Science Ascend
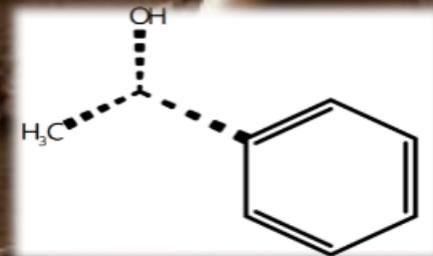
Rising to new heights of discovery with Science!
Every week!

Far more detailed look on the solar system and solar neighborhood, TESS, SOHO and more!

MALDI, biosensors, infrared spectroscopy and more!

Sentinel-1 and 2 fusions, aerial photography, much better neural network bases and more!

Indoor PM measurement insights, air pollution prediction and monitoring and more!

cout<<solutions;

Dimension reduction in categorical and temporal domains, higher performance models and more!

Science Ascend

Rising to New Heights of Discovery!

Science Ascend teleports you to the frontiers of science. It compiles and discuss the scientific research preprints from arXiv, bioRxiv, chemRxiv just from the previous week to be cognizant of the *state-of-the-art* of knowledge in astrophysics, chemistry, environmental chemistry, remote sensing, and applied statistics/data science. Light from the *Science Ascend* will keep brightening the dark horizon beyond the limits of our comprehension. FIRE Araştırma Eğitim Ltd. Şti. guarantees the weekly publication and dissemination of this journal, and make it available for everyone at most fifteen days after its publication freely.

Bilim Yükselişi

Keşfin Yeni Yükseklerine Ulaşmak!

Science Ascend sizi bilimin sınırlarına ışınlar. Astrofizik, kimya, çevre kimyası, uzaktan algılama ve uygulamalı istatistik/veri bilimi alanlarındaki bilgi birikiminin *en son durumu* hakkında bilgi sahibi olmak için arXiv, bioRxiv, chemRxiv'den sadece bir önceki haftaya ait bilimsel araştırma ön baskılarını derler ve tartışır. *Bilim Yükselişi*'nden gelen ışık, kavrayışımızın sınırlarının ötesindeki karanlık ufku aydınlatmaya devam edecektir. FIRE Araştırma Eğitim Ltd. Şti. bu derginin haftalık olarak yayınlanmasını, dağıtılmasını ve yayınlandıktan en geç on beş gün sonra ücretsiz olarak herkesin erişimine açılmasını garanti eder.

Contents

Foreword and Summary

Greetings everyone!

Here is the fifth issue of the *Science Ascend* with several changes from the previous issues formats, and of course, wiith the reports of many novel scientific research outputs.

The format is still very similar to a bibliographic note in many chapters. I am working out a new way to compile recent preprints and read them so the following issues will have more insights and critical thinking, as well as theoretical and empirical details and discussion on the recent studies, and maybe other chapters or other authors, too!

The preprints from the **Astrophysics** part had more Solar and Mars-related studies in the October 1-7, 2024 date range. Other studies were also mostly in solar neighborhoods. This was so even considering the bias of specifically considering astro-ph.EP, which was the case in all issues. Several new instruments were also reported for exoplanet detections.

For **Analytical Chemistry**, quick assays were examined, and studies utilizing mass and infrared spectroscopy in-depth were also present. Computational, as well as machine learning applications took place more in these studies.

In **Remote Sensing**, more emphasis was on the segmentation and object detection where they initially combined several different datasets and squeezed out important factors from them, then, they used these for better predictions in a computationally efficient way. U-Net and MAMBA architectures were extensively used and improved upon.

The **Environmental Chemistry** mostly had monitoring sensor placement-based studies, on air pollution by particulate matter smaller than 2.5 micrometers in diameter.

A major change took place in the last chapter, where the data science was narrowed down to the **Data Decomposition/Transformation**. This is easier to follow and more practical, and closer to the domains where I have been studying and working with algorithms on simulated and real datasets. There were intriguing developments, from the application of SVD in the temporal domain to Graph Fourier Kernels for simulating partial differential equations.

See you at the next issue with more in-depth reviews!

Güray Hatipoğlu

Last week in Astrophysics

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between October 1 - October 7, 2024. These are from arXiv's astro.EP cross-fields without high-energy main cross-list papers.

Stellar Systems - Populations - Clusters

Single Star System (Star, Exoplanet)

Olmschenk et al.[1] constructed a convolutional neural network to look for short-period variables in Transiting Exoplanet Survey Satellite (TESS) mission 30-minute cadence lightcurves. They generated lightcurves from the Full Frame Images of TESS. The package "tess-point" is used to select the targets from the TESS Input Catalog (TIC), and eleanor Python module created lightcurves for each TESS Sector. A total of 67 million light curves were in the study dataset. They separated the data into 80-10-10 training-validation-test sets, respectively. The structure of the neural network[1] was rather explained in Olmschenk's another paper[2]. The NN structure is as follows: 1D convolutional neural network, with 9 convolutional blocks (1D convolution-activation-spatial dropout-pooling (in first 6 blocks)-batch normalization), 3 dense (dense-activation-dropout-batch normalization), and ending with a sigmoid. The training data had 1) real TESS lightcurve (LC), 2) real TESS LC with injected short-period synthetic signals, 3) real TESS LC with injected long-period signals, 4) real TESS LC injected with uniform noise. After the lightcurve creation and NN variable prediction steps, they post-process the data with the Lomb-Scargle periodogram within 1 hour to 10 days period. The highest frequency within 10 % of the maximum power in the Lomg-Scargle periodogram was chosen, and ¿5 hour periods were discarded. The lightcurve, then, was binned to 25 bins in phase space and the median value of the bin was used to determine the minimums and maximums. The false positives with unaligned photometric centers of variability were removed since the origin of variability was assumed to be elsewhere. TESSCut was used to generate time series raw images and 10 x 10 pixels lightcurve and after the similar steps above, the centroid of the variability was estimated. If it was one pixel (21 arcseconds) different than the

---

[1]GitHub links for their NN structure can be found here. Furthermore, the repository for this study is here.

target position, it was discarded. Their sample at the end included mostly main sequence stars and δ-Scuti variables. There was also a human-vetting step on 500 random lightcurves from the algorithm's output, where they considered that 492 out of 500 had an obvious periodic signal in their periodogram. The results were successful with their method and they further compared their results to an existing catalog, analyzed binaries further, and divided the δ-Scuti sample in two, too.

Angelo et al.[3] utilized The Cannon tool[2], which took Gaia Data Release 3 Radial Velocity Spectrometer (RVS) data and generate spectra. In their study, the Cannon used T_{eff}, $\log g$, [Fe/H], [α/Fe], v_{broad} labels to predict the flux at a given wavelength linearly. Among its assumptions are 1) the same label, the same per-pixel flux, 2) the flux at a particular pixel[3] is a polynomial function of the labels, 3) a flux at a given pixel is independent of the neighboring fluxes. After getting trained with both labels and the spectra, the test step attempts to produce the labels from the spectra using the maximum-likelihood Cannon model flux. Their training data with the labels and spectra were from the GALAH+ survey's 588571 objects. After quality flags, filters, and iteratively removing the binaries, they were left with only 563 objects. They modified the Cannon code in three Calcium-related peaks (Ca-triplets). They also compared their method's output labels to the Spectroscopic Properties of Cool Stars (SPOCS) with 107 of the 1615 main sequence stars in R 50000 spectral resolution with the Keck High-Resolution Echelle Spectrometer (HIRES). As long as the Gaia RVS signal-to-noise (SNR) ratio is higher than 50, their model's fit is very successful, such as within 85 K T_{eff} agreement. Moreover, they provided use cases of anomalous spectra identification on active stars, unresolved binaries, and evolved stars.

Réveille et al.[4] studied the star-planet magnetic interactions considering the spectral M-type stars, their magnetic winds, and their planets. They used an Alfvén wave-driven solar wind model, WindPredict-AW, and modified it for working on M-type stars. They also tried to model TRAPPIST-1 and Proxima Centauri b cases from the solar system neighborhood. There is a simple power relation and mass ratios to determine the atmospheric loss and stellar x-ray flux, then, the numerical model considered the wave packet origination, and energy losses while traversing the stellar-exoplanet distance. In the end, they esti-

---

[2]The related GitHub repository with tutorial is here.

[3]For Gaia spectra, different pixels get different colors of the same spatial location, so just consider the word pixel as a "picture element" where the element is another color.

mated magnetic interactions in the TRAPPIST-1 down to the sixth planet, where the fifth and sixth ones are in the star's "habitable zone", and a similar result was the case for Proxima Centauri b.

Pope et al.[5] retrieved surface gravities, masses, and radii of the red giant star samples from radial velocity surveys using TESS photometric light curves. They included 22 stars with a known planet and 26 without a known planet. They used the Python package ligthkurve for retrieving and analyzing the TESS lightcurves. They preferred to use the lightcurves produced by the SPOC and TESS-SPOC pipelines. They analyzed the Pre-search data Conditioning Simple Aperture Photometry (PDCSAP) data and TESS Barycentric Julian Day (BTJD) for the following analyses. They used "to_periodogram" class function of the lightcurve and oversampled it when there is less data. They considered the periodogram's part 100 μHz above and below the expected maximum frequency v_{max} according to the spectroscopic logg and temperature. Then, this periodogram was flattened by subtracting the log_{10} moving filter in the frequency space. They considered different filter widths to do this step for different estimated logg, but for longer period oscillators it was still problematic as there was not much cycle going on within the 27.4 day observation period of TESS. In the end, estimate_numax method was used to compute a 2D Autocorrelation[4] and fit a Gaussian to this. Then, frequency separation was found on 1D autocorrelation to the periodogram centered around the ν_{max} computed previously with the Python SciPy package's find_peaks. Few more procedures more were followed in the case of noise periodograms to refine the $\delta\nu$. They also stacked $\delta\nu$ length periodogram slices on top of each other (echelle diagrams) and discarded the ones without an evidence of a vertical structure. They reported more precise results, slight discrepancies for stellar radii and stellar masses, and there were two outliers HD 100065 and HD 18742, indicative of asteroseismology study challenges while working with red giants.

Iskandarli et al.[6] studied the Helix Nebula (NGC[5] 7293) central star with more TESS data from more TESS sectors than before, as well as modeling the lightcurve with the estimated parameters beforehand. This lcurve[6] model recommended either a Jupiter-sized 0.102 solar radius body with one degree of orbital inclination or a

0.021 solar radius exoplanet with an orbital inclination of 25 degrees. Their spectral energy distribution (SED) figure composed of GALEX, Pan-STARRS, 2MASS, and Spitzer IRAC (excluding IRAC 7.9 micrometers) followed the model SED quite well, meaning that any companion should have the characteristics that made it remain undetected in this curve. It was still uncertain that if the lightcurve indicated a substellar companion or just a stellar variation-related phenomenon, which still required further observations, such as from the James Webb Space Telescope (JWST), but they did constrain the possibilities with their outputs.

Couperus et al.[7] studied four fully convective M dwarf twin binaries in the solar neighborhood: GJ 1183 AB, Kx Com A-BC, 2MA 0201+0117 AB, and NLTT 44989 AB. The observation campaigns were as follows: CTIO/SMARTS 0.9 m long-term optical photometry and short-term optical photometry, CHIRON echelle spectrograph on CTIO/SMARTS 1.5 m, Chandra X-ray imaging, and HRCam on the SOAR 4.1 m and QWSSI on the LDT 4.3 m. They found out that they were not twin, and that one of them, KX Com has actually an unresolved radial velocity companion.

Sun

Remeshan et al.[8] examined the interplanetary coronal mass ejection (ICME) detected in October 12, 2016, in L1 Lagrange point and its relation to a trailing high-speed stream (HSS). They modeled this interaction using a drag-based model. The *in situ* data came from OMNI's Global Geospace Science WIND satellite and the Advanced Composition Explorer's geocentric solar ecliptic system. They studied the magnetic obstacle and shock time from these data and reported a MO duration of 33 hours and a sheath (shock to MO start) of 9.4 hours. For the remote sensing part to study the Sun, they used the data from the LASCO, SOHO, COR2 of SECCCHI - STEREO*A, and the extreme ultraviolet images from the Atmospheric Imaging Assembly on the Solar Dynamics Observatory. Then, the drag modelling originates from the magnetohydrodynamic (MHD) drag by the ambient solar wind and is related to the interplanetary magnetic field. The drag formula is $a = -\gamma(v - w)|v - w|$ where γ is the drag coefficient, and v and w is the instantaneous ICME and ambient solar wind speed, respectively. They iteratively ran the DBM from 100 to 215 times the solar radius with an uncertain interaction location. After this two-pronged approach of data and simulation to predict the interaction location, they constrained it down to the 0.81 AU distance from the sun, compatible with the *in situ* data as

---

[4]Autocorrelations are computed segment by segment through the entire power spectrum with a selected window size, and the vertical axis will be the frequency lag in μHz, while the horizontal is the central frequency in μHz for this moving window.

[5]New General Catalogue

[6]The GitHub repository for the lcurve is here.

well.

Kieokaew et al.[9] worked on defining the geomagnetic baseline with data from Chambon-la-Forêt, France, using above-diurnal variation filtered data with 24h, 12h, 8h, and 6h filters, and SHaplet Additive exPlanations identified the factors correlated with these daily variations. The solar quiet time variations were predicted with the long short-term memory neural network (LSTM NN) with at least 11 years of 1-hour cadence data for mid-latitudes. Moreover, they utilized Coordinated Data Analysis Web and received 1-hour merged OMNI data product for solar wind and solar radio flux, utilizing interplanetary magnetic field and plasma parameters for the following analyses. Later, they checked the linear dependency between predictors and the predictand looking at their linear correlations, and after also checking the mutual information and expecting non-linear dependencies, they dropped *Ls and only considered DistSE*. In order to obtain SHAP values they trained XGBoost to predict y-component of the f_D, daily filter from the 24h, 12h, 8h, and 6h explained above. After this, they considered to use LT, SZA, DistSE, and $F_{10.7}$ to quiet period values for geomagnetism. They walk-forward trained the data with the validation only till the last possible step with annual motion, then at the final step, the last year was the test set. They reported that their modeling endeavor was generally successful and having the previos 12-hour data, it can predict the future 1-hour successfully.

Baalmann et al.[10] studied the spectral peaks on the dust particle impacts on the plasma wave instrument of the WIND/WAVES. They aimed to understand if this has anything to do with interplanetary or interstellar dust, co-rotating interaction regions (CIRs), interplanetary magnetic fields or, other effects. They made spectral analyses and considered the locations of the spacecraft, as well as these aforementioned potential factors, too. In the end, interplanetary magnetic field-related stability was not correlated, but CIRs[7] reduced the dust impact detections in general.

Cole et al.[11] reported a novel way to measure the solar radial velocity in higher precision with the *frequency comb calibrated laser heterodyne radiometry* method. They described the instrument modification on top of what was previously described[8], measurement results and the uncertainty estimation.

---

[7]Further information can be retrieved from here.
[8]The previously described part is here[12].

Mars

Zhao et al.[13] hypothesized a different approach in explaining the atmospheric heavy atom loss from Mars, with remnant crustal magnetic fields increasing plasma wave-particle interaction originated losses. They used the 4-second cadence data from the SupraThermal and Thermal Ion Composition (STATIC) instrument on NASA's Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft to monitor ion's pitch angles and gyro phases. Later, they conducted a backward Liouville test particle simulation to reproduce the ongoing interactions. They illustrated that magnetic and gravitational traps from Mars normally hold much of the atmospheric particles, yet the ionized ones can actually easily escape if interacted with ultra-low frequency (ULF) waves. The crustal magnetic field remnants mentioned above can facilitate the resonance with such ULF and the consequent escape.

Bernal et al.[14] made a reference paper on the visual monitoring camera (VMC) on the Mars Express. The paper outlines the capabilities of the instrument, science targets, noises and how they were counteracted, and also images from the calibration work undertaken that was explained in this paper.

Earth - Space relationship

Carr et al.[15] worked on space capsule atmospheric entry detection using the distributed acoustic sensing (DAS) methodology. They specifically focused on the return of Origins, Spectral Interpretation, Resource Identification, and Security Regolith Explorer (OSIRIS-Rex) capsule with two optical-fiber DAS interrogators with six co-located seismometer-infrasound sensor pairs spread across two sites.

Astrochemistry

Giustini et al.[16] made a theoretical study (density functional theory) of metastable doublet sulfur $S(^1D)$ with water and methanol, which is stated to be relevant for solar system chemistry, comets, and similar cases. This excited state can originate from absorbing an ultraviolet-wavelength photon from several sulfur-bearing species, such as OCS, CS_2, and S_2, as well as from H_2S and SO_2. Furthermore, a cluster of 4 water molecules around the reactants were also considered, which changed the expected reactions and products considerably. They used the Gaussian 09 program to calculate potential energy surfaces (PES) and product branching fractions (BFs), as well as Rice-Ramsperger-Kassel-Marcus estimates

(RRKM)[9]. They estimated that additional water molecules around the reactants can stabilize intermediates, such as hydrogen thioperoxide, mercaptomethanol, and several more compounds.

McGeoch and McGeoch[17] analyzed the stromatolites and benefiting from the isotopic fractionation they pointed to their potential cosmic origin. Spacedust contained a glycine polymer hemoglycin, and the 21 potential space infall sea foam samples from the USA North East - Rhode Island in late autumn to early winter were collected. They analyzed them with the matrix-assisted laser desorption-ionization (MALDI) mass spectrometry. They elaborated on 1200 and 1662 m/z series from the mass spectra. These are from the iron oxide-bound 18 glycine molecule polymer and the core hemoglycin polymer with 22 glycine and iron-silicate, respectively.

Instrumentation

Do Ó et al.[18] made a performance characterization work on HNü 240 Electron Multiplying Charge-Coupled Device (EMCDD) sensor of the Gemini Planet Imager version 2.0. They reported that the inverted mode had charge diffusion problems which blurred the images. Inverted Mode is self-explanatory, it first creates a hole in the sensing surface, a location without an electron, then, the dark current electron is combined with it, decreasing the dark current but exacerbating the clock-induced charge problem and the charge diffusion. The charge can diffuse through neighboring pixels owing to the reduced potential barrier between them, hence the blurring and the effect were more detrimental for the shorter wavelength. Their first qualitative test on spatial resolution with the resolution card was clearly much better in the non-inverted mode. The following tests were for the readout noise, dark current, clock-induced charges, and flat field tests. For the readout, the median of the 1000 dark frames was subtracted from each of them and the resulting standard deviation was computed. This was followed by multiplying it with the K-Gain set for each exposure and divided by the EM gain of 5000 to make it in units of electrons, and the median of these was computed. The results were more or less similar for the IMO and NIMO modes and between 0.07 - 0.17 for 8 different outputs. In the dark current, bias-subtracted dark gains at 5000 EM gain at -45 degrees Celcius were obtained. The exposure times were 1,2, 4, 8, and 10 seconds, and the mean for the pixels in the output converted to the electrons. The NIMO mode, expectedly, had a higher dark current impact, approximately 3.8 more than IMO. For clock-induced charge test, the mode is switched to "photon counting" mode, and the median CIC for each detector was slightly or up to 2 times higher in the NIMO mode. The flat field tests contained EM Gain Linearity, Exposure Time and Light Level Linearity, and Light Level Ratios. The results were generally unchanged from the IMO mode. Nevertheless, the expected spatial resolution was found to be sufficiently adequate and the sensor was considered successful.

Nell et al.[19] reported the characteristics of the Suborbital Imaging Spectrograph for Transition region Irradiance from Nearby Exoplanet host stars (SISTINE), a rocket-borne ultraviolet (UV) imaging instrument. This is an intriguing instrument where it is launched to observe specific targets and until now, and it has been launched three times.

---

[9] A chemical kinetics theory including energy of an excited state of a species, Eyring's transition theory, and vibrational and rotational modes impact on the likelihood of a reaction.

Last week in Chemistry

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between October 1 - October 7, 2024. They are more in nature of spectroscopy alone, and hence several studies regarding biochemistry, chromatography, and several other disciplines might be missed here.

Mass Spectroscopy

Salvati et al.[20] generated a solution to detect low-concentration endocannabionids in biospecimens with the laster-induced post-ionization (MALDI-2) method, specifically for 2-arachidonylglycerol and N-acylethanolamines (N-arachidonoylethanolamine - AEA, palmitoylethanolamide - PEA, oleoylethanolamide - OEA). They analyzed images with the flexImaging software, spectra with the SCiLS Lab software. They also applied this method to a mild traumatic brain injury-experienced mouse brain tissue. Their method was even better than the conventional matrix-assisted laser desorption ionization (MALDI).

Infrared Spectroscopy - IR

Giroux et al.[21] examined the impact of adding anti-coagulants (in this study, acid citrate dextrose anticoagulant, ACD-A) on forensic blood samples on their attenuated total reflectance - Fourier transform infrared (ATR-FTIR), visible absorbance, and x-ray fluorescence spectroscopy (XPS) spectra. They utilized principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) to investigate the factors. All three were affected by anticoagulant addition. PCA and PLS-DA were applied with different variables at different times after ACD-A addition or in no-ACD-A case from 2 hours up to 720 hours and the high-variance containing component was plotted against the common logarithm of time-since-deposition in ATR-FTIR and visible spectra. In other words, the columns in PCA-PLS-DA analyses were the spectra in different TSD. There were slight differences for all spectra. At FTIR, 1523 cm^{-1} wavenumber had a delayed increase, and in the visible spectra, and the maximum peak (also known as Soret peak) around 400 nm wavelength in the visible spectra had blueshifted in samples without ACD-A, yet after 48 hours, in both cases, the peak blue-shifted without any statistically significant difference between ACD-A presence or absence.

Lateral Flow Immunoassay

Scott et al.[22] worked on available xylazine test strips (lateral flow immunoassays) to check their performances.
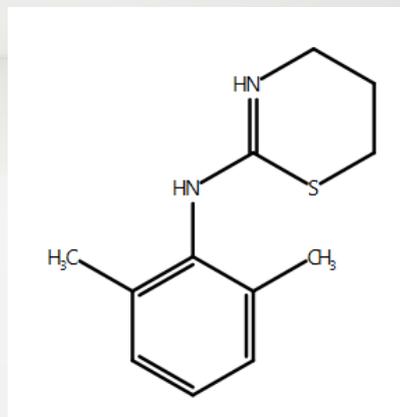

Figure 1: Xylazine molecular structure

Their method was using the test strips, then, imaging them and with te open-access NIH ImageJ and FiJi image processing tools, analyzing them for the control and test bars on the strips. The test strips were BTNx, WHPM, WiseBatch, MD-Bio, 12 PanelNow, WaiveDx, and Medimpex. They separated test results in a semi-quantitative manner: solid, faint, barely visible, no test bands, where the results are inversely proportional (e.g., solid test band means negative). They checked the reported limit of detection (LOD) of the test strips on the xylazine detection and found that many of them showed clear negatives in much higher concentrations than reported LOD. Furthermore, they analyzed the selectivity of these strips using several heterocyclic components, such as ketamine, caffeine, fentanyl, but also naproxen, ibuprofen, acetaminophen, and other aromatic compounds. There were several false-positive test results, especially with the cross-reacting compounds. Later, they tried to find out the optimal waiting time and found out that usually it took 2 minutes for the clear emergence of the bands. Their stability study on seeing the cold temperature or ambient temperature storage impact on the test strips mostly produced satisfactory stability for up to 6 weeks storage duration. Furthermore, pH variance did not impact, which was relevant since urine sample (hence its different pH) was among the usual samples for these test strips.

Biosensors

Bozorgzadeh et al.[23] worked on an immunosensing fully-integrated silicon chip which can simultaneously detect more than one analytes. They reported that boronic acid deposition was the

best method to deposit antibodies of interest onto the silicon chip. More specifically, the chitosan was covalently bound to 4-carboxylphenylboronic acid (CS-CPBA) and after several steps this was the best way to immobilize and make the biosensors ready-to-use. With this method, they bound the antibodies from their bottom, Fragment crystallizable moiety[10]. The analytes were casein and ovalbumin, and no interfering effects were seen, which was another success of this analytical method.

---

[10]This is the bottom part of an antibody, the actual specific part is upper arms, Fragment antigen-binding region (Fab), where it binds the antigen-molecule of interest in this case.

Last week in Remote Sensing

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between October 1 - October 7, 2024. These are generally based on the preprints retrieved when "remote sensing" words are given between quotation marks within arXiv's cs.CV and similar cross-fields.

Segmentation

Dimitrovski et al.[24] proposed the late fusion deep learning model for semantic segmentation using very high-resolution aerial imagery and satellite image time series data. Aerial image information is utilized via a Multi-Axis Vision Transformer backbone UNetFormet, and Sentinel-2 image time series was used by a U-Net with Temporal Attention Encoder. It was found quite successful according to its performance on the FLAIR dataset[11]. They measured the success with intersection over union (IoU) metric, where it simply provides the coverage of model output to the ground truth and divides it to their total area, like if it perfectly covers without spurious assignment it is 100 %. LF-DLM, the augmentation of UNetFormer with aerial image information slightly improved almost all labels.

Rafaeli et al.[25] studied a monocular depth-guided segment anything model (SAM) for the sinkhole detection task (SinkSAM). They highlighted that by incorporating topography, pixel-level sinkhole delineation refinements were possible, and their coherent mathematical prompting made it more robust against unseen data. Moreover, Depth Anything V2 removed the dependency on the LIDAR data via providing monocular depth with DAV2 and ViT-1. Their study area was the Negev Desert, Israel. After DAV2 estimated depressions, the ArcGIS pro fill tool filled them, then this and previous depth estimations were subtracted, and small depressions with ¡2 meters of depth or ¡50 pixels of area were discarded. They compared a Zero-shot SAM and SinkSAM with the conventional accuracy, precision, recall, F1 metrics, and IoU. The results were better compared to the zero-shot version.

Li et al.[26] proposed both the SimFeatUp upsampler and SegEarth-OV models to improve upon CLIP and FeatUp while introducing the open-vocabulary semantic segmentation (OVSS)[12] to the remote sensing. They are making the CLIP and FeatUp approaches much more training-free with their novel methodology. They worked on seventeen different datasets used previously for various segmentation tasks. SegEarth-OV was generally the best in 448 x 448 sized images, or when considered with 896 x 896 size.

Velôso de Souza et al.[27] first provided the MagSet-2 dataset (50-50 train-test), mangrove-annotated Global Mangrove Watch and Sentinel-2 satellite images, then, they compared the convolutional, transformer, and mamba architecture-based models, and reported the superior status of the mamba architecture[13]. Specifically, the tested models were U-Net, MANet, PAN, BEiT, Segformer, and Swin-UMamba.

Yu et al.[28] constructed SpecSAR-Former, a transformer-based network utilizing Sentinel-1 SAR and Sentinel-2 multispectral imager data for global land use land cover mapping. By incorporating Sentinel-1 and -2 data together, they created the Dynamic World+ dataset with 16893 training data images and 299 patches for validation and testing. All images were in 510x510 pixel size. While Sentinel-2 provided multispectral data for the network's spectral perception branch, Sentinel-1 provided SAR data for both augmenting the spectral branch, but primarily for the spatial branch, which is the other branch of the network. The proposed model was the best in accuracy, and also in the number of parameters and floating point operations (not the speed, where CMTFNet was the best). In terms of per-class assessment, their novel approach was best in water, tree, shrub, scrub, flooded vegetation, snow-ice, and in general for the IoU for all classes.

UAV

Spasev et al.[29], a similar group to Dimitrovski et al.[24] above, worked on semantic segmentation but for the UAV. They studied the peformances of SegFormer framework variants (encoder variants) from B0 to B5. with UAVid dataset[14]. The framework's encoder part has multiple transformer blocks including an efficient self-attention, followed by a mix-feed forward network and overlap patch merging step with several different convolutions interspersed in between. The decoder has a multilayer perceptron. They trained the models with a maximum of 100 epochs and early-stopped it if the validation step loss criteria did not improve within 20 epochs. The best method, also considering its performance against U-Net,

---

[11]FLAIR dataset is here. The number of images was 77762 and they have 512 x 512 pixels with 20 cm spatial resolution and contain blue, green, red, and infra-red bands.

[12]The concept is essentially being able to label an image with a vocabulary the model hasn't seen in its training.
[13]Study-related GitHub repository can be found here.
[14]The UAVid dataset and related information is here.
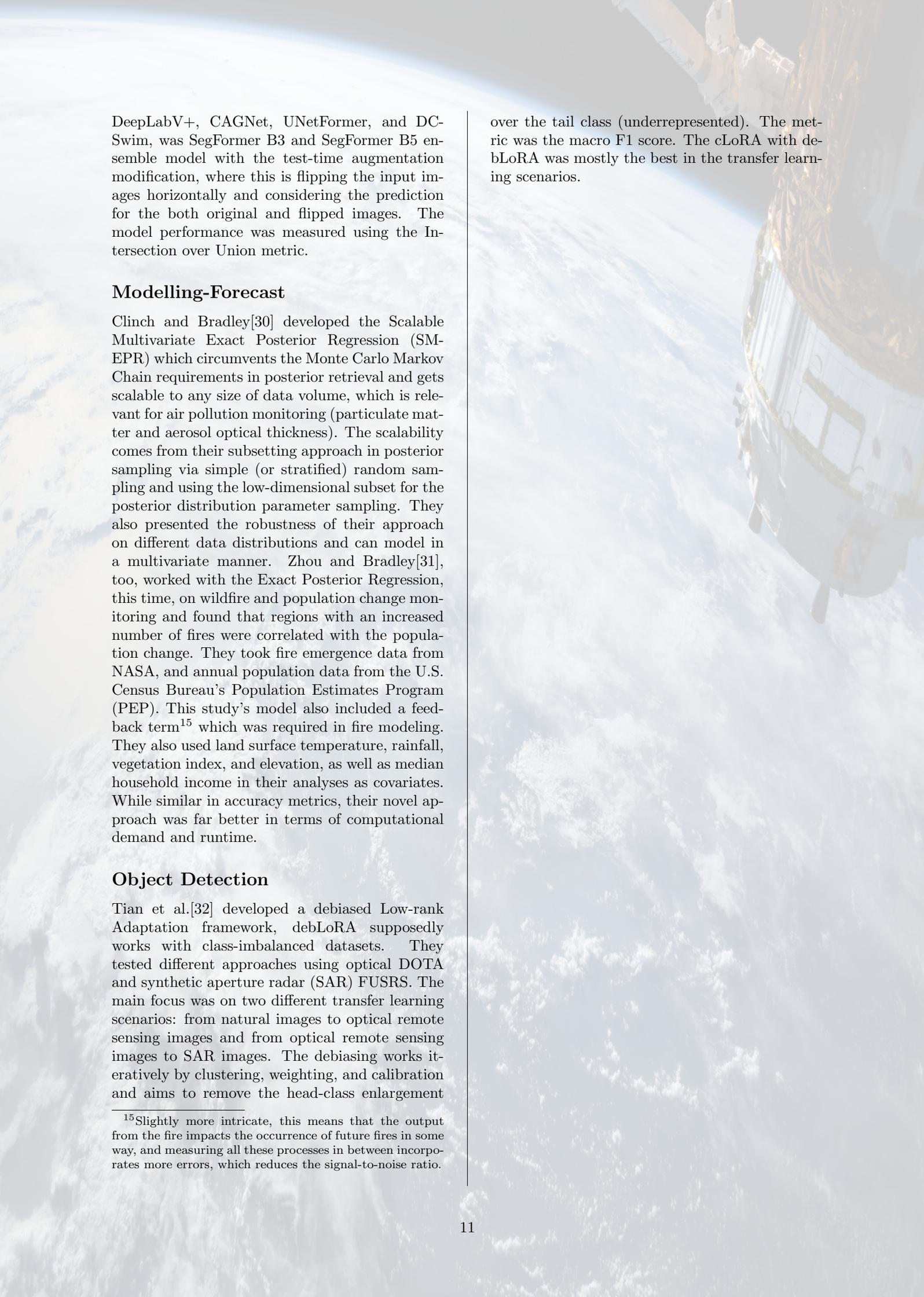
DeepLabV+, CAGNet, UNetFormer, and DC-Swim, was SegFormer B3 and SegFormer B5 ensemble model with the test-time augmentation modification, where this is flipping the input images horizontally and considering the prediction for the both original and flipped images. The model performance was measured using the Intersection over Union metric.

Modelling-Forecast

Clinch and Bradley[30] developed the Scalable Multivariate Exact Posterior Regression (SM-EPR) which circumvents the Monte Carlo Markov Chain requirements in posterior retrieval and gets scalable to any size of data volume, which is relevant for air pollution monitoring (particulate matter and aerosol optical thickness). The scalability comes from their subsetting approach in posterior sampling via simple (or stratified) random sampling and using the low-dimensional subset for the posterior distribution parameter sampling. They also presented the robustness of their approach on different data distributions and can model in a multivariate manner. Zhou and Bradley[31], too, worked with the Exact Posterior Regression, this time, on wildfire and population change monitoring and found that regions with an increased number of fires were correlated with the population change. They took fire emergence data from NASA, and annual population data from the U.S. Census Bureau's Population Estimates Program (PEP). This study's model also included a feedback term[15] which was required in fire modeling. They also used land surface temperature, rainfall, vegetation index, and elevation, as well as median household income in their analyses as covariates. While similar in accuracy metrics, their novel approach was far better in terms of computational demand and runtime.

Object Detection

Tian et al.[32] developed a debiased Low-rank Adaptation framework, debLoRA supposedly works with class-imbalanced datasets. They tested different approaches using optical DOTA and synthetic aperture radar (SAR) FUSRS. The main focus was on two different transfer learning scenarios: from natural images to optical remote sensing images and from optical remote sensing images to SAR images. The debiasing works iteratively by clustering, weighting, and calibration and aims to remove the head-class enlargement

over the tail class (underrepresented). The metric was the macro F1 score. The cLoRA with debLoRA was mostly the best in the transfer learning scenarios.

---

[15]Slightly more intricate, this means that the output from the fire impacts the occurrence of future fires in some way, and measuring all these processes in between incorporates more errors, which reduces the signal-to-noise ratio.

Last week in Environmental Chemistry

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between October 1 - October 7, 2024 in chemRxiv's Earth, Space, and Environmental chemistry preprints are being surveyed, and unfortunately, not many preprints are published under environmental topics in this field.

Wang et al.[33] investigated the sufficiency of the current air quality monitoring stations in the United States for monitoring particulate matter smaller than 2.5 micrometers (PM 2.5). They started with the Core-Based Statistical Areas and then checked with SLAMS to see if there are actually unrepresented hot spots of high PM 2.5 in this monitoring network with partial least squares regressions with universal kriging framework for the 2017-2019 period. There were 988 monitoring sites from the EPA's Air Quality System and they matched this with the Center for Air, Climate, and Energy Solutions (CACES)'s empirical model. They also considered the demographics to see if there were any systematic bias. In the end, they estimated around 2.8 million people-inhabiting areas where the *status quo* considered it as attained the air quality standard, while this study demonstrated that it was not the case.

Zeng et al.[34] studied the viscosity and phase state of biomass-burning organic aerosol (BBOA) originating from large wildfires with a theoretical perspective and further laboratory studies. They started with five different scenarios including varying time after the pyro cumulonimbus event, BBOA proxy, oxygen/carbon atomic ratios, hydrogen/carbon atomic ratios, and $H_2SO_4 : BBOA$ mass ratios. They also utilized AIOMFAC-VISC[16] thermodynamic model, and with Zdanovskii-Stokes-Robinson mixing rule[17] option or not for several of these scenarios for the viscosity estimations. They discussed the implications of different viscosities, and glassy states on stratospheric ozone depletion, too.

Metzger et al.[35] worked on the placement of continuous monitoring systems (CMS) and succeeded in attaining a 92.9 % probability of detecting an emission within 12-hour time blocks with a 3-sensor CMS network. Their input data were the oil & gas facility layouts, facility-specific atmospheric data (wind speed and direction), and financial constraints. They excluded time windows with a wind speed lower than 0.5 m/s to reduce the computational complexity of the plume prediction. The remaining data was split into 75-25 training - test sets. In their simulation, a forward dispersion model simulated the pollution transport as a plume from the cluster center of the emission sources with the Gaussian plume model. Emissions were 1.6 kg/h for each cluster. They only measured detection and non-detection and assumed a tunable diode laser absorption spectroscopy (TDLAS) where it needed to detect 1 ppm. The multiobjective optimization approach maximized the information density (higher with a higher number of detection events for a given node to place the CMS) and minimized the blind time by the CMSs. The methodology for the optimization was the Greedy algorithm.

---

[16]Related releases can be found here.
[17]Further information can be found here.

Last week in Data Decomposition/Transformation

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between October 1 - October 7, 2024. This is generally from arXiv's stat.ML or stat.ME cross-list. The section focuses on preprints heavily worked with or developed data decomposition/transformation techniques, such as principal component analysis (PCA) or Fourier Transformation.

Dimensional Reduction

Ugwu and Kirby[36] combined support vector machine with the novel algorithm of proximal gradient descent and generalized singular value problem. The main idea is reducing the data dimensionality with the smallest beneficial final dimensions. Think of a matrix with n rows and m columns. These n rows can be divided into many other sub-matrices, maybe n/3 and 2n/3. Then, GSVD tries to decompose them into components in the same way so as to minimize the differences between them while divided. The authors reported that combining this with the support vector machine technique resulted in 100 % accurate ovarian cancer classification.

Zhou and Cannings[37] used an ensemble of projections and singular value decomposition to later retrieve a type of ensemble regression. The singular values after the SVD steps illustrated the importance of a specific projection. The randomly produced projections iteratively converge towards the higher explanatory ones through the algorithm runs. For distribution-specific insights, they considered working with the Gaussian rows and Cauchy rows, and reported that in case of an unknown sparsity status, 50-50 Gaussian and Cauchy random projection was the best choice, while the sparse state worked better with the Cauchy, and Gaussian was better in other cases. For their regression model backbone, global linear least squares, global quadratic least squares, Nadaraya-Watson, and Multivariate Adaptive Regression Splines (MARS) were considered. MARS was either the best or a competitive-second. In the studied three models, the first model was the linear function of some of the X parameters, the second was a sinusoidal of one X parameter, and the third one was a quadratic, more complicated model of several X parameters. They also considered the random projection ensemble's second consecutive application to reduce the dimension.

After numerical studies, they applied their methods to the superconductivity, communities and crime, residential building, and the geographical origins of music datasets, where double application had the best case in general with a slight edge.

Time Series

Barigozzi et al.[38] worked on multivariate multi dynamical factor cases and detected change points for them. Their method is based on Moving Sum and they provided a GitHub repository on an implementation here. In simulated data with timestep of 400 and 200 realizations, the MOSUM-diagonal was slightly better than an example approach from the literature. As a real-world example, daily stock prices of 72 US blue chip companies between 2005 - 2022 from the Wharton Research Data Services was used. Here, too, a low number of change point in especially low number of N - realizations was better predicted with the MOSUM.

Tan et al.[39] proposed the functional singular value decomposition (FSVD) and presented its utility in time-series analysis, as well as in COVID-19 case count dataset. One of their aims in developing this FSVD was a function that can work with irregularly-sampled data (more akin to real-world conditions) and non-stationary factor series. They used reproducing Kernel Hilbert Space (RKHS), then, moved toward joint kernel ridge regression. The components are retrieved sequentially, and it can be considered as a minimization/optimization problem or representing a very-high dimensional case with a desired-accuracy low-dimensional result.

Modelling

Loeffler et al.[40] proposed the Graph Fourier Neural Kernels (G-FuNK) for working with non-linear partial differential equation (PDE) to generate solutions with neural operators. They reported better results from this approach while comparing it to FNO, Geo-FNO, and GNN with a neaural ordinary differential equation.

Zhao et al.[41] studied the case with multiple predictor multiple categorical response variables regression with a data parsimonious and interpretable dependence structure-generating method. One key point is their subspace decomposition approach. The model has separate coefficients for different effects, such as conditional or joint independence, hence, their magnitude or absence illustrates the dependency. They could separately study local and global associations via a simple predictor grouping, and the hierarchical

constraints and setting them to zero were achieved by group lasso penalty applications. After giving detailed explanations of the theoretical underpinnings of their method, they also conducted numerical studies.

Hasan and Ahmed[42] studied the behavior of m mixture components with special focus on order of addition (OoA) and component proportions. They provided example designs with minus-one degrees of freedom from all possible permutations (orders of appearance for the mixtures). Then, they provided a study on the joint actions of three hormones on mices, and three components placebo tablet formulation (three diluents, response as crushing strength to retrieve all the interaction terms).

References

[1] Greg Olmschenk, Richard K. Barry, Stela Ishitani Silva, Jeremy D. Schnittman, Agnieszka M. Cieplak, Brian P. Powell, Ethan Kruse, Thomas Barclay, Siddhant Solanki, Bianca Ortega, John Baker, and Mamani Yesenia Helem Salinas. Short-period variables in tess full-frame image light curves identified via convolutional neural networks. *The Astronomical Journal*, 168(2):83, July 2024.

[2] Greg Olmschenk, Stela Ishitani Silva, Gioia Rau, Richard K Barry, Ethan Kruse, Luca Cacciapuoti, Veselin Kostov, Brian P Powell, Edward Wyrwas, Jeremy D Schnittman, et al. Identifying planetary transit candidates in tess full-frame image light curves via convolutional neural networks. *The Astronomical Journal*, 161(6):273, 2021.

[3] Isabel Angelo, Megan Bedell, Erik Petigura, and Melissa Ness. A data-driven spectral model of main-sequence stars in gaia dr3. *The Astrophysical Journal*, 974(1):43, October 2024.

[4] Victor Réville, Jamie M Jasinski, Marco Velli, Antoine Strugarek, Allan Sacha Brun, Neil Murphy, Leonardo H Regoli, Alexis Rouillard, and Jacobo Varela. Magnetized winds of m-type stars and star-planet magnetic interactions: uncertainties and modeling strategy. *arXiv preprint arXiv:2410.01621*, 2024.

[5] Myles Pope, Joleen K Carlberg, Jeff Valenti, and Doug Branton. Tess asteroseismic masses and radii of red giants with (and without) planets. *arXiv preprint arXiv:2410.02051*, 2024.

[6] Leyla Iskandarli, Jay Farihi, Joshua D Lothringer, Steven G Parsons, Orsola De Marco, and Thomas Rauch. Novel constraints on companions to the helix nebula central star. *Monthly Notices of the Royal Astronomical Society*, page stae2286, 2024.

[7] Andrew A Couperus, Todd J Henry, Rachel A Osten, Wei-Chun Jao, Eliot Halley Vrijmoet, Aman Kar, and Elliott Horch. The solar neighborhood lii: M dwarf twin binaries–presumed identical twins appear fraternal in variability, rotation, h α, and x-rays. *arXiv preprint arXiv:2410.04726*, 2024.

[8] Akshay Kumar Remeshan, Mateja Dumbovic, and Manuela Temmer. Deriving the interaction point between a coronal mass ejection and high speed stream: A case study, 2024, 2410.00615.

[9] Rungployphan Kieokaew, Veronika Haberle, Aurélie Marchaudon, Pierre-Louis Blelly, and Aude Chambodut. A novel neural network-based approach to derive a geomagnetic baseline for robust characterization of geomagnetic indices at mid-latitude. *arXiv preprint arXiv:2410.02311*, 2024.

[10] Lennart Robin Baalmann, Silvan Hunziker, Arthur Péronne, James W Kirchner, K-H Glassmeier, David M Malaspina, Lynn B Wilson III, Christoph Strähl, Shivank Chadda, and Veerle Jasmin Sterken. A solar rotation signature in cosmic dust: Frequency analysis of dust particle impacts on the wind spacecraft. *Astronomy & Astrophysics*, 689:A329, 2024.

[11] Ryan K Cole, Connor Fredrick, Winter Parts, Max Kingston, Carolyn Chinatti, Josiah Tusler, Suvrath Mahadevan, Ryan Terrien, and Scott A Diddams. Frequency comb calibrated laser heterodyne radiometry for precision radial velocity measurements. *arXiv preprint arXiv:2410.04588*, 2024.

[12] Ryan K Cole, Connor Fredrick, Newton H Nguyen, and Scott A Diddams. Precision doppler shift measurements with a frequency comb calibrated laser heterodyne radiometer. *Optics Letters*, 48(20):5185–5188, 2023.

[13] J. T. Zhao, Q. G. Zong, Z. Y. Liu, X. Z. Zhou, S. Wang, W. H. Ip, C. Yue, J. H. Li, Y. X. Hao, R. Rankin, A. Degeling, S. Y. Fu, H. Zou, and Y. F. Wang. On energization and loss of the ionized heavy atom and molecule in mars' atmosphere, 2024, 2410.00832.

[14] Jorge From, Jorge Hernández-Bernal, Alejandro Cardesin Moinelo, Ricardo Hueso, Eleni Ravanis, Abel Burgos Sierra, Simon Wood, Marc Costa Sitja, Alfredo Escalante, Emmanuel Grotheer, et al. The visual monitoring camera (vmc) on mars express: a new science instrument made from an old webcam orbiting mars. *arXiv preprint arXiv:2410.02999*, 2024.

[15] Chris G Carr, Carly M Donahue, Loic Viens, Luke B Beardslee, Elisa A McGhee, and Lisa R Danielson. Detection of a space capsule entering earth's atmosphere with distributed acoustic sensing (das). *arXiv preprint arXiv:2410.03034*, 2024.

[16] Andrea Giustini, Gabriella Di Genova, Dimitrios Skouteris, Cecilia Ceccarelli, Marzio Rosi, and Nadia Balucani. Gas-phase and model ice-surface reactions of s (1d) with water and methanol: a computational investigation and implications for cosmochemistry/astrochemistry. 2024.

[17] Julie E. M. McGeoch and Malcolm W. McGeoch. Sea foam contains hemoglycin from cosmic dust, 2024, 2410.01043.

[18] Clarissa R. Do Ó, Saavidra Perera, Jérôme Maire, Jayke S. Nguyen, Vincent Chambouleyron, Quinn M. Konopacky, Jeffrey Chilcote, Joeleff Fitzsimmons, Randall Hamper, Dan Kerley, Bruce Macintosh, Christian Marois, Friedrik Rantakyrö, Dmitry Savransky, Jean-Pierre Véran, Guido Agapito, S. Mark Ammons, Marco Bonaglia, Marc-André Boucher, Jennifer S. Dunn, Simone Esposito, Guillaume Filion, Jean Thomas Landry, Olivier Lardière, Duan Li, Alexander B. Madurowicz, Dillon Peng, Lisa Poyneer, and Eckhart Spalding. Gpi 2.0: exploring the impact of different readout modes on the wavefront sensor's emccd. In Dirk Schmidt, Elise Vernet, and Kathryn J. Jackson, editors, *Adaptive Optics Systems IX*, volume 745, page 293. SPIE, August 2024.

[19] Nicholas Nell, Kevin France, Nicholas Kruczek, Brian Fleming, Stefan Ulrich, Patrick Behr, Manuel A Quijada, Javier Del Hoyo, and John Hennessy. The assembly, characterization, and performance of sistine. *Journal of Astronomical Telescopes, Instruments, and Systems*, 10(3):035003–035003, 2024.

[20] Emanuela Salviati, Francesca Guida, Danila La Gioia, Fabrizio Merciai, Sabatino Maione, Vincenzo Di Marzo, Pietro Campiglia, Fabiana Piscitelli, and Eduardo Sommella. Maldi-2 mass spectrometry imaging enables spatial visualization of endocannabinoids in brain. 2024.

[21] Erin Giroux, Iraklii Ebralidze, and Theresa Stotesbury. Influence of anticoagulant on the spectroscopic analysis of drying blood pools. 2024.

[22] Lena Scott, Katherine Davis, Ju Park, and Saman Majeed. Evaluating the sensitivity, selectivity, and cross-reactivity of lateral flow immunoassay xylazine test strips. 2024.

[23] Somayyeh Bozorgzadeh, Vuslat B Juska, and Alan O'Riordan. Oriented antibody immobilization vs random covalent binding: Improving reproducibility for label-free electrochemical immunosensing of multi-allergens on silicon chips. 2024.

[24] Ivica Dimitrovski, Vlatko Spasev, and Ivan Kitanovski. Deep multimodal fusion for semantic segmentation of remote sensing earth observation data. *arXiv preprint arXiv:2410.00469*, 2024.

[25] Osher Rafaeli, Tal Svoray, and Ariel Nahlieli. Sinksam: A monocular depth-guided sam framework for automatic sinkhole segmentation. *arXiv preprint arXiv:2410.01473*, 2024.

[26] Kaiyu Li, Ruixun Liu, Xiangyong Cao, Deyu Meng, and Zhi Wang. Segearth-ov: Towards traning-free open-vocabulary segmentation for remote sensing images. *arXiv preprint arXiv:2410.01768*, 2024.

[27] Lucas José Velôso de Souza, Ingrid Valverde Reis Zreik, Adrien Salem-Sermanet, Nacéra Seghouani, and Lionel Pourchier. A deep learning-based approach for mangrove monitoring. *arXiv*, 2024.

[28] Hao Yu, Gen Li, Haoyu Liu, Songyan Zhu, Wenquan Dong, and Changjian Li. Specsarformer: A lightweight transformer-based network for global lulc mapping using integrated sentinel-1 and sentinel-2. *arXiv*, 2024.

[29] Vlatko Spasev, Ivica Dimitrovski, Ivan Chorbev, and Ivan Kitanovski. Semantic segmentation of unmanned aerial vehicle remote sensing images using segformer. *arXiv preprint arXiv:2410.01092*, 2024.

[30] Madelyn Clinch and Jonathan R. Bradley. Exact bayesian inference for multivariate spatial data of any size with application to air pollution monitoring. *arXiv*, 2024.

[31] Shijie Zhou and Jonathan R. Bradley. Multiscale multi-type spatial bayesian analysis of wildfires and population change that avoids mcmc and approximating the posterior distribution. *arXiv*, 2024.

[32] Zichen Tian, Zhaozheng Chen, and Qianru Sun. Learning de-biased representations for remote-sensing imagery. *arXiv*, 2024.

[33] Yuzhou Wang, Julian D Marshall, and Joshua Apte. Us ambient air monitoring network has inadequate coverage 1 under new pm2. 5 standard 2. 2024.

[34] Mei Fei Zeng, Andreas Zuend, Nealan GA Gerrebos, Pengfei Yu, Gregory P Schill, Daniel M Murphy, and Allan K Bertram.

Viscosity and phase state of wildfire smoke particles in the stratosphere from pyrocumulonimbus events: an initial assessment. 2024.

[35] Noah Metzger, Ali Lashgari, Umair Ismail, David Ball, and Nathan Eichenlaub. A framework for optimizing continuous methane monitoring system configuration for minimal blind time: Application and insights from over 100 operational oil & gas facilities. 2024.

[36] Ugochukwu O. Ugwu and Michael Kirby. Optimizing sparse generalized singular vectors for feature selection in proximal support vector machines with application to breast and ovarian cancer detection. *arXiv*, 2024.

[37] Wenxing Zhou and Timothy I. Cannings. Random-projection ensemble dimension reduction. *arXiv*, 2024.

[38] Matteo Barigozzi, Haeran Cho, and Lorenzo Trapani. Moving sum procedure for multiple change point detection in large factor models. *arXiv*, 2024.

[39] Jianbin Tan, Pixu Shi, and Anru R. Zhang. Functional singular value decomposition. *arXiv*, 2024.

[40] Shane E. Loeffler, Zan Ahmad, Syed Yusuf Ali, Carolyna Yamamoto, Dan M. Popescu, Alana Yee, Yash Lal, Natalia Trayanova, and Mauro Maggioni. Graph fourier neural kernels (g-funk): Learning solutions of nonlinear diffusive parametric pdes on multiple domains. *arXiv*, 2024.

[41] Hongru Zhao, Aaron J. Molstad, and Adam J. Rothman. Subspace decompositions for association structure learning in multivariate categorical response regression. *arXiv*, 2024.

[42] Taha Hasan and Touqeer Ahmad. Order of addition in mixture-amount experiments. *arXiv*, 2024.