EPSC Abstracts
Vol. 7 EPSC2012-195 2012
European Planetary Science Congress 2012

Exoplanetary spectroscopy using unsupervised machine learning

Ingo P. Waldmann, Giovanna Tinetti (1) University College London (ingo@star.ucl.ac.uk)

Abstract

At this conference, we will present two novel approached to exoplanetary data analysis using unsupervised machine learning. Based on the concepts of Independent Component Analysis (ICA) we demonstrate the de-trending of exoplanetary photometric and spectroscopic data without prior or auxiliary knowledge of the star and the instrument. These types of deconvolution are often referred to as 'blind' techniques as they consider nothing but the data at hand. ICA is inherently limited to high signal-to-noise (SNR) regimes and quickly reaches its limitations at high Gaussian noise levels. Here we will present a natural extension of ICA using wavelet based approaches that make low-SNR data accessible to non-parametric deconvolution.

1. Introduction

The field of transiting extrasolar planets and especially the study of their atmospheres is one of the youngest and most dynamic subjects in current astrophysics. Permanently at the edge of technical feasibility, we are successfully discovering and characterising smaller and smaller planets. To study exoplanetary atmospheres, we typically require a 10^{-4} to 10^{-5} level of accuracy in flux. Achieving such a precision has become the central challenge to exoplanetary research and is often impeded by systematic (nongaussian) noise from either the instrument, stellar activity or both. Dedicated missions, such as Kepler, feature an a priori instrument calibration plan to the required accuracy but nonetheless remain limited by stellar systematics. More generic instruments often lack a sufficiently defined instrument response function, making it very hard to calibrate. In these cases, it becomes crucial to know how well we can calibrate the data without any additional or prior knowledge of the instrument or star.

2. Independent Components

Here we present a non-parametric machine-learning algorithm, based on the concept of independent component analysis (ICA), to de-convolve the systematic noise and all non-Gaussian signals from the desired astrophysical signal. Such a 'blind' signal de-mixing is commonly known as the 'Cocktail Party problem' in signal-processing. We demonstrate the importance and broad applicability of unsupervised machine learning in exoplanetary data analysis of space-based data by showing the removal of time-correlated stellar noise in individual lightcurves observed by the Kepler mission and the de-correlation of instrument systematics common to the Hubble space telescope. The latter is demonstrated in particular with a reanalysis of the Hubble/NICMOS transmission data originally published by Swain et al. (2008). The Swain et al. (2008) analysis heavily relies on auxiliary information of the instrument, so called optical state vectors (OSVs). Figure 1 shows the best-fit ICA result obtained by the blind source deconvolution in comparison with the originally published Swain et al. (2008) spectrum and the spectrum derived by Gibson et al. (2012) using Gaussian Processes. We can see that non-parametric and parametric analysis methods are in good agreement with each other and the science result consistent.

3. The wavelet approach

Furthermore we present a novel approach to the applicability of 'blind-source-separation' techniques in very low signal-to-noise (SNR) conditions often encountered in ground-based efforts of exoplanetary spectroscopy. Whilst such non-parametric deconvolution techniques usually find their limits in low SNR conditions, we present a natural extension of ICA using optimal wavelet filters.

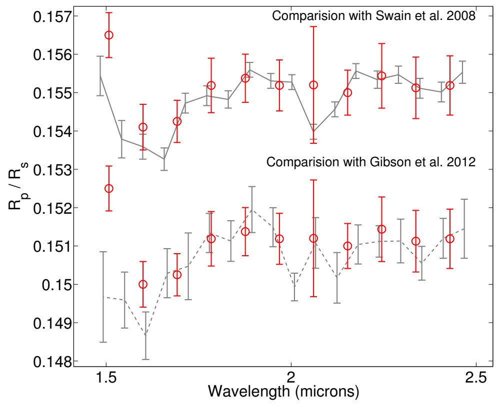

Figure 1: Analysis of the hot-Jupiter HD189733b primary eclipse spectrum obtained by the Hubble/NICMOS instrument. Red: spectrum derived using independent component analysis, compared to Swain et al. (2008) and Gibson et al. (2012). All results are in good agreement with one another and demonstrate the robustness of the exoplanetary spectrum.

4. Summary and Conclusions

At this conference we will present two novel ways of exoplanetary data analysis using unsupervised machine learning algorithms to de-trend data from stellar and instrumental systematics. At a required accuracy of 10^{-3} to 10^{-4} in flux, corrections using auxiliary information of the instrument are often degenerate with the science result for instruments that do not feature an *a priori* calibration plan at the required photometric stability. Using 'blind' deconvolution techniques allows us to de-trend data without any prior or auxiliary assumptions of the instrument with which it was taken and hence allows for an unbiased estimate of the exoplanetary signal.

References

[1] Waldmann, I. P., Tinetti, G., Deroo, P., 2012, ApJ, submitted

[2] Waldmann, I. P., 2012, ApJ, 747, 12

[3] Swain, M. R., Vasisht, G., Tinetti, G., 2008, Nature, 452, 329

[4] Gibson, N. P., Aigrain, S., Roberts, S., Evans, T. M.,Osborne, M.& Pont, F., 2012, MNRAS, 419, 2683