Machine Learning Techniques to Separate the Cosmic from the Telluric

Frederick Dauphin[1*], Andreea Petric[1], Michelle Ntampaka[1], Swara Ravindranath[1], Jennifer Marshall[2], Étienne Artigau[3], Steven Businger[4], Laurie Rousseau-Nepton[5], Andrew W. Stephens[6], Takahiro Morishita[7]

[1]Space Telescope Science Institute, Baltimore, MD 21218,
[2]Texas A&M University, College Station, TX 77843,
[3]Université de Montréal, Montréal, Québec, Canada
[4]University of Hawaii, Honolulu, HI 96822,
[5]Canada-France-Hawaii Telescope, Waimea, HI 96743,
[6]National Optical-Infrared Astronomy Research Laboratory, Tucson, AZ 85719,
[7]California Institute of Technology, Pasadena, CA 91125

ABSTRACT

In the Roman era, wide-field, deep, visible-to-near infrared images will revolutionize our understanding of galaxy evolution (e.g. environments, morphologies, masses, colors). The legacy value of Roman images and low-resolution spectra (with Roman's prism and grism) will be greatly enhanced by massively multiplexed ground-based observations in the near – future and simultaneously allow us to leverage an impressive bounty of archived spectra from Maunakea facilities. We plan to enhance ground-based NIR spectra of astrophysically interesting objects with ground-sky spectra, atmospheric data, HST spectra and images, and machine learning techniques proven to predict galaxy spectra from images (Wu & Peek, 2020)[1].

Keywords: Sky Spectra, Canada-France-Hawaii Telescope, Machine Learning, Time Series Analysis

1. INTRODUCTION

1.1 Scientific Justification

The discovery efficiency of 8-10-m class facilities (e.g. Subaru's Prime Focus Spectrograph - PFS , the Maunakea Spectroscopic Explorer -MSE, the Fiber-Optic Broadband Optical Spectrograph on Keck - FOBOS) is an order of magnitude higher than any other spectroscopic capability currently realized. They will produce datasets equivalent in the number of objects to an SDSS Legacy Survey every few weeks, on a telescope with an aperture 20 times larger and at arguably the best astronomical site on the planet. Therefore, it is critical that we develop an efficient process to examine tens of millions of objects using spectroscopy.

1.2 Need for better than 1% sky subtraction accuracy

Most science cases for these massively multiplexed missions will rely on their ability to detect faint sources and/or measure the widths of narrow lines next to skylines. Both capabilities are dependent on our ability to subtract the contribution of the sky to the observed spectra. Several science cases drive the need for better than 1% subtraction. MSE-like facilities should not waste fiber-hours to look at the sky. Therefore, we need to be able to predict the sky at a specific fiber given observations of sky before, after, or using a small fraction of fibers. We propose to assess the feasibility of such an approach.

*fdauphin@stsci.edu

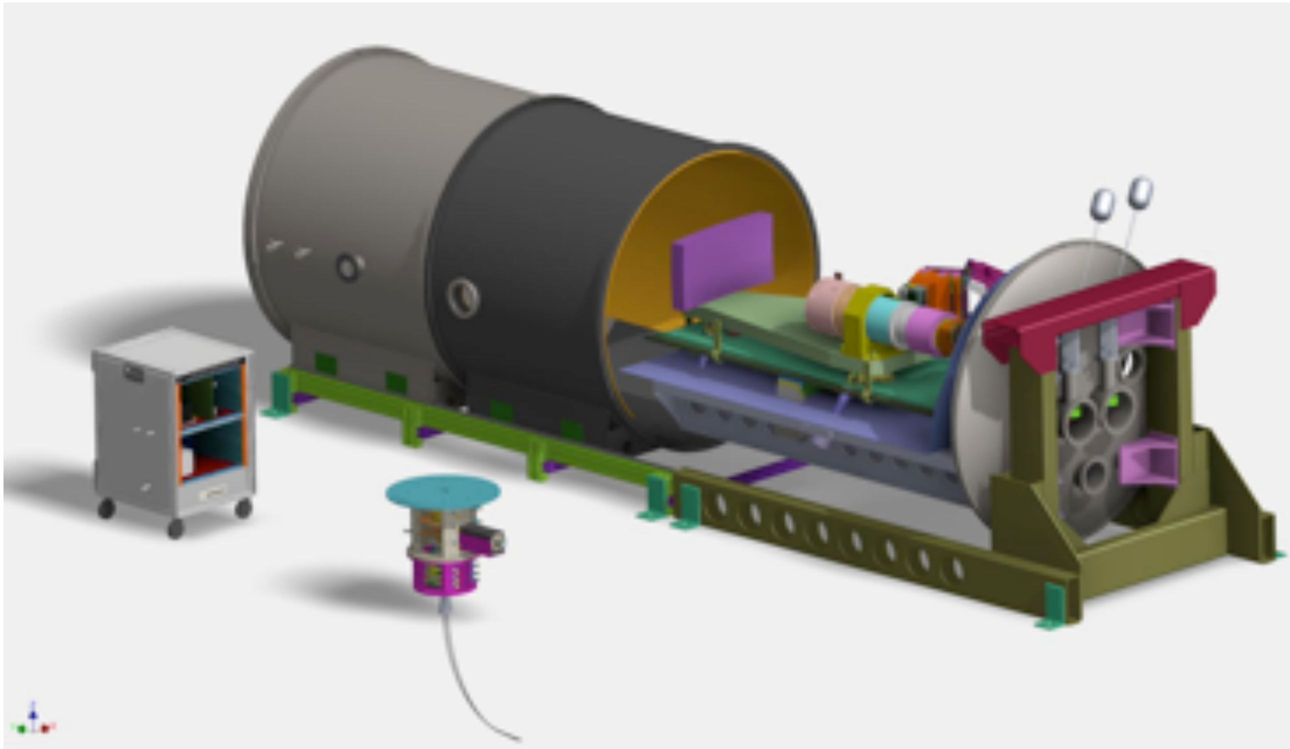


Figure 1: A diagram of the SpectroPolarimètre Infra-Rouge (SPIRou) aboard the Canada-France-Hawaii Telescope (Donati, 2018)[2]

2. DATA AND METHODS

We will focus on 1195 high-resolution (R~30,000) NIR sky spectra taken contiguously (with a cadence of 5 and 10mn) over 3 years with the SpectroPolarimètre Infra-Rouge (SPIRou) - a high precision spectropolarimeter and velocimeter on the Canada-France-Hawaii Telescope (CFHT) on Maunakea. We will then use atmospheric models employed by the ETCs of several facilities, atmospheric data including solar activity from the Maunakea Weather Center and use machine learning techniques to predict the evolution of sky continuum and lines between two sky observations separated by time scales. In addition, we will test this model on a set of archived spectra of non-variable point sources from Gemini's Near-Infrared Spectrograph (GNIRS) also observed by HST (e.g. spectro-photometric calibrators) and develop ways to refine the observed ground based spectra when both space and sky data are available.

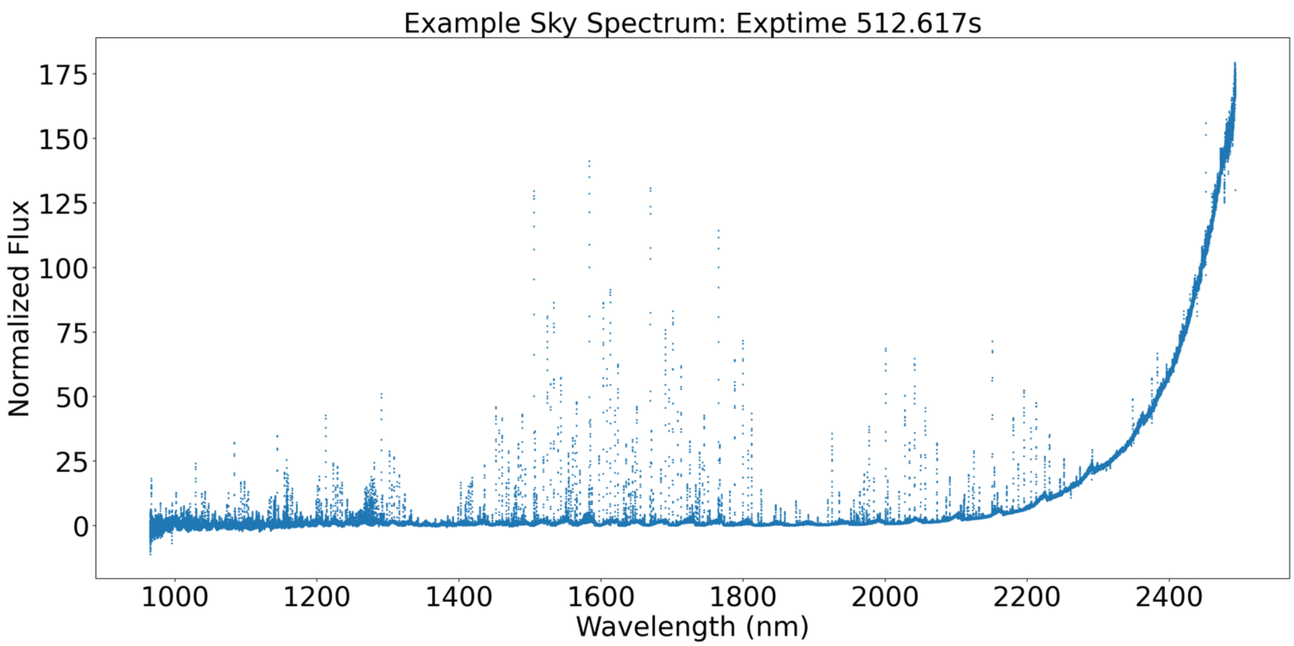


Figure 2: Example spectrum of the sky taken by SPIRou on 21-10-2021. The flux is normalized by the spectrum's median. The steep black body curve starting around 2100nm is the sum of the front end+fiber+telescope emission, plus a tiny bit from the Earth's atmosphere itself (e.g., thermal emission from dust in the line of sight). The spectrum is sampled with a pixel scale of 1 km/s. The 49 diffraction orders of SPIRou have been stitched together. The spectra also show low-level continuum between OH lines shortward of ~2.1µm which may be low-level scattered background within SPIRou. The background between OH lines is extremely faint, and mostly consists of zodiacal light and maybe some Rayleigh scattering from the Moon if it is above horizon. The Moon contribution falls really fast in the IR thanks to the λ^{-4} dependency of Rayleigh scattering.

We start by fitting 328 hydroxyl (OH) sky lines [3-6], which range from 1006.3nm to 2930.0nm and are normalized by the spectrum's median. Below is an example of a Gaussian fit to an OH line at 1290.6nm in the J band.

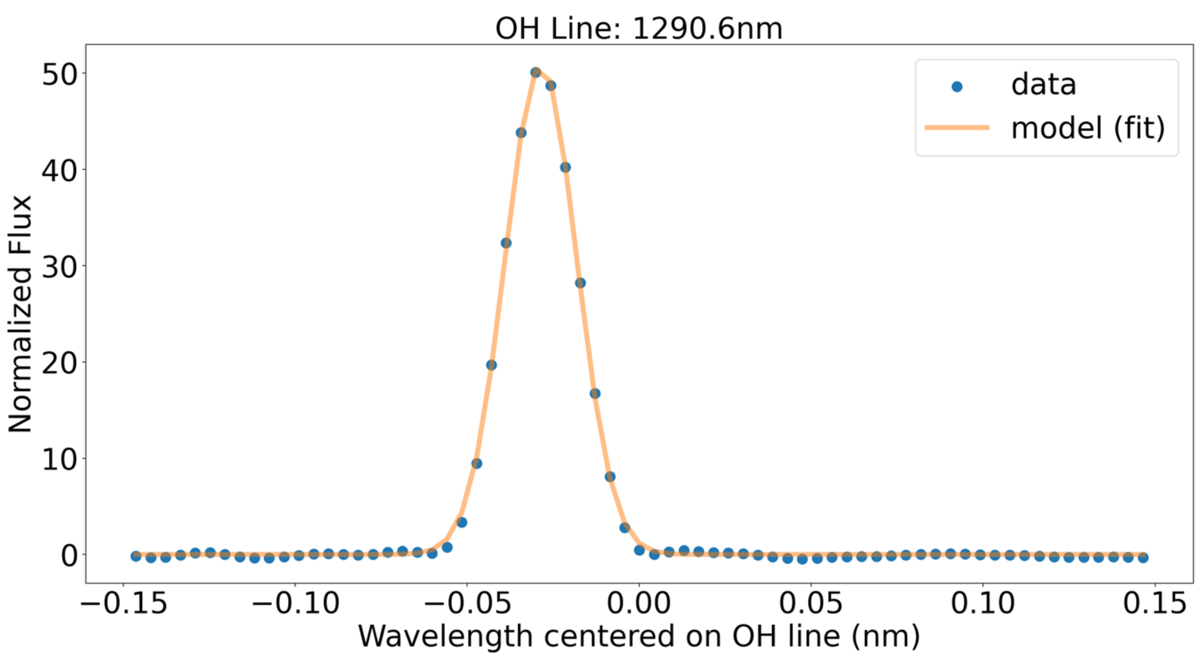
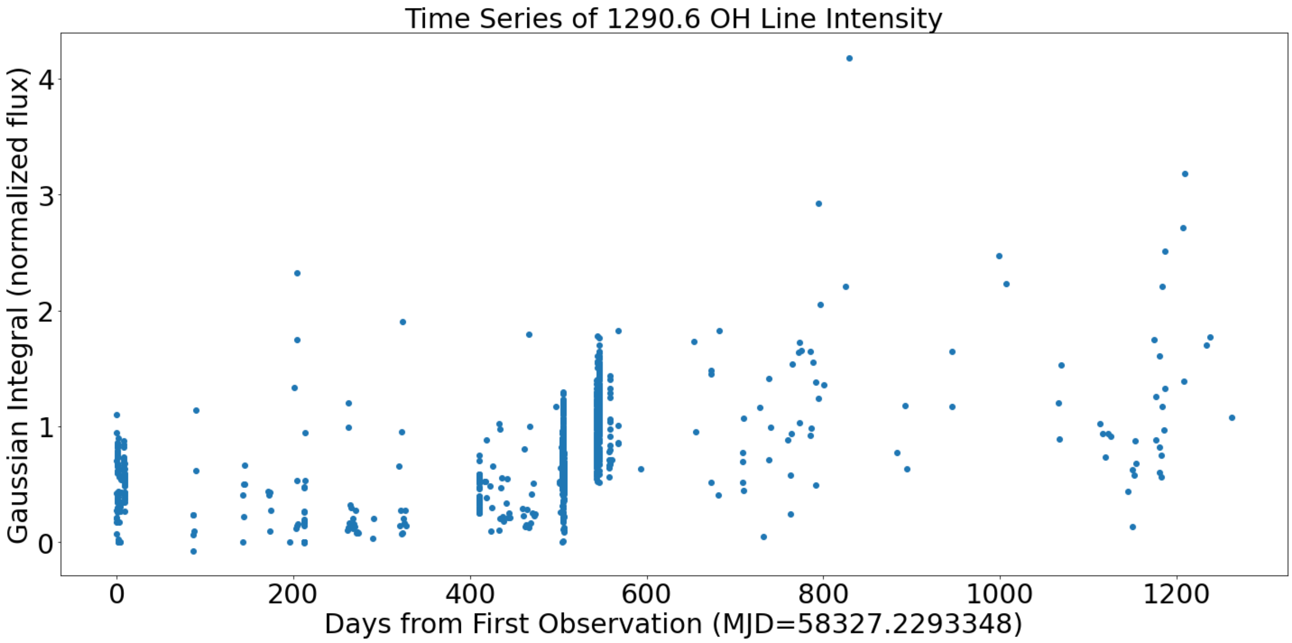


Figure 3: Gaussian fit of a OH line (1290.6 nm) from the example sky spectrum in Figure. 2. This line was described by Rousseleot et al. (2000) [3].

3. RESULTS

After measuring the total OH flux over a few years worth of spectra, we can visualize how the sky changes over time. There is a complex relationship between sky and time, which we aim to model using machine/deep learning algorithms.



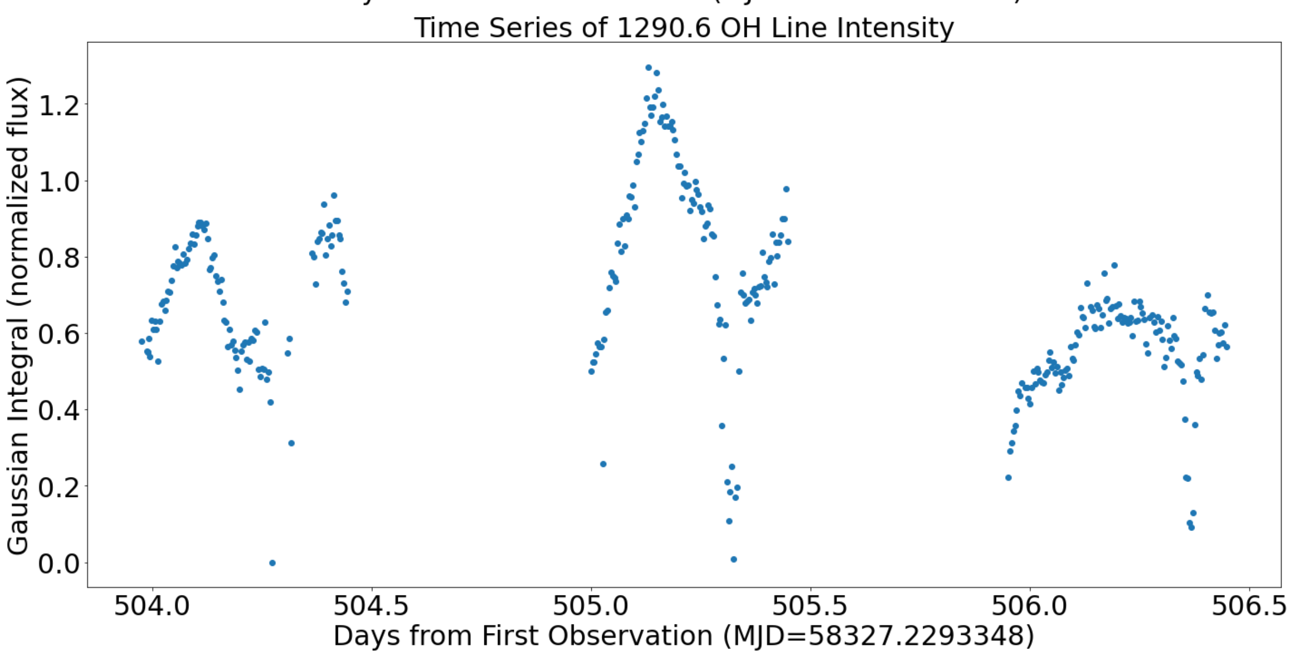


Figure 4: 1290.6 OH line intensity as a function of time. The top and bottom show measurements over ~3 years and ~3 days, respectively. Our first goal is to estimate the errors associated to interpolating sky properties over a range of time-scales, in different observing conditions.

4. CONCLUSIONS

It is critical over the new few years to develop precise sky subtraction to support future missions. We first use OH lines and OH line groups from high resolution R~30K NIR fiber spectra to estimate the typical sky variations on time-scales of minutes to hours. The relationship between sky and time is non-linear, making machine learning an immediate solution to model this variation for accurate sky subtraction.

REFERENCES

[1] Predicting galaxy spectra from images with hybrid convolutional neural networks, Wu, J.F. and Peek, J.E., 2020, *arXiv:2009.12318*

[2] SPIRou: a NIR spectropolarimeter/high-precision velocimeter for the CFHT, Donati, J.F., et al., 2018, *arXiv:1803.08745*

[3] Night-sky spectral atlas of OH emission lines in the near-infrared, Rousselot, P., Lidman, C., Cuby, J.-G., Moreels, G. & Monnet, G., 2000, A&A 354, 1134

[4] The OH airglow spectrum: a calibration source for infrared spectrometers, Oliva, E. and Origlia, L., 1992, A&A, 254, 466

[5] Observations of the OH airglow emission, Maihara, et al., 1993, PASP, 105, 940

[6] Non-thermal emission in the atmosphere above Mauna Kea, Ramsay, S.K., Mountain, C.M. & Geballe, T. R., 2002, *Monthly Notices of the Royal Astronomical Society*, *259*(4), pp.751-760