BEAR

BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

# Case Study Vol. 3

# Copyrights & Acknowledgments


**Published in the United Kingdom by The University of Birmingham, Edgbaston, Birmingham, B15 2TT,UK.**



**Aslam Ghumra,**
**The University of Birmingham,**
**IT Services, Building G5**
**Edgbaston,**
**Birmingham, B15 2TT, UK.**
**email: bearinfo@contacts.bham.ac.uk**

**Acknowledgments**
**We would like to thank everybody who contributed to this publication. Special thanks to Kirsty McCready for her help in editing.**

**Cite**

**DOI : https://doi.org/10.25500/epapers.bham.00004303**

# Foreword


## Professor Andrew Morris

***Chair of Research Computing Management Committee***


**The indispensable role of high-performance computing is more apparent than ever in research and innovation. This book of case studies offers a glimpse into the remarkable exploits of BEAR users, who have harnessed the capabilities of our state-of-the-art facilities and expertise to expand the horizons of knowledge and make substantial contributions to science. From illuminating new insights in the realm of climate science to unraveling DNA, these narratives stand as a testament to the transformative capacity of advanced computational resources at Birmingham and underscore the profound creativity and collaborative ethos of the individuals and teams who make use of these resources.**

# Abstract

**This collection of case studies was brought together to showcase the extent and diversity of research that is supported by the University of Birmingham's Environment for Academic Research (BEAR). BEAR is a collection of contemporary IT resources designed to help research. The following case studies demonstrate how BEAR services such as the Research Data Store (RDS), BEAR software and the University supercomputer BlueBEAR are integral to the progression of important research across disciplines.**

**BlueBEAR is a key component of BEAR, providing compute power and specialist applications for free to enable staff and students to delve deeper into their research. Upgraded in 2023, the cluster includes many large memory nodes and a GPU service alongside standard compute nodes. Alongside BlueBEAR, the RDS is a popular choice amongst researchers to securely store their working research data.**

**As of publication, more than 5,000 researchers across all five colleges were actively using BlueBEAR and/or the RDS. In this volume, we showcase case studies representing diverse research from every college. From estimating snow coverage to modelling second language acquisition, we show how BEAR services are enabling exciting and important research across the university.**

**Left: A look inside the Data Centre that houses BlueBEAR.**
**Right: A closer look: internal workings of a HPC node**

# Contents

# Contributors

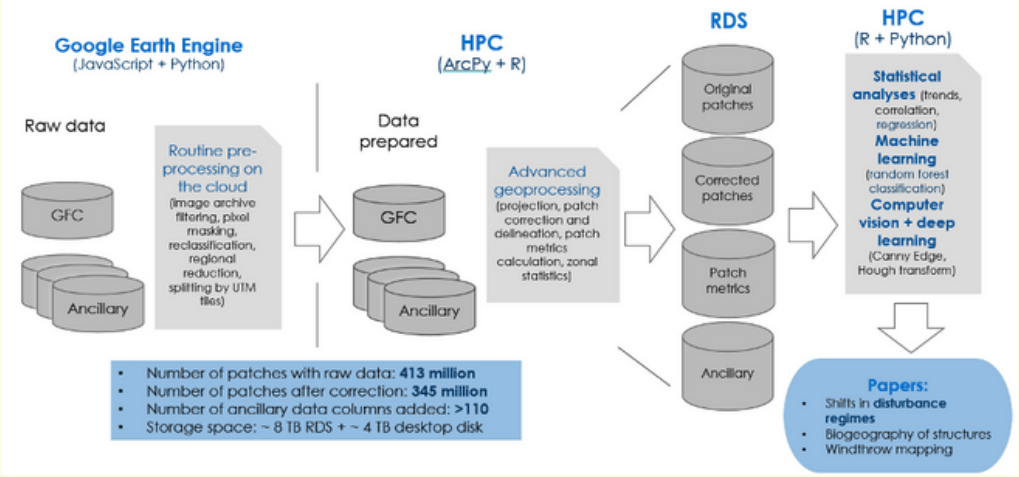
# Database of global forest disturbance data


**Nezha relies on the RDS, BlueBEAR and BEAR training for her work on global forest disturbance.**

My project aims to characterise patterns of forest disturbance at the global scale and identify storm-related damages based on remote sensing data. Considering the sheer amount of data needed to describe the structure, context and properties of forest disturbances at a fine spatial resolution, we had to address a number of technical challenges relating to data storage and processing scalability at the global level.

The project required the use of multiple data hosted online on different repositories. I first used Google Earth Engine (GEE) as a single place to access archives and apply routine preprocessing on the cloud. As I needed more flexibility for customizing algorithms, I transferred GEE outputs to the RDS, so I could use BlueBEAR. This gave me the possibility to use tools from ArcGIS's, Python's and R's libraries, process data with up to 500 GB of RAM and massively parallelise job tasks on clusters.

Throughout my PhD, I have greatly benefited from the BEAR training courses. The introductory courses were a catalyst to grasp how to use Linux in supercomputers. I consolidated my background in R and Python and learned how to sync RStudio projects to GitHub with the Software Carpentries workshops. NVIDIA Deep Learning clarified a direction I am envisaging for my research on the use of artificial intelligence for forest disturbance agent recognition. More recently, I completed the NVIDIA Accelerating Data Engineering Pipelines course, which helped me explore ways to optimize my data processing workflow.



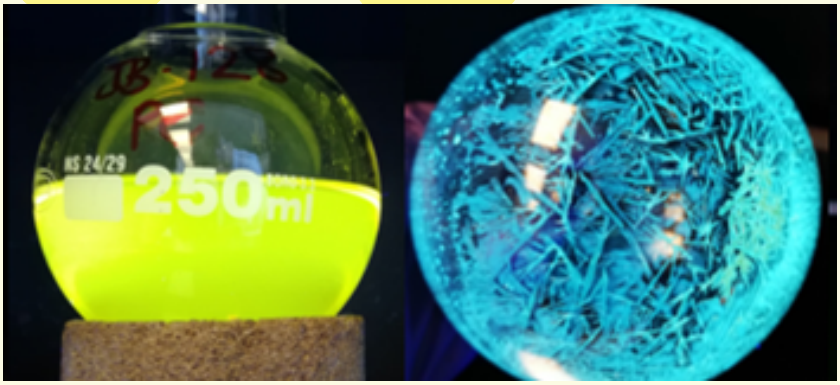

**A summary of the ways BEAR enables Nezha's research**

# The design and synthesis of new fluorescent probes

**Jake uses BlueBEAR to enable his research into the design and synthesis of new fluorescent probes.**


As part of my project, I use computational methods to help design and synthesise new fluorescent probes. Fluorescent probes are a type of fluorophore (a chemical compound that can emit light upon excitation) that is used to detect or measure various biological processes or chemical reactions. These probes are incredibly useful as they can be designed to fit a particular niche or application. We are aiming to develop next-generation fluorophores which can be tuned to exhibit a desired set of characteristics, for example the colors of light the probe absorbs and emits. These characteristics are called photophysical properties. Computational chemistry is a key factor in this.

I use a computational method called Time-Dependent Density Functional Theory (TD-DFT) to screen and design potential fluorophores. This method is the application of quantum mechanical theories, which when applied to a molecule, can be used to probe the dynamics, processes, and properties when the system is in the presence of stimuli like electric or magnetic fields. A standard TD-DFT calculation for a simple molecule of interest on a regular laptop would take around 32 hours, whereas the same calculation using BEAR resources often takes less than 25 minutes, which is an incredible difference! By using BLUEBEAR resources, I have been able to perform over 100 calculations on a range of molecules, which has given me a greater insight and understanding into the workings of my maleimide derivatives. This process has helped inspire new synthesis, streamline my workflow, and save a lot of time from synthesising undesired molecules.




Some fluorescent maleimide derivatives synthesised with the help of BlueBEAR.

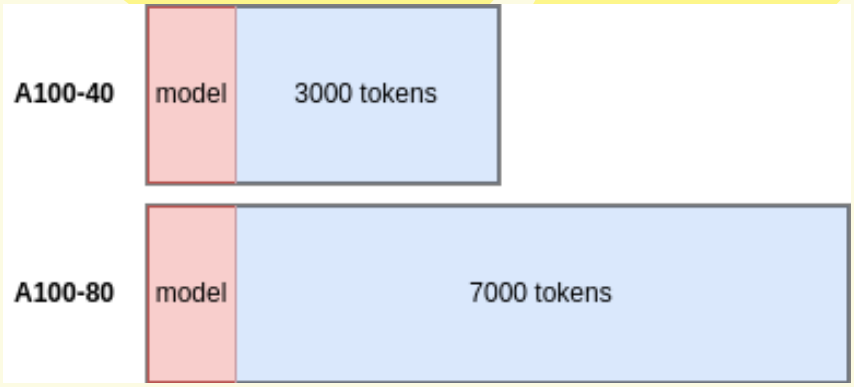

# Baskerville pushes the boundaries

**Christos has been making use of Baskerville to enable his research in text generation using machine learning models.**

Our research is on machine translation (MT), which is the technology used by tools like Google translate to automatically translate text from one language to another. MT is an important tool for many people who want to communicate with others who do not share a common language, for example for those who work or study in another country, for cross-border commerce, or for international relief efforts.

MT works well for languages where there is lots of text available e.g. European languages and Chinese. To improve MT for lesser-known languages, multilingual MT (MMT) can be used. An effective MMT system requires a high capacity neural network, that is one with many parameters, which means it is more resource intensive to train than pairwise MT.

Baskerville's many GPUs with large memory allow us to carry out our work training and improving MMTs. Having A100 GPUs means that training steps run faster than on earlier types of GPUs. The larger memory GPUs enable us to train with the large capacity models required for MMT, without resorting to model parallelism (splitting the model across multiple GPUs), which would slow down training and make it more complex. The A100 GPUs also enable us to use mixed or half-precision (also known as FP16) training, which further significantly reduces memory consumption, thus enabling us to use larger models and batch sizes.



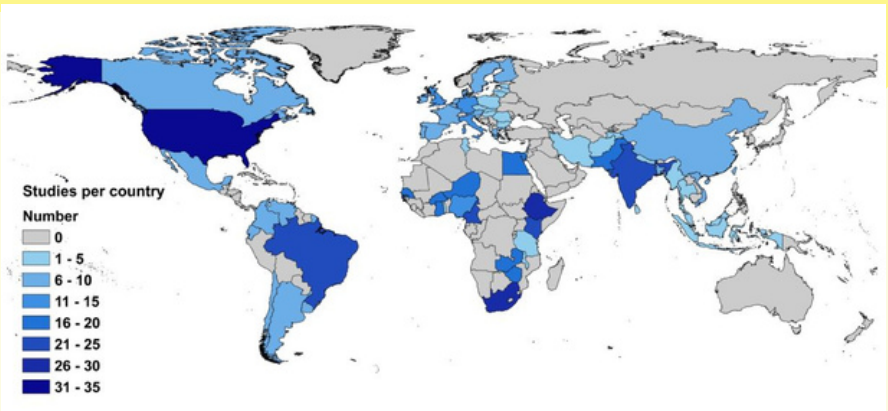

**Baskerville's larger memory GPU enables us to more than double the batch size processed in each iteration**

# Estimating the global impact of climate change on agriculture


**Allan has been making use of BlueBEAR to enable his research into estimating the impact of climate change on agriculture.**

I work with colleagues at Yale University and the National Autonomous University of Mexico to estimate the impact of climate change on agriculture at the global level. This work is based on a sample of 227 studies that rely on the Ricardian Model. The Ricardian Model analyses the impact of climate on land rents and therefore incorporates adaptation to climate change in the decision making of profit maximising agents.

The use of BlueBEAR and the Research Data Store has been essential to carry out this project to the high level of resolution we require. Our estimations involve over 48 million observations and maximum likelihood methods that require large computing power for the estimation and mapping of results. Each of the points representing a square kilometer of cropland contains information on temperature, precipitation, agroecological conditions, elevation and the level of irrigation infrastructure. With the help of BlueBEAR, we are able to get Ricardian estimates of the impact of climate change on agriculture at the global level for the first time.




**227 studies across 71 countries are used in Allan's study**

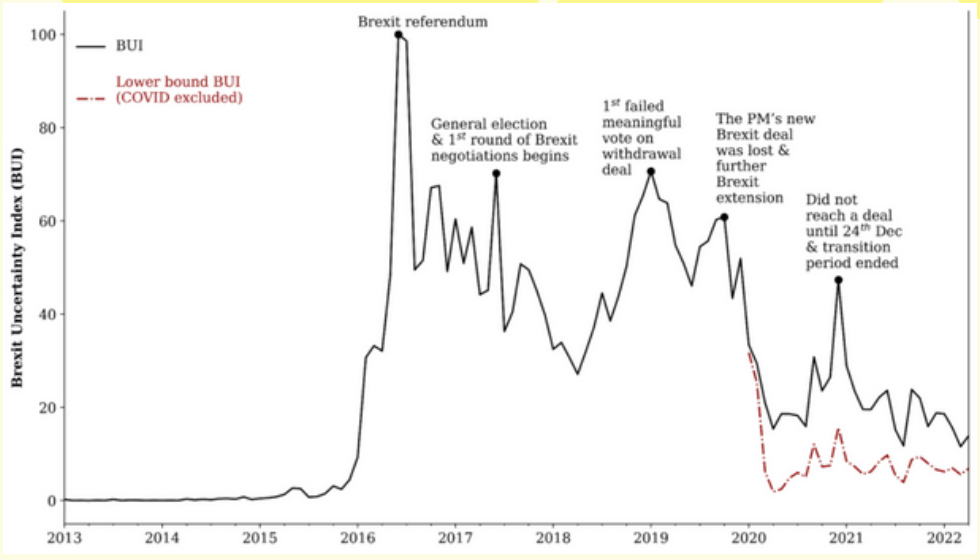
# Measuring Brexit Uncertainty

**BlueBEAR is essential for Claire's research into quantifying Brexit uncertainty, reducing the time spent running her analysis from two months on her own device to just one day!**

Duiyi Dai (Claire) - PhD Student

**Our work focuses on measuring Brexit Uncertainty (BU). We collect topic-specific indices of BU from the news, specifically uncertainty about the UK economy. We need to be able to update these indices in close to real-time and cost-effectively to inform policy-making decisions. Using techniques from computational linguistics we derive our BU indices (BUIs) based on the frequency of news articles that concern BU in eleven leading UK newspapers. The BUIs that are generated in my work (and regularly updated) are freely accessible online at our project website.**

**BlueBEAR has helped me a lot. With BlueBEAR, my program task only takes about 8 hours to run and I don't have to worry about crashes. More specifically, my program is parallelizable. That means, theoretically, using 10 machines can reduce the running time by 10x. BlueBEAR allows us to set up programs in parallel. So, I split the program tasks into 120 blocks. It can be imagined as 120 laptops running my program at once, which greatly enhances the efficiency. Further, researchers can submit multiple tasks at once on BlueBEAR. I, thus, simultaneously run programs with various hyperparameters. BlueBEAR allows me to run my programs in one day, if I used my own device, this would take at least two months!**




**Aggregate Brexit Uncertainty Index from 2013 to December 2022**

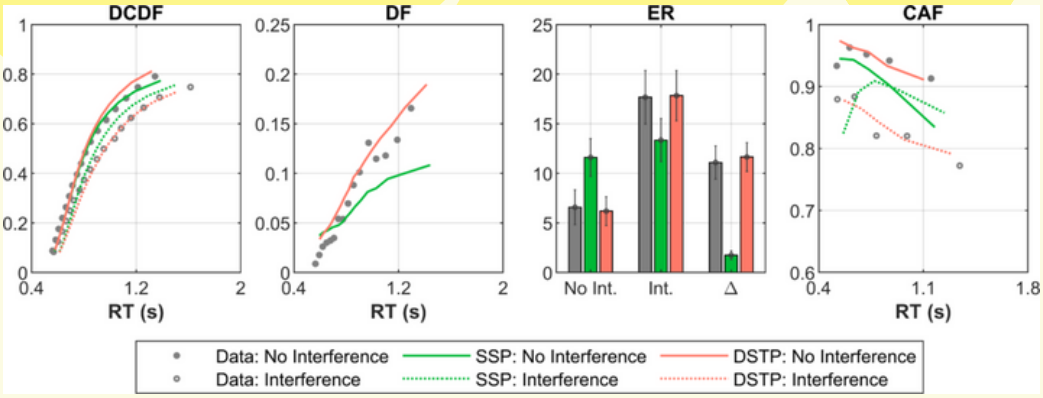
# Beating the Burden of Bayesian Statistics with BEAR!


**Jordan makes use of BlueBEAR and the RDS to enable her research into visual selective attention.**

**My project explores how our ability to focus on important tasks works. Traditional investigations look only at basic reaction times and mistakes. However models exist that analyse the entire range of reactions and accuracy. By using advanced models and fitting them to the data, we can better understand how people ignore distractions and focus on what's important, going beyond just the average reaction time.**

**I was excited to dive into the world of model fitting, before quickly realising just how computationally expensive model fitting can be. However, one day my brilliant supervisor Dr. Dietmar Heinke introduced me to BEAR. With BEAR's HPC resources, I can fit models to full datasets in a matter of hours. Not only is the model fitting procedure extremely quick now, but BEAR allows me to run multiple participants in parallel through array jobs.**

**This has allowed me to turn waiting times of months into hours and my project would simply be impossible without it. Not only this but with thousands of iterations, data files very quickly become huge! BEAR's Research Data Store (RDS) gives me more than enough space to store this data.**




**BEAR has allowed me to compare my models. The data is represented by grey circles, Model 1 is in green, Model 2 is in red. The second model (red) clearly fits my data better.**

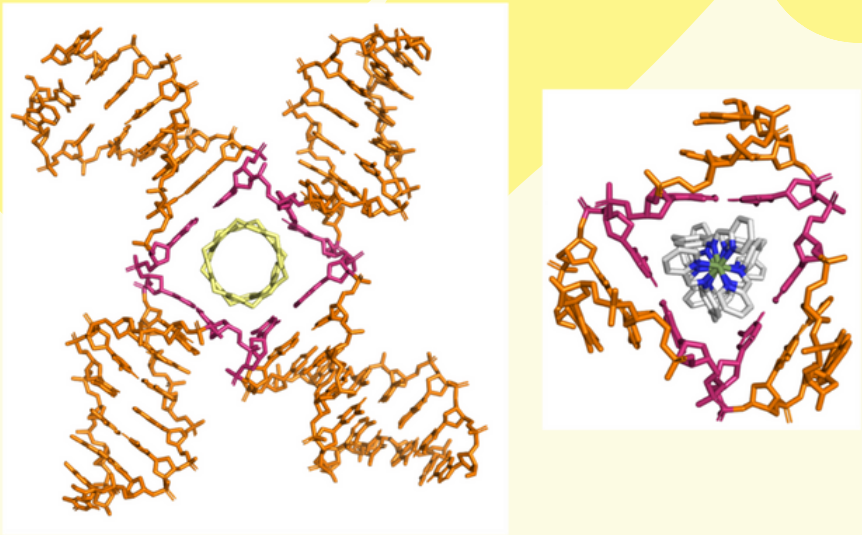
# Simulating DNA and its interaction with metal compounds

Samuel Dettmer - PhD Student

My research project investigates the ability of metal compounds to bind to DNA structures. This is an important area of medicinal chemistry in which we are trying to understand if we can target specific DNA structures that occur during particular cell processes, like DNA replication for example, and achieve a therapeutic effect. We are currently interested in antiviral activity against Sars-CoV-2, which causes COVID-19, and anticancer activity.

A deeper understanding of what is happening at the atomic level can be obtained by using molecular dynamics simulations. These simulations allow me to visualise the movement of DNA and its interactions with water, salt and my compounds over a short timescale (up to 10 microseconds).

So where does BEAR come into this? Well, a standard molecular dynamics simulation of 1 microsecond length typically involves calculating the 3-dimensional movement (and associated energy) of around 50 thousand atoms for 500 million steps! BlueBEAR offers me the resources I need to run up to 8 simulations simultaneously on 100 CPU cores each. Even with these resources, simulations still take up to 4 days to complete, which goes to show just how unreasonably long it would take without BlueBEAR.



Snapshots from molecular dynamics simulations made possible with BlueBEAR

# Detecting malware through AI techniques


**Daniel makes use of BEAR's storage and data processing power, specifically using GPU's to accelerate his research into automated malware detection.**

**The project I'm currently working on is automated malware detection using Graph Neural Networks, which we've needed to create an entire pipeline for.**

**We start by downloading a dataset of Android applications which consists of over 100,000 apps as a dataset to train our mode. All of these apps must undergo some preprocessing to convert them to a form that can be read by a Graph Neural Network, then we need to train our model. Each stage of this is massively computing intensive, and in different ways, with the app preprocessing stage being CPU bound, and model training is GPU bound.**

**The BEAR service has essentially made our project feasible by cutting down the time it takes to perform pre-processing on our dataset. Our dataset consists of 130,000 apps, and for each we need to decompile the apps and convert them to a form which is readable by a Graph Neural Network. Since we need to perform the same process on every app, this is a perfect situation for parallelisation! BlueBEAR allows a user to run 4000 concurrent jobs and this is something we take advantage of. By leveraging all of the compute nodes available to us we can pre-process 4000 apps at once and immediately start the next batch once they are complete. Our entire dataset can be processed in approximately 18 and a half hours. On a standard computer this would take nearly three days!**

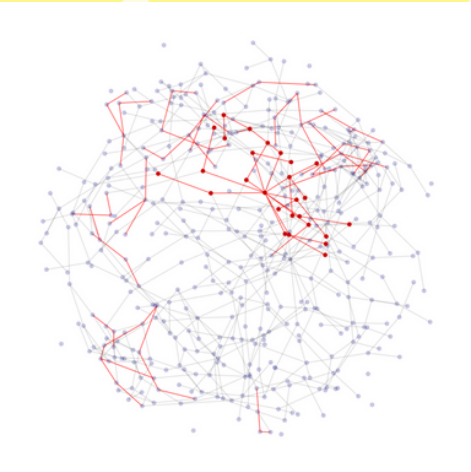


**Image showing the interactions between blocks of code which make up an Android application**

# SNOW much fun: Generating Open Access snow data

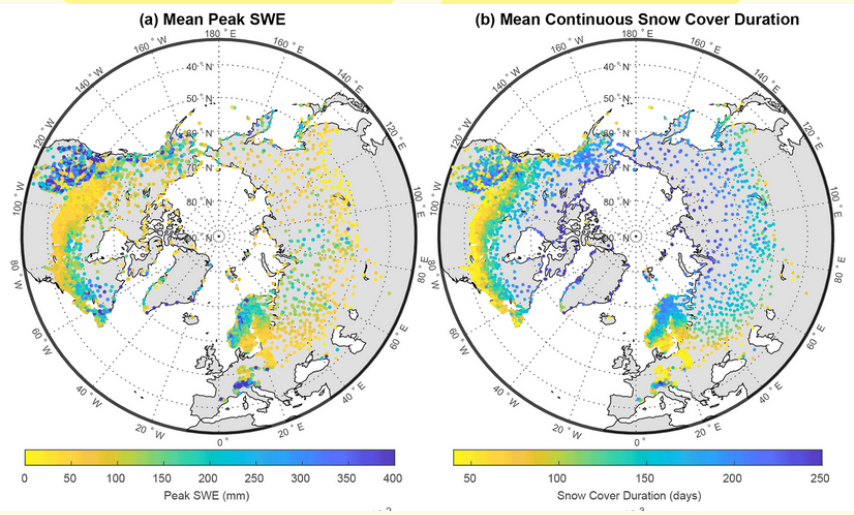
Adrià Fontrodona-Bach - PhD Student

**Adrià created an open access dataset using tools on the BEAR portal**

**"We need data" is a common statement in academic research. However, environmental data is usually scarce, unreliable, or inaccessible. As part of my PhD, I have generated and published an Open Access dataset of snow water equivalent that covers the Northern Hemisphere and provides time series from 1950 to 2022 at the point scale. But why is this snow dataset important, and how did BlueBEAR help produce it?**

**Snow is made of frozen water and air. For certain applications, we need to know how much water there is in the snow, but this can vary from 10% to 60% of the total snow depth, and it is difficult to estimate.Gathering as many of these snow depth observations as possible over the Northern Hemisphere, I used a model to estimate the amount of water in the snowpack, the snow water equivalent (SWE), this resulted in an open access dataset, and a publication that describes all the data and methods used.**

**As you can imagine, this is a lot of data to gather and to work with, and BlueBEAR has certainly made this work possible. Using the MATLAB GUI from the BEAR Portal, I developed a gap-filling method for the snow depth data based on machine learning models called Artificial Neural Networks. These were trained and required some computation time. On a laptop, it would take 115 days! By parallelising the code, using 40 cores in a MATLAB session in the BlueBEAR portal, I left the code running on Friday evening, and it was done on Monday morning!**




**All of the stations available in Adrià's dataset, with time series up to 73 years!**

# Unraveling the Secrets of Viscous Fluids

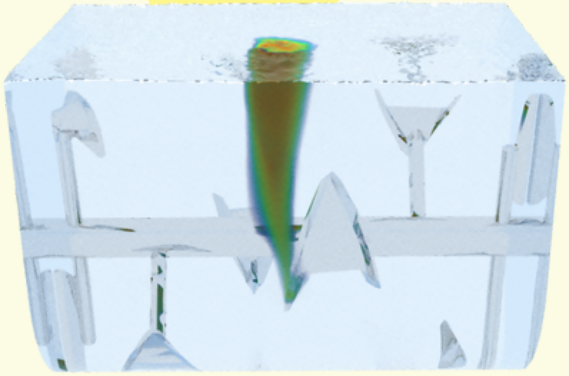
**With the help of BlueBEAR, Roberto has been using cutting-edge simulations to revolutionize product design for different climates.**


Ever packed your favourite cream for a vacation in a tropical paradise, only to discover it's turned into an unrecognisable goo? You're not alone! Temperature sensitivity affects many everyday products, posing challenges for engineers and manufacturers. But fear not, research computing is here to save the day!

Our quest for consistency begins with understanding how viscous fluids like soaps behave under various process conditions. By using Computational Fluid Dynamics (CFD) simulations, researchers can get a detailed look at fluid dynamics, predict mixing performance, and identify potential improvements in agitation, power consumption, and mixing time. Ultimately, this leads to a more consistent product that delights consumers worldwide, regardless of their location.

To make the most of BlueBEAR's capabilities, we designed a workflow that utilizes Acces, an evolutionary optimizer, to automatically calibrate the CFD models. With the help of a slurm wrapper, Python communicates with the slurm job scheduler, sending off batch simulations without human intervention. Comparing these simulations to real experiments, an error function guides Acces to generate new parallel simulation conditions. The process repeats until the model is calibrated and ready for action.

By harnessing the power of BlueBEAR and automating simulations, we're able to speed up the research process significantly. What once took two months on a laptop now takes just a week using BlueBEAR's advanced capabilities. That's the power of research computing!



A CFD simulation of glycerol

# BlueBEAR unlocks the potential of powder simulation


**You probably don't think about how we use powders in manufacturing on a day-to-day basis and I wouldn't blame you! They are however, the second most manipulated material in industry, yet our understanding of them remains in the stone age. Powder simulation has become an extremely promising technology to allow us to model powders, but a major problem is calibration.**

**Fluid processing has been a continuous process for centuries (think of oil refineries). Yet most powder processing remains batch based to this day and currently one of the biggest areas of powder technology research is making the pharmaceutical process continuous.**

**This is where BlueBEAR comes in. By running tens of thousands of simulations, forming a pile of powder and each time slightly changing the microscopic properties of the particles in the pile, we can build up a dataset that links the angle of that pile to the microscopic properties. BlueBEAR allows us to run hundreds of these simulations at a time or if we want to run one big simulation with even more particles then we can use all those cores on one job!**

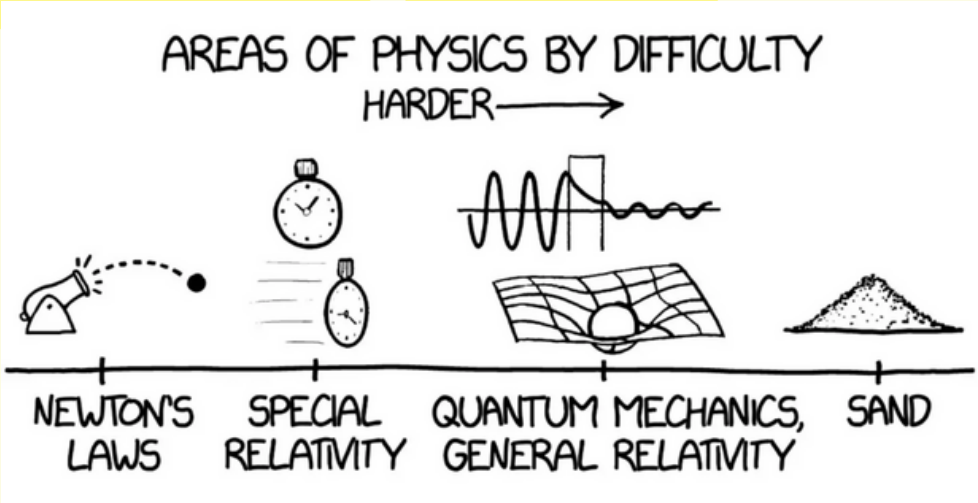



**Image source: What makes sand soft? NY Times**

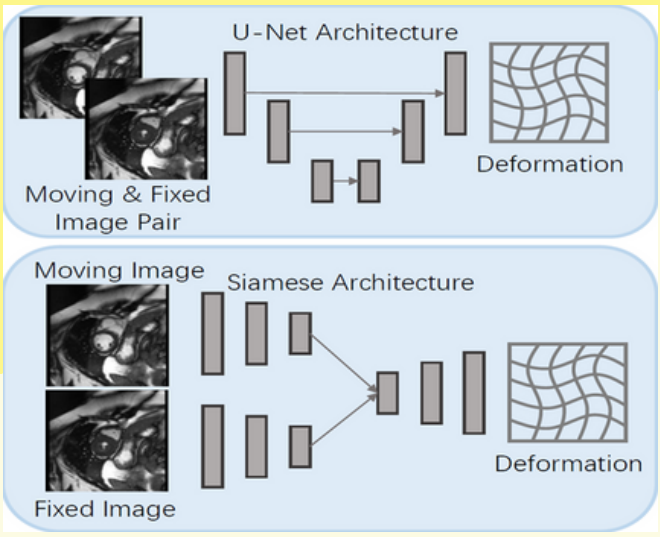
# High-resolution medical scans & Baskerville

Xi Jia - PhD Student

**Xi has been making use of BEAR's Baskerville GPU cluster to dramatically reduce the time it takes to train deep-learning models for medical image registration.**

**In my project I train deep-learning models for medical image registration. Image registration maps a moving image to a fixed image according to its spatial correspondence. I work with high dimensional volumetric data, such as 3D brain MRI scans, 3D Lung CT scans, and 4D cardiac MRI scans. For these high-resolution medical scans, training a basic 3D deep registration model usually requires at least one GPU (with 40G VRAM) and days to optimise, not to mention the tedious hyperparameter tuning that follows. Thanks to the powerful computational resources that Baskerville provided, I was able to train and tune multiple deep registration models in parallel.**

**More importantly, there are 4 A100 GPUs in a Baskerville node, which means I am able to train 4 different models in parallel within 2 days. Thanks to the powerful computational resource provided by Baskerville, our extended method of LKU-Net just won Task 1 of the MICCAI 2022Learn2Reg Challenge, which is one of the most well-known academic contests in the medical image registration field. I am using and benefiting from Baskerville. Why don't you give it a try?**




**How different architectures take moving and fixed images and estimate the optimal deformation**

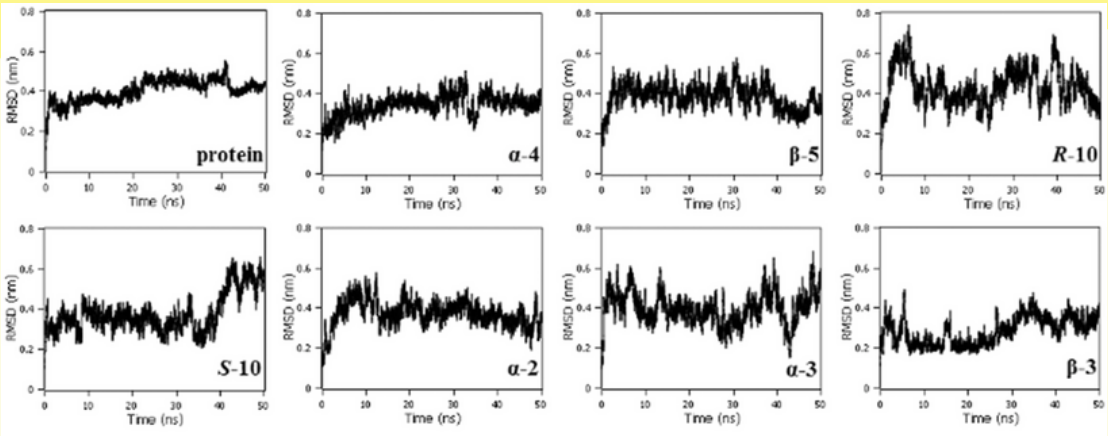
# BlueBEAR aids in drug discovery


**Dr Alan Jones relies on BlueBEAR to enable his research in drug discovery.**

**For several years, our research group has been studying a group of molecules called glycomimetics and their positive impact on experimental models of cardiovascular disease, specifically focusing on atherosclerosis. Atherosclerosis involves various stages of disease, starting with problems in blood vessel lining and ending with artery hardening due to calcium buildup.**

**In our drug discovery project, we've employed common computer-based techniques to study how new glycomimetic compounds interact with our main target, hepatocyte growth factor (HGF). This information helps us develop improved versions of these compounds.**

**Critically for our studies, we were able to perform molecular dynamics simulation using BlueBEAR. Routine molecular docking using commercial software such as AutoDock can be performed on a laptop but the high-performance computing power required to understand the real-time movement of both protein and small molecule for a short time frame, for example, 50 nanoseconds, required the supercomputing power of BlueBEAR.**




**Root Mean Square Deviations of the protein backbone along 50 ns of molecular dynamic simulation in the presence of the glyomimetic. Image Source**

# Ex nihilo discovery of new materials

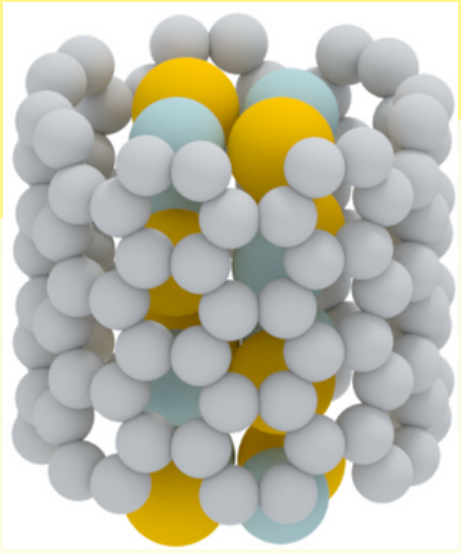
**Professor Morris makes use of BEAR services to predict the the crystal structure of materials.**

Professor Andrew Morris

**"Trial and error" plays a large part in the discovery of new materials. From the initial idea, the material must be synthesised and categorised before it can be tested which is slow, difficult and expensive. High-throughput computation accelerates this process by suggesting, then screening new materials, allowing us to ask "what if?" without the time and expense of manufacturing and categorising samples.**

**Currently, we are interested in new battery materials and their complex structure changes during charging. The structure can be very disordered so many calculations are required for accurate predictions. BEAR allows us to perform many more calculations than we could otherwise do, giving us a better understanding of these complex new battery electrode materials that we discover.**

**We use BEAR in a variety of ways. My students use the standard BlueBEAR HPC and data storage (RDS) offering available to all UoB researchers. We have also bought extra BlueBEAR nodes from research grants, so that these projects have dedicated compute resources. We access the Tier 2 services Sulis and Baskerville to increase our computational abilities. These Tier 2s are accessible to UoB academics through a light-touch internal process.**



**Encapsulation in nanotubes**

# A dive into second language acquisition and corpus linguistics

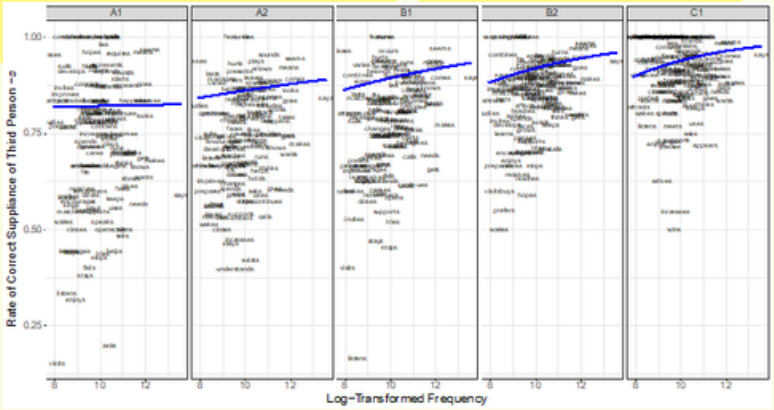
Dr Akira Murakami - Assistant Professor

**Dr. Murakami has been making use of BlueBEAR to enable his research into understanding how people acquire languages other than their native languages.**

My research lies at the crossroads of second language acquisition (SLA) and corpus linguistics. SLA is a discipline devoted to understanding how people acquire languages other than their native languages. In my work, I harness large-scale text data (called corpora), such as collections of written pieces by second language (L2) learners, to delve into the intricacies of this process. Given the computational intensity of these investigations, I rely on BlueBEAR's high-performance computing (HPC) service.

In a recent project, we explored how the arrangement and frequency of specific language elements (distributional factors) influence the accuracy of L2 learners' use of inflectional morphemes (e.g., -ed, -ing) in their writing. BlueBEAR was essential to this work in two ways. First, I used BlueBEAR to examine the distributional factors within a large-scale corpus of American English that served as a proxy for L2 learners' input. Second, the substantial memory capacity of BlueBEAR's HPC was indispensable for the Bayesian statistical modelling used in our project.

In addition, BlueBEAR enabled us to partially parallelise the computation process. Combined with earlier text processing, the entire computation could have taken over two weeks to complete. Using BlueBEAR's HPC significantly reduced the required computation time.




**The relationship between the frequency of inflected form and the accuracy of its use in L2 writing across varying proficiency levels of L2 learners.**

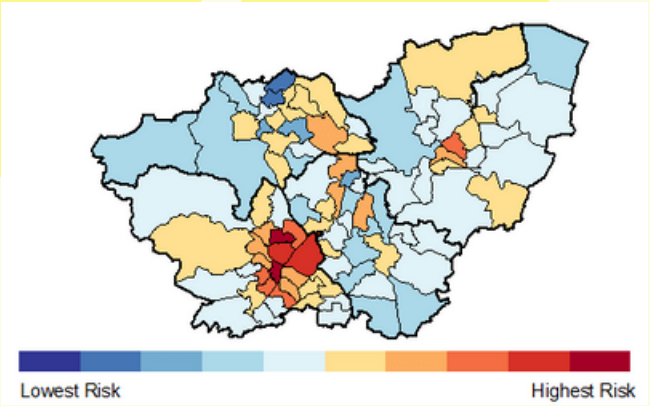
# Statistical modelling to inform safeguarding of women and children


**With assistance from the BEAR team, Dr. Seymour set up an online survey to gather important safeguarding information.**

**My research involves developing statistical tools to map and measure crimes, such as human trafficking and violence against women and girls. Digital research is a valuable tool when designing surveys as it gives you the power to go beyond standard techniques. Well-designed online survey tools make the process of providing data much easier for respondents, as well as reduce cleaning and preparation time for researchers.**

**With the help of the BEAR software group, I used a BEAR hosted website to run a survey to map the risk of Female Genital Mutilation (FGM) in South Yorkshire. FGM is a violent procedure where the female genitals are deliberately cut, injured or changed, but there is no medical reason for this to be done. It is a crime in the UK and people can face a prison sentence for arranging or carrying out FGM.**

**There are many local services in South Yorkshire working towards ending FGM, but no one service has a complete picture of where it is happening. This makes it difficult for services to design local interventions to safeguard those at risk of abuse. BEAR hosted a comparative judgment study I designed to map FGM at the ward level. Safeguarding professionals in the county were shown pairs of wards and asked to say which ward had the higher risk of FGM. From all these comparisons, I was able to estimate the relative risk level in each ward. The BEAR software group were incredibly helpful in making sure the online survey tool was accessible and suitable for use.**




**A map of South Yorkshire showing the risk of FGM in each ward.**

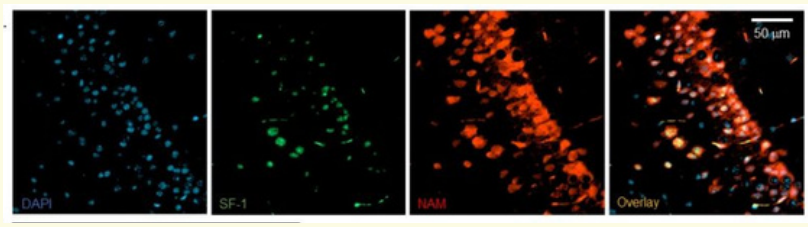
# Covid made me do it....

**Gabriela makes use of BEAR's RDS to store and share microscope images with her research team.**


**My research focuses on how our bodies maintain a balance of energy and what happens when this balance goes wrong, causing health problems like obesity and type 2 diabetes. I want to understand the reasons behind these issues and how they develop.**

**I must confess as a bench scientist I didn't really think very much about my computing needs as these were fairly modest for a big chunk of my life as a researcher. This was mostly changed by the increased use of microscopy techniques for my research – high content images take up a lot of storage space and are not easily shared, so it is not practical to have them stored on a desktop computer. My team's storage and computing needs grew quickly as imaging data was collected and analysed by students and researchers who joined my team. BEAR's Research Data Store (RDS) was a great solution for our storage needs.**

**We frequently 'blind' analysis by asking one operator to take images, label them with a key, and then give the images to another operator for analysis; storing imaging data on BEAR RDS allowed us to share data easily and for us to collaborate effectively. During the COVID-19 lockdown, I was able to share historical imaging data on RDS with students so that they could perform the analysis as part of their final year projects, and they could then share the processed data with me easily. It also allows me to keep track of data and ensure it is appropriately and securely stored, as required for the purposes of publication and grant awards.**




Microscope images of brain sections

# The saving grace ( well on Git )

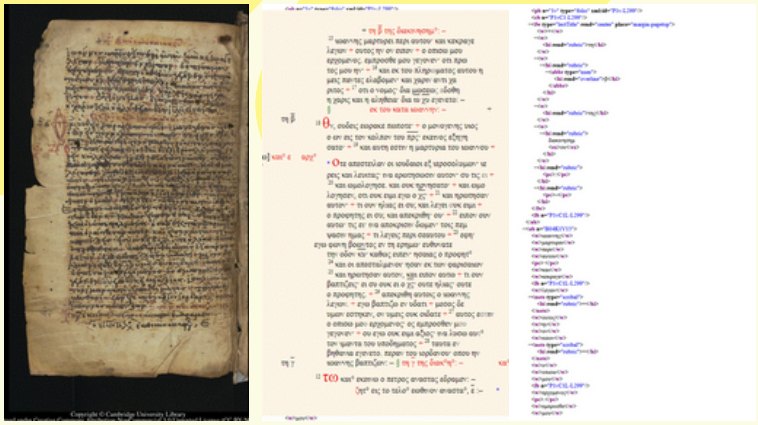
Dr Catherine Smith - Research Fellow

**Catherine has been making use of BEAR's code management platform, BEAR GitLab, to enable her to support researchers using online editing tools to transcribe texts.**

I started working in ITSEE about a week before the first commit was made in GitLab. GitHub already existed but I'm not sure whether private repositories did or not. The existing technical staff in ITSEE were already using git for all of the code development when I arrived. We ran our own remote git installation because it was the best way to have private repositories.

At that time the repository was used exclusively by the two developers and only used for code. It worked fine for us but we had to maintain our own installation. Several years later, non-developer researchers in the team also needed to start working with git. I taught them the basics on the command line and they got along pretty well. We had no GUI for our remote repository, so the researchers just followed my instructions and didn't really understand what was happening behind the scenes. We could, of course, have put all of the code and the data on GitHub but some of the code wasn't really ready to be made public at that time.

As soon as GitLab became part of the BEAR suite of tools we moved everything from our own remote repository into the BEAR GitLab. It gives us the advantage of a better remote repository with a GUI and we don't have to maintain our own system, so it reduces the admin work for me. I think it also makes things more understandable and less intimidating for the researchers using GitLab.



An example of a text transcribed from a manuscript image through to xml code.

# Using BEAR to store and analyse microscope images

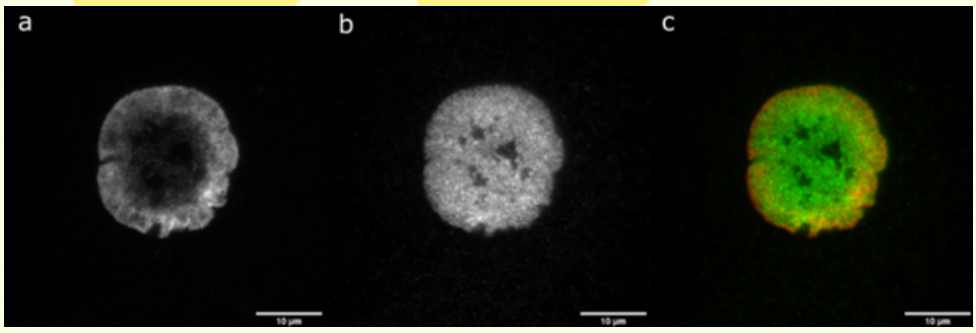

**Stanimir relies on the BEAR portal, Research Data Store and BlueBEAR to carry out his research involving single molecule microscopy.**

**I work in the Institute of Cardiovascular Sciences, where I use microscopes to look at individual molecules and count protein complexes on the surface of cells. These images produce terabytes of data and require hundreds of gigabytes of RAM to process and analyse, which is more than a laptop can handle. This is where BEAR comes in; the RDS, BEAR portal and BlueBEAR all enable me to carry out my project.**

**Microscope imaging generates a lot of data in short amounts of time. The 3TB of free storage space on the RDS enables us to store our data securely and share it with international collaborators via Globus free of charge!**

**We use the software ImageJ and its expansion Fiji to process our images, available on the BEAR portal. This can include simple things such as setting the correct pixel size and adding a scale bar, or it can be more complex such as running a modifiable script or analysing a stack of images.**

**When it comes to analysing our data, BlueBEAR has a quick and intuitive way to run Python scripts. BlueBEAR enables a huge decrease in the time we spend processing data and we are able to perform clustering algorithms to identify clusters of fluorescence emission through an entire image set, which would be impossible on most computers.**




**Microscope images of Jurkat cell line 90 seconds after activation**

# Getting a move on !

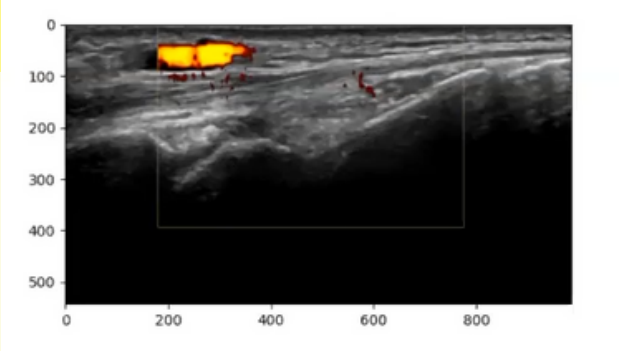

**Jason relies on the BEAR portal, Research Data Store and BlueBEAR to carry out his genomics research, which was helped by the coaching and training BEAR offers.**

**Jason receives data from Genomics Birmingham, which previously had to come via ftp servers meaning data transfer was very slow. Recently his team have started using the BEAR Data Transfer service using Globus. This means they can get the data directly to the BEAR RDS much more quickly.**

**Jason has 9 core projects that he is actively working on, for which he has access to 27TB of free storage via the BEAR RDS. Jason has also recently purchased 100TB of storage to meet the needs of upcoming projects. He stores multiple 10x genomic data sets, bulk RNA-sequencing datasets, ultrasound images as well as general wet lab data and documents.**

**Jason uses thousands of CPU hours across the various projects he is involved in for data processing – his laptop only has 8 cores so can only perform 8 CPU hours and using it for all the data processing would "take a significant amount of time" and prevent him from using the machine for anything else.**

**Coaching from the Research Software Engineering team has enabled Jason to implement automatic grading of ultrasonic images (dicom images), converting the medical grade images to jpegs and tifs, which are more typically used for AI and pulling down the metadata needed to classify the images.**




**700 CPU hours were used for this AI medical image analysis!**

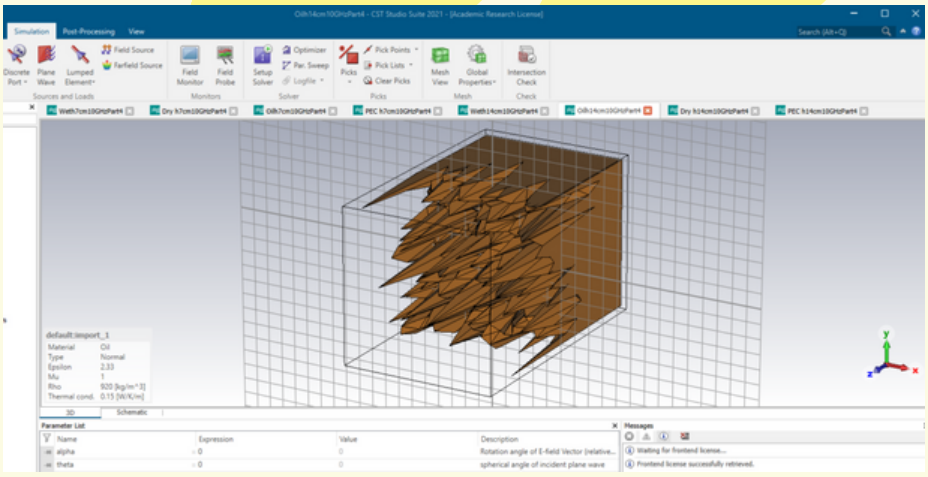
# Using BlueBEAR to calculate reflectivity

**Elijah relies on BlueBEAR to run simulations to detect oil leaks in pipeline infrastructure.**


**The project I am currently working on is the remote monitoring of pipeline infrastructure using RADAR sensors to detect oil leaks. I create rough surface models with various root mean square (RMS) heights and statistics using the software CST. I assign a particular dielectric parameter to each model, for example, sandy soil which could be wet or dry or even a dry sandy soil that has been soaked with oil. Then, from that initial rough surface model with a defined RMS height, I try and fill up the rough surface models with water or oil, until it becomes or appears to become a smooth surface.**

**At each level of filling up the holes or cracks, I upload the model to BlueBEAR and calculate the reflectivity. This takes quite a while to calculate depending on the number of meshes in the rough surface model. Sometimes, the simulation takes at least ten days or even two weeks per model depending on the RMS heights as well as other parameters. And mind you, I have over 200 of these models to run for a single simulation parameter.**

**BlueBEAR has essentially made our project feasible by cutting down the time it takes to perform complex simulations. By leveraging all of the compute nodes available, we can pre-process several apps at once and immediately start the next batch once they are complete. Our entire dataset can be processed in approximately one week. Without parallelisation (which would likely be the case on a standard computer) this would take nearly one month!**



**A rough surface model: sandy soil (50cmx50cm) that has been soaked with oil with a surface roughness of ~14cm**

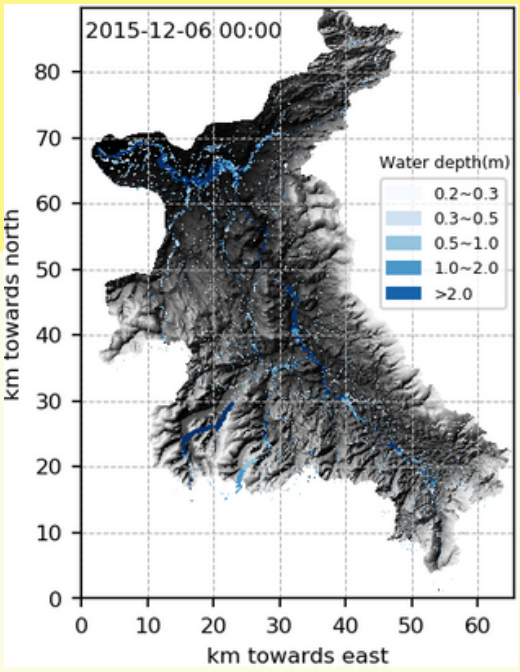
# Baskerville helps to predict floods

**Baskerville is essential to Xilin's research into modelling extreme weather impacts.**

Dr Xilin Xia - Assistant Professor

**Due to Climate Change, we are seeing increasingly more extreme weather around the world. Extreme weather events can cause natural hazards such as flooding, landslides, and debris flows, causing enormous damage each year and posing significant threats to people and infrastructure. My work focuses on developing computational models of natural hazards and trying to understand and predict their impacts so that we can mitigate them. For example, I have developed models that can generate maps showing where there is flooding and the depths and velocities of that flooding, so that we know which place is most at risk.**

**Access to a GPU-based high-spec server is essential to support the work of my team. Luckily Birmingham hosts the new Tier-2 HPC Baskerville, which has 208 latest NVIDIA A100 GPUs! Even better, the university offers free access to all researchers (staff and students). I didn't have to apply for an EPSRC grant or fill in an excessively long application form to use it. Baskerville also offers interactive running mode, in which I can request a node for certain hours to test and debug the code.**




**Simulation of the Storm Desmond flood in the Eden catchment, UK in 2015**

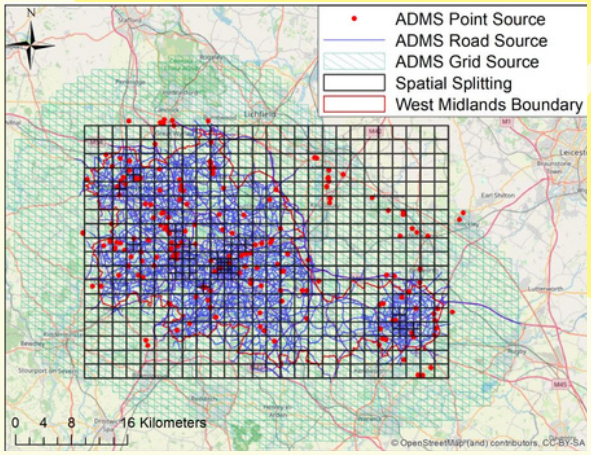
# Street-scale resolution air quality modelling

Dr Jian Zhong - Research Fellow

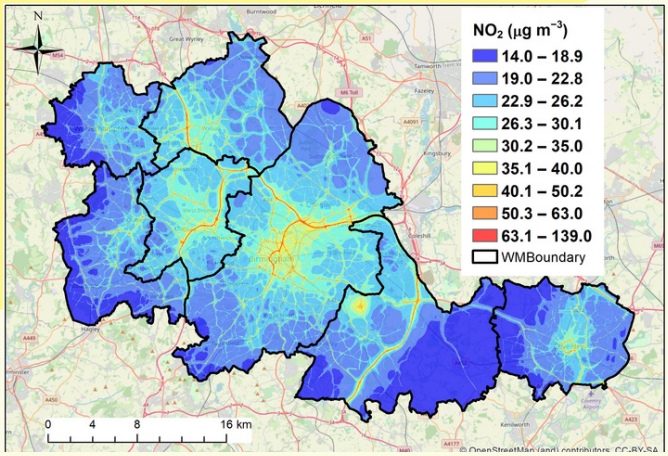
**Dr Zhong has been making use of BlueBEAR to enable his research in air quality modelling.**

**Air pollution has become a significant threat to public health, with approximately 28,000 to 36,000 people in the UK dying each year due to its effects. To combat this issue, it's crucial to gain a deeper understanding of where these pollutants come from and how they behave in the environment. To do this, we need advanced air quality models that consider factors like emissions, chemical reactions, dispersion (how pollutants spread), and dynamic changes in the atmosphere. These models are essential for creating effective policies to improve air quality in cities. We can use air quality modelling to investigate and assess the impact of different actions and interventions on air quality to help us make informed decisions to protect public health and the environment.**

**I used BlueBEAR to apply task farming to my model, which significantly reduced runtime. The task farming approach was achieved by spatially splitting the computational domain. Each smaller sub-domain can then be executed concurrently on BlueBEAR. An array job with 540 cores, each for a single sub-domain, was submitted to BlueBEAR. This optimization process reduced weeks of model execution time to approximately 35 hours for a single model configuration of annual calculations.**




**Model input**
**Emission sources in the model and spatial splitting for the modelling domain over West Midlands**




**Model Output**
**Annual air quality maps for NO2 (in µg m–3) at 10 m × 10 m resolution.**

# Index

## Index 1 – BEAR services



## Index 2 – Colleges