

Science
Ascend

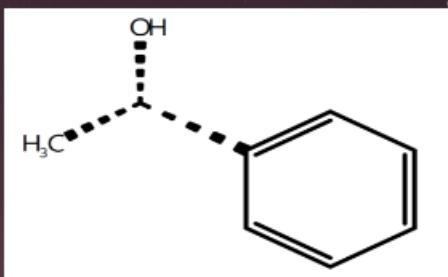
From September 09-
16, 2024!

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



More precise methods,
Higher resolution outputs, ALMA, Keck,
GAIA and more!

Intricacies of glucose analysis,
real-time polymerization in
solid-state NMR, and LMMs in
molecular identification!



Pushing the boundaries in land-use classification via transformers,
anomaly detections, finding looted archaeological sites and more!

Cyclo-dextrin metal organic frameworks, and
updated air pollutant maps for the conterminous
US with land use predictor containing
regressions!



Much better uncertainty
classification, unmatched
performance in image
classification, theoretical
backgrounds and more!

cout<< solutions;



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.

Publisher: FIRE Araştırma Eğitim Ltd. Şti.

Media: Online Journal

Responsible person: Yasin Güray Hatipoğlu

Editor-in-chief: Yasin Güray Hatipoğlu

Editor: Yasin Güray Hatipoğlu

Frequency: Once a week

Address: Yıldızevler Mah. Kişinev Cad. No:10

Çankaya/Ankara/Türkiye

Website: <https://fire-ae.github.io>

This issue: September 16, 2024

Volume: 1

Issue Number: 2

All rights reserved.

Bilim Yükselişi

Kesfin 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 baskalarını 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 yayılmasını, dağıtılmmasını ve yayıldıktan en geç on beş gün sonra ücretsiz olarak herkesin erişimine açılmasını garanti eder.

Yaynıcı: FIRE Araştırma Eğitim Ltd. Şti.

Ortam: Online Journal

Sorumlu Kişi: Yasin Güray Hatipoğlu

Yazı İşleri Müdürü: Yasin Güray Hatipoğlu

Editör: Yasin Güray Hatipoğlu

Yayınlanma Sıklığı: Haftada bir kez

Adres: Yıldızevler Mah. Kişinev Cad. No:10

Çankaya/Ankara/Türkiye

Website: <https://fire-ae.github.io>

Bu sayı: September 16, 2024

Cilt: 1

Sayı Numarası: 2

Tüm hakları saklıdır.

Last week in Astrophysics

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between September 9 - September 16, 2024. These are from arXiv's astro.EP cross-fields.

Space-based

Hesar et al.[1] worked with machine learning algorithms to estimate stellar rotation periods from NASA Kepler mission's corrected lightcurve data. The corrected data comes from "Pre-search Data Conditioning Simple Aperture Photometry¹ from the 25th data release of Kepler. They chose nine Kepler Input Catalog data with clear exoplanet transits and stellar activities. They separated long-term trends via a high-pass filter for all objects. Cosmic rays and outliers were weeded out with sigma-clipping². They applied a Lomb-Scargle periodogram³ to the pre-processed dataset and considered the highest peak in the periodogram as the stellar rotation period. After all these, they employed Decision Tree, Random Forest, k-nearest Neighbors, Gradient Boosting, and *Best Model*⁴ ML methods to determine stellar parameters and transits from the pre-processed Kepler lightcurves. They found that the voting-ensemble best model performed best, even though quick look at RMSE values point to the Random Forest model's superior performance in many cases for both Table 2 and Table 3, model-developed data and other data, respectively.

Steiger et al.[2] simulated exoplanet imaging of Habitable Worlds Observatory (HWO) via energy-resolving detectors. They simulated both an electron multiplying charge-coupled device⁵ and energy-resolving detector technology in dark

¹In addition to the aperture photometry, here in PDC-SAP lightcurve, co-trending basis vectors tried to remove long-term trend in the data. Rather than detrending, co-trending selects several points in an astronomical image and measures the long-term change of flux in irrelevant places, estimating an overall, astronomical signal-independent long-term error. "If these irrelevant fluxes covary, the degree of the covariance must come from instrumental error or similar redundant sources."

²Consider a time-series data, estimate its mean and standard deviation (σ), and consider a specific multiplier of sigma for a threshold to determine if a value is an outlier or not. 3σ clipping means that any value outside of $\pm 3\sigma$ around μ is considered an outlier.

³A graph of the period in the x-axis and power at the y-axis, in which power is in spectral dimension.

⁴Voting-based classifier that gets the results of other predictors and makes a final choice based on them.

⁵Charge-coupled device (CCD) was something everywhere, even in mobile phones, and EMCCD is a modification of it.

hole digging. Their telescope model was High-contrast Imager for Complex Aperture Telescopes test bed, starlight suppression was with apodized-pupil⁶ Lyot coronograph. They had piston and tip/tilt aberrations, pairwise probing for dark hole digging by electric field conjugations, and the exoscene library for stellar flux calibration and a chosen magnitude 8 star. They also simulated the required time to detect water from HWO targets and in both simulations above, ERD-required exposure times were a few factors shorter.

Luhman[3] extensively used Gaia Data Release 3 for improving membership samples from the following moving groups and associates: β Pic Moving group (BPMG), Sco Body, Carina, Columba, χ^1 For, tuc-Hor, IC 2602, IC 2391, NGC 2547. Luhman focused on initial mass functions, X-ray emission, ages, and the presence of circumstellar disks (via checking excess emission with the Wide-field Infrared Survey Explorer (WISE) infrared photometry). There was also a discussion on age-excess relations in different associations and BPMG.

Diaz et al.[4] prepared a guideline-like document for applicants preparing a Gemini Observatory⁷ proposal. They also put a rubric at the end of this document on how the *grading* can be done. The main sections are Scientific Justification, Experimental Design (meaning the sample, strategy, how the data will be analyzed, etc.), technical description (configuration, angular-spectral resolution, conditions, signal-to-noise ratio, total and minimum times to realize the goals of the proposals. There is also a third-priority program, called "Band 3" where if Band 1 and 2 are not available and the programs are still achievable under these conditions. "Plan for Band 3" section is also present in case programs will require modifications under this condition.

Wainer et al.[5] studied dM3⁸ star KIC-8507979/TIC-272272592 object with the 30-min cadence Kepler data and 2-min cadence TESS data. This object was flagged as a star with potentially varying flare activity over 4-year of Kepler mission data. They identified flares by the convolutional neural network flare finding package **stella**, yet the flare identification step was not straightforward. They first filtered the data from ≥ 30 flare probability and then aggregated within-40-min window flares as a single flare event. They also checked **stella**'s flare identification completeness with injection recovery tests. Through these tests, they saw performance variations between

⁶Shape-shifting pupil.

⁷Ground-based twin telescopes, one in Chile and one in Hawaii with 8.1-metre main optical-infrared telescopes. Further info is here.

⁸A spectral subclassification of M-type stars

different TESS sectors. For the TESS sector and Kepler Quarters' point-in-time flare activity, the Flare Frequency Distribution metric was used. One of the main challenges in flare activity estimations was the utilization of two different telescope data in both the short and long term, and the authors were hopeful that TESS alone would be able to provide long-term trends thanks to its higher than 7-year operational period. Another thing that stands out is a few thousand Barycentric Julian Date between Kepler and TESS datasets.

Sha et al.[6] studied Blanco 1 open cluster, within 240 pc⁹ distance from the Sun and approximately 130 million years old. The literature also reported diffuse tidal tails extending up to 60 pc from the cluster center. The researchers aimed to check this by comparing rotation periods of stars in and around Blanco 1 and testing if can be considered with a single-star gyrochronological sequence¹⁰. They started with Gaia Data Release 3, then utilized the lightcurves from TESS full-frame images (FFI) to measure stellar rotation periods in and around Blanco 1. Their procedure to generate lightcurve was forced aperture photometry after a difference imaging according to the Gaia Data Release 2 locations, the pipeline from the *Cluster Difference Imaging Photometric Survey - CDIPS*. The objects fainter than a TESS magnitude¹¹ of 13.3 had a one-pixel radius circular aperture, while the ones with higher TESS magnitude had 1.5 pixels for this value. For rotation period estimation, the algorithm they chose was the Lomb-Scargle Periodogram from the **astropy** package in Python. After comparing the rotation periods, they ruled out the possibility of field contamination by other stars and favored that tidal tails are associated with Blanco 1.

French et al.[8] a brown-dwarf of L1 peculiar type in a white dwarf-brown dwarf binary system. They used time-resolved Hubble Space Telescope Wide Field Camera 3 (HST-WFC3) of WD1032+011 and isolated phase-dependent spectra of WD1032-011B. The data was processed by the CalWFC3 pipeline, for spectral data extraction Amaro et al.[9]'s method was utilized. After 5-sigma clipping of cosmic rays and interpolating over those pixels, they use SourceExtractor to extract sources. For creating lightcurves, they integrated their spectra to one point, and to look for wavelength-dependent intensity changes, subband lightcurves were also generated from these spectra. On their atmospheric fits by non-irradiated models, especially the water absorption feature

⁹One-arcsecond difference in the parallax angle, and one parsec is approximately 3.26 lightyears.

¹⁰Gyro-chronology, age determination of stars from their rotation periods.

¹¹Further details can be found in the Section 2.2.1 of Stassun et al.[7]

was not well defined by either ATMO2020 or Sonora models. They also made a forward model with EGP and generated spectra with PICASO but some parts of the results were not compatible with the observation, especially the blue end of the spectrum.

Maggio et al.[10] examined the youngest known transiting exoplanet system HIP¹² 67522. The data was collected via X-ray Multi-Mission (XMM) - Newton Reflection Grating Spectrograph (RGS, Extreme Ultraviolet) and European Photon Imaging Camera and Optical Monitor (EPIC and OM, Near Ultraviolet) for spectroscopy and Hubble Space Telescope's Cosmic Origin Spectrograph (Far Ultraviolet emission). XMM-SAS reduced the observation data files of XMM-Newton. The spectra were analyzed xspec v. 12.12 using an optically thin plasma emission model VAPEC - [[variable]] astrophysical plasma emission code. They also made a time-resolved spectra analysis after dividing the observation in 7 intervals. Emission lines of the coronal ions in RGS spectra were measured with Pint Of Ale ver. 2.954. HST transition region lines and Pint Of Ale results were combined to reconstruct the plasma Emission Measure Distribution vs. temperature. After these measurements and analyses, they concluded that HIP 67522 is very active with its plasma temperatures reaching 20 million Kelvin.

Hajra et al.[11] focused on April 23-24, 2023 double-peak intense geomagnetic storm. They used 1-min resolution SYM-H index data from World Data Center for Geomagnetism, Auroral substorm activity with the SML index in again 1-min resolution from the SuperMAG project, near-Earth solar-wind and interplanetary conditions from 1-min resolution solar wind plasma and interplanetary magnetic field (IMF) from NASA's OMNIWeb database. Coronal Mass Ejection (CME) information was from the halo CME catalog from the Large Angle and Spectrometric Coronagraph (LASCO) on solar and Heliospheric Observatory (SOHO). They discussed the impacts of this specific geomagnetic storm and presented their results.

Rodgers-Lee[12] estimated the transport of energetic particles with the Monte Carlo energetic particle transport model through a hydrogen-dominated exoplanet atmosphere. For this, a stellar wind model properties velocity, magnetic field strength, and mass loss rate were necessary. The author stressed the level of turbulence's influence, rather than field strength, on particle transfer. Additionally, the exoplanet atmosphere should be modeled, too. One-dimensional energetic particle model was calibrated against *Voyager* and Pay-

¹²Hipparcos Catalogue

load for AntiMatter Exploration and Light-nuclei Astrophysics - PAMELA detections. The exoplanet atmosphere properties were obtained with the Helios model. The study was with the M dwarf star GJ436 and its mini-Neptune GJ436 b with a hydrogen-dominated atmosphere within 0.01 to 0.2 astronomical unit distance from its star, and the result was high ionization rates originating from the stellar energetic particles.

Altinier et al.[13] constructed a simulator pipeline for Roman¹³ Coronagraph Instrument, CAPyBARA - Coronagraph and Aberration Python-Based Algorithm for Roman Analysis. Their simulation is in three levels, 1) optical propagation, 2) environmental simulation, and 3) post-processing. The optical part is how the light propagates through the hardware¹⁴, and also the dark-hole digging part¹⁵. For dark hole digging, the Electric Field Conjugation method measures the electric field and reshapes deformable mirrors to result in destructive interference¹⁶. The environment is modeled by quasi-static aberrations from thermal fluctuations originating from physical movements and insulation of the telescope. Random surface aberrations were introduced with 21 Zernike polynomial modes excluding piston, tip, and tilt-related first three modes. For post-processing, classical Reference-Star Differential Imaging (cRDI) was used. They cautioned that the simulation was on monochromatic light, hence not very realistic, but the methodological sequence did improve the image.

Ground-based

Wanderley et al.[14] used Sloan Digital Sky Survey - SDSS Apache Point Observatory Galactic Evolution Experiment (APOGEE) data to estimate stellar magnetic fields of 43 exoplanets, then they evaluate the minimum planetary magnetic

¹³From Nancy Grace Roman Space Telescope, up-to-date details can be seen via here. One thing that makes the Roman telescope important is it can be critical before the Habitable Worlds Observatory. Very shortly, it has a wide field instrument 100 times greater than the Hubble Space Telescope infrared instrument, and a Coronagraph instrument (a corona that blocks the starlight and permits orbiting exoplanet light more, hence the middle is empty but graphing the corona.) which much better exoplanet contrast.

¹⁴Remember wave diffraction, e.g. Hubble Space Telescope point sources has four spikes, James Webb Space Telescope has 6, etc., because of the nature of photons. Furthermore, other optical elements impact and coronograph result.

¹⁵This step purpose is reducing the direct starlight in the corona more and more, so the image keeps getting darker, in which there was already a dark hole in the middle.

¹⁶Interference is another wave concept, different waves can amplify or destroy each other according to how they interfere, and utilizing this phenomenon the technique aims to remove starlight as much as possible.

field of the Kepler-186f and Tess Object Identifier - TOI-700d, inside habitable zone exoplanets, to shield their atmosphere from their star's magnetic fields. They estimated mean magnetic fields from the four selected Fe I lines, sensitive to magnetic fields. OH lines were used to determine stellar rotational velocities ($v\sin i$). Then with Monte Carlo and Markov Chain methodology with emcee Python package, an ensemble of filling factors (2kG step variation between 2 - 6 kG μ B_T) and metallicity were fitted to these four lines. In the end, they estimated that 0.05-0.024 G magnetic fields from the exoplanets would be sufficient to keep their atmosphere scrapped (where the Earth has a magnetic field between 0.25 and 0.65 G).

Xu et al.[15] made millimeter (as opposed to centimeter) waveband very long baseline interferometry (VLBI) observations with Korean VLBI Network in 22/43/88/132 GHz (K/Q/W/D bands, respectively). Geodetic and astrometric VLBI monitors compact extra-galactic objects that have radio wave emissions with pairs of telescopes, and by measuring the difference of light arrival to different pairs of telescopes, telescope, celestial object positions; as well as earth orientation parameters can be inferred. They presented the details of scheduling, observation programs and projects, and sample selection with Multi-frequency Active Galactic Nuclei Survey with the Korean VLBI Network (MASK), scheduling program SKED and in the end, the study had 82 sources and 485 scans. In correlation and fringe fitting, DiFX software correlator, Haystack Observatory Processing System (HOPS) suite of programs, and National Radio Astronomy Observatory (NRAO) Astronomical Image Processing System, and for atmospheric turbulence consideration in millimeter wavelength, frequency phase transfer (FPT) technique¹⁷ was employed. Geodetic data analysis was performed using nuSolve program. Using HOPS, they implemented FPT in the following steps:*fourfit* initiated the fringe fitting and *alist* tracked the fringe phase, SNR, and generated a related text file from these cases. Then, several-second length segmentation (not too short to have a good signal-to-noise ratio (SNR) and not too long to miss short-duration atmospheric fluctuations) of approx. 2 min scans with *fringex* program. After that, frequency ratios were used to multiply the solutions from the lower frequencies and transfer them to the higher ones. From this step, reference stations was designated as KVN-Yonsei - KYS, and it measured the atmospheric phase from KVN-Ulsan - KUS and KVN Tamma - KTN, then *fourfit* fitted the fringe. They report the remarkable success of FPT in

¹⁷High-frequency observations are calibrated with solutions from lower frequencies.

SNR improvement for fringes, and especially systematic errors in group delay and delay rate for 88 and 132 GHz W and D bands, and with longer baselines, more application areas are expected for millimeter VLBI cases.

Blind et al.[16] present RISTRETTO, a Very Large Telescope Extreme Adaptive Optics (VLT XAO) design for visible monitoring of Proxima Centauri b reflected light analysis, with simulations and estimated positive working weather conditions. They first present the XAO requirements and error budget, then explained wavefront sensing (WFS) in the near-infrared (NIR) with a dual WFS architecture with “red” and “blue”¹⁸. Wind effect correction is another challenge, where the local effect must be accounted and corrected. Then, they moved to deformable mirrors. Deformable mirror case had MEMS technology-related issues on dead actuator management and deformable mirror complex dynamics. Then, they gave the XAO architecture and provided details on the simulation they did for the end-to-end performance: XAO simulation was from the OOMAO Matlab library, and then, residual phase screens were injected into the PIAA-Nuller simulations from Hcipy Python library. DM actuator count was 41x41 and pupil or both WFSs were oversampled to 60x60 subapertures. The flux was from Proxima Centauri for this test. They cautioned on the small dynamical range of zWFS.

Artigau et al.[17] also worked on stellar temperatures with spectroscopic data. One of the most intriguing parts is that their approach utilizes not specific lines, but the entire spectrum and generates sub-Kelvin accuracy in temperature variation detections. Their inspiration was from line-by-line technique¹⁹ from radial velocity measurements. SPIRou (SpectroPolarimètre Infra-ROUge) and HARPS (High Accuracy Radial Velocity Planet Searcher) instruments-generated datasets were used to construct the template. They then applied this to precision radial velocity (RV) datasets, incidentally, these stars are generally nearer to the Solar System. They discussed the results one-by-one for three stars from this precision RV dataset. Additionally, they used this precise temperature change estimation method to transiting HD²⁰ 189733 b, and for a spectral li-

¹⁸red here is higher wavelength than 1400 nm, while blue is 850-1000, hence both are in NIR region.

¹⁹Simply put, the difference between the spectrum and ideal, noiseless spectrum at a given point is the derivative of that spectral value against the infinitesimal change in wavelength, which then, multiplied by wavelength change. They corresponded this wavelength change to radial velocity, or more properly, the velocity shift, and this study here works for temperature variation.

²⁰Henry Draper catalogue

brary, they chose PHOENIX²¹. They stated that disk-averaged temperature measurement via this method was generally successful and promising.

Theoretical

Hayashi et al.[18] made a three-body problem study with one central massive body and two lighter bodies revolving it as a binary in TSUNAMI[19]. They aimed to improve the Hill-type stability criterion in these systems and checked the fate of Hill-unstable systems in much longer time scales with longer simulations. They defined improved stability boundaries for hierarchical triple satellite type (HT-S) systems. For the robustness of this new criterion, the inner eccentricity dependence, the outer eccentricity dependence, and the mass-ratio dependence of the previous study by Grishin et al.[20] with numerical analyses (similar to a sensitivity analysis).

Kan et al.[21] made a fluid-dynamics study and presented the rescaled incompressible Navier Stokes equations (RiNSE). They reported interesting results with low Ekman number²² asymptotics, and they were also relevant for the outer core of the Earth and the convection zone of the Sun.

Practical-Algorithm Development

Hutchison and Koepferl[22] developed an interactive contour analysis tool with Python²³ for background and morphology analysis-segmentation purposes. The algorithm has a suite of statistical smoothing and image operation functions, as well as specific background removal techniques in-built, and can work both raw images and pre-processed images in a modular way. On contours, lowest closed contour (LCC), Otsu thresholding, and average thresholding were coded and explained in the paper. There is also a masking utility. They made several tests to see the utility and benefit of their algorithm. One major advantage is it eases the implementation by not requiring an extensive coding background, and instead, providing a highly interactive user interface.

Kim et al.[23] worked on photonic lantern spectroastrometry²⁴. The authors first started by discussing the drawbacks of using the long-slit spectrographs and integral field spectrometers. In

²¹PHOENIX spectral library can be found [here](#)

²²viscous force to Coriolis force ratio

²³The code is [here](#).

²⁴Estimating object morphology via wavelength-different measurements for object sizes smaller than the point-spread function. The point spread function is, very roughly, the diffraction of the light coming from a point-like source on the mirrors of the monitoring instrument and the resulting spread. For example, check Section 5.2 of this document on JWST PSF

this lantern, few-mode or multi-mode fibers (FMF or MMF) gradually become single-mode fibers (SMFs). A critical phenomenon is that these fibers will have the spatial structure information embedded in their outputs so a 2D spectroastrometry is possible, and the different modes are being differentiated to each SMF, which can feed the spectrometers. The paper also contains simulated spectroastrometric signals and mock observations and a discussion on the considerations on photonic lantern design.

Last week in Chemistry

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between September 9 - September 16, 2024. They are more in nature of spectroscopy alone, and hence several studies regarding biochemistry, chromatography, and several other disciplines might be missed here.

Hooper, Silva, and Sakellariou[24] studied n-butyl acrylate UV-polymerization *in situ* with the photo-irradiation magic angle spinning²⁵ Nuclear Magnetic Resonance (MAS NMR). In this way, they can see the shifts in peaks from the ¹H and ¹³C NMR spectra in real time, and this hints at the reaction mechanism pathways. They show their custom-made rotor that permits the irradiation of the sample while spinning at the magic angle and monitoring it via NMR. Their setup replaced a commercial pencil rotor with an in-house manufactured tapered glass rod from quartz glass, the narrow end of the rod is held tight by the polytetrafluoroethylene (PTFE, also known as Teflon) insert, and the rod itself stays in place with friction, removing the glue requirement and resulting spectral interference.

Kret et al.[25] focused on glucose testing with Matrix Assisted Laser Desorption/Ionization Imaging Mass Spectrometry (MALDI IMS) with two matrices: N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC) and α -Cyano-4-hydroxycinnamic acid (CHCA). They have vastly different MS/MS results from these matrices, which were surprising results. They further checked a betaine aldehyde derivatisation or again NEDC matrix of hexoses (here galactose, mannose, fructose, and glucose). Similarly, intensities were one order of magnitude lower in NEDC, and noisy structures were much apparent in the spectra, especially from the glucose from the brain and liver, and galactose. Their NEDC matrix also still gave tissue-specific maps with isotopic maps, and potential contamination/interfering effects were discussed but remained open.

Kawchak[26] utilized large multimodal models to generate text regarding the chemical from given spectrum information and then judged the results by another model, ScholarGPT. Prompts and related tests were provided in detail. The researcher stated that these models are now capable of synthesizing information from many different spectra

and come up with detailed answers.

²⁵This 54.74-degree angle in solid-state NMR spectroscopy reduces the issues coming from nuclear dipole interaction, chemical shift anisotropy, and quadrupolar interaction.

Last week in Remote Sensing

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between September 9 - September 16, 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.

Li et al.[27] worked on a transferrable anomaly detector for earth observation remote sensing images. They focused on image-independent deviation metrics to overcome the need to train their anomaly detector for each image separately. Their method is on a deviation metric Lipschitz continuous²⁶ and satisfying large margin condition²⁷ to be transferable to the unseen images. Pixel level and feature level losses were defined with the area under the ROC (receiver operating characteristic) curve metric and separate loss for feature level. Then, spatial and spectral anomaly tests were conducted with iSAID dataset²⁸ and compared their unified anomaly detector in remote sensing (UniADRS) with the global RX detector (GRX), the abundance and dictionary-based low-rank decomposition(ADLR), the collaborative representation (CRD), the spectral constraint autoencoder (SC_AAE), the eep low-rank prior-based detector (DeepLR), and the transferred direct detection model. The algorithm was successful in attaining the detection ability in different datasets: hyperspectral, visible light, synthetic aperture radar, infrared, and low light²⁹.

Xu et al.[28] leveraged the panchromatic images to denoise hyperspectral image bands. The name of their method is Panchromatic Weighted Representation Coefficient Total Variation (PWRCTV). Representative Coefficient Total Variation, in short, assumes much lower rank in the image matrix than the number of the hyperspectral bands, and expects mostly locally-smooth, similar neighboring pixels. However, rather than completely smoothing the image or losing edges and feature margins, panchromatic band with lower noise was also used to guide the algorithm where there can be noise artefacts and where there are indeed real features. They provided the algorithm in detail and provided code and data here. Their algorithm mostly out-

²⁶In a continuous function, the *Lipschitz constant* is the largest the slope of the corresponding function can take. The function is almost everywhere differentiable.

²⁷Referring to the Lipschitz constant, large margin is defined to be 2 times the Lipschitz constant or higher

²⁸The iSAID dataset is here

²⁹Night-time remote sensing

performed other internal information and deep learning-based algorithms on various datasets.

Qin et al.[29] used Sentinel-2 Satellite Image Time Series data with Mamba structure³⁰ included neural network³¹ for crop classification. They used PASTIS³² and MTLCC datasets in comparing their novel architecture to the state-of-the-art. SITSMamba generally outperformed other models in both datasets in most crop classes. MacDonald, Jacoby, and Coady[30] also worked on satellite image time series segmentation and utilized PASTIS and MTLCC crop-type segmentation benchmarks. Especially on MTLCC dataset, their algorithm VistaFormer with Multi-Head Self-Attention had better results than the state-of-the-art. Another preprint utilizing the Mamba structure considered the pyramid pooling local auxiliary SSM-based model, PPMamba[31]. They wanted to ameliorate the problem of Mamba structures in keeping the local semantic information, and they tested their proposed algorithm on ISPRS Vaihingen and LoveDA Urban against convolutional-based, transform-based, and Mamba-based algorithms. PPMamba performed better in the highest number of the classes and was competitive in terms of computational complexity.

Vincent et al[32] provided a looted archaeological site dataset from 675 Afghan archaeological sites, including 55480 images over 8 years³³. From red-green-blue (RGB) images of this dataset, several single-frame, pixel-wise multi-frame, and whole-image multi-frame methods were evaluated by them. Their best algorithm was the Dynamic One-for-All (DOFA) model and Light Temporal Attention Encoder (LTAE).

Kim et al. kim2024nbboxnoisyboundingbox argued that bounding box transformations can be beneficial in model regularization in object detection tasks utilizing aerial imagery. Their NBBOX (Noise Injection into Bounding Box) algorithm was applied to DOTA³⁴ and DIOR-R³⁵ dataset’s aerial images. They utilized acnchor-based³⁶ two-stage, anchor-based one-stage, and anchor-free one-stage detection models as Faster R-CNN, RetinaNet, and FCOS, respectively. They rotated, scaled, and translated the objects and monitored the performances of the algorithms after these operations. NBBOX was found to be

³⁰A selective state-space modelling block.

³¹SITSMamba code can be retrieved from here.

³²PASTIS dataset link.

³³DAFA-LS, looted archaeological dataset and the related code is here.

³⁴DOTA dataset can be retrieved from here.

³⁵DIOR-R dataset can be found here.

³⁶First, there are specific bounding boxes, anchors that parts of the image are analyzed for a given class(es). Then, with class probabilities, the anchors got aggregated on objects.

the best among the state-of-the-art augmentation methods.

Liu, Prasad, and Crawford[33] worked on hyperspectral imagery classification with vision Transformers. They used Houston 2013, Botswana, and Pavia University datasets (involving different classes in different numbers, and the Pavia University dataset contained much more data than others. Their framework has a unified architecture with different mixer blocks. All components in this structure are spatial-self attention (SSA), channel self-attention (CSA)³⁷, convolutional neural network (CNN), linear normalization (LN), and multi-layer perceptron (MLP) elements. They stressed the importance of the network assemblage instead of the specific self-attention module, and their developed network surpassed the classical available models on these three datasets.

Vu et al.[34] studied masked image modeling for remote sensing object detection. They created a cross-attention masked image modeling approach where the masked area was not defined by a single token as in the conventional approach, but its relationships with other unmasked areas are also considered. They used VEDAI (vehicle detection in aerial imagery), DIOR, and AVIID (aerial-visible to infrared image translation, data link) datasets. Different YOLO models were compared and the developed model was mostly successful, and the best except car, camping, and tractor vehicles.

In terms of querying the change in remote sensing images with time, Deng, Zhou, and Wu[35] developed ChangeChat that surpasses other state-of-the-art approaches and it is the first specifically designed bitemporal vision-language model for remote sensing change analysis.

Cao et al.[36] worked on open-vocabulary remote sensing image *semantic* segmentation and reported improvements over available methods.

Josephson et al.[37] first discussed the implications of the use of remote sensing earth observation-generated weather products in an economic context providing a case from the Africa content with a limited *in situ* observations. Then, they utilize publicly available socioeconomic data with obfuscated Global Positioning System (GPS) coordinates. The authors cautioned that selecting a different earth observation data can significantly alter the results, and the robustness checks should be in place.

Cha, Yu, and Seo[38] followed a different path to increase the success of vision-language models with 9.6 million vision-language paired datasets from the very high resolution imagery domain.

³⁷Inter-channel relations on feature detections, where the spatial dimension is squeezed.

Their model is RSCLIP, utilizing the benefits of InstructBLIP and CLIP together.

Li et al.[39] extensively discussed the problem in *instance segmentation*³⁸ and pointed out the scale, translation problems, as well as the ratio of pixels between the object of interest and the background. They proposed an additional local prompt encoder³⁹ and transformer element into the instance segmentation model architecture. Another global-to-local prompt gives a global texture insight into these local cases. For their tests, two optical (ISAID and NWPU VHR-10) and two synthetic aperture radar (SSDD and HRSID) datasets were chosen. They reported that the added local and global-to-local prompts were successful. Their model was generally more successful than other types of relevant algorithms.

In another preprint, researchers[40] constructed the AnomalyCD (anomaly change detection) algorithm and benchmarked it with custom-developed dataset AnomalyCDD. Their unsupervised way of approaching this problem is interesting, as you don't need to label specific instances and you can feed the algorithm with millions of data and check its results later. Their anomaly categories in the datasets were explosions, collapses, landslides, fire, dam break, and various other cases under the umbrella term "other". Their model was compared to fully convolutional change detection with generative adversarial network (FCD-GAN), and the zero-shot change detection models (segment anything model (SAM)-multimodal change detection). AnomalyCD reduced the false positive anomalies in almost all cases reported.

Zhang et al[41] proposed a knowledge discovery network for efficient feature extraction with the *renormalized connection*. In addition to renormalized connections, they also illustrate multi-branch detector architectures. They shared the IPIU, MS-COCO-TOD datasets, and Renormalized-Connections code in the following link. Essentially, different tasks of detections in different branches (like tiny or medium size, square or irregular shapes) were to be renormalized with 421 coefficient ratios.

³⁸Identifying objects in an image.

³⁹This local prompt module crops the instances (object of interest) from the image and generates a token for this specific cropped part and feeds it to this module.

Last week in Environmental Chemistry

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between September 9 - September 16, 2024. chemRxiv's Earth, Space, and Environmental chemistry preprints are being surveyed, and unfortunately, not many preprints are published under environmental topics in this field.

Lim et al.[42] proposed that a metal-organic framework with cyclodextrin basis was better than current most-used amine scrubbers for carbon dioxide removal from the dilute streams of air or natural gas combined cycle power plant flue gas. They first discuss the conventional sorbents inability to effectively capture carbon dioxide and then safely cycle it back owing to low-partial pressure challenge in adsorption and high-temperature requirement to desorp the carbon dioxide. The latter can damage the sorbent, too. They create Rubidium (III) carbonate cyclodextrin-metal organic framework and characterized it with powder X-ray diffraction and scanning electron microscope. Thermogravimetric analysis was conducted to determine the optimal activation temperature for each CD-MOF. Carbon dioxide selectivity in the presence of nitrogen or oxygen molecules was also experimented. They conducted isothermal cycling experiments as well, where under which temperature and humidity conditions the cycling is better. They also hypothesized an irreversible carbon dioxide reaction with hydroxyl sites under elevated temperatures in dry conditions. To understand this further, they conducted diffuse reflectance infrared Fourier transform spectroscopy (DRIFT) and confirmed their hypothesis.

Lu et al.[43] generated a high-resolution updated geospatial database of population-weighted annual-average concentration of six air pollutants for the contiguous U.S. for 2016-2020 period. They also developed land-use regression models with partial least-square universal kriging framework. Their data and model can be reached via this link.

Last week in Data Science-Applied Statistics

Author: Yasin Güray Hatipoğlu

The preprints summarized here were published between September 9 - September 16, 2024. This is generally from arXiv's stat.ML (and correspondingly cs.LG) cross-list. Large-language model-related, or text-mining and similar studies are omitted, also several less-application oriented studies (they, too, are important, but currently *Science Ascend* can't accomodate to review them within the time constraints.

Calvo-Ordoñez, Palla, and Ciosek[44] worked on epistemic uncertainty⁴⁰ and observation noise on the neural tangent kernel. The idea in the paper is to make the neural network more interpretable. They start by citing another study that training a wide neural network with gradient descent on a standard loss is *formally* equivalent to the posterior mean of a Gaussian Process (GP)⁴¹, where neural tangent kernel (NTK) with zero aleatoric noise. Neural Tangent Kernel is a Jacobian of function value differential via infinitesimal parameter change and converges to a deterministic, constant kernel in the limit (remember *tangent*). The authors wanted to extend this case with non-zero aleatoric noise as the real world is not noiseless, and they also wanted to know more about the model outcome. They defined two separate algorithms for posterior mean and posterior covariance (employing partial singular value decomposition) in NTK-GP, provided proofs, and experimented with it in a toy dataset with successful results.

Agrawal et al.[45] focused on the estimation of asymptotic variance of a function to assist in the inference of the stationary mean by Markov Chain⁴² Variance Estimation. "Under appropriate assumptions" and according to the *Central Limit Theorem*⁴³. The asymptotic part comes into play when it is long enough for this con-

⁴⁰Epistemic uncertainty comes from the model itself, while *aleatoric* noise or uncertainty comes directly from data.

⁴¹Gaussian Processes defines the input-output pairs and uncertainty by assuming a jointly-Gaussian covariance structure, and in addition to input-output consideration, it also has a specific parameter for observational-aleatoric noise.

⁴²A series of data in which an element's value is only dependent on the previous element's value, but not others. Furthermore, if a memory "m" is defined for dependence, it will be the Markov Chain in the order of m where the element is dependent on m previous elements preceding it.

⁴³Roughly, a mixture of any kind of distribution will add up to something very close to the normal distribution.

vergence, and expecting a specific variance, and in this paper, that variance $\kappa(f)$ is being estimated. The challenge is that literature generally estimated this variance with independent and identically-distributed assumptions (no auto-correlation) while here, obviously, there is a dependence between the current and previous element. Compared to the literature, this study employed a linear statistical approximation for this estimation, with $O(1)$ memory requirement and $O(1/n)$ mean squared error estimation in terms of complexity. They established algorithms formally for asymptotic and stationary variances and presented applications in reinforcement learning frameworks. Specifically for temporal difference learning⁴⁴, they provided algorithms to calculate these variances, too, again with reduced complexity in $O()$ notation.

Boopathy et al.[46] considered the theoretical basis for the better performance and scalability of modular neural networks compared to nonmodular ones. They considered an analytically tractable regression task for setting up a model network. They first showed the generalization trend of Neural Networks via linear regression and then moved to modularity for the generalization. An important property of this modular approach is how they can be independent of dimensions, hence the superior scalability part. They showed this on sine-wave regression task and compositional image labeling CIFAR-10⁴⁵. In another relevant preprint, Boopathy and Fiete[47] constructed unified NN scaling laws with model size, training time, and data volume. Moreover, there is a "double descent" concept, where initially the model learns and error descends and then it overfits to the training data and sharply increase in error, and then after learning better, its performance increases for a while (second descend) and slowly starts to decrease later. Even though for data volume and model size this double descent was defined well, training time had differing explanations. The authors made both theoretical explanations and empirical validations, and in empirical validations they see a clear inversely proportional relation between the logarithm of epochs⁴⁶ and scale (by varying network widths and dataset size with subsampling). Their datasets were MNIST⁴⁷, CIFAR-10, and SVHN

⁴⁴A reinforcement learning where the algorithm updates its weight to better predict a future outcome. "Tabular" TD is when TD creates a table of values and updates as appropriate in the following states.

⁴⁵You can reach CIFAR-10 and related datasets from here.

⁴⁶Directly related to the training time, maybe a good analogy is how many laps a runner traverses on a running track.

⁴⁷You can reach this Modified National Institute of Standards and Technology database from here

(Street View House Numbers) with multi-layer perceptron, 7-layer convolutional neural network with stochastic gradient descent (SGD).

Scheuerer et al.[48] created a machine learning method to estimate seasonal precipitation for East Africa region. Their method was taking the observed precipitation fields and performing an empirical orthogonal function analysis⁴⁸ to reduce dimensions, predict with the factor loadings. They employed Climate Hazards group InfraRed Precipitation with a Station (CHIRPS) data and for analyzing longer training data impact the Global Precipitation Climatology Centre (GPCC) precipitation data. They used above-normal and below-normal for features and *tercile* seasonal approach of below-seasonal normal, above-seasonal normal and seasonal-normal for the outcomes. They found that the data-driven approach is generally on par with the ECMWF forecast, and for the sake of better data quality, CHIRPS data was found more appropriate, as well as higher degree of the spatial aggregation improved the forecast accuracy.

Qu et al.[49] proposed a novel approach for the uncertainty quantification in seismic inversion with the integration of the deep ensemble methods and the importance sampling⁴⁹. In the seismic inversion case, the uncertainty has both aleatoric and epistemic sources, and a specific set of seismic observations can be generated by more than one model, hence the non-uniqueness phenomenon. Their network was in the U-net architecture. Increasing the ensemble size and training duration decreased the uncertainties in the deep ensemble method comparatively better than the Monte Carlo dropout alone. One of the emergent challenges with UQ using the deep ensemble and importance sampling, though, is the higher computational cost.

Jayalal et al.[50] worked on making the Online Robust Principal Component Analysis a tuning-free method with an implicit regularization. Simply, OR-PCA has the observed data as $Z = X + E$, where E is the sparse outlier, and X is bilinearly decomposed to $X = LR^T$, with L basis and R as representation in this basis. They also applied their proposed method to PETS2006, Pedestrians, and bungalows. Their algorithms separately estimate the error in positive and negative components, u and v for the coefficient vector r_t with the momentum gradient descent, and a similar route was followed for L . Their proposed algorithm was on par with the other ways to solve these problems, and it was tuning-free.

⁴⁸Employing Principal Component Analysis in geographical domain

⁴⁹The weights are not sampled from the posterior distribution, but from an alternative distribution

Chen et al.[51] created the approximate Fisher Information Ratio Active Learning - ApproxFIRAL with orders of magnitude faster results compared to the FIRAL on the same tasks. Their novel algorithm had obvious lower complexities in FIRAL steps on the storage and computation. For accuracy tests, MNIST, CIFAR-10, Caltech-101, and ImageNet datasets were used. After several instances, ApproxFIRAL and FIRAL had more or less the same accuracy, and even in some cases of the ImageNet dataset, ApproxFIRAL was better. They also made sensitivity analyses and scalability tests for parallel computing.

Denoising Diffusion Probabilistic Models is another venue of research in machine learning[52]. Li and others worked on link prediction with their proposed Sub-graph Based Diffusion model (SGDiff), in their graph learning task they employed generative models for predicting the perturbation, unlearning the perturbation, and making an estimate over the links.

Cavallo et al.[53] provided the Fair coVariance Neural Networks (FVNNs) algorithm, which overcomes the limitations of covariance-based data processing and low-sample instability of fair principal component analysis. The “fairness” for this case it sounds, tries to resolve the problem coming from the biased representation of different groups, underrepresented groups, and similar inherent problems in the dataset. For fair covariance matrix, for two groups, separate covariance calculations for each groups are also conducted and a balancing term is used that can be 1 with zero balancing, the usual covariance matrix calculation, and 0 corresponds to a completely separate covariance calculation, and subtracting the prominent covariance from the disadvantaged group’s covariance. Another tweak is in the penalties in the training loss to mitigate biases. They theoretically and numerically evaluated these conditions. They both considered a synthetic data with two groups, and then, real regression data and real classification data. In the end, even though FVNN improved bias, it increased the error with the training loss-related balancing term for the debiasing operation.

Pandey et al.[54] proposed a new halo-creation method, CHARM⁵⁰ - Creating Halos with Auto-Regressive Multi-stage networks, to reduce the computational cost in cosmological calculations and creating mock halo catalogs to match dark matter density field-related effects. They used N-body Quijote project⁵¹ simulation with a volume of 1000Mpc/h. They had 1800 Latin hypercube simulations for the training and left the 200 LH for the test part. For the dark matter field inputs, the

⁵⁰The code of CHARM can be reached herer.

⁵¹Quijote data can be retrieved from here.

FastPM particle mesh algorithm was used. The Halo catalog was obtained with the Rockstar halo finder. They gave the results of this FastPM to the two stacked 3D ResNet (residual network) layers. They also calculated the power spectrum, bispectrum, and wavelets which are sensitive to the correlations among the galaxies. In the end, they presented the success of their approach to sort of bypass the computationally prohibitive N-body and dark matter halo calculations.

Theoretical underpinnings

Ichikawa and Hukushima[55] worked on min-max problems, introduced a statistical-mechanical formalism, tested it on a bilinear min-max game in their sensitivity analysis, and finally made further analyses on generative adversarial network’s generalization performance. Pointing out the usefulness of statistical-mechanical approaches in optimization and constraint-satisfaction, they chose to use a virtual two-temperature system. They reported that they had further discovered the mix-max optimization concept.

Rozenfeld[56] examined the Shapley value calculation methods from the causal perspective since two common approaches conditional and marginal approaches contradicted each other. The analogy is the features of a machine learning model are players here and we are measuring their coalition on explaining an outcome or *explicand*. For the model explanation and the exploration of relationships within the data, correlated features result in quite distinct outcomes. In this causal case, Rozenfeld argued that the correlations among the features resulted in spurring causal estimations and violated the fundamental principle of statistics (and in the end, the marginal approach was recommended).

Ziemann[57] gave an information theory-related note on the consistency of Gaussian maximum likelihood estimated in linear auto-regressive models in a frequentist statistics setting.

Aria et al.[58] used the R package e2tree, and extended the explainable ensemble trees approach to the regression analysis by Random Forest⁵² and tested it with freely available datasets like Iris. They both show a dissimilarity matrix between these trees and also associations between predictors and response values.

⁵²Generally 500-1000 decision trees, randomly attaining a piecewise value to assign classes or values according to a part of a feature, but weakly. It is one of the most commonly used ML techniques.



References

- [1] Fatemeh Fazel Hesar, Bernard Foing, Ana M. Heras, Mojtaba Raouf, Victoria Foing, Shima Javanmardi, and Fons J. Verbeek. Advancing machine learning for stellar activity and exoplanet period rotation, 2024, 2409.05482.
- [2] Sarah Steiger, Laurent Pueyo, Emiel H. Por, Pin Chen, Rémi Soummer, Raphaël Pourcelot, Iva Laginja, and Vanessa P. Bailey. Simulated performance of energy-resolving detectors towards exoplanet imaging with the habitable worlds observatory. In Laura E. Coyle, Marshall D. Perrin, and Shuji Matsuura, editors, *Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave*, page 68. SPIE, August 2024.
- [3] K. L. Luhman. A census of the beta pic moving group and other nearby associations with gaia, 2024, 2409.06092.
- [4] C. G. Díaz, R. Petrucci, L. V. Ferrero, and E. Jofré. Valuation system for telescope observation time proposals for the gemini observatory, 2024, 2409.05826.
- [5] Tobin M. Wainer, James R. A. Davenport, Guadalupe Tovar Mendoza, Adina D. Feinstein, and Tom Wagg. Searching for stellar activity cycles using flares: The short and long timescale activity variations of tic-272272592, 2024, 2409.06631.
- [6] Lizhou Sha, Andrew M. Vanderburg, Luke G. Bouma, and Chelsea X. Huang. Confirming the tidal tails of the young open cluster blanco 1 with tess rotation periods, 2024, 2409.07550.
- [7] Keivan G. Stassun, Ryan J. Oelkers, Joshua Pepper, Martin Paegert, Nathan De Lee, Guillermo Torres, David W. Latham, Stéphane Charpinet, Courtney D. Dressing, Daniel Huber, Stephen R. Kane, Sébastien Lépine, Andrew Mann, Philip S. Muirhead, Bárbara Rojas-Ayala, Roberto Silvotti, Scott W. Fleming, Al Levine, and Peter Plavchan. The tess input catalog and candidate target list. *The Astronomical Journal*, 156(3):102, August 2018.
- [8] Jenni R. French, Sarah L. Casewell, Rachael C. Amaro, Joshua D. Lothringer, L. C. Mayorga, Stuart P. Littlefair, Ben W. P. Lew, Yifan Zhou, Daniel Apai, Mark S. Marley, Vivien Parmentier, and Xianyu Tan. The only inflated brown dwarf in an eclipsing white dwarf-brown dwarf binary: Wd1032+011b, 2024, 2409.06874.
- [9] Rachael C. Amaro, Dániel Apai, Yifan Zhou, Ben W. P. Lew, Sarah L. Casewell, L. C. Mayorga, Mark S. Marley, Xianyu Tan, Joshua D. Lothringer, Vivien Parmentier, and Travis Barman. Hotter than Expected: Hubble Space Telescope (HST)/WFC3 Phase-resolved Spectroscopy of a Rare Irradiated Brown Dwarf with Strong Internal Heat Flux. *apj*, 948(2):129, May 2023, 2303.07420.
- [10] A. Maggio, I. Pillitteri, C. Argiroffi, D. Locci, S. Benatti, and G. Micela. Xuv irradiation of young planetary atmospheres. results from a joint xmm-newton and hst observation of hip67522, 2024, 2409.07229.
- [11] Rajkumar Hajra, Bruce Tsatnam Tsurutani, Quanming Lu, Richard B. Horne, Gurba Singh Lakhina, Xu Yang, Pierre Henri, Aimin Du, Xingliang Gao, Rongsheng Wang, and San Lu. The april 2023 sym-h = -233 nt geomagnetic storm: A classical event, 2024, 2409.08118.
- [12] D. Rodgers-Lee. Stellar energetic particle and cosmic ray effects in exoplanetary atmospheres, 2024, 2409.07274.
- [13] Lisa Altinier, Élodie Choquet, Arthur Vigyan, Nicolás Godoy, and Alexis Lau. Escape project. capybara: a roman coronagraph simulator for post-processing methods development. In Laura E. Coyle, Marshall D. Perrin, and Shuji Matsuura, editors, *Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave*, page 130. SPIE, August 2024.
- [14] Fábio Wanderley, Katia Cunha, Verne Smith, Oleg Kochukhov, Diogo Souto, Carlos Allende Prieto, Suvrath Mahadevan, Steven Majewski, Philip Muirhead, Marc Pinsonneault, and Ryan Terrien. Magnetic fields in a sample of planet-hosting m dwarf stars from kepler, k2, and tess observed by apogee, 2024, 2409.06637.
- [15] Shuangjing Xu, Taehyun Jung, Bo Zhang, Ming Hui Xu, Do-Young Byun, Xuan He, Nobuyuki Sakai, Oleg Titov, Fengchun Shu, Hyo-Ryoung Kim, Jungho Cho, Sung-Moon Yoo, Byung-Kyu Choi, Woo Kyoung Lee, Yan Sun, Xiaofeng Mai, and Guangli Wang. A geodetic and astrometric vlbi experiment at 22/43/88/132 ghz, 2024, 2409.07309.

- [16] Nicolas Blind, Muskan Shinde, Isaac Bernardino Dinis, Nathanaël Restori, Bruno Chazelas, Thierry Fusco, Olivier Guyon, Jonas Kuehn, Christophe Lovis, Patrice Martinez, Mathieu Motte, Jean-François Sauvage, and Alain Spang. Ristretto: a vlt xao design to reach proxima cen b in the visible. In Dirk Schmidt, Elise Vernet, and Kathryn J. Jackson, editors, *Adaptive Optics Systems IX*, page 213. SPIE, August 2024.
- [17] Étienne Artigau, Charles Cadieux, Neil J. Cook, René Doyon, Laurie Dauplaise, Luc Arnold, Maya Cadieux, Jean-François Donati, Paul Cristofari, Xavier Delfosse, Pascal Fouqué, Claire Moutou, Pierre Larue, and Romain Allart. Measuring sub-kelvin variations in stellar temperature with high-resolution spectroscopy, 2024, 2409.07260.
- [18] Toshinori Hayashi, Alessandro A. Trani, and Yasushi Suto. Stability and fate of hierarchical triples comprising a central massive body and a tight binary in eccentric orbits, 2024, 2409.06253.
- [19] Alessandro A. Trani and Mario Spera. Modeling gravitational few-body problems with tsunami and okinami. In Dmitry Bisikalo, Dmitri Wiebe, and Christian Boily, editors, *The Predictive Power of Computational Astrophysics as a Discover Tool*, volume 362 of *IAU Symposium*, pages 404–409, January 2023, 2206.10583.
- [20] Evgeni Grishin, Hagai B. Perets, Yossef Zenati, and Erez Michaely. Generalized Hill-stability criteria for hierarchical three-body systems at arbitrary inclinations. *Monthly Notices of the Royal Astronomical Society*, 466(1):276–285, 11 2016, <https://academic.oup.com/mnras/article-pdf/466/1/276/10865455/stw3096.pdf>.
- [21] Adrian van Kan, Keith Julien, Benjamin Miquel, and Edgar Knobloch. Bridging the rossby number gap in rapidly rotating thermal convection, 2024, 2409.08536.
- [22] Mark A. Hutchison and Christine M. Koepferl. Contour Analysis Tool: an interactive tool for background and morphology analysis. *arXiv e-prints*, page arXiv:2409.06421, September 2024, 2409.06421.
- [23] Yoo Jung Kim, Michael P. Fitzgerald, Jonathan Lin, Yinzi Xin, Daniel Levinstein, Steph Sallum, Nemanja Jovanovic, and Sergio Leon-Saval. On the potential of spectroastrometry with photonic lanterns, 2024, 2409.09120.
- [24] Thomas Hooper, Rodrigo de Oliveira Silva, and Dimitrios Sakellariou. High-resolution in-situ photo-irradiation mas nmr: Application to the uv-polymerization of n-butyl acrylate, 2024.
- [25] Paulina Kret, Paulina Kret, Przemysław Mielczarek, Giuseppe Grasso, Piotr Suder, and Anna Bodzoń-Kułakowska. Different ion maps for the same substance? glucose analysis by ms imaging with betaine aldehyde derivatization and nedc matrix., 2024.
- [26] Kevin Kawchak. High dimensional and complex spectrometric data analysis of an organic compound using large multimodal models and chained outputs, 2024.
- [27] Jingtao Li, Xinyu Wang, Hengwei Zhao, Liangpei Zhang, and Yanfei Zhong. Learning a cross-modality anomaly detector for remote sensing imagery, 2024, 2310.07511.
- [28] Shuang Xu, Qiao Ke, Jiangjun Peng, Xiangyong Cao, and Zixiang Zhao. Pan-denoising: Guided hyperspectral image denoising via weighted represent coefficient total variation, 2024, 2407.06064.
- [29] Xiaolei Qin, Xin Su, and Liangpei Zhang. Sitsmamba for crop classification based on satellite image time series, 2024, 2409.09673.
- [30] Ezra MacDonald, Derek Jacoby, and Yvonne Coady. Vistaformer: Scalable vision transformers for satellite image time series segmentation, 2024, 2409.08461.
- [31] Yin Hu, Xianping Ma, Jialu Sui, and Man-On Pun. Ppmamba: A pyramid pooling local auxiliary ssm-based model for remote sensing image semantic segmentation, 2024, 2409.06309.
- [32] Elliot Vincent, Mehraïl Saroufim, Jonathan Chemla, Yves Ubelmann, Philippe Marquis, Jean Ponce, and Mathieu Aubry. Detecting looted archaeological sites from satellite image time series, 2024, 2409.09432.
- [33] Wei Liu, Saurabh Prasad, and Melba Crawford. Investigation of hierarchical spectral vision transformer architecture for classification of hyperspectral imagery, 2024, 2409.09244.
- [34] Minh-Duc Vu, Zuheng Ming, Fangchen Feng, Bissmella Bahaduri, and Anissa Mokraoui. Interactive masked image modeling for multimodal object detection in remote sensing, 2024, 2409.08885.

- [35] Pei Deng, Wenqian Zhou, and Hanlin Wu. Changechat: An interactive model for remote sensing change analysis via multimodal instruction tuning, 2024, 2409.08582.
- [36] Qinglong Cao, Yuntian Chen, Chao Ma, and Xiaokang Yang. Open-vocabulary remote sensing image semantic segmentation, 2024, 2409.07683.
- [37] Anna Josephson, Jeffrey D. Michler, Talip Kilic, and Siobhan Murray. The mismeasure of weather: Using remotely sensed earth observation data in economic context, 2024, 2409.07506.
- [38] Keumgang Cha, Donggeun Yu, and Junghoon Seo. Pushing the limits of vision-language models in remote sensing without human annotations, 2024, 2409.07048.
- [39] Xuexue Li. Insight any instance: Promptable instance segmentation for remote sensing images, 2024, 2409.07022.
- [40] Jingtao Li, Qian Zhu, Xinyu Wang, Hengwei Zhao, and Yanfei Zhong. Anomalycd: A benchmark for earth anomaly change detection with high-resolution and time-series observations, 2024, 2409.05679.
- [41] Fan Zhang, Lingling Li, Licheng Jiao, Xu Liu, Fang Liu, Shuyuan Yang, and Biao Hou. Renormalized connection for scale-preferred object detection in satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–23, 2024.
- [42] Carbon capture from natural gas flue emissions and air via (bi)carbonate formation in a cyclodextrin-based metal-organic framework, 2024.
- [43] High-resolution geospatial database: national criteria-air-pollutant concentrations in the contiguous u.s., 2016 – 2020, 2024.
- [44] Sergio Calvo-Ordoñez, Konstantina Palla, and Kamil Ciosek. Epistemic uncertainty and observation noise with the neural tangent kernel, 2024, 2409.03953.
- [45] Shubhada Agrawal, Prashanth L. A., and Siva Theja Maguluri. Markov chain variance estimation: A stochastic approximation approach, 2024, 2409.05733.
- [46] Akhilan Boopathy, Sunshine Jiang, William Yue, Jaedong Hwang, Abhiram Iyer, and Ila Fiete. Breaking neural network scaling laws with modularity, 2024, 2409.05780.
- [47] Akhilan Boopathy and Ila Fiete. Unified neural network scaling laws and scale-time equivalence, 2024, 2409.05782.
- [48] Michael Scheuerer, Claudio Heinrich-Mertsching, Titike K. Bahaga, Masilin Gudoshava, and Thordis L. Thorarinsdottir. Applications of machine learning to predict seasonal precipitation for east africa, 2024, 2409.06238.
- [49] Luping Qu, Mauricio Araya-Polo, and Laurent Demanet. Uncertainty quantification in seismic inversion through integrated importance sampling and ensemble methods, 2024, 2409.06840.
- [50] Lakshmi Jayalal, Gokularam Muthukrishnan, and Sheetal Kalyani. Tuning-free online robust principal component analysis through implicit regularization, 2024, 2409.07275.
- [51] Youguang Chen, Zheyu Wen, and George Biros. A scalable algorithm for active learning, 2024, 2409.07392.
- [52] Hang Li, Wei Jin, Geri Skenderi, Harry Shomer, Wenzhuo Tang, Wenqi Fan, and Jiliang Tang. Sub-graph based diffusion model for link prediction, 2024, 2409.08487.
- [53] Andrea Cavallo, Madeline Navarro, Santiago Segarra, and Elvin Isufi. Fair covariance neural networks, 2024, 2409.08558.
- [54] Shivam Pandey, Chirag Modi, Benjamin D. Wandelt, Deaglan J. Bartlett, Adrian E. Bayer, Greg L. Bryan, Matthew Ho, Guillemin Lavaux, T. Lucas Mäkinen, and Francisco Villaescusa-Navarro. Charm: Creating halos with auto-regressive multi-stage networks, 2024, 2409.09124.
- [55] Yuma Ichikawa and Koji Hukushima. Statistical mechanics of min-max problems, 2024, 2409.06053.
- [56] Ilya Rozenfeld. Causal analysis of shapley values: Conditional vs. marginal, 2024, 2409.06157.
- [57] Ingvar Ziemann. A short information-theoretic analysis of linear auto-regressive learning, 2024, 2409.06437.
- [58] Massimo Aria, Agostino Gnasso, Carmela Iorio, and Marjolein Fokkema. Extending explainable ensemble trees (e2tree) to regression contexts, 2024, 2409.06439.