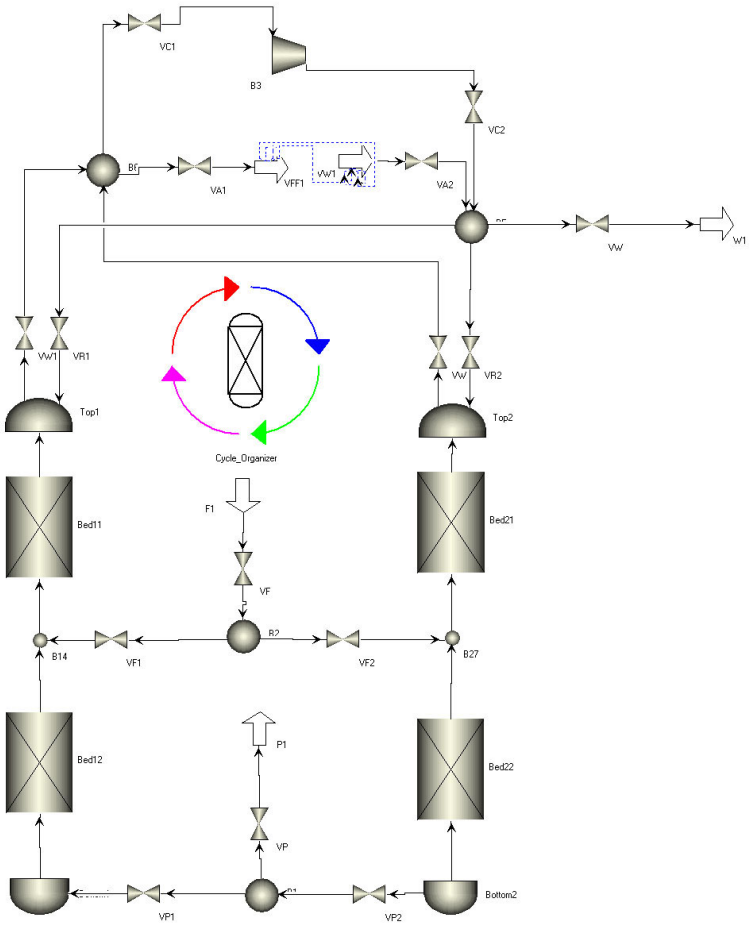


Figure 1: *The model built in the Aspen Adsorption for simulation*

Figure 2: The experimental 2-column setup

This research is performed in collaboration with the University of Western Australia, where they perform the experiments. Figure 3 below shows the results of a temperature distribution comparison between experiment and simulation. There is a very good agreement, which could only be obtained thanks to the computational power provided by the VM.

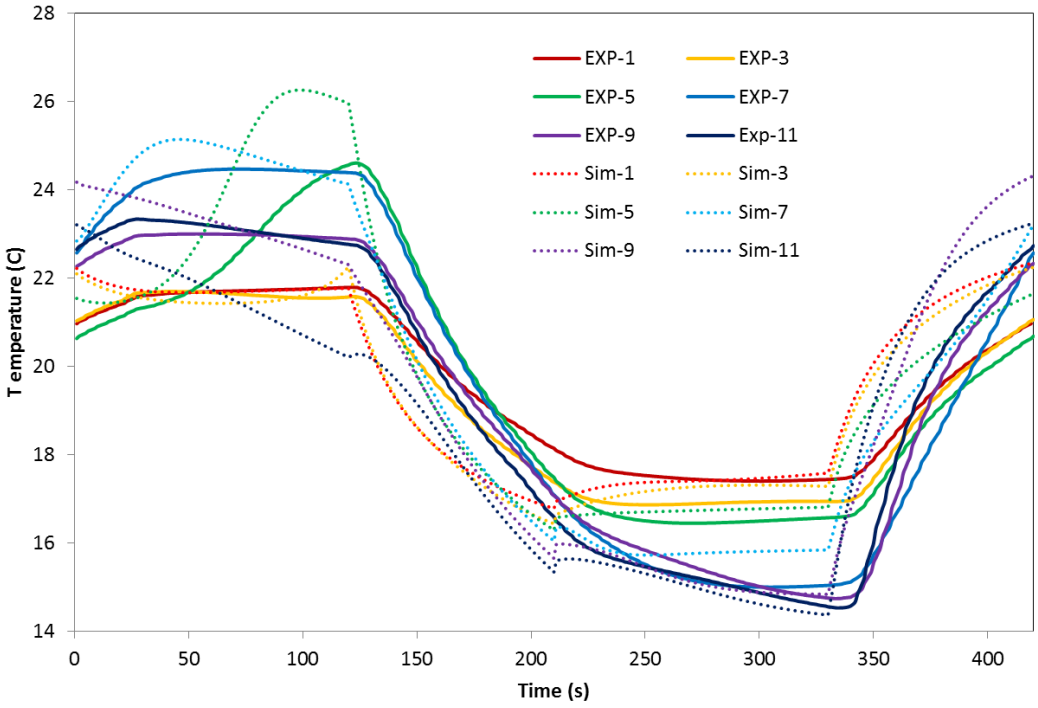


Figure 3: Temperature profile comparison – Vinlet spec

Case study 12

Multiscale modelling of saliva secretion

James Sneyd, Department of Mathematics, University of Auckland
David Yule, School of Medicine and Dentistry, University of Rochester
John Rugis, New Zealand eScience Infrastructure

This interdisciplinary project encompasses a range of activities targeting anatomical data based structural modelling of individual salivary cell clusters, solution of cellular calcium dynamics function in full 3D simulations, interactive visualisation of resultant calcium waves and validation of results by comparison to experimental data. The model will be used to test duct cell function and for the testing of pathological conditions. The overall project is funded by the National Institutes of Health, USA (Sneyd, Yule).

Current activity and results

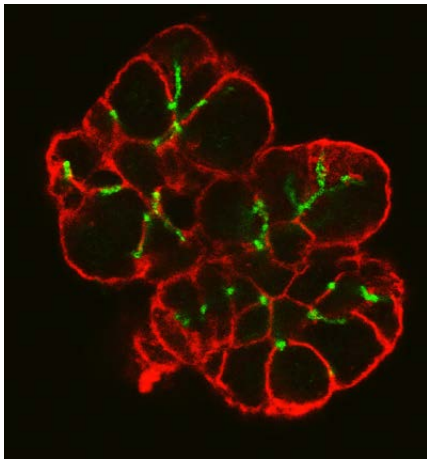


Figure 1: Colour coded digitised image slice.

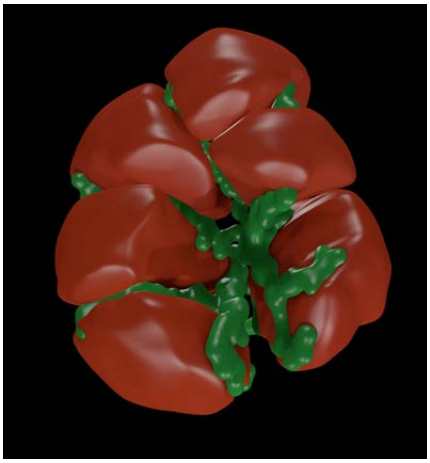


Figure 2: Full 3D mesh model of a cluster of cells.

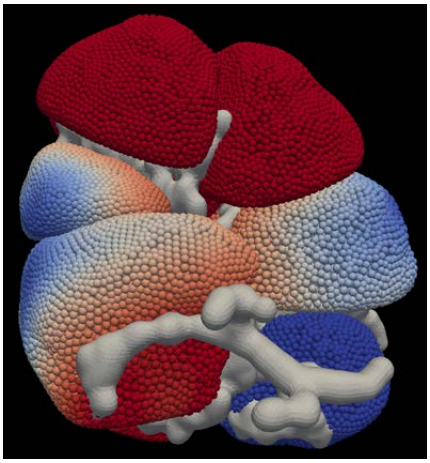


Figure 3: Simulation output snapshot.

Real biological samples where digitized using fluorescent markers and confocal microscopy. A sample image slice in which individual cell outlines can be seen is shown in Figure 1. The cell membranes are colour coded red and the interconnecting lumen is colour coded green. Note that, in living beings, the saliva secreted from the cells is transported through the assumed tube-like lumen structure. The full set of images slices was used as the basis for a full 3D graphics model reconstruction of one cluster of cells as shown in Figure 2. The tube-like structure of the lumen can now be clearly seen. This anatomically correct model was used in turn as the basis for the creation of a 3D tetrahedral mesh suitable for finite element simulations.

The same underlying 3D graphics mesh was used in the animated visualisation of the calcium concentration simulation time series results. One time series frame is shown in Figure 3. Through the NeSI Pan cluster was used for both graphics model rendering and running the finite element simulations. Thanks to the NeSI, we were able to render higher quality images and run many more simulation variants than would have been possible on a desktop computer. This facility will also enable us to scale up our model to include many more cells.

What's next

As expected, simulation results for each of the cells differ somewhat. Further work will include a detailed analysis of how cell geometry effects the generation and propagation of calcium waves within each cell. We also plan to construct a larger model based on new digitisations using refined microscopy techniques.

About the authors

Case study 1:

Ben Davies, PhD candidate
School of Social Sciences
Email: b.davies@auckland.ac.nz

Case study 2:

Alexei Drummond, PhD candidate
Bioinformatics
Department of Computer Science
Email: alexei@cs.auckland.ac.nz
BEAST 2 developer team:
https://github.com/CompEvol/beast2/graphs/contributors

Case Study 3:

Assoc Prof Klaus Lehnert
School of Biological Sciences
Email: klaus.lehnert@auckland.ac.nz

Prof Russell Snell
School of Biological Sciences
The Centre for Brain Research, FMHS
Email: r.snell@auckland.ac.nz

Dr Jessie Jocobsen, Research Fellow
School of Biological Sciences
Email: j.jacobsen@auckland.ac.nz

Brendan Swan, PhD candidate
School of Biological Sciences
Email: b.swan@auckland.ac.nz

Case study 4:

Alexei Drummond (see above)

Dong (Walter) Xie
Research Programmer
Department of Computer Science
Email: w.xie@auckland.ac.nz

Andrew Dopheide, PhD candidate
School of Biological Sciences
Email: a.dopheide@auckland.ac.nz

Case study 5:

Layne Price, PhD candidate
Department of Physics
Email: lpri691@aucklanduni.ac.nz

Case study 6:

Steen Wilhelm Knudsen, PhD candidate
School of Biological Sciences
Email: sknu003@aucklanduni.ac.nz

Case study 7:

Dr Stephane Guindon, Senior Lecturer
Department of Statistics
Email: s.guidon@auckland.ac.nz

Case study 8:

Dr Shi Bai, Postdoctoral Fellow
Department of Mathematics
Email: s.bai@auckland.ac.nz

Steven D. Galbraith, PhD candidate
Department of Mathematics
Email: sgeb522@aucklanduni.ac.nz

Case study 9:

Sarah Knight, PhD candidate
School of Biological Sciences
Email: s.knight@auckland.ac.nz

Case study 10:

Susann Beier, PhD candidate
Faculty of Medical and Health Science
Email: b.beier@auckland.ac.nz

Assoc Prof Brett Cowan
Director of MRI Centre
Department of Anatomy with Radiology
Email: b.cowan@auckland.ac.nz

Case study 11:

George Zhang, PhD candidate
Department of Chemical and Materials Engineering
Email: yechun.zhang@auckland.ac.nz

Prof Brent Young, Supervisor
Department of Chemical and Materials Engineering
Email: b.young@auckland.ac.nz

Case study 12:

Dr John Rugis, Computational Scientist
New Zealand eScience Infrastructure (NeSI)
Email: j.rugis@nesi.org.nz

Centre for eResearch and NeSI computational and technical support team

Aaron Hicks
Software and Systems Engineer, NeSI
Email: HicksA@landcareresearch.co.nz

Bart Verleye
eResearch Support, Faculty of Engineering
Email: b.verleye@auckland.ac.nz

Ben Roberts
HPC Application Scientist, NeSI
Email: b.roberts@auckland.ac.nz

Daniela Dunn
Tuakiri Representative, NeSI
Email: dm.dunn@auckland.ac.nz

Gene Soudlenkov
Technical Lead, HPC Services, NeSI
Email: g.soudlenkov@auckland.ac.nz

Grant Wooding
Technical Writer, NeSI
Email:grant.wooding@auckland.ac.nz

Greg Hall
System Engineer, NeSI
Email: g.hall@auckland.ac.nz

John Rugis
HPC Application Scientist, NeSI
Email: j.rugis@auckland.ac.nz

Jordi Blasco
HPC Application Scientist, NeSI
Email: j.blasco@auckland.ac.nz

Marcus Gustafsson
Services Delivery Manager, NeSI
Email: m.gustafsson@auckland.ac.nz

Markus Binsteiner
Software Developer, NeSI
Email: m.binsteiner@auckland.ac.nz

Martin Feller
eResearch Support, Faculty of Science
Email: m.feller@auckland.ac.nz

Matthew Healey
Project Coordinator, NeSI
Email: matthew.healey@otago.ac.nz

Nick Young
Web Developer, NeSI
Email: nick.young@auckland.ac.nz

Peter Higbee
Software and Systems Engineer, NeSI
Email: peter.higbee@otago.ac.nz

Peter Maxwell
HPC Application Scientist, NeSI
Email: peter.maxwell@nesi.org.nz

Robert Burrowes
Technical Lead, Research Data Services
Faculty of Science
Email: r.burrowes@auckland.ac.nz

Sina Masoud-Ansari
eResearch Support, Faculty of Science
Email: s.ansari@auckland.ac.nz

Yuriy Halytskyy
Software and Systems Engineer, NeSI
Email: y.halytskyy@auckland.ac.nz

Testimonies from the University of Auckland researchers

The Centre has conducted an annual survey regarding our support for researchers within the University. We concentrate on those active users on their experience in using High Performance Computing, RVM and Visualisation facilities. There were 226 surveys sent out during 2014 with 89% response rate from the University of Auckland researchers. Below were some of the feedback that we received from them.

Department of Chemical and Materials Engineering

The virtual machine is extremely powerful and the connectivity is brilliant. I never encountered any problems when I need to access the machine. It would be better if the user can have a bit few more options - for example user can reboot the VM whenever they need. The memory allocation is more than needed - I required 16G but was allocated 64G and I think it can be better used somewhere else. (rvmf00002)

Department of Civil and Environmental Engineering

The guys at the Centre for eResearch are lovely and professional. (uoa00056)

Excellent help from the support staff. (uoa00074)

I am very impressed by the level of support and dedication provided by the cluster support team. (uoa00174)

Department of Electrical and Computer Engineering

I have found access to NeSI extremely valuable in being able to run the very large electromagnetic simulations of building-scale environments. I look forward to using this facility in the future as this work is continuing and our techniques are evolving. We have a number of interesting problems to look at. (uoa00191)

Department of Engineering Science

Without NeSI resources this project would not be possible within the limited constraints of departmental computational resources, which provide little more than modest configured desktop compute facility. (uoa00224)

Department of Mechanical Engineering

At the moment my overall compute time is heavily dependent on queue time and licensing restrictions for LS-DYNA. My jobs run somewhat faster on the cluster but the non-SMP version of LS-DYNA we have, doesn't parallelize very well. Since my jobs take <30 hours in real time to complete on my desktop, a long wait on the cluster can negate its advantages. My future jobs will benefit from larger available memory on the cluster as my current jobs come close to running out of memory on my desktop. (uoa00080)

School of Medical Sciences

An essential aspect to our HPC access is the number of project that can be started. It must be noted that these are projects that use integrated computer and wet laboratory methods. Our main use is in the discovery of new molecules that inhibit the function of disease causing proteins. In this regard if the compounds investigated do not show any effect in the laboratory experiments, it will take longer for the project to show an outcome of the types listed. Although this is the case, access to HPC is vital to the project whether it is ultimately successful or not. This provides a unique post graduate student training opportunity for students in biomedical and biological sciences where they can develop multi-disciplinary projects that expose them to computational aspects. Success here is the completion of the project. For example team member Grace Gong is investigating the potential for inhibiting enzyme function by different mechanisms and has completed a number of virtual screens to help this. Without access this would have been impossible to do and the project would not have been offered. (uoa00040)

Bioinformatics Institute

The availability of the services is fantastic. In general, there is a problem with stability and reporting on the progress of long jobs. We have had to re-submit several jobs as (parts of the) cluster crashed, and this happened at times when the jobs had been running for 10 days or so and were anticipated to finish within the next couple of days. It is also difficult to tell when a job is not progressing, but that might be very difficult to report, too. In our case, we had our research programmer do all the work. Not having such a person available would basically mean that researchers would have to fiddle around with the submission of jobs, which is basically infeasible. I think that investing in support or investing in easier accessible submissions of jobs would increase the quality of the service by a large margin. (uoa00175)

Department of Mathematics

We appreciate very much the staff at the Centre for eResearch for help and support. (uoa00183)

Department of Physics

All is going well and with my students we are progressing towards creating finalized papers as research outputs. The ability to have large amounts of disk space has been a great help in allowing me to construct the synthetic populations and base stellare evolution models. (uoa00094)

Resolution of issues and assistance with getting jobs onto the cluster has been excellent! I have attended one SLURM training session and a one-on-one session which were both excellent! (uoa00255)

Department of Psychology

The cluster has been great for this project. I have been able to run a range of analyses simultaneously. This work is currently being prepared for publication. (uoa00163)

The cluster has been amazing for exploring a range of analyses quickly. We are writing up this work now and submitting it to PNAS. (uoa00164)

Department of Sport and Exercise Science

This is a great service and should be supported by the University. (uoa00006)

Department of Statistics

Pan is really useful for me to implement my large size simulation, the employees of eResearch are very supportive and helpful, I really appreciate all the helps that I could get, especially I do not have much experience and background in the computer science. (uoa00250)

School of Biological Sciences

This is a very useful project - it allows the project members (all have roles both in the University of Auckland and with LIC) to establish research projects that frequently develop into compute jobs that are executed under LIC's 'commercial' relationship with CeRES. This will likely increase with our migration to SLURM (same scheduler in both organisations). (uoa00192)

*We are great fans of the **unbureaucratic** support provided by the NeSI - ranging from solution to small newbie problems to support of our storage needs. The support for the recent migration to SLURM was especially helpful. This is invaluable to our projects and allows us to focus our efforts to our scientific problems without getting distracted by IT issues. Many thanks!* (uoa00001)

I am very grateful for the fast and helpful support I have been receiving from the e-Research group. I have had very helpful answers on both submission of jobs, installation of software and developing scripts for submission. I hope I will be able to get equally great help in the future as this has been extremely valuable to my research. Even though I have finished my PhD now and am residing in Europe I still collaborate with my PhD supervisor Kendall Clements on this project, and can work on the cluster remotely, which I think is a fantastic service and very supportive. I hope that I will be able to run both more jobs and run them faster with more jobs distributed across multiple CPUs. (uoa00029)

School of Chemical Sciences

I think the migration to SLURM has been great, and there has been huge support from the NeSI support team in helping to iron out the kinks in the job scripts. Also, having people who answer your requests for help in the middle of the night (which is when I'm most productive) has been a godsend. Running my calculations on the cluster has immensely improved my research capability and has allowed me to conduct a benchmark analysis of porphyrin-fullerene supramolecular interactions. With this data, I will be able to publish a paper which is of equal or better quality to those who are leaders in this field of chemistry. The calculations I run on the cluster require huge computing time and resources, and without access to the cluster, this research would be near impossible. (uoa00046)

School of Environment

Having access to this facility has helped immensely to advance my research, and to consider more ambitious modelling projects than I might have otherwise. The staff have been extremely helpful, and have been quite generous with their time and in lending their expertise. I have taken advantage of any seminars on GPU programming or parallelisation that have come along. Keep up the great work! (uoa00035)

Note: The end codes represent projects. Feedback provided by project team leaders.

Appendices

Journal Articles

Paul Hadwin. "Domain Truncation and Boundary Uncertainties in Electrical Impedance Tomorgaphy". Journal article in submission 2014.

Paul Hadwin. "Domain Truncation in Three Dimensional Eelectrical Impedance Tomorgaphy". Journal article in submission 2014.

Jessie Wu, Bouckaert, Remco, et al. "BEAST 2: a Software Platform for Bayesian Evolutionary Analysis." PLoS Computational Biology 2014.

Jessie Wu, Chieh-Hsi, Marc A. Suchard, and Alexei J. Drummond. "Bayesian Selection of Nucleotide Substitution Models and Their Site Assignments". Molecular Biology and Evolution 30.3, pp 669-688. 2013.

Buckleton, J., Bright, J.A., Taylor, D., Evett, I., Hicks, T., Jackson, G., Curran, J. M., "Helping Formulate Propositions in Forensic DNA Analysis. Science and Justice, 54 (4), 258-261, 2014.

Bright, J. A., Curran, J. M., Buckleton, J. S., "The Effect of The Uncertainty in The Number of Contributors to Mixed DNA Profiles on Profile Interpretation". Forensic Sci Int Genet, 12C, 208-214, 06.09. 2014.

Curran, J. M., & Buckleton, J., "Uncertainty in The Number of Contributors for The European Standard Set of loci". Forensic Sci Int Genet, 11, 205-206, 06.09. 2014.

Bright, J. A., Curran, J. M., & Buckleton, J. S., "Modelling PowerPlex Y Stutter and Artefacts. Forensic Sci Int Genet". 11, 126-136, 03.07. 2014.

Buckleton, J., Bright, J.A., Taylor, D. Evett, I., Hicks, T., Jackson, G., Curran, J. M. "Helping Formulate Propositions in Forensic DNA Analysis". Science and Justice. 02.07. 2014.

Taylor, D., Bright, J. A., Buckleton, J., & Curran, J. (2014). "An Illustration of The Effect of Various Sources of Uncertainty on DNA Likelihood Ratio Calculations". Forensic Sci Int Genet, 11, 56-63. 02.03. 2014.

Bright, J. A., Curran, J. M., & Buckleton, J. S., Erratum: Corrigendum to "Relatedness Calculations for Linked loci Incorporating Subpopulation Effects". Forensic Science International: Genetics, 7 (5), 565. 2013.

Bright, J. A., Curran, J. M., & Buckleton, J. S., Erratum: Corrigendum to "Investigation into The Performance of Different Models For Predicting Stutter". Forensic Science International: Genetics, 7 (5), 564. 2013.

Bright, J. A., Taylor, D., Curran, J., & Buckleton, J., "Searching Mixed DNA Profiles Directly Against Profile Databases". Forensic Science International: Genetics, 9 (1), 102-110. 2014.

Buckleton, J., Kelly, H., Bright, J.-A., Taylor, D., Tvedebrink, T., & Curran, J. M., "Utilising Allelic Dropout Probabilities Estimated by Logistic Regression in Casework". FSI: Genetics, 9(1), 9-11. 2014.

Bright, J. A., Taylor, D., Curran, J. M., & Buckleton, J. S., "Degradation of Forensic DNA Profiles. Australian". Journal of Forensic Sciences, 45 (4), 445-449. 2013.

Bright, J. A., Curran, J. M., & Buckleton, J. S., "Investigation into Stutter Ratio Variance". AJFS 2013.

Morrison, G. S., Lindh, J., & Curran, J. M., "Likelihood Ratio Calculation for a Disputed-Utterance Analysis with Limited Available Data". Speech Communication. Vol 58, Mar., 2014.

Bright, J. A., Neville, S., Curran, J. M., & Buckleton, J. S., "Variability of Mixed DNA Profiles Separated on a 3130 and 3500 Capillary Electrophoresis Instrument". AJFS Vol 46, 2014.

Bright, J. A., Stevenson, K. E., Coble, M. D., Hill, C. R., Curran, J. M., & Buckleton, J. S., "Characterising The STR Locus D6S1043 and Examination of Its Effect on Stutter Rates". FSI Genetics, 8(1), 20-23. Jan. 2014.

Bright, J. A., Curran, J. M., Hopwood, A. J., Puch-Solis, R., & Buckleton, J. S., "Consideration of The Probative Value of Single Donor 15-Plex STR Profiles in UK Populations and Its Presentation in UK Courts II". Science and Justice, 53 (3), 371. Sept 2013.

Kelly, H., Bright, J. A., Buckleton, J. S., & Curran, J. M.,"A Comparison of Statistical Models for The Analysis of Complex Forensic DNA Profiles". Science and Justice. 07.03.2013.

Puch-Solis, R., Rodgers, L., Mazumder, A., Pope, S., Evett, I., Curran, J., Balding, D., "Evaluating Forensic DNA Profiles Using Peak Heights, Allowing for Multiple Donors, Allelic Dropout and Stutters". FSI Genetics, 7 (5), 555-563. Sept 2013.

Curran, J. M. "Is Forensic Science The Last Bastion of Resistance Against Statistics?" Science and Justice. 53 (3) 251-252. Sept 2013.

Kelly, H., Curran, J. M., Bright, J.-A., & Buckleton, J. S., "Identifying and Modelling The Drivers of Stutter in Forensic DNA Profiles". AJFS. Vol 46 (2), 2014.

Booth, M. A., Vogel, R., Curran, J. M., Harbison, S., & Travas-Sejdic, J., "Detection of Target-Probe Oligonucleotide Hybridization Using Synthetic Nanopore Resistive Pulse Sensing". Biosens Bioelectron, 45, 136-140, July 2013.

Grand R., Pichugina T., Gehlen L., Jones M.; Tsai P., AllisoN J., Martienssen R., O'Sullivan J., "Global Analyses of Cell Cycle Dependent Changes in Fission Yeast Chromosome Organization Reveal The Sub-Nuclear Structure of The Schizosaccharomyces Pombe Genome". NAR (under review) 2014.

Eylem K., M. J. O'Sullivan, A. Yeh., " Three-Dimensional Model of The Deep Geothermal Resources in The Taupo-Reporoa Basin, New Zealand". JVGR, Vol 284, pp 46-6, Sept 2014.

C. Huang, M. Roncoroni, R. C. Gardner. "MET2 Affects Production of Hydrogen Sulfide During Wine Fermentation". Applied Microbiology and Biotechnology May 2014.

A Lu, S Guindon. "Performance of Standard and Stochastic Branch-Site Models for Detecting Positive Selection Amongst Coding Sequences". Molecular Biology and Evolution 31 (2), 484-495, Jan 2014.

L Ranjard, D Welch, M Paturel, S Guindon. "Modelling Competition and Dispersal in a Statistical Phylogeographic Framework". Systematic Biology, In Press. Mar. 2014.

Nathaniel J Burbery; Raj Das; W. G. Ferguson." Modelling with Variability: Dislocation Nucleation from Symmetric Tilt Grain Boundaries in Aluminium". Submitted, Computational Materials Science Jun 24, 2014.

Easther R, Frazer J, Peiris HV, and Price LC. "Simple Predictions from Multifield Inflationary Models". Phys.Review Letters,112, 2014.

Easther R, Price LC, and Rasero J., "Inflating an Inhomogeneous Universe", JCAP, Aug. 2014.

R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond "BEAST 2: a Software Platform for Bayesian Evolutionary Analysis". PLoS Computational Biology. Vol 10 (4) 2014.

J. Heled, R. Bouckaert, A. Drummond,W. Xie. "Estimating Species Trees from Multilocus Data", Schnolar July 2013.

R. G. Snell, E. W. Chua, S. Cree, N. Helsby, K. Lehnert, M. Kennedy. "A Next-generation-Sequencing Amplicon Assay for Multiplexed Analysis of The Pharmacogenes CYP2D6 and CYP2C19". (under review) 2014.

R. G. Snell:, E. W. Chua, S. Cree, M. L. Barclay, K. Doudney, K. Lehnert, A. Aitchison and M. Kennedy. "Exome Sequencing and Array Comparative Genomic Hybridisation Analysis of Preferential 6-Methylmercaptopurine Producers". (under review) 2014.

McIntosh, G. J., & Russell, D. K., "Molecular Mechanisms in The Pyrolysis of Unsaturated Chlorinated Hydrocarbons: Formation of Benzene Rings". 1. Quantum chemical studies. J Phys Chem A, 117(20), 4183-4197. 2013.

McIntosh, G. J., & Russell, D. K., "Molecular Mechanisms in The Pyrolysis of Unsaturated Chlorinated Hydrocarbons: Formation of Benzene Rings". 2. Experimental and kinetic modeling studies. J Phys Chem A, 117(20), 4198-4213. 2013.

McIntosh, G. J., "Theoretical Investigations into The Nucleation of Silica Growth in Basic Solution Part II–Derivation and Benchmarking of a First Principles Kinetic Model of Solution Chemistry". Phys Chem , 15 (40), 17496-17509, 2013.

Chen X., Iremonger K., Herbison A., Kirk V., and Sneyd J., " Regulation of Electrical Bursting in a Spatiotemporal Model of a GnRH Neuron". Bulletin of Mathematical Biology, Vol 75 (10), pp1941-1960, 2013.

Storey, R., Norris, S., & Cater, J. E. (2014). "An Actuator Sector Method for Efficient Transient Wind Turbine Simulation". Wind Energy, Mar 2014.

Grigor Aslanyan, Layne C. Price, Kevork N. Abazajian, Richard Easther, "The Knotted Sky I: Planck Constraints on The Primordial Power Spectrum", JCAP 08 (2014) 052.

K. N. Abazajian, G. Aslanyan, R. Easther, L. C. Price, "The Knotted Sky II: Does BICEP2 Require a Nontrivial Primordial Power Spectrum?". JCAP 08 (2014) 053.

G. Aslanyan, "Cosmo++: An Object-Oriented C++ Library for Cosmology", CPC Aug. 2014.

Petoe MA, Byblow WD, de Vries EJ, Krishnamurthy V, Zhong CS, Barber PA, and Stinear CM. "A Template-Based Procedure for Determining White Matter Integrity in The Internal Capsule Early After Stroke". NC 4, 2014.

Stinear CM, Petoe MA, Anwar S, Barber PA, and Byblow WD. "Bilateral Priming Accelerates Recovery of Upper Limb Function After Stroke: A Randomized Controlled Trial". Stroke 45: 205-210,2014.

Nelson, W.A., Sutherland, J.E., Farr, T.J., Hart, D.R., Neill, K.F., Kim, H.J., Yoon, H.S., "Multigene Analyses of New Zealand Coralline Algae: Corallinapetra Novaezelandiae". Gen.Et Sp. Nov. and recognition of The Hapalidiales Ord. Nov. Journal of Phycology - in review

Ward RM, Drover DR, Hammer GB, Stemland CJ, Kern S, Tristani-Firouzi M, Lugo RA, Satterfield K, Anderson BJ. "The Pharmacokinetics of Methadone and Its Metabolites in Neonates, Infants and Children". Pediatr Anesth 591-601, 2014.

Hannam J, Anderson BJ, Mahadevam M, Holford NHG. " Postoperative Analgesia Using Diclofenac and Acetaminophen in Children". Pediatr Anesth, EPub, May 2014.

Eaton MP, Keyvanjah K, Alfieris GM, Sweeney DM, Angona RE, Cholette JM, Venuto C, Forrest A, Anderson BJ. "The Pharmacokinetics of Epsilon-Aminocaproic Acid in Neonates Undergoing Cardiac Surgery with Cardiopulmonary Bypass". Anesthesiology 2014 (submitted)

Sherwin CMT, Stockmann C, Grimsrud K, Herd DW, Anderson BJ, Spigarelli MG. "Development of An Optimal Sampling Schedule for Children Receiving Ketamine for Short-Term Procedural Sedation and Analgesia". Pediatr Anesth Sept. 2014.

Allegaert K, Olkkola KT, Owens KH, van de Velde M, de Maat MM, Anderson BJ. "Covariates of Intravenous Paracetamol Pharmacokinetics in Adults". BMC Sept 2014.

Storset E, Holford N, Hennig S, Bergmann TK, Bergan S, Bremer S, et al. "Improved Prediction of Tacrolimus Concentrations Early after Kidney Transplantation Using Theory-Based Pharmacokinetic Modelling". BJCP Aug. 2014.

C. Walter, A.G. Jones and G.A. Ryan. "The Effects of Model Regularization in 3D Magnetotelluricinversions: An Example of The Taupo Volcanic Zone / New Zealand", Submitted to GJI, 2014.

Edwards, M., Meyer, R., Christensen, N.L. "Bayesian Parameter Estimation of Core Collapse Supernovae Using Gravitational Wave Simulations". Inverse Problem 30. Submitted July 2014.

Gao J, Sathar S, O'Grady G, Archer R, Cheng LK. "A Stochastic Algorithm for Generating Realistic Virtual Interstitial Cell of Cajal Networks". IEEE Trans Biomed Eng, 2014. (In submission)

Adrian Baddeley, Rolf Turner. "Bias Correction for Parameter Estimates of Spatial Point Process Models". Journal of Statistical Computation and Simulation. Published online 7 Jan. 2013.

Rolf Turner, Celeste Jeffs:. "A Note on Exact Monte Carlo Hypothesis Tests". Submitted for publication. 2014.

H. J. Carmichael, "Breakdown of photon blockade. '"A Dissipative Quantum Phase Transiton in Zero Dimensions", New Journal of Physics, in prepartion. 2014.

Richard Easther, Layne C. Price, Javier Rasero. "Inflating an Inhomogeneous Universe". JCAP, Jun 2014.

Kevork N. Abazajian, Grigor Aslanyan, Richard Easther, Layne C. Price. "The Knotted Sky II: Does BICEP2 Require A Nontrivial Primordial Power Spectrum?". JCAP Aug. 2014.

Grigor Aslanyan, Layne C. Price, Kevork N. Abazajian, Richard Easther. "The Knotted Sky I: Planck Constraints on The Primordial Power Spectrum". JACP, Aug, 2014.

Richard Easther, Jonathan Frazer, Hiranya V. Peiris, Layne C. Price. "Simple Predictions from Multifield Inflationary Models". Phys.Rev. Lett. 11 (2014)

G.Aslanyan, R. Easther. "Large Scale Anomalies in The Microwave Background: Causation and Correlation". Phys.Rev.Lett.111 (2013) 261301, v.2 Jan. 2014.

R. Easther. R. Flauger. "Planck Constraints on Monodromy Inflation". IOP Feb 2014.

R. A. Galvez. "Cosmo++: An Object-Oriented C++ Library for Cosmology". CPC July 2014.

R. Devendra, N.R. Edmonds, T. Sohnel. "Document Computational and Experimental Investigations of The Urethane Formation Mechanism in The Presence of Organotin (IV) Carboxylate Catalysts", Journal of Molecular Catalysis A: Chemical366, pp. 126-139, (2013).

R. Devendra, N.R. Edmonds, T. Sohnel. "Interaction Between Trialkyltin Alkoxide and Phenyl Isocyanate in The Formation of Tin Carbamate: A Computational and Experimental Study", Journal of Molecular Catalysis A: Chemical in press. Aug 2014.

Marie Fitch, M. Beatrix Jones, Helene Massam. "The Performance of Covariance Selection Methods That Consider Decomposable Models Only". Project Euclid, posted online 2014-03-14.

Tercel, M.; Pruijn, F. B.; O'Connor, P. D.;Liyanage, H. D. S.; Atwell, G. J.; Alix, S. M., "Mechanism of Action of AminoCBIs: Highly Reactive but Highly Cytotoxic Analogues of The Duocarmycins". Chem. Biochem. 2014.

Djanali V, Armfield SW, Kirkpatrick M, Norris. "Parallel Speed-up of Preconditioned Fractional StepNavier-Stokes Solvers" SEApplied Mechanics and Materials Jan. 2014.

S. E. Norris, M. P. Kirkpatrick and S. W. ArmfeldJ. "Prandtl Number Dependence and Instability Mechanism of The Near-Field Flow in a Planar Thermal Plumetae Hattori'. Journal of Fluid Mech. Oct 2013.

Kauermann, G., & Meyer, R. (2013). "Penalized Marginal Likelihood Estimation of Finite Mixtures of Archimedean Copulas". Computational Statistics, 1-24. Sept 2013.

Popinga, A., Vaughan, T., Stadler, T., & Drummond, Aj. (2014). "Bayesian Coalescent Epidemic Inference: Comparison of Stochastic and Deterministic SIR Population Dynamics". Submitted July 2014.

Denise Kuehnert: Denise Kuehnert, Tanja Stadler, Timothy G. Vaughan, Alexei J. Drummond. "Simultaneous Reconstruction of Evolutionary History Znd Epidemiological Dynamics from Viral Sequences with The Birth-Death SIR Model". Journal of the Royal Society, Interface Vol 11. May 2014.

M. Radjainia, Phillips et al., "Peroxiredoxin Is a Versatile Self-Assembling Tecton for Protein Nanotechnology". Biomacromolecules, Apr. 2014

Zhang S, G. Austin, L.Sutherland-Stacey. " The Implementation of Reverse Kessler Warm Rain Scheme for Radar Reflectivity Assimilation Using a Nudging Approach in New Zealand". AP JAS. Vol 50, 2014.

Gao, J., Sathar, S., O'Grady, G., Han, J., & Cheng, L. K. "Developmental Changes in Postnatal Murine Intestinal Interstitial Cell of Cajal Network Structure and Function". Ann Biomed Eng. Aug 2014.

Sathar, S., Trew, M. L., Du, P., O'Grady, G., & Cheng, L. K. "A Biophysically Based Finite-State Machine Model for Analyzing Gastric Experimental Entrainment and Pacing Recordings". Ann Biomed Eng. Apr. 2014.

Sathar, S., Grady, G. O., Trew, M. L., & Cheng, L. K. (2014). "A Biophysically-Based Tissue Model for Optimizing Gastric Pacing". 15th International Conference on Biomedi Eng. Vol. 43, pp. 16-19, 2014.

S. Wilhelm Knudsen, K. D. Clements, "Revision of the Fish Family Kyphosidae (Teleostei: Perciformes)"Zootaxa 3751 Dec. 2013.

Jon Francis Tunnicliffe. "1-Dimensional Simulation of Postglacial Valley-Fill Erosion" submitted in JOG Research: Earth Surface. 2014.

Shi Bai, Richard Brent, Emmanuel Thome. "Root Optimization of Polynomials in The Number Field Sieve". Mathematics of Computation, 2014

Shi Bai and Steven D. Galbraith. "Lattice Decoding Attacks on Binary LWE". Cryptology ePrint, 2014.

Shi Bai and P. Zimmermann. "Size Optimization of Sextic Polynomial in The Number Field Sieve". Math of Comp preprint 2014.

S. Gavryushkina, D. Welch, T. Stadler and A. Drummond. "Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration". Submitted to PLoS Computational Biology. 2014.

Hoang M. Nguyen, Burkhard C. Wnsche, Patrice Delmas, Christof Lutteroth, Wannes van der Mark. "A Robust System for High-Quality Reconstruction of 3D Objects from Photographs". Human-Computer Systems Interaction. Backgrounds and Applications 3, Springer, 2013.

Conference Proceedings

Ben Davies. "Pattern From Process: A Spatial Simulation Approach to Time- Averaged Surface Deposits". CAA Paris 2014.

K. Norris, Watson. "A Statistical Exploitation Module for Texas Hold'em and It's Benefits When Used With an Approximate Nash Equilibrium Strategy". Computational Intelligence in Games (CIG), 2013 IEEE Conference.

Richards, PJ and S. Norris, SE, "LES Modelling of Unsteady Flow Around The Silsoe Cube", 6th International Symposium on Computational Wind Engineering, June 8 - 12, 2014 Hamburg, Germany.

Richards, PJ, Hoxey, R, S. Norris, SE, Haspels, M, "Unsteady Flows Around a Cubic Building", UK Wind Engineering Conference, Sept 2014, Birmingham, UK.

K. B. Lehnert, R. Snell, et al, "Loss of Function Mutations Reveal that DGAT1 Is Essential for Gastrointestinal Homoeostasis, Lipid Absorption and Triglyceride Deposition in Humans and Cows, But Not Rodents". Annual Conference of the American Society of Human Genetics San Diego, Oct 2014.

Clarke, R. J., Calabretto, S., Walbran, S., Denier, J., Mattner, T. and Cater, J,., "Receptivity Mechanism in a Rotating Torus: Experiments and Simulations". Pittsburgh. US. Bull. Am. Phys. Soc., 2013.

H. Abbas, S. R. Alkhafaji, M. Farid. "CFD Investigation on Pulsed Electric Fields Treatment". iFOOD, Hanover, Germany Oct 2013. Abstract submitted to ISEF12 Qubic, Canada July 2014.

Nathaniel James Burbery, "Understanding The Threshold Conditions for Dislocation Transmission from Tilt Grain Boundaries in FCC Metals under Uniaxial Loading " , ACCM 2013 and Applied Mechanics and Materials 553: 28-34. Burbery N. J., D. R., Po G., Ghoniem N. (2014).

M. J. Neve, M. Leung, and J. Cater, "Indoor Wireless Communications - An Electromagnetic Compatibility Challenge", Proc. International Workshop on Antenna Technology (iWAT), Sydney, Australia, March 2014.

M. J. Neve, M. Leung, and J. Cater, "Inter-Building Propagation Modelling for Indoor Wireless Communications System Deployment", EuCAP, The Hague, Netherlands, Apr. 2014.

Bachtiar, Luqman R., Charles P. Unsworth, and Richard D. Newcomb. "Application of Artificial Neural Networks on Mosquito Olfactory Receptor Neurons for An Olfactory Biosensor." IEEE, 2013.

Carroll, Emma; O'Rorke, Richard; Sewell, Mary; Scheltema, Emma; Dassanyayke, Asela; Ross, Howard; Zeldis, John; Constantine, Rochelle. "Plankton to Pooh: Ecosystem Approach to Investigating Whale Diet", 20th Biennial Conference on Biology of Marine Mammals, Dunedin, New Zealand, Sept 2013.

H. Koehler, U. Leck, S. Link and H. Prade. "Logical Foundations of Possibilistic Keys". Proceedings,14th European Conference, JELIA 2014, Portugal, 2014.

H. Koehler, S. Link, H. Prade, X. Zhou,"Cardinality Constraints for Uncertain Data". 33rd International Conference on Conceptual Modeling (ER 2014), ranked A by ERA. To appear as a volume in Springer's Lecture Notes in Computer Science.

Dominik Vogt, Jessienta Anthony, Rainer Leonhardt. " Broadband THz Guidance in Helical Waveguides", 39th IRMMW-THz Sept 14-19, 2014, Tucson, Arizona, USA.

B. Ragued, L.W. Wotherspoon & J.M. Ingham "Seismic Response of a Typical New Zealand Pile-supported Wharf Configuration" Technical Conference and AGM 21-23 Mar, 2014, Aotea Centre, Auckland.

C. Rivera, T. Lumey, "Countermatched Design as a Two-phase Sampling Design". Mandurah, Australia. Oral presentation, IBS Conference 1-5 Dec 2013, W. Australia.

H. J. Carmichael, "Breakdown of Photon Blockade: A Dissipative Quantum Phase Transition in Zero Dimensions", 10th Rochester Conference on Coherence and Quantum Optics, New York, USA, June, 2013.

Williamson N, Armfield SW, Kirkpatrick MP, Norris. "A Canonical Model for Stratified Flow in Estuaries and Rivers". ANZIAM Journal, V52 2013.

Yusuf B.O., Naresh S., Simon S., L. James W. and Jhannes R. "Transformation of Bisphenol A, Triclosan and Nonylphenol by The Fe-TAML/H2O2 System". The 248th ACS National Meeting USA, Aug, 2014.

S. Beier, J. Ormiston, M. Webster, J. Cater, P. Medrano-Gracia, A. Young, B. Cowan. "Coronary Artery Bifurcation Haemodynamics - Comparison between Phase ContrastMR and Computational Fluid Dynamics" SCM, USA, 2014.

S. Beier, J.Ormiston, M. Webster, J. Cater, S. Norris, P. Medrano-Gracia, A. Young, B. Cowan. "Ex-Vivo Stented Coronary Artery Haemodynamics Using 4D Flow Measurements and Computational Fluid Dynamics (CFD)". CSANZ Conferences, Australia, 2013.

Youming Chen, Raj Das* and Mark Battley. "Modelling of Closed-cell Foams Incorporating Cell Size and Cell Wall Thickness Variations". Proceedings, 11th World Congress on Computational Mechanics, 20-25 July 2014 Barcelona, Spain.

Miguel Roncoroni. "Mapping Genes of Oenological Importance in Commercial Winemaking Yeast". 26th International Conference on Yeast Genetics and Molecular Biology (Frankfurt) 2013. Also, in Yeast: Products, Discovery (Auckland) 2013 Winner of best student presentation award.

Miguel Roncoroni. "Identifying Genes of Oenological Relevance in Winemaking Yeast". 15th Australian Wine Technical conference 2013.

Miguel Roncoroni. "A Sequenced Yeast Cross for QTL Mapping". 5th NZ NGS Conference 2013.

Shi Bai: Shi Bai, Steven Galbraith. "An Improved Compression Techniquefor Signatures Based on Learning with Errors". CT-RSA, 2014.

John Benjamin Weber: Abstract accepted "Effect of Curvature On Structural Loading of Prismatic Bodies Subjected to Water Slamming". 8th ACAM Nov 2014.

Oral presentations

Laith Hurmez. "Comparison Between The Lattice Boltzmann and Finite Element Method for Two Capillary Driven Flow Problems" EMAC2013.

Suzanne Jean Reid. Result on "Comparison Between The Lattice Boltzmann and Finite Element Method for Two Capillary Driven Flow Problems" presented in Lab Group Meetings 2014.

Ben Davies. "The Virtual Archaeologist: Using Spatial Simulation to Connect Behavioral Models with Data in Arid Australia". University of Vermont, Mar 2014

Ben Davies. "Modeling Archaeology: Using Spatial Simulation to Connect Behavioral Models with Data in Arid Australia". Boston University, Apr 2014

Ben Davies: "Anomalous Country: Contrasting Prehistories in The Australian Outback". Plymouth State University, Apr, 2014.

Victoria Twort, Richard Newcomb, Howard Ross and Thomas Buckley. "Comparative Genomics and Transcriptomics of New Zealand Weta". Evolution June 2014, North Carolina.

Victoria Twort, Richard Newcomb, Howard Ross and Thomas Buckley. "Draft Sequencing and Assembly of The New Zealand Giant Weta". Genome.Genetics Society of AustralAsia Conference, Sydney, 2013

Denier, J., Clarke, R. J., Hewitt, R. and Hazel, A. "The Unsteady Flow within a Rotating Torus". European Turbulence Conference 14, Lyon. France, Sept 2013

Russell Grant Snell,J. Jacobsen. Oratia Primary School: Autism and science 2014.

Russell Grant Snell, J. Jacobsen. "Genetics and Genomics in Neurodegenerative Conditions" Newmarket Rotary Club 2014.

Russell Grant Snell, J. Jacobsen:: "Advanced Genomics Research to Understand Neurodegenerative Conditions". Rotary Science and technology Forum 2014.

Sarah Jane Knignt. "Yeasts: Products, Discovery and More" YPD meeting 2013.

Sarah Jane Knight. "Advancing Analyses of Microbial Population Structure: Quantifying Gene-Flow and Connectivity in New Zealand's Saccharomyces Cerevisiae". EcoTas Conference 2013.

Grigor Aslanyan." Te Knotted Sky" papers at Tufts/MIT joint Cosmology Seminar, CFA at Harvard University, IAS, Yale University, Stony Brook University, UC San Diego 2014.

Tet Chuan Lee. "Modelling The Microcirculation Using Boundary Element Methods" Fluids in New Zealand (FiNZ) 2014.

Philip Sharp: Sharp, P.W., "Numerical Experiments with The N-body Code GENGA", New Zealand eResearch HPC applications workshop, University of Waikato, June, 2014.

Philip Sharp: Sharp, P.W., "Dynamical Delivery of Volatiles to The Outer Main Belt", ANZIAM, Jan 2014.

Philip Sharp: Sharp, P.W., "Dynamical Delivery of Water Ice and Organic Compounds to The Outer Asteroid Belt", SAWONA conference, KL, Malaysia, Apr, 2014.

C. Walter, A.G. Jones. "Comparing Multiple 3D Magnetotelluric Inversions of The Same Dataset". AGU Fall Meeting 2013

Grace Gong "Structure-Based Design of P110a-Specific PI3K Inhibitors, and Inhibitors That Are Proposed to Target an Alternative Pocket Present in The Kinase Domain of Pi3K Alpha". ACSRC seminars 2014.

Gwynn Sturdevant. Related to project of "Statistical Modelling of Carryover Effects after Cessation of Treatment" at Australasian Epidemiological Association 2014

Gwynn Sturdevant. Related to project of "Statistical Modelling of Carryover Effects after Cessation of Treatment" at New Zealand Statistical Association 2014

Gwynn Sturdevant. Related to project of "Statistical Modelling of Carryover Effects after Cessation of Treatment" at International Biometric Society 2014

Yevhen Mohylevskyy. "2-Dimensional Voter Model Chordal Interface", NZ Probability Workshop, , Te Anau, Jan 2014

H. J. Carmichael and S. Whalen, "Quantum trajectories Wthout Lindblad", Frontiers of Quantum and Mesoscopic Thermodynamics, Prague, Czech Republic, 28 July - 3 August, 2013.

H. J. Carmichael, "Breakdown of Photon Blockade: A Dissipative Quantum Phase Transition in Zero Dimensions", 7th Annual Dodd-Walls Symposium, Dunedin, New Zealand, 11-13 November, 2013.

H. J. Carmichael, "Dissipative Quantum Phase Transitions for Photons", ANZ Conference on Optics and Photonics, Perth, Australia, Dec 2013.

H. J. Carmichael, "Dissipative Quantum Phase Transitions for Photons", NZIAS Symposium: Physics of Complex Systems, Albany, New Zealand, Dec 2013.

H. J. Carmichael, "Open Quantum Systems I-IV", Victorian Summer School in Ultracold Physics, Swinburne University, Melbourne, Australia, Jan 2014.

V. S. C. Canela and H. J. Carmichael, "Manipulation of The Photon Number Dynamics in a Micromaser", 23rd Annual International Laser Physics Workshop, Sophia, Bulgaria, July 2014.

H. J. Carmichael, "Breakdown of Photon Blockade: A Quantum Phase Transition of The Driven Jaynes-Cummings Model", Weizmann Institute of Science, Rehovot, Israel, Jun 2014.

Meyer, R., & Kirch, C., "Bayesian Semiparametric Likelihood Approximations for Stationary Time Series". eJoint NZSA+ORSNZ Conference, Hamilton, Nov 2013.

Renate Meyer. "Bayesian Semiparametric Likelihood Approximations for Stationary Time Series". Seminar, University of Dortmund, Germany, August 2013.

Meyer, R., & Kirch, C. "Likelihood Approximations for a Bayesian Semiparametric Analysis of Stationary Time Series". Seminar, University of Sydney, Sept 2014.

P. Wameyo, E. Aboud, M. F. Abdelwahed, J. Cherrington, J. Hoeberechts, C. L. Kenedi, Jan M. Lindsay, M. R. H. Moufti "Electromagnetic Imaging of Harrat Rahat" VORISA Scientific Meeting Nov. 2013 Saudi Arabia.

Susann Beier. "Stented Coronary Hemodynamics Using 4D Flow Measurements and Computational Fluid Dynamics (CFD)". HealtheX, FMHS UoA, 2013.

Mazdak Radjainia et al. "The Structure of a Human Peroxiredoxin Filament". Queenstown Molecular Biology Meeting, 2014.

Sasha Gavryushkina. "Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration". Epidemics 2014, Amsterdam, the Netherlands.

Sasha Gavryushkina. "Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration". The18th Annual New Zealand Phylogenomics Meeting, Waiheke 2014.

Attique Ur Rehman. Project related to "Modelling The Long-Term Evolution of The Large Satellites of Uranus", NZMS Colloquium 2013

Attique Ur Rehman. Project related to "Modelling The Long-Term Evolution of The Large Satellites of Uranus", Conference Talk at ANZIAM 2014.

Attique Ur Rehman. Project related to "Modelling The Long-Term Evolution of The Large Satellites of Uranus", Conference Talk at WCE, UK 2014.

Attique Ur Rehman. Project related to "Modelling The Long-Term Evolution of The Large Satellites of Uranus", Conference Talk at SIAM Annual Meeting USA 2014.

Patent

Wannes van der Mark. French Patent Application 1457862 (c. 18 August 2014)

Posters

Renee Marie Miller. Poster at the Auckland Bioengineering Research Forum. 2014.

R. Meyer. The best Poster Award at the 33rd International Workshop for Bayesian Inference and Maximum Entropy Methods in Science and Engineering, in Canberra, Australia. Dec. 2013.

Pichugina T., Grand R., Allison J., Gehlen L., O'Sullivan J. "Coarse-grained3D Genome Model of Schizosaccharomyces Pombe Created Using Chromosome Interaction Data". CellML workshop, Auckland 2014.

Ransi Devendra. "Catalytic Effect of Organotin Carboxylate on Urethane Formation in The Presence of Acetic Acid: An Experimental and Computational Investigation". Chemistry Showcase, UoA, 2014.

Levi Li. Poster presentation " Proteins & Nanoparticles at Membranes" Research Center Juelich, 2014.

Andy Xindi Wang. Poster presentation at RACI Physical Division 2013.

S. Gavryushkina, D. Welch, and A. Drummond. "Bayesian MCMC for Trees with Sampled Ancestors". The 4th International Conference on Infectious Disease Dynamics, Nov 2013, Amsterdam, The Netherlands.

Sarah Jane Knight: Poster at Wine Industry Conference, Bragato 2013

Mazdak Radjainia et al. "The Structure of a Human Peroxiredoxin Filament". Presented at Gordon Research Conference on 3DEM in Girona, 2014.

J.C.Jacobsen, H. Tsang, T. Love, R. Taylor, Snell, Lehnert et al. About the mutation identification in New Zealand. "A Cohort for Researching Autism Genetics in New Zealand". ASHG San Diego Oct 2014.

Nick Holford. "Power and Type 1 Error of Tumour Size Metrics Used to Predict Survival". The Annual Meeting of The Population Approach Group in Europe. ISSN 1871-6032.

Book

Howard John Carmichael: H. J. Carmichael and S. Whalen, "Quantum Trajectories Without Lindblad", Frontiers of Quantum and Mesoscopic Thermodynamics, Prague, Czech Republic, 29 July - 3 August, 2013.

Grants

Jon Francis Tunnicliffe: Vice-Chancellor's Learning Enhancement Grant. Proposal has been submitted 2014.

John Edward Cater: MBIE Smart Ideas. Proposal no. PROP-37861-SIP1-UOA 2014 is successfully funded.

Howard John Carmichael: Marsden Fund (2014-16). H. J. Carmichael (PI), A. S. Parkins (AI), and L. A. Orozco (AI), "New Directions in the Quantum Theory of Photo-Emissive Sources",

Jack Flanagan: Perry J. " Inhibiting The Human GH Receptor with Small Molecule Antagonists" HRC 14/708 (2013-2014)

Russell Grant Snell: Faculty Research Development Fund for "Sequencing of Clinical Genomes and Exomes"

Russell Grant Snell: Analysis supports the Rutherford Fellowship to Jessie Jacobsen. "Using High Performance Computing to Investigate Human Genomic Information for Human Health" 2014.

Maarten Hoogerland, Donald Hylton White: "Behaviour of Ultracold Atoms", Marsden fund 2013-2015.

Software Packages

Grigor Aslanyan: Cosmo++ publicly available at http://grigoraslanyan.com/cosmopp

Holford NHG. Wings for NONMEM Version 733 for NONMEM 7.3 and NONMEM 7.2. http://wfn.sourceforge.net

Theses

Paul Hadwin: Applied the Bayesian approximation error (BAE) approach to treat unknowns present in biomedical electrical impedance tomography (EIT).

Olga Perederieieva: Algorithms for multiobjective network equilibrium problems.

Steffen Klaere: Influence measures site-by-site in phylogenetic inference and exploring rate heterogeneity in genomes.

cheng liu: Simulate gentrification from both supply and demand side perspectives

Sina Masoud-Ansari: Ground motion simulation and time series prediction using feature analysis.

Jessie Wu: Bayesian approaches to model uncertainty in phylogenetics.

Chris David Mathieson: The outcome of physical testing of a novel cold-formed steel connection and finite element modelling of the specimens tested.

Tai-Ying Tom Tu: Numerical studies of dynamic stall on the thick-root S814 turbine foil.

Howard Ross: Comparison of metagenomic analysis pipelines including sequence pre-processing and clustering techniques

Howard Ross: A comparison of metabarcoding data analysis pipelines.

Steen Knudsen: Phylogeny and phylogeography of the family Kyphosidae (Perciformes: Teleostei)

Tom Allen: Mechanics of flexible composite hull panels subjected to water impacts.

Anna Maria Matuszek: Defining known drug space by DFT based molecular descriptors. Virtual screening for novel Atg5-Atg16 complex inhibitors for autophagy modulation.

Eylem Kaya: Desy caesary title: Natural state model calibration for a deep natural convective model of part of the taupo volcanic zone (tvz).

Tet Chuan Lee: Near-wall microfluidics in the microcirculation.

Pavel Petrovich Sumetc: Modelling vascular vessel walls.

Levi Li: Electrostatic modelling of antimicrobial proteins.

Brendan swan, PhD thesis is supported by project uoa00001 using HPC to investigate human genomic Info for human health.

Howard Ross: Reconstructing the demographic history of adlie penguins (pygoscelis adeliae) using mtdna and coalescent methods.

Chris David Mathieson. "Novel Pin-Jointed Connection for Cold-Formed Steel Trusses" A summary of the thesis research.

Other research outcomes

Jim Denier: refereed archival journal publications in server leading journals.

Suzanne Jean Reid: Presented results from PLINK analysis of genotyping data in lab group meetings.

John Eldridge: A new set of stellar evolution tracks that have been combined to create synthetic stellar populations including binaries.

Klaus Bernd Lehnert: Ongoing collaboration with LIC and attracted LIC as a commercial client.

Johannes Reynisson: Project 1 Computational studies of metallabenzenes completed. Two manuscripts in preparation. Projects 2. New computational studies to find stable intermediates in the reaction mechanism that could be isolated experimentally. Two maunscripts in preparation.

Stuart Norris: The research output at this stage is just two Engineering Part 4 Project Reports

Badar Al-Nasri: Investigation of the effects of coil geometry on heating of a hot water cylinder, Mechanical Engineering Project Report2013-ME-1

Mathew Peacock: Investigation of the effects of coil geometry on heating of a hot water cylinder, Mechanical Engineering Project Report2013-ME-49

Russell Grant Snell: Project uoa00001- Using High Performance Computing to investigate human genomic information for human health. Collaborations with researchers at University of Otago, Harvard University, and Garvan Institute (Sydney).

Russell Grant Snell: Collaboration with several clinical practitioners at Auckland Hospital to sequence and analyse the exome and genome sequences of undiagnosed patients.

Russell Grant Snell: Community engagement: Newsletter to "Minds for Minds" registrands (>1,000 individuals across New Zealand).

Russell Grant Snell: Community awards: Finalist in Westpac "Women of Influence" (Jessie Jacobsen - uoa00001)

Stuart Norris: Gatland, A. "Heat Transfer from a Porous PlateMechanical". Engineering Part 4 Project ReportME29-2013.

Stuart Norris: Ware, M. "Heat Transfer from a Porous PlateMechanical" Engineering Part 4 Project ReportME74-2013, 2013

John Edward Cater: Supported two Part IV Engineering Science student projects (Michael Hanks & James Hou). Collaborative research using the vocal tract between R. Clarke, C. Watson and J. Cater. A related PhD thesis is expected to be submitted at the end of 2014.

Pierre-Marie Bonnemason: Internship report

Cenanning Li: Project related to simulation of information theoretic criteria for least-squares trees for thesis expected to be published in 2015.

Benjamin Lu: Models for Part 4 Project related to CFD analysis using ANSYS, Abaqus and OpenFOAM.

Rolf Turner: Monte Carlo Methods of Testing Hypotheses about the Parameters of Point Process Models. Celeste Jeffs, Summer Scholarship project, University of Auckland, 2011/2012.

Rolf Turner: Article in preparation; to be submitted to the Journal of Statistical Computation and Simulation;article title "Adjusted Composite Likelihood Ratio Test for Gibbs Point Processes". Authors: Adrian Baddeley, Rolf Turner, Ege Rubak.

Richard Anthony Galvez: Layne Price has given a conference presentation at Cosmo2013, University of Cambridge (September, 2013); seminars at: University of Nottingham University of Sussex University of Valencia University of Barcelona Cornell University MIT.

Richard Anthony Galvez: Seminars and Colloquia (Richard Easther) 2013 August Victoria University; Invited Plenary September Cosmo 2013; Keynote speaker at New Zealand Institute of Physics October Massey, Albaby 2013; Keynote at Colloquium, Tokyo University January 2014; Monash University, Queensland, Apr 2014.

Richard Anthony Galvez, Grigor Aslanyan. "Primordial Power Spectrum pre- and post- BICEP"', Tufts/MIT Joint Cosmology Seminar Harvard University, Apr 2014; Institute for Advanced Study, Apr 2014; Yale University, Apr 2014; Stony Brook University, Apr 2014, UC San Diego, May 2014.

Ewan Tempero: The Qualitas Corpus is a curated collection of software systems intended to be used for empirical studies of code artefacts

Susann Beier: Runner-Up Award for Emerging Researcher of the Year, FMHS University of Auckland, 2013

Rajnish Narayan Sharma: Intern report

Matthieu Detienne: Intern report by Matthieu Detienne

Rui Gong: Conference paper is in progress.

THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE

Postal address
Centre for eResearch
Computer Science
University of Auckland, Private Bag 92019
Auckland 1142, New Zealand

Physical address
Centre for eResearch
Room G021, LG, Building 409
24 Symonds Street, Auckland 1010
New Zealand

Mark Gahegan, Director
Phone: +64 9 373 7599 ext 81891 | **DDI**: +64 9 923 1891
Email: m.gahegan@auckland.ac.nz

www.eresearch.auckland.ac.nz