Virtual Conference on Applications of Statistical Methods and Machine Learning in the Space Sciences

17 – 21 May 2021

hosted by
Space Science Institute, Boulder, Colorado

Applications of Statistical Methods and Machine Learning in the Space Sciences

ABSTRACTS OF E-POSTERS[1]

1 **Yasser Abduallah**, New Jersey Institute of Technology, Newark, New Jersey, USA

Deep Learning-Based Reconstruction of Solar Irradiance

Yasser Abduallah, Jason T. L. Wang, Khalid A. Alobaid, Serena Criscuoli, Haimin Wang

The Earth's primary source of energy is the radiant energy generated by the Sun, which is referred to as solar irradiance, or total solar irradiance (TSI) when all of the radiation is measured. A minor change in the solar irradiance can have a significant impact on the Earth's climate and atmosphere. As a result, studying and measuring solar irradiance is crucial in understanding climate changes and solar variability. Several methods have been developed to reconstruct total solar irradiance for long and short periods of time; however, they are physics-based and rely on the availability of data, which does not go beyond 9,000 years. Here we present a new method, dubbed TSInet, to reconstruct total solar irradiance by deep learning for short and long periods of time that span beyond the physical models' data availability. On the data that are available, our method agrees well with the state-of-the-art physics-based reconstruction models. To our knowledge, this is the first time that deep learning has been used to reconstruct total solar irradiance for more than 9,000 years.

8 **Laura A. Balmaceda**, George Mason University, Fairfax, Virginia, USA

Probabilistic cross-matching of CME catalogs

Laura A. Balmaceda (presenter), Fernando Mut (GMU)

Coronal mass ejections (CMEs) are huge amounts of plasma expelled from the Sun into the interplanetary medium. They play the most important role in Sun-Earth relationship as they cause severe perturbations in the Earth's space environment. These include large solar energetic particle (SEP) events and major geomagnetic storms, both representing hazards to humans and their technology in space and ground. Although they were discovered in the 70's, it was not until 1995 that their systematic observation was possible. The rapid increase of the CME research with the advent of the Solar and Heliospheric Observatory mission in 1995 and Solar TErrestrial RElations Observatory in 2006 enabled the compilation of CME catalogs which present differences in the techniques used for CME detection, the delivered products, and the number of events reported, among other things. In order to advance in the knowledge of CMEs it is crucial to obtain a unified list of CMEs. The cross-identification among the entries of different catalogs is, however, a difficult task scientifically, computationally and statistically.

In this work we explore the applicability of a probabilistic approach that uses Bayesian hypothesis and has already been proven successful in other fields in Astrophysics, to this problem.

13 **Shreya Bhattacharya**, Royal Observatory of Belgium, Belgium

Quality Assessment of Sunspot data using various catalogs

Shreya Bhattacharya, Maarten Jansen, Laure Lefèvre, Frédéric Clette

The SN series is one of the longest and most detailed available series in astrophysics. The series was first constructed in 1849 by Prof. Rudolf Wolf and a time series is built in real time since then, involving a lot of observers who differ from each other in terms of their way of counting sunspots, different telescopes and eye sights, which demands proper calibration.

We present a novel time-dependent error determination on the sunspot number (SN) based on non-parametric statistical techniques in smoothing. In particular we propose a generalized linear regression model with overdispersed count data as response variables in the estimation of a time varying calibration of different sunspot time series with overlapping periods. The nonparametric regression

[1] The numbers on the left hand side are the abstract numbers given in the Abstract Book

takes place through a local polynomial smoothing procedure.
This data assimilation model does not restrict to SN only but is applicable to most sunspots' parameters such as area covered on the Sun, position etc... Many catalogs such as the one from Catania Observatory, the Royal Greenwich Observatory, the US Air Force solar observations and others, include the above mentioned parameters whose overall quality assessment is lacking as of yet. The time dependence criteria of our model allows us to access the quality of daily observations with respect to other catalogs thus adding an error bar. We focus this study on one of the stable stations (Mathieu et al,2019) of the World Data Center SILSO network, the Uccle Solar Equatorial table station in Brussels (USET).
The study we present aims at a further refinement of earlier work in Mandal et al. (2020), adding a realistic statistical model and a time dependent calibration factor. Applying regression before computing the calibration avoids the presence of outliers and biased estimation.

15 **Téo Bloch**, University of Reading, UK

Deep-Ensemble Modelling of Electron Flux at the Radiation Belt's Outer Boundary With Bayesian Neural Networks

T. Bloch, P. Tigas, C. E. J. Watt, M. J. Owens

As space-based infrastructure (and society's dependence on it) becomes more ubiquitous, it is evermore important to be able to accurately model the environment within which spacecraft will pass their lifetime. For spacecraft in geosynchronous orbits or those which utilise electric orbit raising, specifically, understanding the outer (electron) radiation belt is critical.
There are a variety state-of-the-art radiation belt models each taking different approaches to understanding the radiation belts. One commonality between them is the importance of correctly quantifying the outer boundary – which acts as a time dependent source for the simulations. Previous work (T. Bloch et al., 2021, under review) quantified one aspect of this boundary, it's location – nominally located at 8.25 RE, further out that most models currently use. This leaves the other curial aspect to be characterised - the electron distribution function at the boundary location. Our work addresses this latter aspect.
Given the relatively distant location of the boundary (well beyond geosynchronous orbit, or the apogee of RBSP), we use THEMIS SST data obtained in its vicinity. As the THEMIS spacecraft cannot always be in the correct location, we create a deep-ensemble Bayesian neural network model to map from fluxes measured from geosynchronous orbit by GOES to the boundary location. The model is additionally parameterised using solar wind and geomagnetic index data from OMNI. Our approach not only allows us to predict fluxes for 11 THEMIS SST energy channels simultaneously, but also produces probabilistic outputs with associated uncertainties.
On average, our model predicts the fluxes within a factor of 1.5 for the lower energies and with a factor of 2.5 for the higher energies. The correlation between our predictions and the measured values is 0.5-0.8 across the energy channels.

19 **Luke Bowden**, Cornell University and the SULI program at SLAC

Cosmological Evolution of the Formation Rate of Short Gamma-ray Bursts With and Without Extended Emission

Dainotti, M. G., Petrosian, V., and Bowden, Luke

Originating from neutron star-neutron star (NS-NS) or neutron star-black hole (NS-BH) mergers, short gamma-ray bursts (SGRBs) are the first electromagnetic emitters associated with gravitational waves. This association makes the determination of SGRBs formation rate (FR) a critical issue. We determine the true SGRB FR and its relation to the cosmic star formation rate (SFR). This can help in determining the expected GW rate involving small mass mergers. We present non-parametric methods for the determination of the evolutions of the luminosity function (LF) and the FR using SGRBs observed by Swift, without any assumptions. These are powerful tools for small samples, such as our sample of 68 SGRBs. We combine SGRBs with and without extended emission (SEE), assuming that both descend from the same progenitor. To overcome the incompleteness introduced by red shift measurements we use the Kolmogorov-Smirnov (KS) test to find flux thresholds yielding a sample of sources with a redshift drawn from the parent sample including all sources. Using two subsamples of SGRBs with flux limits of 4.57×10^{-7} and 2.15×10^{-7}erg cm^{-2} s^{-1} with respective KSp = (1, 0.9), we find a 3σevidence for luminosity evolution (LE), a broken power-law LF with significant steepening at L $\sim$ 1050erg s^{-1}, and a FR evolution that decreases monotonically with redshift (independent of LE and the thresholds). Thus, SGRBs may have been more luminous in the past with a FR delayed relative to the SFR as expected in the merger scenario.

23 **Elena Garca Broock**, Instituto Astrofsico de Canarias, La Laguna, Tenerife, Spain

Performance of solar far-side active regions neural detection

Tobas Felipe (1,2) and Andrés Asensio Ramos (1,2)

(1) Instituto Astrofsico de Canarias, 38205, C/ Va Láctea, s/n, La Laguna, Tenerife, Spain

(2) Departamento de Astrofsica, Universidad de La Laguna, 38205, La Laguna, Tenerife, Spain

Far-side helioseismology is a technique used to detect activity signatures in the far hemisphere of the Sun, based on near-side wave field interpretation. We evaluated the performance of a new neural network approach, developed to improve the sensitivity of the seismic maps to the presence of far-side active regions, and thoroughly compared it with the standard method commonly applied to predict far-side active regions from seismic measurements, using STEREO extreme ultraviolet observations of the far hemisphere as a proxy of activity.

We have confirmed the improved sensitivity of the neural network to the presence of far-side active regions. Approximately 96% of the active regions identified by the standard method with strength above the threshold commonly employed by previous analyses are related to locations with enhanced extreme ultraviolet emission. For the same amount of false positives, the neural network can provide a 28% increase in the number of far-side active region detections confirmed by their extreme ultraviolet brightness. Weaker active regions can be detected by relaxing the threshold in their seismic signature. For almost all the range of thresholds, the neural network delivers a higher number of confirmed detections and a lower rate of false positives.

The neural network is a promising approach to improve the interpretation of the seismic maps provided by local helioseismic techniques, which can lead to improvements in space weather forecasting.

24 **Giovanni Bruno**, INAF - Catania Astrophysical Observatory, Italy

Filtering stellar activity out from exoplanet observations with Gaussian processe

Giovanni Bruno, Vikash Singh, Luca Malavolta, Isabella Pagano, Gaetano Scandariato, Daniela Sicilia, Flavia Calderone

Dynamo-driven stellar magnetic activity is one of the most problematic sources of contamination on the exoplanet signal, because of the stochastic behaviour of its manifestations, such as starspots and faculae. Photometric and spectroscopic spurious signals with a similar amplitude as the planetary signal hamper the detection and characterization of Earth-like planets.

Machine-learning Gaussian processes (GP) regression algorithms have become a standard approach to tackle at least part of this issue. The flexibility of this method allows the treatment of stellar noise as correlated signal.

Our team is active in the application of this technique to exoplanet detection and characterization.

25 **Andrea Bulgarelli**, INAF/OAS Bologna, Italy

The AGILE on-ground event filtering

A. Bulgarelli, A. Argan, A. Chen, V. Fioretti, A. Giuliani, F. Longo, N. Parmiggiani, A. Pellizzoni, M. Tavani, A. Trois, S. Vercellone, N. Parmiggiani, L. Baroncelli, A. Addis, A. Di Piano

AGILE is an ASI (Italian Space Agency) Small Scientific Mission dedicated to high-energy astrophysics, which was successfully launched on April 23, 2007. The AGILE/GRID instrument is com-

posed of three main detectors: a Tungsten-Silicon Tracker designed to detect and image photons in the 30 MeV - 50 GeV energy band, a Mini-Calorimeter that detects gamma-rays and charged particles energy deposits between 300 keV and 100 MeV, and an anti-coincidence (AC) system that surrounds the two detectors.
This paper presents the AGILE/GRID event filter executed by the on-ground analysis software to discriminate the background events. This background event filter called FM3 is currently used for the scientific analysis of the AGILE/GRID data. It is based on a Boosted Decision Tree (BDT) technique and selects events of interest (signal events) out of numerous background events.
The BDT technique maximizes the signal-to-background ratio, suppressing the background events and keeping a high signal detection efficiency. Each event is described by a set of parameters used as input for the BDT. The FM3 filter is developed by training the BDT with a Monte Carlo training sample of events and then tested with an independent Monte Carlo testing sample. The track reconstruction is performed with a Kalman filtering technique and provides the event incident direction and energy. The track reconstruction and other methods are used to provide input parameters to the Multivariate analysis method.

31 **Luisa Capannolo**, Boston University, USA

Deep Learning Techniques to Identify the Drivers of Relativistic Electron Precipitation

L. Capannolo, W. Li, and S. Huang

Energetic ($>$ 10s keV) electrons that populate the Earth's radiation belts can precipitate into the Earth's atmosphere. Electron precipitation is not only one of the key mechanisms of causing radiation belt electron loss, but also affects the Earth's atmosphere. Thus, identifying the drivers and quantifying this process is important to understand the dynamics of the radiation belts and the impact on the Earth's atmosphere.
Previous studies found that relativistic electron precipitation is typically associated with various plasma waves. However, electron precipitation can also be caused by another process called current sheet scattering (CSS), which occurs when magnetic field lines are stretched away from Earth. These two drivers can be distinguished by the shape of precipitation fluxes observed at Low-Earth-Orbit (LEO).
In this study, we use deep learning techniques to identify the driver of electron precipitation (waves or CSS) by classifying the electron precipitation observed by the NOAA POES satellites at LEO. POES satellites provide measurements of electron fluxes in 4 integral channels, from $>$ 30 keV up to $>$ 700 keV, over several magnetic local time (MLT) sectors and L shells. To train the models, we use a dataset of wave-driven events together with a dataset of CSS-driven events that occur over 22-02 MLT. These events were classified by visual inspection in the two categories depending on their precipitation shape. We use data augmentation techniques to improve the dataset, and we compare the performance of the various deep learning models. Our ultimate goal is to obtain a dataset of events associated either with waves or with CSS at all MLTs, which can be used in future studies to understand how the distribution of wave-driven electron precipitation compares to that driven by CSS, and to quantify their relative contribution to the overall observed electron precipitation.

37 **Yaxue Dong**, Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado, USA

Identifying Fundamental Drivers of Martian Ion Escape Using an Artificial Neural Network Model

Y. Dong[1], R. Ramstad[1], D. Brain[1], X. Fang[1], S. Ruhunusiri[2], J. McFadden[3], J. Halekas[2], J. Espley[4], F. Eparvier[1], and B. Jakosky[1]

[1]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO; [2]Department of Physics and Astronomy, University of Iowa, Iowa City, IA; [3]Space Science Laboratory, University of California, Berkeley, CA; [4]Goddard Space Flight Center, NASA, Greenbelt, MD

Understanding the dependence of Martian ion escape with upstream solar radiation and solar wind conditions is crucial for the studies of long-term atmospheric loss and evolution of Mars, and more generally, unmagnetized planets. However, quantifying the dependence of the Martian ion escape rate on multiple upstream drivers (solar Extreme UltraViolet flux, solar wind, and interplanetary magnetic field) from observations can be challenging. With limited data and correlated upstream parameters, it is difficult to distinguish the effect of each individual parameter on the ion escape. Besides, manually constructing any empirical model of the ion escape rate as a function of multiple upstream parameters risks introducing subjective bias. In order to avoid these limitations in traditional data analysis methods, we apply machine learning methods to study the Martian ion escape variation with multiple upstream drivers. We train neural network models using data from the NASA Mars Atmosphere and Volatile EvolutioN mission, with positions and multiple upstream parameters as input and planetary ion densities and velocities as output. The successfully trained model will return spatial distributions of the escaping ion density and velocity with any upstream condition input, and thus can be used to quantify the ion escape rate dependence on multiple upstream drivers and to study the underlying mechanism of these variations. We will also compare the results from machine learning methods with those from traditional data analysis methods to assess the advantages and limitations of both methods in studying Martian ion escape variation.

38 **Luiz Fernando Guedes dos Santos**, NASA GSFC/CUA

How flux rope signatures are impacted by magnetic field fluctuations? A machine learning approach

Luiz F. G. dos Santos, Ayris Narock, Thomas Narock, Teresa Nieves-Chinchilla, Vadim Uritsky

Identifying a coherent magnetic configuration of internal magnetic structures inside Interplanetary CMEs (ICMEs) caused by a spacecraft crossing a large flux rope with helical magnetic field lines topology is critical for predicting these structures' geomagnetic effect reaching Earth. The increased number of space- and ground-based capabilities increased the amount of data available, necessitating more people and methods to analyze this additional data. Dos Santos et al. (2020, Solar Physics) recently conducted a study to implement machine learning (ML) techniques and expand our understanding of the Space Weather hazard's main drivers. This study successfully interpreted the ICME in situ magnetic field observations using ML. It gained a thorough understanding of what to expect from in situ magnetic field observations when a spacecraft crosses flux ropes with different trajectories. We use a pre-existing Deep Neural Network handwriting model that has been trained with synthetic data with high accuracy in well-behaved events and then tested it against observed ICMEs from WIND from 1995 to 2015.

We correctly identified flux rope signatures in 84 percent of simple real-world cases and had a 76 percent success rate when applied to a more extensive set. Moreover, we use this established model to investigate magnetic field fluctuations in flux rope identification by analyzing observed fluctuations in magnetic components and constructing more realistic synthetic data ones.

39 **James "Andy" Edmond**, graduate student at University of New Hampshire, USA

Magnetospheric Plasma Region Classification From THEMIS Data Using Machine Learning

James Edmond, Joachim Raeder, Banafsheh Ferdousi, Maria Elena Innocenti

The volume of data generated by magnetospheric instruments accelerates with our computational progress, but this data becomes voluminous enough that parsing it is time-consuming for the scientist. Using machine learning to automate the classification of magnetospheric data into different regions would reduce the human time spent. To this end, we have used the KMeans clustering algorithm to classify 2 minute-averaged THEMIS data into different plasma regions of the magnetosphere. Among the regions we can generally identify are, (1) pristine solar wind / magnetosheath plasma, (2) lobe plasma, (3) inner magnetospheric plasma, and (4) the plasma sheet. We also discuss how these algorithmically-determined regions change using different styles of input (e.g. the classifications found when using all components of the magnetic field vector measurements vs simply the vector magnitude).

40 **Haroun El Mir**, Cranfield University, UK

Certification Approach for Physics Informed Machine Learning and its Application in Landing Gear Life Assessment

Suresh Perinpanayagam, Haroun El Mir, Cranfield University

Landing Gear Systems on Aircraft undergo a multitude of forces during their life cycle, leading to the eventual replacement of this system based on a "safe life" approach that majorly underestimates the component's remaining life due to factors such as generalizing the impact cycle to all Landing Gear systems. The efficacy of fatigue life approximation methodologies is studied and compared to the ongoing Structural Health Monitoring techniques being researched, which will forecast failures based on the system?s specific life and withstanding abilities, ranging from creating a digital twin to applying neural network technologies, in order to simulate and approximate locations and levels of failure along the structure. Explainable Artificial Intelligence allows for the ease-of-integration of Deep Neural Network data into Predictive Maintenance, which is a procedure focused on the health of a system and its efficient upkeep via the use of sensor-based data. Test data from a flight includes a multitude of conditions and varying parameters such as the surface of the landing strip as well as the aircraft itself, requiring the use of Deep Neural Network models for damage assessment and failure anticipation, where compliance to standards is a major question raised, as the EASA AI roadmap is followed, as well as the ICAO and FAA. This paper additionally discusses the challenges faced with respect to standardizing the Explainable AI methodologies and their parameters specifically for the case of Landing Gear.

42 **Divyam Goel**, University of California Berkeley, USA

Exploringthe effectsof geomagnetic storms in the ionosphere using Principal Component Analysis

Divyam Goel,Thomas Immel, Yen-Jung Wu, Brian Harding, Colin Triplett, Chihoko Cullens

Understanding the short term variability of the ionosphere is a key puzzle in our understanding of space physics. The recently launch GOLD satellite provides a consistent view of the Earth due to its Geostationary orbit positioned above the SAA. This view creates geospatial-temporal data which allows us to observe the trends in the ionosphere over the same location. This is especially useful when studying the effects of geomagnetic storms on the ionosphere by allowing us to draw comparisons with quiet days. We perform Principal Component Analysis to GOLD's O/N2 disk data to study the disturbances of the O/N2 ratio due to geomagnetic storms. We plot the eigenvectors from the PCA to study the major modes of variation in O/N2 data in the month of April. We also compare the weights of the principal components for every day of April. We observe that the first principal component in April 2020 shows a significantly large weight on the 20th April storm day. We then verify the influence of the storm on the monthly principal component by comparing the eigenvectors from the PCs with and without the storm day. The weights of the third principal component also hint at a quasi 6-day variation in the month of April. We intend on further investigating the quasi 6-day variation and its relationship with the geomagnetic storm. We also intend on applying more machine learning techniques to try and find daily O/N2 variations in the ionosphere.

43 **Jeremy Grajeda**, Klipsch School of Electrical & Computer Engineering, New Mexico State University, USA

Exploring Stability of Coronal Hole Detection to Intensity, Spatial Resolution, and Short Timescales

Jeremy Grajeda & Laura E. Boucheron (NMSU), Michael Kirk (ASTRA LLC, NASA GSFC)

Accurate detection and segmentation (delineation) of coronal holes is an important input to solar wind models. Many algorithms for automated coronal hole detection, however, rely on intensity thresholds to segment solar images. Relying on intensity thresholds can produce vastly different segmentations between and within datasets as a result of changes in overall image intensity due to intrinsic and extrinsic factors. Rather than relying on intensity thresholds, the Active Contours Without Edges (ACWE) algorithm relies on the relative homogeneity of quiet sun and coronal hole regions. The use of the ACWE algorithm for coronal hole segmentation was validated in Boucheron, Valluri, & McAteer (2016). Here, we quantify the stability of ACWE-based coronal hole segmentation to variation in image intensity and resolution such as those introduced in common image formats. In this work we study 193 A images from SDO for August 1 through September 27, 2010. We use a suite of metrics including Intersection over Union (Jaccard index), Structural Similarity Index Measure, and Global and Local Consistency Error. We find that the ACWE algorithm is stable for aggressive intensity scaling, displays no significant artifacts for spatially scaled images, and is consistent in segmentation across small timescales. Through this study we are able to study the systematic versus physical changes within a coronal hole.

49 **Sheng Huang**, Boston University, USA

Hiss in the Plasmasphere and Plumes: Global Distribution From Machine Learning Technique and Their Effects on Global Loss of Energetic Electrons

Sheng Huang, Boston University; Wen Li, Boston University; Xiaochen Shen, Boston University; Qianli Ma, Boston University; Xiangning Chu, University of California Los Angeles; Luisa Capannolo, Boston University

Whistler mode hiss waves are typically observed inside the plasmasphere and plumes, and are known to play an important role in energetic electron loss processes in the Earth's inner magnetosphere. In particular, hiss in plumes is previously shown to be stronger than the waves inside the plasmasphere; however, it has been challenging to achieve the dynamic evolution of hiss inside the plumes on a global scale. We use machine learning technique, more specifically, artificial neural network (ANN) to construct the global evolution of total electron density and hiss wave amplitude inside the plasmasphere and plume and the associated hiss waves therein. These constructed hiss wave models are used to quantify the effects of hiss on the global electron loss at $L < 6$ using the 3D Fokker Planck simulation. We demonstrate that neural network is able to reconstruct the dynamic evolution of total electron density and hiss inside the plasmasphere and plume. Moreover, the simulation result indicates that plume hiss can cause an efficient loss of energetic electrons in the outer radiation belt.

52 **Maria Elena Innocenti**, Ruhr-Universitaet Bochum, Germany

Unsupervised classification of simulated magnetospheric regions

Maria Elena Innocenti, Jorge Amaya, Joachim Raeder, Romain Dupuis, Banafsheh Ferdousi, Giovanni Lapenta

In this work, we apply an unsupervised classification method based on Self Organizing Maps (SOMs) to data points from global magnetospheric simulations obtained with the OpenGGCM-CTIM-RCM code, as a first step towards classification of magnetospheric observations.

The classification relies exclusively on local plasma properties at the selected data points, without information on their neighborhood or on their temporal evolution, a characteristic that makes the method particularly appealing for spacecraft on-board data analysis. We classify the SOM nodes into an automatically selected number of classes, and we obtain clusters that map to well defined magnetospheric regions. For the sake of result interpretability, we examine the SOM feature maps, and we use them to unlock information on the clusters. We repeat the classification experiments using different sets of features, and we obtain insights on which magnetospheric variables make more effective features for unsupervised classification.

55 **Sudha Kapali**, Computational Physics Inc., Massachusetts, USA

Data Validation Framework for Scientific Instruments: A platform for positive feedback between scientific expertise and machine-learning based validation

Sudha Kapali, Jennifer L Gannon, John Noto, Charles Poole, Robert Kerr, Juanita Riccobono, and Michael A. Migliozzi, Computational Physics Inc.

Technological advancements in sensors and instrumentation have made available a large volume of data for space weather analysis, modeling and forecasting. However, the ready availability of validated data sets that can be immediately absorbed for scientific analysis has been unable to keep up with the rapid growth in the volume of data. It requires expert knowledge to tease out true physical events from subtle non-physical artifacts, however, active human involvement in performing the categorization is often not possible when fielding a large array of remote instruments. Further, non-physical artifacts such as system drifts and environmental perturbances are often subtle as data is streamed in from the instrument.
The above characteristics of sensor data from scientific instruments points to the realization that it takes both – domain-specific scientific expertise and the use of advanced statistical, signal processing and machine learning algorithms to produce a solution for the data validation problem.
We present the design of a validation framework and a set of machine-learning based validation tools that are deployed within this framework. The analytics-based framework creates a pathway to capture and utilize domain-specific scientific expertise, provides mechanisms for analyzing science-informed trends in data and takes advantage of co-located sensors and distributed sensor networks to handle complex constraints. Further, incorporating the validation algorithms within a framework rather than implementing them as standalone operations streamlines the process of evaluating new algorithms. Trends and patterns in data validation results provide valuable insights back to the scientific expert thus completing the feedback loop between technology and science.

61 **Thurmon Lockhart** - School of Biological and Health Systems Engineering, Arizona State University, USA

Dynamic Stability of Exoplanetary Systems using the Largest Lyapunov Exponent

Thurmon Lockhart and Douglas Walker

Empirical methods to measure dynamic stability are available from the field of nonlinear dynamics. These have been applied to quantify dynamic stability of exoplanetary systems and to identify chaotic motion (Toth, and Nagy, 2014) as well as musculoskeletal stability during dynamic human locomotion (Lockhart and Liu, 2008). It is reasonable to assume that every walking stride could be similar to every other stride. Similarly, we assume that every orbital rotation of any planetary systems is similar to every other rotation. Natural variance observed in empirical data is therefore attributed to random mechanical disturbances or in humans, neuromuscular control errors. These disturbances are attenuated in time by the gravity (or friction) in order to maintain a stable orbital pattern resembling and exhibiting a fractal and chaotic structure. Disturbance to the orbital trajectory is an ongoing process so the attenuation of these variability is continually manifest in order to remain stable in the planetary system. Thus, stability can be characterized from the time-dependent attenuation or expansion of variability. This is measured by the Lyapunov exponent, λ. One Lyapunov exponent exists for every rotation or partial rotation and has the dimension, λi. A system is stable when the sum of λi is negative, i.e. the rate of kinematic convergence is greater than the rate of divergence. However, calculation of the full Lyapunov spectrum from experimental data is exceedingly difficult. These calculations may be greatly simplified by realizing that two randomly selected initial trajectories diverge, on average, at a rate determined by the largest Lyapunov exponent, λMax. Presentation will include the demonstration of acquiring λMax from using human gait data. We will also investigate the precision of radial velocity data needed to establish λMax for examples of known exoplanet systems.

Lockhart, T., and Liu, J., Differentiating fall-prone and healthy adults using local dynamic stability. Ergonomics, 51(12), 1860-1872.

Toth, Z. and Nagy, I., 2014. Dynamical stability of the Gliese 581 exoplanetary system. MNRAS 442, 454-461 (2014)

62 **Luning Li**, School of Aerospace Transport and Manufacturing, Cranfield University, UK

Digital Twin in Aerospace Industry: The Concept, Applications and Challenges

Luning Li, Suresh Perinpanayagam, Andrew Wileman

Digital Twin (DT), a virtual replica of any conceivable entity, is a highly transformative technology with implications that are profound in nature. For instance, by leveraging the tetra-drivers of innovation (namely, cloud, internet of things, big data, and artificial intelligence), DT can be made to undertake real-time performance assessments and diagnostics more precisely than presently possible; repairs can be affected instantly; and innovations can be catapulted. This, however, is marred with a number of challenges and uncertainties, such as an unobstructed transfer and assimilation of heterogenous data from various source-streams including design, manufacturing, and operations. This paper highlights such challenges in the adaptation of DTs to an aerospace environment whilst also providing a way ahead and future directions. Besides presenting the concept and the growing significance of Digital Twin across a wider spectrum of industries, the paper puts forth several critical future challenges within the aerospace industry vis--vis key applications, such as for prognostics and health management, manufacturing, data quality and quantification, visualization, cybersecurity, etc. These aspects of DT for aerospace (aero-DT) have not been previously elaborated in the literature, thus leaving a void in DTs' optimal implementation. This paper has filled this void by providing a critical analysis of challenges that pose difficult questions for the aerospace industry and suggesting ways along with a roadmap to steer away from them and multiply their potential for innovation in aircraft design, manufacture, and operation.

70 **Jorge H. Namour**, Facultad de Ciencias Exactas y Tecnologa (FACET), Universidad Nacional de Tucumán (UNT), Argentina

Ionosphere F2 critical frequency forecasting using deep learning

Namour, Jorge H (1); Molina, Mara Graciela (1, 2) ; Ise, Juan (1)

(1) Facultad de Ciencias Exactas y Tecnologa (FACET), Universidad Nacional de Tucumán (UNT);

(2) CONICET

In this work, we present preliminary results obtained for the forecasting of the F2 ionospheric layer using deep learning. The forecasting model has been done using 12 years data set (almost a complete solar cycle) of measurements corresponding to the foF2 parameter (F2 ionospheric layer critical frequency) obtained from the AIS-INGV ionospheric sounder deployed at Tucumn Space Weather Center (https://spaceweather.facet.unt.edu.ar/) in Argentina. The mentioned low latitude ionospheric station ($26°51'$ S, $65°12'$ W) is located near the south crest of the Equatorial Ionization Anomaly posing great interest for the space weather community.

We present the preliminary results of forecasting 3 hours in advance using a Long-Short-Term Memory (LSTM) model. This recurrent neural network model is univariate, multistep, and uses 3 hours backward to perform the forecast. Often, data available is highly unbalanced due to the domain characteristics where few cases of sudden ionospheric irregularities caused by space weather events occur within a certain period (e.g. during a year). Thus the training set has been selected to comply with the following criteria: it has to consider as much as possible the different scenarios of the domain, minimize missing values effects (e.g. due to instrument failure), and cover a great portion of a solar cycle. Then, the more intense work was done in selecting and cleaning the training, validation, and testing sets.

In order to evaluate the performance, we used RMSE. The results for the testing sets showed a good agreement with the actual measurements. We also tested the model using extended data sets to stress the model and test how it behaves against new data. Finally, we discuss the results, the possibility of an operative implementation and we propose further approaches to be implemented in the near future.

79 **Mariano Poisson**, Instituto de Astronoma y Fsica del Espacio (IAFE), Argentina

Modeling the photospheric magnetic field distribution of emerging solar active regions

M. Poisson[1], F. Grings[1], M. López Fuentes[1], C.H. Mandrini[1,2], P. Démoulin[3]

[1] Instituto de Astronoma y Fsica del Espacio (UBA-CONICET), Argentina; [2] Universidad de Buenos Aires,

Argentina; [3] Observatoire de Meudon, France.

Active regions (ARs) appear in the solar atmosphere as a consequence of the emergence of magnetic flux tubes. It has been shown that in order to survive their transit through the convective zone, these flux tubes should be twisted, forming so-called magnetic flux-ropes (FR) and transporting with them magnetic helicity from the solar interior. This quantity plays a key role in the production of energetic events such as flares and coronal mass ejections. Photospheric observations of the line-of-sight (LOS) component of the magnetic field reveal that during the emergence of ARs the presence of twisted field lines around the FR axis produces an elongation of the main AR magnetic polarities known as magnetic tongues. In this work we aim to estimate the intrinsic geometrical properties of FRs, by comparing LOS magnetograms of emerging ARs with synthetic magnetograms derived from a toroidal magnetic flux tube model. In particular, we estimate the tilt angle, defined as the inclination of the AR main bipolar axis with respect to the solar equator, which is strongly affected by magnetic tongues. Our method uses a probabilistic scheme based on the Bayes theorem to infer the most probable intrinsic parameters of an emerging flux tube, assuming a normal departure between the model and the observations. Open access python 3 packages, pymc3 and Theano are used to optimize the sampling of the model 8-dimensional parameterco space. Analyzing a sequence of 66 magnetograms, we are able to infer the most probable values of the intrinsic parameters for AR 10268, including its magnetic helicity content and its intrinsic tilt angle.

82 **Gouri Ramesh**; Cranfield University, UK

Digital Fault Simulation and Identification in an Electric Braking System

Gouri Ramesh, Pablo Eduardo Garza, Suresh Perinpanayagam

In recent years, the aerospace industry has tried to introduce technologies such as More Electric Aircraft (MEA) and Integrated Vehicle Health Management (IVHM) to achieve various benefits such as weight reduction and lesser fuel consumption and a decrease in unexpected failures and in turn unavailability of the aircraft. In this paper, data from digital model of an electric braking system of an aircraft is used to identify faults that are critical in nature, namely Open Circuit fault, Intermittent Open Circuit, and Jamming. The variation of characteristic parameters during normal working conditions, and when faults are encountered are analysed qualitatively, and training features for the reasoner are selected. A data-driven approach is adopted for the development of the reasoner due to the availability of clean data. The paper discusses the performance of univariate and multivariate models of Long Short - Term Memory Neural Network time series classifier. Its performance is also studied w.r.t two other models - K Nearest Neighbour time series and Time Series Forest classifiers. The comparison of the reasoners is then carried out in terms of accuracy, precision, recall and F1-score and the relevancy of choosing the correct metrics for evaluation is also discussed.

88 **Anastasia Marie Seifert**, Institute of Space Sciences and Astronomy, Malta

Mask R-CNN based FRB Detection in Noisy Environments

Anastasia Marie Seifert, Andrea De Marco, and Kristian Zarb Adami

The automatic detection and segmentation of fast radio bursts (FRBs) from radio telescopes is extremely useful for instruments with high data throughput such as the Square Kilometre Array. In this paper, we apply a recent deep learning technique to scan channelised spectra for the detection and segmentation of FRBs in simulated highly noisy backgrounds, as well as with injected RFI contamination. Our dataset consists of simulated dispersed FRBs and Green Bank Telescope (GBT) telescope observations were used for backgrounds including RFI, with the simulated events injected at different Signal-to-Noise ratios. This work evaluates the use of Mask R-CNN to be able to detect FRBs in harsh conditions whilst avoiding the need to correct for dispersion before running a more standard FRB detection pipeline. We evaluate our model using the Average Precision (AP) metric at select Intersection of Union (IoU) for object detection and instance segmentation tasks. This paper will also describe the optimal parameters for the task of FRB detection for a Mask R-CNN implementation. The implementation is publicly available at https://github.com/radastro/FRBMaskRCNN/.

92 **Zena Stevenson**, Klipsch School of Electrical & Computer Engineering, New Mexico State University, USA
Using Fully Convolutional Neural Networks to Infer Solar Magnetic Structure from Extreme Ultraviolet Image
Zena Stevenson, Laura E. Boucheron

Recent applications for fully convolutional neural networks (FCNNs) have included the prediction of an image. This has been used extensively for image segmentation, style transfer, and translation between image modalities. Here we study the use of a UNet-based FCNN originally proposed to predict fluorescence microscopy images from transmitted light images of cells. Using data from NASA's Solar Dynamics Observatory, we train a similar FCNN to translate extreme ultraviolet (304 A) images of active regions on the Sun into predicted unsigned line-of-sight solar magnetograms. These SDO data comprise a larger volume and variety than used in the cellular microscopy application. We first present small-scale experiments with similar training data volume to the cellular microscopy images, using 30 training pairs from a single active region at a time. Using normalized mean-squared error and structural similarity index to compare predictions, we find these small scale experiments demonstrate similar success to the cellular microscopy experiments. We next present preliminary results from larger scale experiments using several thousand image pairs for training, first drawn from 10 active regions and then from 52 active regions. We find that this is a promising application for this type of neural network and expect that future work will be able to train a model that can produce quality predictions when working with even higher volumes of data.

95 **Sujitra Sutthithatip**, Cranfield University, UK
The current stage of AI in aerospace applications
Sujitra Sutthithatip, Suresh Perinpanayagam, Sohaib Aslam

In this era, Artificial Intelligence (AI) has been well-known as a machine that is capable of doing things like humans. The growth of AI technology can benefit humans in unsafe and time-cost situations such as aerospace tasks. This technology advancement facilitates human decision-making in crisis events like flight control operation, and with the pros of dealing with massive data, it helps to predict subsequent events in aircraft maintenance. Likewise, AI is also used as a tool to increase passenger satisfaction. The exploiting AI in aerospace arises in a state of applying machine learning models as a simple classifier, and neural network models in sophisticated problems. Whilst the current AI development state carries out the ability of self-driving, motion recognition, and ethical and privacy issues. These technologies come to be a key role to improve in autonomous scenarios like autonomous navigation, real-time vehicle health monitoring, flight prediction, diagnostic and prognostic. However, At the data level, other learnings are recommended to present better performance; for example, Distributed and parallel learning are selected to reduce the suffering of the data volume of learning models in either large or tiny scales of data which might cost infinite computing time. Also, representation learning is chosen when the variety of different types of data such as structured, semi-structured, and unstructured sources manipulate non-linearity, high-dimensionality, and heterogeneous characteristics for the dataset. When the occurrence of a feature changes over time, the feature selection which relies on time range should be considered. Moreover, in case of uncertainty and incompleteness data where the feature is controlled by various values, deep learning is preferably exploited.

98 **Ajay K Tiwari**, Centrum Wiskunde and Informatica, Amsterdam
Predicting arrival time for CMEs: Machine learning and ensemble methods
Ajay K Tiwari, Enrico Camporeale, Jannis Teunissen, Raffaello Foldes, Gianluca Napoletano, and Dario Del Moro

Coronal mass ejections (CMEs) are arguably one of the most violent explosions in our solar system. CMEs are also one of the most important drivers for space weather. CMEs can have direct adverse effects on several human activities. Reliable and fast prediction of the CMEs arrival time is crucial to minimize such damage from a CME. We present a new pipeline combining machine learning (ML) with a physical drag-based model of CME propagation to predict the arrival time of the CME. We evaluate both standard ML approaches and a combination of ML + probabilistic drag based model (PDBM, Napoletano et al. 2018) on the database of more than 200 CMEs. We analyzed and compared various machine learning algorithms to identify the best performing algorithm for this database of the CMEs. We also examine the relative importance of various features such as mass, CME propagation speed, and height above the solar limb of the observed CMEs in the prediction of the arrival time. The model is able to accurately predict the arrival times of the CMEs with a mean square error of 8.8 hours. We also explore the differences in prediction from ML models and ensemble predictions method namely P-DBM model. We also present a way to predict the distribution of the parameters of the PDBM model just using the physical properties of the CMEs.

100 **Thorold Tronrud**, Universidad Andres Bello, in Santiago, Chile

Machine Learning for Galactic Archaeology

Patricia Tissera (Pontificia Universidad Catlica), and Facundo Gomez (Universidad de La Serena)

Wednesday 19 May 2021 3:04PM

Present-day surveys are providing exquisite information about stellar population distributions in the MW, allowing us to reconstruct its formation history, however it can be difficult to separate stars formed in-situ from those that have been accreted over the course of our galaxy's evolution. We introduce a probability-based binary classification method to identify accreted stars in galactic discs using a neural network trained on the age and chemical properties of disc and halo stars. We implement this method across multiple suites of simulations, and quantify the performance of the model in both the recovery of accreted stars, and rates of false positives to build a robust understanding of the model's strengths and weaknesses.

This method is designed to be applied to data collected from observational surveys to both eliminate spurious stellar data from the disc for modelling of galactic chemical evolution. More importantly,it can be applied to flag and isolate accreted stars in the Galactic disc in order to elucidate the accretion history of the Milky Way.