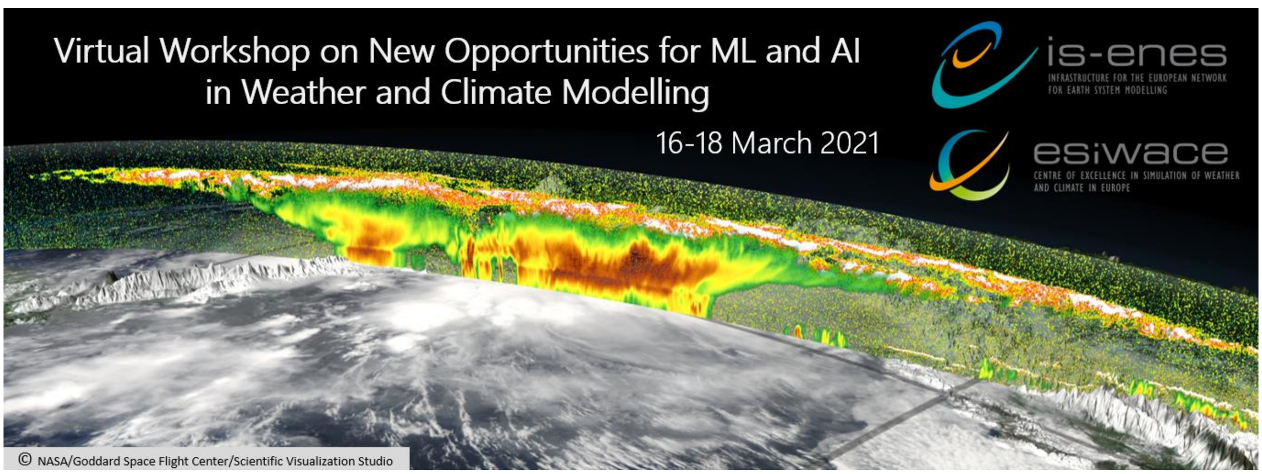
IS-ENES3 Milestone MS16 (M4.2) v1.0

New Technical Opportunities workshop in ML and AI

Reporting period: 01/07/2020 to 31/12/2021

Authors: G.D. Riley (UNIMAN), G. Aloisio (CMCC) & D. Elia (CMCC)

Release date: 28th April 2021



ABSTRACT

This document reports on the joint ISENES3/ESiWACE2 Virtual workshop entitled « New Opportunities in ML/AI for Weather and Climate Modelling » held on 16-18 March 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Table of contents

1. Objectives

This report describes the virtual workshop entitled "New Opportunities for ML/AI in Weather and Climate Modelling" held over the three days of 16-18 March 2021.

The aim of the workshop was to bring together climate scientists and experts from academia and industry to share knowledge and experience and to identify new opportunities in the areas of machine learning, artificial intelligence and big data techniques for Weather and Climate Modelling.

The workshop was sponsored jointly by the IS-ENES3 project and the ESiWACE2 Centre of Excellence. Each project had complimentary interests in the workshop. ESiWACE2 essentially having a focus *inward* to the Weather and Climate modelling community while the focus of IS-ENES3 is more *outwards* toward other communities and industrial partners, as reflected in the deliverable documents related to the workshop described below.

2. Description of work: Methodology and Results

The workshop was organized around three three-and-a-half hour sessions, each commencing at 3pm CET on consecutive days:

- (Day 1): Views from Domain Science
- (Day 2): ML/AI Software Technologies
- (Day 3): High performance, Infrastructure and Big data Challenges

The first session, "Views from Domain Science" is particularly relevant to the ESiWACE2 Centre of Excellence and further details can be found in the associated ESiWACE2 deliverable, D2.10, "*Machine Learning Workshop*". Days 2 and 3 have some interest for the general Weather and Climate community and the aspects relevant to the IS-ENES3 project are explored in the associated sections of IS-ENES3 deliverable, D4.5, "*White Paper on Innovation, Tools, Platforms and Techniques*". This document gives a brief overview of the content of each session only.

The workshop website can be viewed at:
Joint IS-ENES3/ESiWACE2 Virtual Workshop on New Opportunities for ML and AI in Weather and Climate Modelling

The website provides links to the workshop agenda as well as links to the presentations and videos of the talks presented in each session.
The following IS-ENES3 link is to google docs ("IS-ENES3/ESiWACE ML/AI in Weather and Climate modelling workshop – March 2021") where the documents related to the workshop (including links to the Q&A documents of each session, and copies of the presentations) can be found:
https://drive.google.com/drive/u/1/folders/1zSh7OlNZXXhRoovW9msKxRJ3TvjauZTP

The Q&A documents associated with each session summarize the discussions following each talk; these documents provide a further useful resource to the community. The use of this Q&A documents proved to be effective in fostering a virtual discussion among the presenters and the participants of the workshop.

The workshop sessions on each day were structured with seven talks and a panel session. The first talk was a 40-minute keynote, with the other six talks being 20 minutes each, including questions. The panel sessions lasted approximately half-an-hour.

The organizing committee consisted of:

- (Chair) Graham Riley (University of Manchester, UK)
- (Co-chair) Giovanni Aloisio (CMCC, University of Salento, Italy)
- Jean-Claude Andre (France)

- (Session 2 chair) Caroline Arnold (DKRZ, Germany)
- Balaji (Princeton, USA)
- Peter Dueben (ECMWF, UK)
- (Session 1 chair) Marlene Kretschmer (University of Reading, UK)
- (Session 3 chair) Carlos Gomez Gonzalez (BSc, Spain)
- Donatello Elia (CMCC, University of Salento, Italy)

Of the total of 21 speakers, 14 were from academic institutions and Weather and Climate centres around the world and 7 were from industry, representing both software and hardware technology developers as well as service providers. In particular, those from industry were:

- Luke Madaus & Steve Sain from Jupiter Intelligence,
- Stephan Rasp from ClimateAI,
- Akshay Subramaniam and Thorsten Kurth from two parts of NVIDIA,
- Pete Warden from Google,
- Phil Ridley from Arm,
- Jonathan Wayne, Microsoft Research/U. of Washington.

The workshop was very well attended. During the workshop there was a total of 206 unique participants taking part in the three sessions (out of a total of 287 registrations). These were spread over the three sessions as follows: Day 1: 173 participants, Day 2: 136 participants, Day 3: 97 participants.

The participants had a range of backgrounds from across academia, weather and climate centres and industry around the world. There were representatives from the following companies registered for the workshop: Ramboll, Arm, Google, Cervest Ltd., Wikilimo, GEOMAR, NVIDIA Ltd., Predictia Intelligent Data, SISTMMA GmbH, ARCADIS, Airbus, Benchmark Labs, Climate Scale, Descartes Labs, Lobelia Earth, CIEMET, Randbee Consultants, Eötvös Loránd, ClimateAi, ClimaCell, NUS, Kyrgyzhydromet, Arpae, ULB Brussels Physics, Verisk Maplecroft, Simula, Jupiter Intelligence, Microsoft.

List of talks

(Day 1): Views from Domain Science

- Emily Shuckburgh (University of Cambridge): *New approaches based on ML for a range of climate prediction problems*
- Luke Madaus & Steve Sain (Jupiter Intelligence): *Philosophy and Targeted Applications of ML/AI Techniques for Climate Risk Analytics at Jupiter*
- Stephan Rasp (ClimateAI): *The optimization dichotomy: Why is it so hard to improve climate models with machine learning*

- Zhaoyi Shen (Caltech): *Improving convection parameterizations with a library of large-eddy simulations*
- Rachel Prudden (MetOffice Informatics Lab): *Stochastic Super-Resolution for Convective Regimes using Gaussian Random Fields*
- Kirsten Mayer (Colorado State University): *Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network*
- Pedram Hassanzadeh (Rice University): *Using transfer learning and backscattering analysis to build stable, generalizable data-driven subgrid-scale models: A 2D turbulence test case*

(Day 2): ML/AI Software Technologies

- Jussi Leinonen (MeteoSwiss): *Stochastic machine learning for atmospheric fields with generative adversarial networks*
- Andreas Gerhardus (DLR Jena): *Causal discovery in time series with unobserved confounders*
- Jinlong Wu (Caltech): *Estimating stochastic closures using sparsity-promoting ensemble Kalman inversion*
- Gionata Ghiggi and Michaël Defferrard (EPFL): *Deep Learning on the sphere for weather/climate applications*
- Thomas Chen (AMSE): *Deep learning-based remote sensing for infrastructure damage assessment*
- Akshay Subramaniam (NVIDIA): *Leveraging physics information in neural networks for fluid flow problems*
- Jonathan Weyn (Microsoft Research/University of Washington): *Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models*

(Day 3): High performance, Infrastructure and Big data Challenges

- Tal Ben-Nun (ETHZ): *Scaling Up Deep Learning Workloads - A Data-Centric View*
- Micheal Simpson (NOAA): *Radar QPE and Machine Learning*
- Pete Warden (Google): *Using ML at the Edge to Improve Data Gathering*
- Phil Ridley (Arm Ltd.): *An Overview of ML and AI on Arm Based HPC Systems for Weather and Climate Applications*
- Jan Ackmann (Oxford University): *Machine-Learned Preconditioners for Linear Solvers in Geophysical Fluid Flows*
- Theo McCaie (MetOffice Infrmatics Lab): *You do you. How next-gen data platforms can stop weather and climate scientists from being software engineers and other perversions*
- Torsten Kurth (NVIDIA): *3D bias correction with deep learning in the Integrated Forecasting System*

Session 1: View from Domain Science, Chair: Marlene Kretschmer

Further details of the scientific application topics addressed in the talks of this session, and session 2, can be found in the ESiWACE2 deliverable, D2.10, “Machine Learning Workshop”. The Q&A documents associated with the sessions are also useful sources of further information.

Keynote: Emily Shuckburgh (University of Cambridge), Keynote

The keynote speaker was Emily Shuckburgh whose talk title was *New approaches based on ML for a range of climate prediction problems*. Emily discussed four examples of recent approaches to using ML:

- Seasonal sea ice forecasting using data-driven techniques (Convolutional Neural Networks (CNNs)).
- Heatwave prediction using a Gaussian Process (GP) approach to bias existing physical models based on the use of historical information to learn the implicit function governing the process.
- Understanding the physical basis of a data-driven ocean model being developed.
- Developing a data-driven model of a flood damage function as an example from hydrological modelling of climate impacts.

Luke Madaus (Jupiter Intelligence)

Luke’s title was *Philosophy and Targeted Applications of ML/AI Techniques for Climate Risk Analytics at Jupiter*. Jupiter is a climate services and risk analytics provider, providing services in the areas of global- and urban-scale quantification of weather and climate-related risks and uncertainty over various time frames, but typically decadal. Luke’s aim was to present an overview of the ML and AI tools and methods currently used at Jupiter. Luke described the cautious approach Jupiter take to the use of ML and AI techniques using examples from climate downscaling and dynamic model emulation. The talk emphasized some of the issues related to the *interpretability* (explainability and transparency) of the results from using ML techniques from both their own science teams as well as from customers and users of their data, who, for example, have to provide validation guarantees for their clients. One example was that of a hotel chain portfolio owner who wished to know which of their 1000 hotels around the world were most likely to flood as a consequence of climate change.

Stephan Rasp (ClimateAI)

Stephan’s title was: *The optimization dichotomy: Why is it so hard to improve climate models with machine learning*. ClimateAI is a company that provide predictive weather analytics data to a range of clients. Stephan started by pointing out that one of the key contributors to uncertainty in climate predictions arises from subgrid parameterizations, especially those of clouds. The issues were discussed in the context of developing ML-based emulators trained from model data.

Zhaoyi Shen (Caltech)
Zhaoyi's title was: *Improving convection parameterizations with a library of large-eddy simulations*. This was another talk focussing of the issues related to improving physics parametrizations, this time focussing on large-eddy simulations (LES) of turbulence, clouds, and convection in the context of global climate models (GCMs). The talk presented some early results on the calibration of convective parameterizations using the library.

Rachel Prudden (MetOffice Informatics Lab)
Rachel's title was: *Stochastic Super-Resolution for Convective Regimes using Gaussian Random Fields*. This talk also addressed the issue of sub-grid scale parameterization, this time focussing on convection, the simulation of which is currently at the edge of current resolutions. The approach was based on the use of Gaussian Random Fields (i.e., a Gaussian Process with 2D fields).

Kirsten Mayer (Colorado State University)
Kirsten's title was: *Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network*. Kirsten described an example prediction related to the Madden-Julia Oscillation (MJO) concerning subseasonal (2 weeks to 2 months) prediction in midlatitudes which is difficult due to the chaotic nature of the atmosphere. There are occasions when the atmospheric conditions are favourable to prediction which can lead to enhanced skill in forecasting; these are known as "forecasts of opportunity". The aim is to use NNs to classify the confidence in predictions related to the MJO (based on the physical features in known categories of MJO). The classification supports explanation of the observed behaviour.

Pedram Hassanzadeh (Rice University)
Pedram's title was: *Using transfer learning and backscattering analysis to build stable, generalizable data-driven subgrid-scale models: A 2D turbulence test case*. This was another talk focusing on the difficult issue of turbulence modelling and focused on techniques to learn tricky parameterization terms in the equation linking low-level turbulence (requiring a very high-resolution grid and very small timesteps) to LES models (on coarse grids, around 1/8 of the size, and using a longer timestep, around x10 larger).

Session 2 – ML/AI Software technologies, Chair: Caroline Arnold

Further details of these talks can be found in IS-ENES3 Deliverable, D4.5, "White paper on innovation, tools, platforms and techniques" and the ESiWACE2 deliverable, D2.10, Machine Learning Workshop.

Jussi Leinonen (MeteoSwiss), Keynote

Jussi's title was: *Stochastic machine learning for atmospheric fields with generative adversarial networks*. In the first part of his keynote talk Jussi gave an overview of the development and relationships between many of the current ML technologies, from Neural Networks through CNNs and Encoder-Decoder architectures, followed by extensions leading to Autoencoders and then Residual Networks and Recurrent Nets, concluding with Generative models and Generative Adversarial Networks (GANs) and their application to Conditional probability problems in the form of Conditional GANs (CGANs).

In the second part of the talk, Jussi described work using CGANs as an approach to manage uncertainty in prediction; CGANs can learn to generate the conditional distribution of solutions since they can generate multiple solutions for a given set of predictors. GANs appear to have wide applicability since conditional probability problems are very common in weather and climate modelling (and physics generally).

Andreas Gerhardus (DLR Jena)

Andreas' title was: *Causal discovery in time series with unobserved confounders*. The aim of this talk was to try to develop a mathematical basis for understanding cause and effect applied to Earth's complex climate system: an example of a system on which it is not possible to perform real experiments (to explore climate phenomena) but for which there is an increasing wealth of observational data. Simulation, however, does provide a basis for experimentation but with uncertainty related to the fidelity of models. The key mathematics is contained in frameworks for *causal inference* and *causal discovery*. A working definition of causality was presented in the context of causal inference: "X causes Y if an experiment changes only X and Y is seen to change". These ideas can be used to produce a Structural Causal Model based on causal graphs – under appropriate assumptions. Causal understanding can lead to physical insight (scientific understanding) and can lead to robust and forecasting and also supports *attribution* ("Why did this happen?", "Is this due to climate change?"). The talk went on to discuss the recent application of causal discovery to time-series data. Encouraging results were reported based on their novel LPCMCI causal discovery algorithm for learning the cause-and-effect relationships in multivariate time series, which are a common form of climate data. LPCMCI represents a promising technology in the search for understanding of the Earth's climate system.

Jinlong Wu (Caltech)

Jinlong's title was: *Estimating stochastic closures using sparsity-promoting ensemble Kalman inversion*. This was a talk about a technique/technology to produce better parameterizations (or closure models). There are, of course, many of these required in weather and climate simulations: turbulence being a key example for which numerical simulation is simply too expensive (at the scale required to resolve it sufficiently). The point of a good closure model is that it is sufficient enough to correctly reproduce time-averaged statistics. Jinlong described a new technique to quantify model error in the closures of dynamical systems, based on Ensemble Kalman Inversion (EKI), and demonstrated the merit of introducing stochastic processes to quantify model error for certain systems. The topic of replacing existing closures with purely data-driven closures (i.e., emulators) using the proposed methodology was also discussed. The methodology promises to provide a systematic approach to estimate model error in many closures used in climate modelling.

Gionata Ghiggi and Michaël Defferrard (EPFL)

Gionata gave this talk, and the title was: *Deep Learning on the sphere for weather/climate applications*. This talk presented a Deep Learning (DL) approach (making use of convolution operations) to reducing the execution time of model components. Many models compute on grids which hold planar projections of data but because of the resulting non uniformity of area in these grids, they are unsuitable for performing convolutions on the data. Gionata described a new technique for computing convolution and pooling operations, common in DL solutions, directly on several common grids used in climate modelling that are defined on the sphere: e.g., Gauss-Legendre, icosahedral, cubed-sphere. In addition, their solution allows the mixing of data from different grids and scales linearly with the number of grid points (pixels), allowing it to ingest millions of inputs from 3D spherical fields. Their results show that the technique improves the prediction performance ("skill") of data-driven weather forecasting with no impact on computational overhead. Results were shown in examples training autoregressive ResUNets on five common spherical grids used in weather and climate modelling. The technique should be applicable to several areas, including, for example: post-processing (e.g., bias correction and downscaling), model error corrections, linear solver pre-conditioning, model components emulation and sub-grid parametrizations.

Thomas Chen (AMSE)

Thomas' title was: *Deep learning-based remote sensing for infrastructure damage assessment*.

This talk presented a number of ways that are being explored to apply Deep Learning to rapidly assess damage caused by extreme weather events and natural disasters. One example uses Deep learning to classify levels of building damage (in terms of a standard 4 category terminology) from satellite imagery by processing images of an area taken before and after a damage event using, for example, ResNet18 trained on ImageNet data, using an improved ordinal cross-entropy loss function. This approach can be supplemented by image segmentation. Natural Language

Processing (NLP) has also been used to analyse the content of social media to determine the impact of damage events. Images from mobile phone posted on social media (pre- and post- incident) can also be processed to gather information on the damage caused. To support future work, better labelled data is needed for training and a cleaner dataset with more distinct differences between damage categories is being developed.

Akshay Subramaniam (NVIDIA)

Akshay's title was: *Leveraging physics information in neural networks for fluid flow problems*. Akshay's talk had two parts both related to physics-informed Generative Adversarial Networks (GANs). In the first part he discussed another approach to the problem of modelling turbulence in the context of Large Eddy Simulation (LES) where the issue of representation at different scales is crucial. Akshay addressed the problem of how to recover a high-resolution, point-based picture from a low resolution LES. The approach proposed uses the technology of GANs, which were also discussed by Jussi Leinonen in another context in the keynote talk of this session of the workshop. The idea of a GAN is that a *generator* component generates a forgery of some input which is iterated until the *discriminator* component of the GAN cannot distinguish the forgery from the input. An example of the new GAN technology called TEGAN (inspired by a model called SRGAN) was presented where a ResNet-like CNN was used as the generator and more standard CNN was used as the discriminator. The test case presented had three velocity fields and a pressure field, at low Reynolds number, on a 64x64x64 grid. This data was downsampled to a range of different scales to provide the input which TEGAN then attempted to reconstruct. The physics informed aspect of the system arises from the need for the results to respect conservation laws and to do this, the generator is trained using a loss function that is based on the equations governing the flow, applied to the generated fields. Comparison of results with a more traditional learning system (TEResNet) shows that TEGAN performs better and there is also evidence of better generalization.

The second part of the talk focussed on a technology to learn solutions of PDEs using Physics-informed Neural Nets (PiNNs) using only the governing equations and given some boundary conditions, i.e., *without any data*. The solver methodology was described and a number of increasing complex examples discussed. The approach can be applied systems like the lid driven cavity but also to inverse problems (such as finding the unknown coefficients of a PDE describing a heat sink in a processor), parameterized PDEs. The technology is available in the NVIDIA SimNet toolkit.

Jonathan Weyn (Microsoft Research/University of Washington)

Jonathan's title was: *Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models*. This talk acknowledged the growing use of ML in areas such as post-processing, extreme weather prediction and replacing physics-based models with emulators, among others. The focus of this talk was on whether CNNs could be used to predict the evolution of the whole atmosphere. The system developed is called Deep Learning Weather Prediction (DLWP) and currently handles only a few (but key) variables and uses a cubed-sphere grid. The aim is to predict at a two-to-six-week timescale (i.e., subseasonal to seasonal). Training data is acquired from data assimilation programs. Current forecast systems have low skill in predicting one-or 2-week-average weather patterns at these timescales. An ensemble approach to prediction is taken with ensemble spread primarily produced by randomizing the CNN training process to create a set of 32 DLWP models with slightly different learned weights. A DLWP model recursively predicts key atmospheric variables with six-hour time resolution. The approach is computationally efficient, requiring just three minutes on a single GPU to produce a 320-member set of six-week forecasts at 1.4-degree resolution. Examples of use were then discussed. Although the DLWP model does not forecast precipitation, it does forecast total column water vapor, and it gives a reasonable 4.5-day deterministic forecast of Hurricane Irma. In addition to simulating mid-latitude weather systems, it spontaneously generates tropical cyclones in a one-year free-running simulation. At the subseasonal to seasonal (S2S) scale the DLWP ensemble is only modestly inferior in performance to the ECMWF S2S ensemble over land at lead times of 4 and 5-6 weeks. However, at shorter lead times, the ECMWF ensemble performs better than DLWP.

Session 3 – High performance, Infrastructure and Big data challenges, Chair: Carlos Gomez Gonzalez

Tal Ben-Nun (ETHZ), Keynote

Tal's title was: *Scaling Up Deep Learning Workloads - A Data-Centric View*. With datasets in the weather and climate domain currently being in the region of 100s of GBs and rapidly heading to beyond TB size, and current DL DNN models already requiring more than a single GPU, technologies to scale to larger machines are required. One promising approach is to exploit more concurrency in the DL pipeline, for example, by passing gradient information on as soon as it becomes available). By default, exploiting more concurrency will lead to lots of data movement on a large distributed system. This problem has led to the development of DNN compilers which, exploiting compiler construction techniques, take the description of a DNN, convert it to a graph representation and translate to an IR (internal representation) which typically treats DNN operators as black boxes. The IR can then be manipulated by applying transformations (including merge steps to fuse operator steps) before generating a mapping of the resultant DNN to hardware. Examples of DNN compilers include IntelNGraph and the TVM Stack. Tal then described their DACE system which is a recent development in the space of DNN compilers. DACE "goes inside" black box operators and works at that lower level. DACE has a focus on data movement optimization, including 4D data transformations such as transposition and provides support for the global optimization of data use. DACE also supports the exploitation of data parallelism (enabled by duplication of parameters), model parallelism, e.g., over operators (implying communication to back propagation) and pipeline parallelism across layers. Pipelining can work well with static NNs but node stalling can be an issue. The DACE system also has some optimizations for the training pipeline and the techniques can also be applied to GANs. In training there is a focus on providing an efficient communication substrate including support for efficient global communication algorithms and asynchronous execution. Examples showing how data movement volumes can be significantly reduced were presented using two case studies on transformer neural networks executing on the Piz Daint supercomputer. The DACE versions outperform existing frameworks by better utilizing both local compute and distributed I/O resources. DACE is available as open source on github.

Micheal Simpson (NOAA)

Micheal's title was: *Radar QPE and Machine Learning*. This was a talk related to the impact and mitigation of extreme events, in this case rainfall and flooding. Accurate, rapid, prediction of precipitation (quantitative precipitation estimation (QPE)) is very difficult in areas of extreme terrain which have their own microclimates. The main examples used in the talk was Taiwan and the US. Currently, QPE is based on a mix of real-time radar data, rainfall gauge measurements, information about lightening and satellite data, from their Multi-Radar Multi-sensor (MRMS) system, all blended into physics-based prediction models. Predictions are validated against independent sets of gauge data. There is a huge amount of data available (~1TB daily) and the question is whether ML techniques can improve predictions. NOAA have investigated using a CNN model and a LSTM (Long Short-term Memory), which both show good results for both Taiwan and the US and may well improve on prediction in areas with low radar coverage. Training is, as always, an issues – currently 18 days training were required when applying the models to Taiwan, but new hardware should improve this. Future models will include combining a CNN with the LSTM, but other approaches will also be explored (Unet, GANs etc.).

Pete Warden (Google)

Pete's title was: *Using ML at the Edge to Improve Data Gathering*. Pete's focus is on ML in edge devices which are at the opposite end of the scale from supercomputers and large data centre-based computing. Edge computing is beginning to allow the deployment of large numbers of cheap, independent processing devices into the environment which provides another, flexible, way to gather new, or different, data in areas which are currently sparsely covered; for example, lots of sensors could be deployed in areas of sea to gather detailed information on temperature, tides etc. Developments in ML at the edge, for example, TensorFlow-lite-micro, provide the possibility of processing gathered data in the devices in the environment, thus reducing the volume and frequency of data returned to a central station. This contrasts with current practice where data gathered by remote devices is directly returned to the cloud and would also conserve power in the distributed devices. Pete then gave several examples of environmental-related work which might inspire further use cases in the weather and climate community. Images from small cameras distributed around a city are being used to estimate air quality using DL on the device (a classification exercise based on visibility). Images from cameras can also be processed to make human-like judgement calls on processes such as precipitation and plant sprouting. Another example discussed the use of old Android phones with solar panels attached to trees in the Amazon rain forest which are being used to detect illegal tree felling by listening for the sound of saws and trucks. Sensors other than cameras are also becoming available with ML processing being able to be executed on devices like the RasberryPi and Arduino (MobileNet runs on an Arduino at 1 frame per second). Small, cheap devices allow deployment at scale. Another example involved a cheap way to provide smart metering of power use in homes. Rather than installing expensive equipment, cheap cameras can be focused on the dials of existing meters and use ML to interpret their readings. Pete posed the question of what could be done with a near infinite number of people texting local information regularly? The possibilities opened up by the new capabilities of machine learning on cheap, embedded systems, and the smart sensors they produce, should find interesting use cases in weather and climate modelling.

Phil Ridley (Arm Ltd.)

Phil's title was: *An Overview of ML and AI on Arm Based HPC Systems for Weather and Climate Applications*. Phil presented aspects of Arm's developing support for ML and AI applications in three areas: support for containerization of workflows through standard interfaces; developments in processors and library support; and Instruction set developments. Arm are part of the Open Container Initiative (OCI) and support the open-source Kubernetes and Docker technologies for cloud/edge applications, which are well supported on Arm. There are several images available for direct download with tools targeting Arm-based Aarch64 architectures (e.g., TensorFlow, DeepBench, Torch, Mahout, Weka, Caffe, Theano, which are all tuned for Arm architectures). In terms of processors, the Int8 GEMM performance has improved by a factor of nearly 6 over that of an A72 on their recent N1 (Ares) processor and will be nearly 26x faster in their upcoming Zeus processor. N1 has reduced half-precision support. Processors will continue to have more cores/CPU, support for vector operations (Scalable Vector Extensions (SVE)), intrinsics for MMA (matrix multiply accumulate), dot product and other common operations, and Bfloat16 support (in Zeus). Arm has a number of ML-related libraries with their Arm Compute library (supported by a team of 8 specialists in Manchester) including ArmNN and CMSIS-NN. Processing elements include CPU, Mali GPU, NPU (Neural Processing Unit) and support for integrating with partner IP, including FPGAs.

Jan Ackmann (Oxford University)

Jan's title was: *Machine-Learned Preconditioners for Linear Solvers in Geophysical Fluid Flows*. The focus of this talk was on using ML techniques to accelerate the solvers used in semi-implicit schemes. Semi-implicit schemes are used in a number of atmosphere and ocean dynamical cores, one of their key advantages being that they enable large model time-steps. However, these schemes involve a costly linear solve of the pressure equation for which preconditioning techniques, which are now standard, have been widely researched. ML offers a new route to finding good preconditioners. Jan's work has looked at incorporating ML into the step controlling convergence of an iterative solver. The idea is that the residual at each iteration can be linked to a change in the pressure field at each step. A Neural Network can be trained by giving it local stencils of values in the coefficient matrix along with residual values from an iterative solver and the NN can output the pressure increment at grid points. Jan gave the example of using a modified version of the Semi-implicit Richardson preconditioner to solve the MPDATA shallow water model using various ML approaches to predict field increments at grid points as the method progresses, rather than solving the (typically stiff) system in the usual way. Variants of NNs with different numbers of layers and nodes-per-layer were tested and complex NNs were found not to give an edge over linear regression, which is cheap to compute. Convergence was found to be good, though not as good as the original SI Richardson solver, which is much more costly to compute. Hybrid methods using SI Richardson and the NN applied in different latitude region give a x2 improvement overall over pure SI Richardson. Future work will be to apply the technique to more realistic models.

Theo McCaie (MetOffice Infrmatics Lab)
Theo's title was: *You do you. How next-gen data platforms can stop weather and climate scientists from being software engineers and other perversions*. Theo's talk identified the software engineering problems faced by data scientists as the amount of data needing to be processed has escalated in recent years. The talk discussed some of the common issues data scientists face in using and analysing their data and pointed to useful tools and growing communities to interact with that can help improve efficiency. Essentially, data science effort appears to consist of around 80% "data wrangling" to get the data into a sufficient shape to analyse, and only 20% performing analyses that generate scientific insight. So any attempt to improve efficiency should focus on the data wrangling. Theo stressed the need for models to produce "Analysis-ready data" which would be supported by a standard interface to make datasets trivial to load and analyse. Approaches are emerging providing such common ways to data wrangling which will benefit many scientists. One useful project is the Spatio-temporal Asset Catalogues (STAC). This is a common language for describing and discovering datasets. Use of such an approach would eliminate individual scientists having to develop their own, dataset-specific, pipelines for this. Theo recommends that scientists should use STAC in their tools and in the data they publish. There are also cloud-based solutions emerging, such as ARCO which help integrate the thousands of files typically produced nowadays into a coherent dataset. The Pangeo community is an Open source community supporting big data geoscience research. Pangeo has an ML group and supports cluster computing and manages low-level scheduling and fault tolerance. Pangeo also has funds to provide entry-level training materials and looks to be a good community to engage with (for those who are not already). Alternative communities include DISCOURSE.

Torsten Kurth (NVIDIA)

Torsten's title was: *3D bias correction with deep learning in the Integrated Forecasting System*. Torsten reported work directed at extracting bias information from ESMWF's IFS model using satellite data. This has led to improving the bias in 4D-Var data assimilation through ML techniques using a DL system which is an extended version of DeepLabV3. Errors in numerical weather prediction (NWP) arise from two main sources: incorrect initial conditions and deficiencies in the forecast model. To correct initial errors, 4D-Var adjusts the initial state of the atmosphere to fit the most recent meteorological observations. The forecasts produced by IFS are known to have large stratospheric temperature biases. Torsten's approach takes a deep learning approach for offline bias correction based on satellite temperature retrievals from Radio Occultation (RO) measurements. Satellites usually provide a good spatial and temporal coverage, but they will never observe the different physical variables for all the model grid points. This problem of data sparsity can be reduced by averaging, but this also impacts the resolution of the bias. The talk focussed on how to incorporate support for data sparsity into the deep learning model and discussed possible approaches, showing preliminary results which compare favourably to those obtained from traditional approaches.