ON SIGNALS FAINT AND SPARSE: THE ACICA ALGORITHM FOR BLIND DE-TRENDING OF EXOPLANETARY TRANSITS WITH LOW SIGNAL-TO-NOISE

I. P. Waldmann
University College London, Gower Street, WC1E 6BT, UK; ingo@star.ucl.ac.uk
Received 2013 February 26; accepted 2013 October 24; published 2013 December 10

ABSTRACT

Independent component analysis (ICA) has recently been shown to be a promising new path in data analysis and de-trending of exoplanetary time series signals. Such approaches do not require or assume any prior or auxiliary knowledge about the data or instrument in order to de-convolve the astrophysical light curve signal from instrument noise. These methods are often known as “blind-source separation” (BSS) algorithms. Unfortunately, all BSS methods suffer from an amplitude and sign ambiguity of their de-convolved components, which severely limits these methods in low signal-to-noise (S/N) observations where their scalings cannot be determined otherwise. Here we present a novel approach to calibrate ICA using sparse wavelet calibrators. The Amplitude Calibrated Independent Component Analysis (ACICA) allows for the direct retrieval of the independent components’ scalings and the robust de-trending of low S/N data. Such an approach gives us an unique and unprecedented insight in the underlying morphology of a data set, which makes this method a powerful tool for exoplanetary data de-trending and signal diagnostics.

Key words: methods: data analysis – methods: statistical – techniques: photometric – techniques: spectroscopic

Online-only material: color figures

1. INTRODUCTION

As we explore smaller and smaller extrasolar planets around ever fainter stars, it is unsurprising that the need for more accurate data-calibration and de-trending techniques is growing. In the recent past, there has been a notable emergence of so-called non-parametric data de-trending algorithms in the fields of transiting extrasolar planets and time-resolved exoplanetary spectroscopy (Carter & Winn 2009; Thatte et al. 2010; Gibson et al. 2012; Waldmann 2012; Waldmann et al. 2013). The use of such “non-parametric” algorithms is a reactionary response to the difficulties of calibrating and de-trending time series observations when the instrument response function is not known at the precision of the science signal to be extracted. Previous, more conventional “parametric” de-trending approaches relied on the auxiliary information of the instrument (e.g., temperature sensor readings, telescope tilt, drifts in x and y positions of the science signal across the detector, etc.). Such methods have the disadvantage of being heavily reliant on the signal-to-noise (S/N) of the auxiliary information used to de-trend the science data, as well as suffering from a degree of arbitrariness in their definition of the instrument response function.

In Waldmann (2012) and Waldmann et al. (2013), we have demonstrated independent component analysis (ICA) as novel de-correlation strategy for exoplanetary time series. ICA belongs to the class of blind-source separation (BSS) algorithms that attempt to de-correlate an observed mixture of signals into its individual source components without prior knowledge of the original signals or the way they were mixed together. Such an approach requires the least amount of information on a given system and hence ensures a high degree of objectivity in the de-trending of data.

1.1. Limitations of Conventional ICA

Whilst it has been shown that ICA is well suited to the de-correlation of non-Gaussian signals in simultaneously observed exoplanetary time series, it has two major limitations, which will be addressed in this paper. These are as follows.

1. Susceptibility to Gaussian noise. To de-correlate non-Gaussian signals, ICA is inherently limited to a low degree of Gaussian white noise in the observed time series observations. So far, this has posed a significant limitation on the types of data that can be de-correlated. Waldmann et al. (2013) showed that medium- to high-S/N space-based observations are somewhat permissible but noisier ground-based observations of exoplanetary time series are often out of reach for conventional ICA algorithms.

2. Amplitude and sign ambiguity. Like all BSS algorithms, ICA can de-correlate signals up to an amplitude and sign ambiguity. As explained in Section 1.2, the algorithm attempts to simultaneously estimate the source signals, \( s \), as well as their respective mixings (the mixing matrix), \( A \), which represent our observations, \( x = A^{-1}s \). Given that both \( s \) and \( A \) are unknown, a scalar multiplication of either can be canceled by an equal division of the other. Hence no BSS algorithms attempt to retrieve the scalar amplitudes of the source signals \( s \). Waldmann et al. (2013) resolved this by iteratively fitting components of \( s \) onto observed out-of-transit (OOT) data to retrieve the lost scaling factors. Whereas this is a valid approach, it again limits us to high-S/N observations as too much scatter in the observed time-series inhibits a good convergence of such a scaling factor regression.

In this paper we will address both these limitations by defining ICA in orthogonal wavelet space. In the wavelet domain, as explained in later sections, we can threshold Gaussian wavelet coefficients and increase the signal’s sparsity, making the ICA algorithm more robust in difficult S/N conditions. We can furthermore inject a sparse wavelet coefficient calibration signal (CS), which allows us to directly calibrate the amplitudes of the mixing matrix \( A \) without the need of any post-ICA scaling factor regression.
Similar wavelet-ICA approaches have been studied in the context of medical physics (e.g., La Foresta et al. 2006; Sánchez et al. 2006; Inuso et al. 2007; Boroomand et al. 2007; Mammone et al. 2012) and engineering (e.g., Lin & Zhang 2005).

A quick introduction to BSS and wavelets are given in Sections 1.2 and 1.3, a description of the wavelet-ICA and noise thresholding in Section 2. Section 2.1 describes the amplitude calibration algorithm, which is demonstrated in Sections 3.1 and 3.2 using simulations and Spitzer/IRS data, respectively.

### 1.2. Blind-source Separation

Besides ICA, other BSS algorithms include principal component analysis, factor analysis, projection pursuit, non-negative matrix factorization, stationary subspace analysis, and morphological component analysis, among others. For an extensive review of these algorithms, we refer the interested reader to Comon & Jutten (2010) as well as relevant ICA literature (Oja 1992; Hyvärinen 1999; Hyvärinen & Oja 2000, 2001; Stone 2004; Koldovský et al. 2006; Yeredor 2000; Tichavský et al. 2008). Where the underlying statistical assumptions differ significantly, all these algorithms take $N$ simultaneously observed signals $x_k(t)$, where $k$ is the observed signal index, and de-correlate these into the source signals $s_l(t)$, where $l$ is the source signal index. They all follow the functional form

$$x_k(t) = a_{k,1}s_1(t) + a_{k,2}s_2(t) + a_{k,3}s_3(t) + \cdots + a_{k,l}s_l(t),$$

where $a_{k,l}$ are scaling factors. Assuming that the exoplanetary observation consists of a mixture of astrophysical signal, $s_1(t)$, instrument or stellar systematic noise, $s_2(t)$, and the white noise signal, $s_3(t)$, from a Gaussian process $w(t)$, we can express Equation (1) as the sum of vectors (the time-dependence has been dropped for clarity):

$$x_k = a_{k,1}s_1 + a_{k,2}s_2 + a_{k,3}s_3 + \cdots + a_{k,l}s_l,$$

where $N_m$ is the number of systematic noise sources. Finally this can also be expressed as column vectors $x = [x_1(t), x_2(t), \ldots, x_k(t)]^T$, $s = [s_1(t), s_2(t), \ldots, s_l(t)]^T$ and the mixing matrix $A$,

$$x = As.$$  

We can furthermore define the “de-mixing matrix” $W$ as the inverse of the “mixing matrix”

$$W = A^{-1}.$$  

For a perfect de-correlation of the observed data $x$ into its source components, the original mixing matrix is the perfect inverse of the estimated de-mixing matrix $W$ and $WA = I$, where $I$ is the identity matrix.

ICA attempts to estimate both $s$ and the de-mixing matrix $W$ by assuming that all components of $s$ are statistically independent of one another. This is achieved by iteratively maximizing the non-Gaussianity of each signal component by estimating their respective Shannon entropies (Shannon 1948; Hyvärinen & Oja 2000). For more information, we refer the reader to Waldmann (2012) and the standard literature.

### 1.3. Introduction to Wavelets

Readers familiar with wavelet decompositions may jump to Section 2.

Similar to a Fourier transform (FT), the wavelet transform (WT) decomposes a given time series signal into its frequency components. Where the FT uses sine and cosine functions that extend over the full range of the data, the WT uses highly localized impulses. These impulses or “wavelets” scan through the time series and, much like a tuning fork to an instrument, “resonate” with localized features in the time series that are akin to the wavelet’s shape and scaling. The individual wavelet basis functions are derived from a single mother wavelet $\psi(t)$ through translation and dilatation of the mother wavelet (Percival & Walden 2000). Different wavelets exist with different analytical properties, here we use the orthogonal basis wavelets of the Daubechies (dB) family (Daubechies 1988). The wavelet analogue to the FT of the times series $x(t)$ is given by

$$c_{\tau,\varphi} = \int_{\mathbb{R}} x(t)\psi_{\tau,\varphi}(t)dt,$$

where $\psi_{\tau,\varphi}(t)$ is called the “mother wavelet” for a given scaling $\varphi$ and translation $\tau$ and $e$ is the wavelet coefficient with respect to $\tau$ and $\varphi$. We define the mother wavelet for the continuous wavelet transform (CWT) as

$$\psi_{\tau,\varphi} = \frac{1}{\sqrt{2}} \psi \left( \frac{t - \tau}{\varphi} \right).$$

The wavelet base is orthogonal, which means we can reconstruct the data by taking the sum of the product of all coefficients for a given scale and translation, $c_{\tau,\varphi}$, with the respectively scaled and translated mother wavelet

$$x(t) = \sum_{\varphi \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} c_{\tau,\varphi} \psi_{\tau,\varphi}(t).$$

For a more in-depth definition of wavelets and their respective properties we refer the interested reader to Daubechies (1992) and Percival & Walden (2000).

#### 1.3.1. Multi-resolution Analysis

The above equations apply to the CWT case. The wavelet coefficients describe the correlation between the wavelet at varying scales (or frequencies). These can be calculated by changing the scale of the wavelet (i.e., the analysis window). Thus, we can speak of a multi-resolution decomposition, where each scaling of the mother wavelet denotes a given resolution. Here, the analogy to the FT would be band-pass filters of varying size. It is often more sensible to exploit the discrete nature of the data and to define the orthogonal discrete wavelet transform (DWT). The DWT is significantly easier to implement and faster to compute. Similarly to the continuous case, in the DWT we have a “mother” wavelet and a scaling function, also known as the “father” wavelet. Here, the “mother” wavelet is denoted by $h(t)$ and the “father” by $g(t)$ (Daubechies 1992; Percival & Walden 2000; Press et al. 2007). It is important to note that unlike in the CWT case, where the “mother” wavelet itself is scaled to represent different frequencies in the data, this is not the case in the DWT. In the DWT, in analogy with the FT, $h(t)$ and $g(t)$ can be thought of as high-pass and low-pass frequency filters, respectively. Different scalings are then achieved by progressively “down-sampling” the data.

The DWT is best understood by following the individual steps of the algorithm that compute the transform.
1. The observed, discrete time series, \(x(t)\), is convolved with the “mother” wavelet \(h(t)\)

\[
cD_\phi(t) = (x * h)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau) \cdot h(t - \tau),
\]

where \((x * h)\) denotes the convolution of \(x\) with \(h\) and \(cD_\phi\) represents the “mother” wavelet coefficients for a given scale \(\phi\). As mentioned earlier, the “mother” wavelet, \(h(t)\), acts as a high-pass filter, sensitive to the high frequencies or details of the time series. We hence refer to the coefficients of \(h(t)\) as detail coefficients.

2. The next step is to convolve the same time series, \(x(t)\), with the scaling function or “father” wavelet

\[
cA_\phi(t) = (x * g)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau) \cdot g(t - \tau).
\]

As opposed to the “mother” wavelet, the “father” wavelet acts as a low-pass filter of the time series, and its coefficients can be viewed as a moving average of the underlying trend of \(x(t)\). We hence refer to its coefficients as average coefficients and denote them with \(cA_\phi\). Furthermore, the low-pass filter \(g(t)\) is related to the high-pass filter by

\[
g(L - 1 - t) = (-1)^L \cdot h(t),
\]

where \(L\) is the filter length and corresponds to the number of points in the time series \(x(t)\).

3. We now have two sets of time series, a low-pass filtered, moving-average time series, \(cA_\phi\), and a high-pass filtered time series, \(cD_\phi\). We record the \(cD_\phi\) coefficients and proceed with our analysis of the average coefficients, \(cA_\phi\). Because half of the frequencies in \(cA_\phi\) (namely, the high-pass ones) have been removed by Equation (9), the Nyquist theorem tells us that we are oversampled by a factor of two. We can hence remove every second coefficient in \(cA_\phi\) without losing information. This operation is termed “down-sampling” and abbreviated by \(\downarrow 2\), leaving us with \(N/2\) coefficients to describe \(cA_\phi\). A similar process applies to the detailed coefficients \(cD_\phi\).

4. The Nyquist down-sampling introduces the concept of scaling or multiple resolutions. If we repeat steps 1–3 on the down-sampled average coefficients, \(cA_\phi\), we obtain a new set of coefficients \((cA_{\phi+1}\) and \(cD_{\phi+1})\) on a scale that is double the size of the previous decomposition. This is illustrated in Figure 1 as a flow chart and a data example is given in Figure 2.

For a given scale, \(\phi\), the data can now be reconstructed by reversing the above process

\[
x_{\phi}(t) = (cA_{\phi=0}(\phi) \cdot g(-t + 2\tau)) + \sum_{\phi} \sum_{\tau=-\infty}^{\infty} (cD_{\phi} \cdot h(-t + 2\tau)),
\]

where \(\Phi\) is the total number of decompositions.

For an algorithmic implementation using quadrature mirror filters (QMFs) see the Appendix and Press et al. (2007).

2. WAVELET ICA

We now perform the DWT on each observed time series, \(x_k\), and obtain a series of average coefficients, \(cA_k\), and detail coefficients for a given scale, \(cD_{k,\phi}\). For our ICA decomposition, we use these series of coefficients instead of the time domain series, \(x_k\), and define our observed data as

\[
\hat{x}_k = \sum_{\tau} cA_k(\tau) + \sum_{\psi} \sum_{\tau} cD_{k,\phi}(\tau)
\]

and similarly we can express our source signals, \(s_l\), as the wavelet equivalent

\[
\hat{s}_l = \sum_{\tau} cA_l(\tau) + \sum_{\psi} \sum_{\tau} cD_{l,\phi}(\tau).
\]

From here onward we will use \(\hat{x}\) to denote the wavelet domain presentation of the time-domain signal \(x\). We hence restate Equation (3) as

\[
\hat{x} = \hat{\Lambda} \hat{s}.
\]

There are several important considerations to note here.

As we are dealing with a multi-resolution analysis, it becomes possible to perform the BSS on individual scales (or bandpasses) only or to actively exclude some frequency ranges from the analysis. Such an approach may be advantageous if one has prior knowledge of the signals’ frequency bandwidths and wishes to restrict the impact of high-frequency noise or other signals on the BSS of a given signal. An example of this is given by Lin & Zhang (2005). In this paper, we do not take this approach for reasons described in Section 2.1.

Alternatively, rather than excluding individual scales, \(\psi\), we can use individual bandpasses as inputs to the ICA algorithm (i.e., \(\hat{x}_{k,\phi}\)). This increases the redundancy of the data and can be advantageous in the case of an over-complete ICA, where more source signals are present in the data than the time series observed \((N_{\text{observations}} < N_{\text{signals}})\). This over-completeness leads to an improper separation of the independent source signals in the data. In the case of the systematic noise being constrained to specific bandpasses we can increase the redundancy of the data by taking individual levels, \(\psi\), of the DWT as input to the ICA (i.e., \([N_{\text{observations}} + N_{\text{scales}}] > N_{\text{signals}}\)). This has been shown to alleviate the problem of over-completeness and allows for the efficient use of ICA in very restricted data ranges (Inuso et al. 2007; Mammone et al. 2012).
Figure 2. Example of a discrete wavelet transform (DWT). Top: sinusoidal time series with Gaussian noise and saw-tooth functions superimposed. Bottom: four level DWT decomposition of a noisy sinusoid using symlet-5 wavelets. It can be seen that the average coefficients, \( c_A \), represent a “moving average” of the data, whereas the detailed coefficients, \( c_{D\phi} \), represent bands of higher frequencies. (A color version of this figure is available in the online journal.)

Furthermore, as pointed out in Section 1.1, ICA is inefficient in the presence of high-frequency scatter. This is alleviated in the wavelet space, as, by the virtue of the multi-resolution analysis, most high-frequency scatter is contained in the \( c_{D1} \) coefficients and a better degree of separation can be achieved for lower frequency systematics. We will demonstrate this property in Section 3.2. It is also possible in wavelet space to selectively suppress Gaussian noise via soft or hard thresholding (e.g., Stein 1981; Donoho 1995) and thus increase the robustness of the BSS algorithm in low S/N conditions.

2.1. Amplitude Calibrated ICA (ACICA)

Arguably the central problem with an exoplanetary data detrending algorithm based on ICA is the scale and sign ambiguity of the de-correlated, independent components (ICs).

While we do not know the scaling of individual source components in \( s \), we know by definition of the ICA pre-processing step (whitening) that each source signal is normalized to unit variance, \( E[s_i^2] = 1 \), and that \( x = As \). Hence the elements of \( A \) do not need to preserve the absolute, but the relative scalings of \( s \). If we know the absolute scaling of one source signal, we can then solve the scaling degeneracy of all signals. This is usually possible for simulations where we control the inputs, but not for real life examples where the source signals and their scalings are a priori unknown. One way to solve this predicament is via the introduction of a CS. Such a CS must have the following properties.

1. **Minimal to no impact on the data.** The introduction of a CS should not distort any underlying signals or their amplitudes.
2. **Temporal localization.** The CS should not be located in in-transit (INT) regions and not take up too much OOT data.
3. **Stability to noise.** The CS should not be affected by the noise of the data (be it Gaussian or otherwise).
4. **Non-Gaussianity.** The signal should be non-Gaussian enough for the ICA to recognize it, but not more non-Gaussian than the science signal itself. A too prominent CS could bias the ICA toward the retrieval of the CS and could impair the retrieval of other non-Gaussian signals.

In the time domain, it is difficult to implement property 4 without violating property 1. In order to implement a distinct enough non-Gaussian signal in the time domain, one needs to significantly alter large sections of the data. In addition, the treatment of noise (property 3) is problematic. In the frequency domain, property 2 is violated, as the FT does not contain temporal information and the CS would superimpose the science signal. However, all four criteria can be met in the wavelet domain.

2.1.1. Injecting the Calibration Signal

In the wavelet domain, we can introduce a much sparser CS than in the time domain and have control over its temporal location (unlike in the frequency domain) via the lag term \( \tau \).
Given that \( \hat{x} \) is more redundant than \( x \), we can minimize the impact on the observed data or avoid any alterations to the data altogether by selecting wavelet coefficients of \( \hat{x} \) with zero or near-zero amplitudes to be used for our CS. We can now re-define Equation (2) as

\[
\hat{x}^c = \hat{x} + \hat{b} \hat{x} + \sum_{i=2}^N \hat{a}_{k,i} \hat{s}_i,
\]

where \( \hat{x}^c \) denotes the observed data with CS, \( \hat{x} \) is the CS with the same dimensions as \( s_k \), and \( \hat{b} \) is a random but known scaling constant. We define \( \hat{s}_i \) as

\[
\hat{s}_i = \begin{cases} 
\frac{cD_\phi(\tau) = 1}{\sum_{\tau} cA(\tau) + \sum_{\tau} cD_\phi(\tau)} & \text{if } \tau = \tau^c \\
0 & \text{otherwise},
\end{cases}
\]

where \( \tau^c \) are pre-selected lags for a given scale that correspond to a section of the OOT data. Note the following.

1. No average coefficients, \( cA \), are used in \( \hat{s} \), because these are not redundant enough.
2. As Equation (17) suggests, only \( N \) number of ICs can be extracted. Adding the CS in the observed data \( \hat{x} \) may render the ICA over-complete as one source signal will not be retrieved in favor of the CS. For large data sets \( (N_{\text{observations}} > N_{\text{signals}}) \) this is generally not a problem.
3. We choose \( 0 < \hat{b} < (1/2) \max |\hat{s}_k| \) to avoid having the CS be the most dominant feature in the data.
4. In Equation (18) we have chosen to define \( \hat{s}_i \) to contain one non-zero coefficient per scale \( \phi \) and lags corresponding to the same area of OOT data in the time-domain. This allows for efficient use of OOT data and a high sparsity in the wavelet domain, but is entirely arbitrary otherwise.

2.1.2. Retrieving the Scaling Information

Having injected the CS into our observations, we can now perform the ICA deconvolution (Section 1.2) on the data with CS in the wavelet domain, \( \hat{x}^c \). We identify the CS in the retrieved source signals and denote their respective elements of the mixing matrix, \( A \) as \( \hat{a}_{k,i} \). By measuring the average amplitude of the wavelet coefficients comprising the retrieved CS, \( \langle \hat{s}_i \rangle \), and knowing the original CS amplitude, \( \hat{b}_k \),

\[
\langle \hat{s}_i \rangle = \sum_{\phi} \frac{\sum_{\tau} cD_\phi(\tau)}{\Phi},
\]

we can retrieve the scaling of the CS as well as those of the other signals contained in \( \hat{s} \). We denote this calibrated mixing matrix as \( \hat{G} \) with its elements given by

\[
\hat{a}_{k,l} = \frac{\hat{a}_{k,l}}{\hat{s}_i} \times \langle \hat{s}_i \rangle.
\]

2.1.3. Calibration Error

Using the above calibration approach we consider the total error on the ICs to be a combination of source separation error (SSE) of the individual IC and the SSE of the CS components, \( s_i \). The SSE for the \( f \)th source signal can be estimated by

\[
\sigma_f = E \left[ \sum_{i=1}^N \frac{G_{kl}^2}{E[G_{kl}]} \right], \quad k, l = 1, 2, \ldots, N.
\]

where \( G \) is called the “gain matrix” and defined as \( G = WA \). For perfectly separated sources, we now have \( G = WA = I \), where \( I \) is the identity matrix. The ICA algorithms employed here (EFICA and WASOBI; Koldovsky et al. 2006; Yeredor 2000; Tichavský et al. 2006b) can be shown to be asymptotically efficient and converge to the correct solution in the limit of \( N_{\text{iter}} \rightarrow \infty \) iterations. However, in real world scenarios this is not the case and we find that the estimated mixing (or de-mixing) matrix is only approximately equal to the true underlying mixing matrix \( A \) (i.e., \( W \approx A^{-1} \)). To estimate the SSE, we can inspect the variance of the matrix \( G \). Whilst it is possible to directly calculate \( G \) for simulations, the true mixing of the signals is usually unknown in real data applications. Tichavský et al. (2006a, 2008) and Koldovsky et al. (2006) have shown that an asymptotic estimate of \( G \) is nonetheless possible. For a derivation we refer the reader to the cited literature.

We define the SSE for the calibrated components in \( s^c \) as the quadrature error of the individual source components’ error and the SSE of the CS

\[
\sigma^c_{f \phi} = \sqrt{\sigma^2_{f \phi} + \sigma^2_c}.
\]

3. APPLICATION EXAMPLES

3.1. Simulations

In this section we illustrate the ICA calibration approach using simulations.

1. For this we generated five input signals: a Mandel & Agol (2002) secondary eclipse curve of HD189733b, a Gaussian noise at with FWHM of \( 2 \times 10^{-3} \) normalized flux.

![Figure 3. Input signals before mixing. From top to bottom: three systematic noise components of the HST/NICMOS instrument (Waldmann et al. 2013); a Mandel & Agol (2002) lightcurve of the secondary eclipse of HD189733b; Gaussian noise at with FWHM of \( 2 \times 10^{-3} \) normalized flux.](http://www.mathworks.co.uk)
for four scales (Φ = 4) of each time series in x to obtain \( \hat{x} \) (Figure 5, black and blue curves). In the scope of this simulation we found the choice of wavelet and number of vanishing moments (VMs) not to matter greatly. However, the choice of wavelet and decomposition depth may vary from data set to data set.

4. We generated the CS, \( s_k \), to consist of one wavelet coefficient per scale \( \Phi \) giving \( \Phi \) the total number of non-zero coefficients. For each series, \( \hat{x}_k \), the CS was multiplied with the random scaling factor \( b_k \) (i.e., \( s_{\hat{x},k} = b_k s_k \)). Figure 5 shows the scaled CS for each \( \hat{x}_k \) (red peaks). Note that we chose the lags of the non-zero coefficient to correspond to the 95th percentile of each scale. This guarantees the CS to be located in the post-eclipse OOT data. It also overlaps the differently scaled wavelet impulses and hence minimizes the impact on the time-domain data representation. See Figure 6 for a time-domain representation of \( x_{k=3} \) (third from top in Figure 5).

5. The BSS via ICA as described in Waldmann (2012) was performed and the time-domain representation of the retrieved ICs are shown in Figure 7.

6. Following Section 2.1.2 we calculate the scaled mixing matrix \( O \) and apply the scalings to individual source signals. Figure 8 shows the observed data \( x_{k=3} \) from Figure 6 (blue circles) and the re-constructed data using the scaling matrix \( O_{k=3,l} \) (green crosses). The scaled ICs are shown offset underneath.

3.2. Spitzer/IRS: HD189733b Secondary Eclipse

We test the proposed algorithm on Spitzer/IRS (Houck et al. 2004) observations of a secondary eclipse of HD189733b. These observations were obtained in 2006 November (program ID: 30473) in low resolution mode ranging from 7.46 to 14.29 \( \mu \)m. The secondary eclipse was followed for 5:48 hr with integration times of 14.7 s per ramp. The Spitzer pipeline
calibrated data were reduced using the Spice2 spectral reduction software. Examples of the resulting time series are shown in Figures 9 and 10 (blue crosses). Similar to Section 3.1, we take these time series as inputs to our algorithm. Figure 11 shows the DWT, using db4 wavelets, of the time series at 7.6971 μm (black lines) with the injected CS over-plotted in red. Note that no binning in wavelength was performed before, which marks a significant difference to the HST/NICMOS analysis in Waldmann et al. (2013), where a relatively coarse binning was necessary to reduce the Gaussian noise.

We now follow each individual step as described in the previous sections and obtain the scaled ICs of the data set. Figures 9 and 10 show two observed time series (blue crosses) with the correctly scaled individual ICs (black dots) underneath. The first and second ICs comprise the secondary eclipse signal and the CS, respectively. These are also shown in Figure 12.
remaining ICs are instrument or stellar systematic noise. The amplitudes of these systematic components can be seen to be lower toward the blue end of the spectrum (Figure 9), which is also evident in a cleaner observed time series, and more pronounced toward the red end of the spectrum (Figure 10). In Figure 12 we overplotted the best fit Mandel & Agol (2002) model using a MCMC fitting routine (Waldmann et al. 2013; Haario et al. 2006) with the transit depth as only free parameter. The transit depth posterior distribution is shown in Figure 13. The total error is the quadrature sum of $\sigma_{\Delta}$ and the MCMC derived standard error $\sigma_{mc}$. We obtain at 7.6971 $\mu$m and 11.6285 $\mu$m a planet/star contrast ratio of $F_p/F_s = 4.15 \times 10^{-3} \pm 1.5 \times 10^{-3}$ and $F_p/F_s = 2.60 \times 10^{-3} \pm 1.5 \times 10^{-3}$, respectively, which we find consistent with the measurement by Grillmair et al. (2007). The computation and interpretation of the full spectrum is outside the scope of this publication and will be discussed in I. P. Waldmann et al. (in preparation). Figure 12 also shows the retrieved CS (black dots) with the original CS overplotted in red. The excellent match between both indicates an adequate signal separation and the correct scaling of the ICs.

In order to demonstrate the increased efficiency of the proposed ACICA algorithm in the presence of noise, we also show the ICs derived by performing the more “traditional” ICA in the time domain only (Figure 14). The components in Figure 14 are poorly separated and the standard ICA analysis did not converged with traces of the secondary eclipse feature present in three separate components.

4. DISCUSSION

Using ACICA, we can directly retrieve the signal amplitudes of the source components comprising our data. It not only allows us to non-parametrically de-trend low S/N data sets, but also allows for a unique insight into the structural make-up of the observations. A study of the systematic components and their amplitudes offers a powerful diagnostic on systematic noise behavior across a detector. As shown in the Spitzer/IRS example, amplitudes of systematic components in the data can vary significantly across the dispersion of a grism. A study of these systematic components may allow, among others, for an optimization of residual flat-fielding errors across a field or the characterization of wavelength dependent slit-losses of the instrument. As data diagnostic we can thus define the S/N of a time series $k$ in terms of its systematic noise,

$$S/N_k = \frac{\Delta F_k}{\sqrt{\sum_{l,k=1}^{N_v} \sigma^2 (s^c_{l,k})}}$$

where $\sigma^2 (s^c_{l,k})$ is the variance of a given systematic noise component for the $k$th time series and $\Delta F_k$ is the respective measured transit depth.

There is a great wealth of orthogonal (and biorthogonal) wavelet bases existing in the literature that can be used for our DWT. The most popular choices are the Haar wavelet (or db1), the Daubechies, Symlets, and Mexican Hat series. Although the ACICA algorithm can be run with any of these wavelets, the choice of the wavelet basis can be important given the morphology of the data.

Should the data’s systematics be dominated by flux offsets in the time series, such as those often induced by nodding (e.g., Grillmair et al. 2007), we find a Haar wavelet to give the sparsest representation of the data due to its intrinsic “box shaped” morphology. However, we note that we found this “box” shape morphology to be of little use to directly describe the lightcurve feature itself. Here smoothness induced by limb-darkening or the intrinsic trapezoidal shape of eclipse curves render Haar wavelets less efficient than other transforms.

For data sets that are dominated by smoothly varying trends we find high order Daubechies wavelets or symmetric wavelets (Symlets) to yield the sparsest representation of the data. For noisy data we find a Daubechies-4 (db4) wavelet base to be the best compromise between slowly varying continues functions and step-like, discrete offsets in the data. Out of the various wavelet sets available, the Daubechies series is the most versatile as it stretches from the Haar wavelet, a Daubechies with one VM, to very smooth wavelets with a high number of VMs. The number of VMs should be kept above four to describe a degree of

---

3 Vanishing moments (VMs) are statistical moments of the wavelet series that tend to zero (e.g., a wavelet with one VM has zero mean). The number of VMs determines the order of a polynomial a wavelet can describe exactly (i.e., a db4 wavelet contains two VMs and can perfectly describe all polynomials up to second order, or a db8 can describe a fourth-oder polynomial).
smoothness, but it is otherwise arbitrary. A compromise between
smoothness and number of coefficients is usually sought (e.g.,
Carter & Winn 2009).

Furthermore, the effect of autocorrelative noise (e.g., corre-
related Gaussian noise) can manifest itself as a weak correlation
in wavelet coefficients. This effect can be amplified or diminished
by the choice of wavelets, depending on whether the user wants
to extract this autoregressive structure or not.

The choice of wavelet family can be an important factor in
the convergence properties of the code, as an increased sparsity
aids the convergence of the ICA algorithm. In the absence of
any information on the data, one can choose the optimal wavelet
base by inspecting the SSE (Equation (21)) for various wavelets.

As mentioned in Section 2, Gaussian noise can be suppressed
in wavelet space by the selective thresholding of detailed
coefficients. While true in theory, we found it difficult to
implement in practice and great care needs to be taken in
choosing the optimal wavelet by the selective thresholding of
detailed Gaussian noise) can manifest itself as a weak correlation
in wavelet space alone and further cleaning was (in the scope of
this analysis) unnecessary.

5. CONCLUSION

In this paper we present a novel approach to the amplitude
calibration of ICA (the ACICA algorithm) via the introduction of
a sparse CS in orthogonal wavelet space. By transforming the
observed time series into a series of wavelet coefficients, we are
able to overcome two main limitations of BSS algorithms such as
ICA: (1) the convergence of ICA algorithms is strongly impaired
by the presence of Gaussian noise in the data. By transforming
data to a sparser and more redundant basis we can significantly
improve the performance of the ICA de-convolution without
otherwise altering the data. (2) In the wavelet domain it is
possible to inject an artificial CS of known amplitude into the
data without significant or any impact to the original data.

With the use of such a CS we can directly determine the
individual absolute amplitudes of the derived ICs. This marks a
significant improvement over the methods that were discussed
in Waldmann et al. (2013), where component amplitudes were
derived through a regression analysis to existing OOT baseline
data.

ACICA allows us to de-trend otherwise inaccessible data sets
non-parametrically. We demonstrated this using simulations and
archival Spitzer/IRS data. It furthermore offers us an
unprecedented and unique insight into the morphology of a data
set by allowing us to directly map out temporal/wavelength
dependent variations of instrumental or stellar noise in the data
set.

Together, these attributes make the algorithm proposed here
a highly versatile and powerful tool for exoplanetary time series
analysis.

APPENDIX

THE QUADRATURE MIRROR FILTER

A very fast and simple implementation of the DWT for multi-
resolution decomposition is by constructing the quadrature
mirror transformation matrix. For a Daubechies-4 wavelet we
obtain four coefficients comprising \( h(t) \) and similarly \( g(t) \).

Rather than convolving the time series, \( x(t) \) with both filters
separately and then down-sample, we can also construct a matrix
where each odd row contains \( h(t) \) and each even row contains
\( g(t) \) coefficients. This automatically down-samples the data to
the new resolution \( q + 1 \). Such a matrix is called a QMF and
Equation (A1) is an example of such (Press et al. 2007),

\[
\begin{bmatrix}
c_0 & c_1 & c_2 & c_3 \\
c_3 & -c_2 & c_1 & -c_0 \\
c_0 & c_1 & c_2 & c_3 \\
c_3 & -c_2 & c_1 & -c_0 \\
\end{bmatrix}
\]

where the empty spaces denote zeros. To obtain a DWT
using this QMF, we multiply the QMF with the column vector
containing the time-series on the right. Note the circular
behavior of the matrix at the bottom, where the wavelet
coefficients wrap around to the beginning. This has important
consequences as it indicates that the DWT wraps around
the data and the transform at the end of the time-series is sensitive
to data at the beginning of the time series. This effect can be
avoided by adding sufficient zero-valued points to the time series
at its beginning and end. This process is also known as “zero-
padding.”

REFERENCES

Boroomand, A., Ahmadian, A., Oghabian, M. A., Alirezaie, J., & Beckman,
C. 2007, in IEEE International Symposium on Signal Processing and
Information Technology, ed. A.-J. van der Even, 408

Carter, J. A., & Winn, J. N. 2009, ApJ, 704, 51

Comon, P., & Jutten, C. 2010, Handbook of Blind Source Separation: Indepen-
dent Component Analysis and Applications (New York: Academic)

Daubechies, I. 1988, CPAM, 41, 909

Daubechies, I. 1992, Ten Lectures on Wavelets (SIAM)

Donoho, D. L. 1995, ITIT, 41, 613

Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683

Grillmair, C. J., Charbonneau, D., Burrows, A., et al. 2007, ApJL, 658, L115

Haario, H., Laine, L., Mira, A., & Saksman, E. 2006, Stat. Comput., 16, 339

Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, ApJS, 154, 18

Hyvärinen, A. 1999, ITNN, 10, 626

Hyvärinen, A., Karhunen, J., & Oja, E. 2001, Independent Component Analysis
(New York: Wiley)

Hyvärinen, A., & Oja, E. 2000, NN, 13, 411

Inuso, G., La Foresta, F., Mammonne, N., & Morabito, F. C. 2007, Int. J. Conf.
Neural Netw., 20, 1524

Koldovský, Z., Tichavský, P., & Oja, E. 2006, ITNN, 17, 1265

La Foresta, F., Mammonne, N., & Morabito, F. C., 2006, in Neural Nets, Vol.
3931, ed. B. Apolloni (Berlin: Springer), 78

Lin, J., & Zhang, A. 2005, NDTI, 38, 421

Mammonne, N., La Foresta, F., & Morabito, F. C. 2012, IEEE Sensors Journal,
12, 533

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Oja, E. 1992, NN, 5, 927

Percival, D. B., & Walden, A. T. 2000, Wavelet Methods for Time Series
Analysis (Cambridge: Cambridge Univ. Press)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007,
Numerical Recipes (Cambridge: Cambridge Univ. Press)

Sánchez, C., Reta, J. J., Vayó, C., Moratal Perez, D., Zangróniz, R., & Millet,
J. 2006, LNCS, 3889, 486
Shannon, C. 1948, Bell System Tech. J., 27, 379
Stein, C. M. 1981, AnSta, 9, 1135
Stone, J. V. 2004, Independent Component Analysis: A Tutorial Introduction (Cambridge, MA: Bradford)
Thatte, A., Deroo, P., & Swain, M. R. 2010, A&A, 523, 35
Tichavský, P., Doron, E., Yeredor, A., & Gomez-Herrero, G. 2006a, in Proc. 14th European Signal Processing Conf., ed. A. Gershman (Waltham, MA: Academic), 1568981712
Tichavský, P., Doron, E., Yeredor, A., & Nielsen, J. 2006b, in Proc. 14th European Signal Processing Conf., ed. A. Gershman (Waltham, MA: Academic), 1568981867
Tichavsky, P., Koldovsky, Z., Yeredor, A., Gomez-Herrero, G., & Doron, E. 2008, ITNN, 19, 421
Waldmann, I. P. 2012, ApJ, 747, 12
Waldmann, I. P., Tinetti, G., Deroo, P., et al. 2013, ApJ, 766, 7
Yeredor, A. 2000, ISPL, 7, 197