Finding exoplanets orbiting young active stars – I. Technique

V. E. Moulds, C. A. Watson, X. Bonfils, S. P. Littlefair and E. K. Simpson

1 Department of Physics and Astronomy, Queens University Belfast, Belfast BT7 1NN, UK
2 UJF-Grenoble 1/CNRS-INSU, Institut de Planétologie et d’Astrophysique de Grenoble (IPAG)UMR 5274, F-38041 Grenoble, France
3 Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK

ABSTRACT
Stellar activity, such as starspots, can induce radial velocity (RV) variations that can mask or even mimic the RV signature of orbiting exoplanets. For this reason RV exoplanet surveys have been unsuccessful when searching for planets around young, active stars and are therefore failing to explore an important regime which can help to reveal how planets form and migrate. This paper describes a new technique to remove spot signatures from the stellar line-profiles of moderately rotating, active stars (\( v \sin i \) ranging from 10 to 50 km s\(^{-1}\)). By doing so it allows planetary RV signals to be uncovered. We used simulated models of a G5V type star with differing dark spots on its surface along with archive data of the known active star HD 49933 to validate our method. The results showed that starspots could be effectively cleaned from the line-profiles so that the stellar RV jitter was reduced by more than 80 per cent. Applying this procedure to the same models and HD 49933 data, but with fake planets injected, enabled the effective removal of starspots so that Jupiter mass planets on short orbital periods were successfully recovered. These results show that this approach can be useful in the search for hot-Jupiter planets that orbit around young, active stars with a \( v \sin i \) of \( \sim 10\)–50 km s\(^{-1}\).

Key words: line: profiles – methods: data analysis – planets and satellites: detection – planets and satellites: general – stars: activity – stars: spots.

1 INTRODUCTION
Currently over 800 exoplanets have now been found through a variety of techniques, e.g. via transits, radial velocity (RV) variations and microlensing. The most fruitful of these is the RV method which has been responsible for discovering approximately 70 per cent of exoplanets to date. High-precision spectrographs with an accuracy of a few ms\(^{-1}\) can easily detect the hundreds of ms\(^{-1}\) amplitude of the RV signature of short period hot-Jupiter planets. At this level of accuracy, the amplitude of the signal can only be hidden by important stellar noise.

Desort et al. (2007) showed that stellar noise, such as starspots, can produce RV shifts that can mimic or mask out planet signals. The limiting parameter for finding planets in the presence of such activity is the ratio between the amplitude of the planetary and activity signal. This means that stellar activity on stars similar to the Sun, which exhibit an RV jitter of several ms\(^{-1}\), will only affect the detection of low-mass, long orbital period planets (e.g. Lagrange, Desort & Meunier 2010). Whereas, for young, active stars where stellar jitter can be of the order of km s\(^{-1}\), the search for planets is limited to high-mass planets and in particular those on short orbital periods (e.g. Paulson, Cochran & Hatzes 2004; Paulson & Yelda 2006).

There have been several cases where planets have been announced and then subsequently retracted after starspots were discovered to be the source of the RV signature. Notable examples of this are TW Hydrae (Huélamo et al. 2008; Setiawan et al. 2008), BD+20 1790 (Figueira et al. 2010; Hernán-Obispo et al. 2010) and HD 166435 (Queloz et al. 2001). Due to these problems, RV surveys often exclude active stars. Young stars with convective outer envelopes tend to be faster rotating than older stars and therefore more active. According to the data available on the Extrasolar Planets Encyclopaedia established in 1995 by Jean Schneider, there have been over 50 planets with ages less than 1 Gyr known to date.

Targeting young planets is important for understanding the formation and evolution of planetary systems. Since the core accretion model (Pollack et al. 1996) predicts longer formation time-scales than the disc instability model (Boss 1997), the detection of a young planet would provide information about the formation mechanism itself. Several people have already conducted surveys for young planets (e.g. Paulson et al. 2004; Huerta et al. 2008) but without any
success due to small sample sizes as well as the high stellar noise problem.

Searching for companions around young stars is also important for understanding the lack of substellar objects with masses greater than 20\(M_J\) and in orbits closer than 3 au, i.e. the brown dwarf desert. Grether & Lineweaver (2006) found that 16 per cent of nearby Sun-like stars had companions with orbital periods less than 5 years and of these less than 1 per cent were brown dwarfs. It has been suggested by Armitage & Bonnell (2002) that the brown dwarf desert could be the result of orbital migration. Since the time-scale for catastrophic migration is of the order of 1 Myr, they predict that young stars would not have destroyed their brown dwarf companions and have an order of magnitude more brown dwarf companions than main-sequence stars. A survey for substellar companions around young stars could help with this theory and our understanding of the brown dwarf desert.

Finding a solution to the stellar activity problem is not only important for conducting young planet RV surveys but also for transiting missions. The Kepler space mission was launched in 2009 and monitors over 150 000 stars with early data showing numerous transiting planets to be of low mass (Borucki et al. 2011). With future planned missions such as Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2010), which will monitor over 2 million stars in its lifetime, the number of exoplanets, especially small ones, will continue to increase. However transiting surveys, unlike RV surveys, do not initially exclude active stars from their missions. In addition, early Kepler results show that approximately 50 per cent of the stars that Kepler observed during the first month of the mission are more active than the Sun (Basri et al. 2010). This could result in difficulties in the RV follow-up work to confirm their planetary nature particularly around more active stars.

Indeed, this has already been found to be the case for the Super-Earth planet transiting the active star CoRoT-7 in 2009. The activity of the star completely masks out the planetary signatures which has led to problems when analysing the RV data with separate groups finding different planetary solutions. Queloz et al. (2009) found two planetary signals giving masses of 4.8\(M_{\text{Earth}}\) and 8.4\(M_{\text{Earth}}\) for the planets, whereas Hatzes et al. (2010), in the analysis of the same data, suggest the star has three planets with masses of 6.9\(M_{\text{Earth}}\), 12.4\(M_{\text{Earth}}\) and 16.7\(M_{\text{Earth}}\).

Over the past couple of years a lot of effort has been put into finding a solution to this problem. One solution is to use the relationship between bisector-span and RVs in order to remove activity signatures from the data; e.g. Boisse et al. (2009) used this technique on HD 189733 to improve the RV jitter from 9.1 ms\(^{-1}\) to 3.7 ms\(^{-1}\). However, this cannot be applied to all systems. Desert et al. (2007) showed that this correlation breaks down for stars with \(v\sin i\) that is lower than the resolution of the spectrograph, due to the fact the spot cannot be resolved easily. There are also problems when stars have multiple spots, due to spot distortions compensating for each other. If other effects contribute more significantly to the RV variations (such as a planet signal) then the correlation may in fact be absent.

Other methods, such as pre-whitening techniques (Queloz et al. 2009), Fourier Analysis (Hatzes et al. 2010) and Harmonic decomposition (Boisse et al. 2011) have also been used to uncover planet signals from stellar jitter. These methods are all based on analysing the RV data in order to identify spurious RV signals due to activity and remove them. There are, however, some limitations when using these techniques. They require accurate knowledge of the stellar rotation period and in the case when the planetary orbit is close to the stellar rotation period then separating out the signal can be extremely difficult. Generally a long time series of RV points is required in order to measure the rotation period and so considerable observational time is important. However, when all these criteria are met, harmonic decomposition has been found to pull out planetary signals that are a third smaller in amplitude than the activity signal (Boisse et al. 2011).

We have developed a technique to complement these existing methods. Instead of removing the activity signal from the RV points we remove the spot signatures from the actual line-profiles directly. This method does not require the knowledge of the stellar rotation period or a long time series of data, provided the planetary timescale is short and the jitter/planetary signal ratio is small. This is the first paper of a series, in which we shall outline this novel spot removal technique. A forthcoming paper will apply this technique to Fibre-fed Echelle Spectrograph (FIES) observations of a sample of young, moderately fast, rotating stars.

Section 2 of this paper provides a detailed description of this spot removal technique. We then tested this technique on model stars with varying spots and planets. In Section 3, we describe the construction of these models, and the results of our simulations are presented in Section 4. In Section 5, we apply our spot removal technique to archival data of the known active star HD 49933. A discussion of this technique in comparison to other spot removal methods is provided in Section 6, with future work detailed in Section 7. Finally, we summarize our findings in Section 8.

2 SPOT REMOVAL TECHNIQUE

Our technique is based on the fact that a planet and a spot have different effects on the line-profile. A planet causes the line-profile to shift in wavelength or velocity whereas a spot distorts the shape of the line-profile resulting in an apparent RV shift. Spot bumps in the lines can be resolved when rotational broadening dominates the shape of the line profile, typically for \(v\sin i\) greater than the resolution of the spectrograph. We have written a code, Clearing Activity Signals In Line-profiles (CLEARASIL), to assess the stellar absorption line-profiles and remove any spot features, effectively cleaning the RVs to allow any planet signatures to be uncovered. CLEARASIL is based on the method used by Collier Cameron, Donati & Semel (2002) on AB Doradus, to directly track starspots in order to assess stellar differential rotation.

2.1 Least-squares deconvolution

In order to resolve the typically small spot bumps, an absorption line with a signal-to-noise ratio (SNR) of several hundred is required. Although large format CCDs and modern échelle spectrographs enable high-resolution spectra to be obtained spanning a large wavelength range (typically 3500 to 7000 Å for optical spectrographs such as FIES) the SNR of a single line is often not adequate. This is due to the requirement of short exposure times in order to reduce the effects of blurring of the spot features, and is further compounded for faint targets.

Each of the thousands of photospheric lines contained in an échelle spectrum is assumed to be affected in a similar way by the presence of spot features in the stellar photosphere. This assumption enables the application of multiline techniques to extract common physical information from many spectral lines and determine an average profile with a high SNR. Least-squares deconvolution (LSD) is the method used in this paper to compute a high SNR line-profile from the thousands of spectral lines available in a single échelle spectrum. It was first implemented for use on polarimetric Stokes
V spectra by Donati et al. (1997) and has since been successfully employed in other areas of astronomy, such as Doppler imaging.

However, the simplifying assumption that all spectral lines have the same shape, i.e. are independent of wavelength, is not valid for spectral lines formed at different heights which will be impacted by spots differently and also for bluer line-profiles where spot features will appear larger due to the increased temperature contrast. Therefore, the LSD technique is only applied after spectral regions containing chromospherically sensitive and strong spectral lines were removed from the data in this paper.

In this paper the LSD code used was written by C.A. Watson (Watson, Dhillon & Shahbaz 2006). Using a stellar line list from the Vienna Atomic Line Database (VALD; Kupka et al. 2000) and the method of least squares, the average profile is the one that gives the optimal fit to the spectra when convolved with the line list. The resulting line-profile has a SNR that is typically 20 to 30 times higher than a single isolated line.

2.2 Radial velocity calculation

The LSD line-profiles can then be used to measure the RVs for the data, and this is achieved by fitting a Gaussian to the LSD line-profiles and taking the peak of this Gaussian. This technique for calculating the RVs corresponds to the widely used cross-correlation method whereby a template spectrum is cross correlated with the target spectrum and the peak of a Gaussian fitted to this cross-correlation function is measured.

The LSD line-profile can be approximated by a Gaussian of the form

\[
p_i = A \exp \left( \frac{(v_i - \mu)^2}{2\sigma_i^2} \right) - y_{off}
\]

where \(p_i\) is the model flux value at the velocity value \(v_i\). The four parameters which define the Gaussian are the area under the curve, \(A\), the variance (i.e. the width of the line), \(\sigma\), the mean (i.e. the peak position), \(\mu\), and the continuum flux level, \(y_{off}\).

A least squares minimization technique is used to determine the model parameter values that give the best fit to the data. The measure of the goodness of fit between the model and observed data is given by the following \(\chi^2\) statistic:

\[
\chi^2 = \sum_i \left( \frac{f_i - p_i}{\sigma_i} \right)^2
\]

where \(f_i\) and \(p_i\) correspond to the observational and model flux at the \(i\)th data point and \(\sigma_i\) corresponds to the observational error for the \(i\)th data point.

The best-fitting model is found by minimizing the \(\chi^2\) statistic using a non-linear least-squares technique, specifically the Levenberg–Marquardt algorithm. This is an iterative process that rapidly converges on a numerical solution to the minimization of the function. As with all least-squares techniques, an initial guess of the parameters is required. This must be relatively close to the true value in order to avoid the results returning a local minima. The initial parameters are chosen by the user based upon the shape of the line-profile. In this case, the C version of the IDL MPFIT routine (Markwardt 2009) was employed.

This method for measuring the RVs using the LSD line-profiles was tested on our model data as detailed in Section 3.2 and the results showed this method to be accurate.

2.3 CLEARASIL

The high SNR mean line-profiles (in this case obtained using the LSD technique) are then input into CLEARASIL. The code operates on the basis that RV variations of a time series of line-profiles are either due to spots only or due to both spots and an orbiting planet. CLEARASIL processes the line-profiles according to both these theories and then tests to see which assumption best removes the spot features. This is a three-stage decision process that is outlined in more detail below. The code iterates over this process using the resulting average line-profile from the best spot removal method in the next iteration. At the end of each iteration a goodness-of-fit test (\(\chi^2\)) between the resulting line-profiles from each fitting method and their average is calculated to determine the correct fitting technique. When further iterations do not show any improvements to this goodness-of-fit test then CLEARASIL stops.

2.4 Step 1 – planet fitting technique

This method is only employed when CLEARASIL is testing the assumption that the data contain both a spot and a planet. If a planet is present then the line centres will be shifted relative to one another. If we correctly remove this shift then the line centres should lie close to one another and the average profile of these shifted lines should have a similar width and central velocity peak compared to each of the individual lines. On the other hand if we shift the lines by the wrong amount then their average will be orbitally smeared and hence appear broader when compared to the individual lines. A goodness-of-fit test between the average line and each individual orbitally corrected line-profile is therefore an appropriate measure to find the best planet signal.

If the planet has zero orbital eccentricity then the RV variation of the star over time will be a smooth sinusoid. The majority of hot-Jupiter planets have been found in circular orbits which is thought to be a result of short circularization time-scales for these close orbiting planets (Jackson, Greenberg & Barnes 2008). As the work in this paper concentrates on detecting hot-Jupiter planets, a valid approach in our method is to assume zero orbital eccentricity. With this assumption in mind the code generates sinewaves with various periods, amplitudes and phase offsets. Shifting the line-profiles according to each sinewave and measuring the goodness of fit between these shifted lines and the average determine the best model planet for the data. It is important to remove the planet signature at this stage of the fitting procedure so as to enable accurate removal of any spot features present in the subsequent steps.

2.5 Step 2 – spot fitting technique

The observable signature of a dark spot on the surface of the star is a bright bump in the photospheric line-profile. To isolate these bright bumps in the LSD profiles we employed the technique of line-profile subtraction.

Subtracting a line-profile with no spot features from each individual line-profile would result in residuals that contained noise with any spot bumps that were present superimposed, similar to Fig. 1. Unfortunately, it is impossible to obtain a completely immaculate line-profile from the data and so we generated a pseudo-immaculate profile by averaging all the line-profiles over the course of the observations. Averaging the lines will cause any spot features present to be smeared out and diluted. So although the average line-profile will not be completely free from spot features compared to the individual line-profiles, spot features present will be severely diminished.
This method utilizes the approximation that spot features present in the residuals are Gaussian-like in shape as discussed by Collier Cameron et al. (2002). The residuals are scanned for any peaks above this noise level and these features are fitted with a Gaussian using the Levenberg–Marquardt technique. The code also checks the width of these features to ensure spikes in noisy data are not wrongly treated as spot features. This enabled the spot bumps to be isolated from the noise features. These isolated spot bumps were then subtracted from the individual line-profiles to generate a cleaner data set.

2.6 Step 3 – the optimal fitting procedure

Having fitted the line-profiles assuming that either spots and a planet are both present (i.e. fitting the lines using both steps 1 and 2 given above) or that only spots are present (i.e. fitting the lines using only step 2 given above) we now determine which of these techniques best models the data. This is achieved by comparing the average of the cleaned line-profiles to the individual cleaned line-profiles for each method.

In the next iteration the average cleaned line-profile resulting from the best spot fitting procedure is used as a new pseudo-immaculate profile along with the original observed line-profiles to generate the residuals in step 2. Hence with each iteration we are improving our ‘immaculate’ profile enabling better spot identification and removal. As more iterations are conducted the average line becomes cleaner and in our simulations (see Section 4) very closely resembles the ‘true’ immaculate profile. The RV of the line-profiles also becomes less noisy and, if any planet is present, it should be easily detected. CLEARASIL stops iterating when there is no longer any improvement to the average cleaned line-profile with subsequent iterations. The RVs of the resulting cleaned line-profiles at this iteration are then searched for any planet signatures.

3 SIMULATIONS

To validate our spot removal method we generated a variety of model stars with differing spot and planet configurations. This section gives a description of how these models were constructed.

3.1 Description of simulation models

We first generated a model star defined by seven parameters: mass ($M_*$), radius ($R_*$), effective temperature ($T_{\text{eff}}$), rotational period ($P_{\text{rot}}$), inclination ($i$), microturbulent velocity ($v_{\text{micro}}$) and macroturbulent velocity ($v_{\text{macro}}$). We set the microturbulence to 1.5 km s$^{-1}$ for all the models used in this paper, while the macroturbulent velocity varies as a function of $T_{\text{eff}}$ and so has been calculated for each model using equation (1) in Valenti & Fischer (2005). Including radial–tangential macroturbulence (as described in Gray 2008) resulted in a more realistic model line-profile making it harder to resolve spot bumps. Along with these parameters, the non-linear limb-darkening law of Claret (2000) was used.

We then placed circular spots on the model star with free parameters given by the spot latitude ($\phi$) and longitude ($\theta$) on the stellar surface, the temperature of the spot ($T_{\text{spot}}$) and the spot size described by the fraction of the visible hemisphere covered by the spot ($f$). We then added planetary reflex motion according to the planets mass ($M_p$) and orbital period ($P_{\text{orb}}$). We assume the hot-Jupiter planet has zero orbital eccentricity given that most hot-Jupiter planets have been found to have circular orbits.

We model the stellar surface as a series of quadrilateral tiles of approximately equal area. Each tile is assigned a copy of the local (intrinsic) specific intensity profile, which includes micro- and macroturbulence effects convolved with the instrumental resolution. A Gaussian profile is used for the local (intrinsic) specific intensity profile and the equations in chapter 17 of Gray (2008) are used to model the micro- and macroturbulence effects. For the simulations in this paper we use an instrumental resolution of 4.6 km s$^{-1}$, which is equivalent to the FIES on the Nordic Optical Telescope (NOT). These profiles are scaled to take into account the projected area, limb-darkening, obscuration and the presence of a spot when relevant. The intensity profile for each tile is then Doppler shifted according to the RV of the surface element at a particular stellar rotation phase. Summing up the contributions from each element gives the rotationally broadened profile at that particular stellar rotation phase.

Any tiles with a spot feature present will have a lower temperature than the surrounding tiles and hence a lower intensity. We employ Planck’s radiation law which gives the energy radiated by an ideal blackbody in a particular wavelength interval (the central wavelength of the model star line-profile is set to 5800 Å). This law depends on the temperature of the blackbody and so if the tile has no spot feature then the energy radiated will be proportional to the effective temperature of the star. Whereas if the tile has a spot then the intensity will be proportional to the spot temperature which is lower than the effective temperature of the star. This results in an emission bump appearing in the rotationally broadened line-profile with a shape and position that are dependent on the temperature, size and position of the spot as shown in Fig. 2.

We generate the resulting line-profiles for different epochs over the rotational phase. Typically, we take 19 epochs to cover the rotational period. If there is a planet orbiting the star then the rotationally broadened profiles are shifted in velocity according to the RV amplitude of the planet at that particular orbital phase. The resulting line-profiles are then wavelength binned so that they are
all on a common scale. Each rotationally broadened profile is then convolved with an appropriate stellar line list, taken from VALD, with photon noise added in order to obtain a model spectrum as shown in Fig. 3. The model spectrum covers the wavelength range 4500 to 7000 Å, similar to the typical usable wavelength range of spectra taken with an optical spectrograph such as HARPS or FIES.

We have added photon noise assuming that the SNR is proportional to the square root of the flux. Since our model spectra are normalized, in order to determine the flux as a function of wavelength we have performed a cubic spline fit to a synthetic spectrum. The Pollux database (Palacios et al. 2010) provided the synthetic spectrum of similar spectral type to our model star. This enabled us to correct relative photon noise to be calculated. All the models described in this paper have been given a peak SNR of 100.

3.2 Testing the model

We checked how accurate our models were by measuring the RVs for the line-profiles generated over one rotational period for a G2V model star and compared the resulting RV jitter to that expected using equation (1) in Saar & Donahue (1997). Again, we calculated the RVs by fitting a Gaussian to each individual line-profile and measuring the peak position (see Section 2.2). Fig. 4 shows the RVs for a G2V model star ($T_{\text{eff}} = 5800$ K) with $v \sin i = 7$ km s$^{-1}$ seen at an inclination of 90° with a spot placed at a latitude of 0°, covering 1.05 per cent of the stellar surface and given a $T_{\text{spot}}$ of 0 K. Saar & Donahue (1997) predict a semi-amplitude of 47.54 ms$^{-1}$ for these parameters. Our model produced a RV semi-amplitude of 52 ms$^{-1}$. This gives us confidence in our model line-profiles and also in our method for calculating the RVs.

Satisfied with the method of generating our model line-profiles we computed a series of stellar models with varying spots and planets in order to validate CLEARASIL. The model star was given $M_\ast = 1.05 M_\odot, R_\ast = 0.95 R_\odot, T_{\text{eff}} = 5657$ K, $P_{\text{rot}} = 3.2$ d and $i = 51.8$. The spot and planet configurations as well as the results for these models are discussed in the following sections.

4 MODEL RESULTS

4.1 No spot model

For the first test case a 1 $M_J$ planet on a 4 d circular orbit was placed around our model star. No spots were present on the surface of the star and so the line-profiles should show an RV shift solely due to the injected planet. A total of 19 line-profiles covering one stellar rotation period (i.e. 3.2 d) were generated using the method...
Figure 3. Example portion of a synthetic spectrum generated by our model, assuming a G5V type star ($M_*=1.05\,M_{\odot}$, $R_*=0.95\,R_{\odot}$, $T_{\text{eff}}=5657\,\text{K}$, $P_{\text{rot}}=3.2\,\text{d}$ and $i=51.8^\circ$) with a 5 per cent spot having a peak SNR of 100 and a $v\sin i=15\,\text{km}\,\text{s}^{-1}$.

4.2 One-spot model

In the next test case we placed one spot on our model star at a latitude $=20^\circ$ with $T_{\text{spot}}=4600\,\text{K}$ and covering 5 per cent of the visible stellar surface, as shown in the top of Fig. 6. A total of 19 line-profiles covering one stellar rotation period (i.e. 3.2 d) were generated using the method described in Section 3. The RVs were found to have a semi-amplitude of $265\pm5.7\,\text{ms}^{-1}$ as shown in the bottom left panel of Fig. 6. The amplitude of the RV signal is approximately the same as that caused by a $5M_J$ planet orbiting this star on a period equal to that of the stellar rotation period.

CLEARASIL was then applied to these line-profiles and the results are shown in the bottom right panel of Fig. 6. As described in Section 2, CLEARASIL fits the signal with a combined planet and spot model, as well as with a spot only model to assess which best fits the data. For this model star the code found correctly that the best fit for the line-profiles was the spot only fit and was able to effectively clean the true spot features from the line-profiles.

The $\chi^2$ levelled off at the sixth iteration and the resulting RVs showed no realistic planet signature. The resulting RVs had an rms of $7.45\pm3\,\text{m}\,\text{s}^{-1}$, showing that a 96 per cent reduction in the stellar noise was achieved.

Next we tested the case when a relatively small planet signature (compared to the stellar jitter) is present. We placed a $1M_J$ planet on a 4 d circular orbit around the same model star and again generated 19 line-profiles that evenly covered one stellar rotation period. The RV semi-amplitude caused by this planet is $48.7\,\text{ms}^{-1}$, which given the previous test should be discernible if our spot removal algorithm does not overly affect the planet signature. We note that, in this test, the stellar rotation period and planetary orbital period are similar. Thus, for a significant fraction of our simulated observations, the RV jitter due to the spot dominates over the weaker planet signature and is phased similarly (as can be seen in Fig. 7).

For the first iteration of CLEARASIL the spot only fit was found to best remove any spot features from these line-profiles. However, in subsequent iterations, the code cleaned the line-profiles using the combined planet and spot fit. The $\chi^2$ reached a minimum at the 10th iteration and the results showed there to be a periodic RV variation with a semi-amplitude of $49.8\pm1.4\,\text{ms}^{-1}$ which closely
matches the RV signal of the injected planet (see right-hand panel of Fig. 7). Fitting these RVs with a sinewave reveals a $1.03 \pm 0.05 M_\text{J}$ on a $4.08 \pm 0.115 d$ orbital period, indeed confirming the results are consistent with the same as the planet injected into this model. This simple test case shows that CLEARASIL not only is capable of removing the majority of a large spot from the line-profiles but also has the ability of revealing a planet signal that is partially hidden in the RV data.

Having shown that the code was able to impressively uncover the $1 M_\text{J}$ mass signal hidden in the stellar noise we decided to further test the ability of CLEARASIL. In the next case we investigated the impact the phase of the injected planet signal had on the planet signature that could be uncovered from the noise. Again, we used the same one-spot model star with a $1 M_\text{J}$ planet on a 4 d circular orbit. The phase of the planet compared to the stellar rotation phase was varied between 0 and 1 in steps of 0.1.

These models were then processed through CLEARASIL and the impact the phase had on the mass and period of the planet detected is shown in Fig. 8. When the phase of the planet is less than 0.5 the code is able to successfully remove the spot signatures to reveal the $1 M_\text{J}$ planet with a 4 d orbital period. However, when the planet phase is set to 0.5 or greater the code struggles to uncover the correct planet signal. This is due to the interplay between the planet signal and the spots causing the code to struggle when fitting the RV data. Increasing the data size would help the code determine the correct planet signal in cases like this.

This next test case is to show the ability of the code of uncovering a planet signal that is the same as the stellar rotation period. Again, the same model star with one spot covering 5 per cent of the stellar surface was used but this time the injected planet signal had a mass of $1 M_\text{J}$ and a period of 3.2 d (i.e. the same as the stellar rotation period). The initial semi-amplitude of the RV signal was $299 \pm 5.1 \text{ ms}^{-1}$ which completely masks the RV signal of the planet (see Fig. 9).

The spot removal code was able to successfully reduce the stellar jitter to reveal a periodic RV variation with a semi-amplitude of $51.5 \pm 1.4 \text{ ms}^{-1}$ which closely matches the RV signal of the injected planet (see right-hand panel of Fig. 9). Fitting these RVs with a sinewave reveals a $0.98 \pm 0.04 M_\text{J}$ on a $3.2 \pm 0.04 d$ orbital period, indeed confirming the results are consistent with the same as the planet injected into this model. This highlights the ability of CLEARASIL to uncover the correct planetary parameters even if the planetary orbital period is the same as the stellar rotational period.

### 4.3 Four-spot model

The previous one-spot model represents a simple test case. In this section we investigate the ability of CLEARASIL to cope with a more realistic multiple spot model. Four spots were placed on the model G5V star (see Section 3) and the details of the latitude, size and temperature of these spots can be found in Table 1. As for the previous model, 19 line-profiles were generated covering one stellar rotation period. The RVs resulting from this model are not very periodic due to the interplay of multiple spots crossing the stellar disc and have an rms of $30 \text{ ms}^{-1}$ and a peak amplitude of $96 \text{ ms}^{-1}$, as shown in Fig. 10.

The line-profiles were put through CLEARASIL, which correctly found the spot only fit to be best for all iterations of the code. The $\chi^2$ levels out at the 13th iteration and the resulting line-profiles reveal an RV signal with an rms of $13 \text{ ms}^{-1}$ (Fig. 10). In this case CLEARASIL has effectively reduced the stellar jitter by 57 per cent. The peak-to-peak amplitude is reduced from $96 \pm 6.3 \text{ ms}^{-1}$ to $24 \pm 6.2 \text{ ms}^{-1}$.
Figure 8. Results showing the impact of varying the planet signal phase on
the ability of CLEARASIL to uncover the correct planet signature. In this case a
G5V model star with a dark spot ($T_{\text{spot}} = 4600$ K) covering 5 per cent of the
visible surface at a latitude of $20^\circ$ and a 1M$_J$ planet on a 4 d orbit was used.
The phase of the planet was varied from 0 to 1 in steps of 0.1. The top panel
displays how the planetary period uncovered by CLEARASIL varies with the
injected planetary phase, while the bottom panel shows the planetary phase
against the mass of the uncovered planet signal.

which is a 75 per cent reduction in amplitude. Although this is a
significant result, we do note it is smaller than the RV improvement
seen in the one-spot model case under the same sampling and SNR
conditions. CLEARASIL assumes nothing about the number of spots
present on the star and fits the line-profiles for any spot-like features
according to the fitting technique outlined in Section 2.5. There is
no limit on the number of spots that CLEARASIL will try and fit and
remove from the line-profiles. Therefore if CLEARASIL detects 10
spots in the line-profile then it will fit for 10 spots or if it detects
20 spots then it will fit for 20, etc. Multiple, smaller spots are
difficult to fit using this Gaussian-based fitting technique and so the
line-profiles are not cleaned as effectively in this four-spot model.

We also note that in this instance the resultant RVs appear to
have a systemic velocity of approximately $-13$ m s$^{-1}$ which does
not match the systemic velocity of 0 m s$^{-1}$ that was given to the
model data. When dealing with multiple spots rotating across the
star it becomes more likely that a spot feature will appear in
the same place in all the line-profiles. If this does happen then
when the line-profiles are averaged together to create the pseudo-
immaculate profile as described in Section 2.5 this feature will also
appear in the average profile. Subtracting this pseudo-immaculate
profile from each individual profile will have the result of cancelling
this spot feature out and so it will never appear in the residuals. This
results in the code being unable to isolate this spot feature and
remove it from of the line-profiles.

Table 1. A description of the size, position and temperatures of the spots
placed on the four-spot model G5V star.

<table>
<thead>
<tr>
<th>Spot number</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Size (per cent of visible surface)</th>
<th>Temperature in K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>4600</td>
</tr>
<tr>
<td>2</td>
<td>–20</td>
<td>90</td>
<td>1</td>
<td>4600</td>
</tr>
<tr>
<td>3</td>
<td>–5</td>
<td>–90</td>
<td>1</td>
<td>4600</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>120</td>
<td>1</td>
<td>4600</td>
</tr>
</tbody>
</table>
was once again able to successfully reduce the stellar signature injected into the model.

4.4 Variable spot model

The previous models discussed have shown CLEARASIL works for single- and multiple-spot scenarios. However, in both cases the line-profiles have been evenly spaced over one stellar rotational period. In reality, observations will be spread out due to limitations caused by telescope availability and the stars’ visibility. So we decided to further test the code on unevenly sampled data.

In this case, the model star ($M_*=1.05M_{\odot}$, $R_*=0.95R_{\odot}$, $T_{\text{eff}}=5657\ K$, $P_{\text{rot}}=3.2\ d$ and $t_i=51.8$) was used to generate 16 spectra with varying spot features that covered the time-span of the RV observations taken to confirm the transiting nature of WASP-39b (Faedi et al. 2011). Note that we generate one model star for each of the 16 spectra with the spot features set to correspond to that particular observation. WASP-39b was chosen because it has similar properties to the models so far run in this paper and is therefore a good representation of a typical RV follow-up strategy.

The 16 line-profiles span over 76 d with a gap of 20 d in the middle of the observations.

As with the previous one-spot case, we decided to test the code further by inserting a $1M_{J}$ planet on a 4 d circular orbit into this multiple spot model. Fig. 11 shows the RV curve for this model, which has a semi-amplitude of $87.14 \pm 6.3\ms$. There is a hint of a planetary signal in the RV jitter due to the fact that both the RV amplitude and orbital period of the planet are similar to that of the spots, and they both have similar phase. Fitting a sinewave to this original ‘noisy’ data reveals a $1.8M_{J}$ planet on a 4.7 d orbital period.

CLEARASIL was once again able to successfully reduce the stellar jitter and clean the injected planetary signal. The lowest $\chi^2$ occurred at the 21st iteration with the resulting RV curve showing a periodic shape and semi-amplitude of $63 \pm 6.3\ms$. Fitting a sinewave to the RVs shows a $1.2 \pm 0.15M_{J}$ planet on a $4.15 \pm 0.083\ d$ orbital period. This is an improvement in the orbital period by 20 per cent and the planet mass by 30 per cent. Due to the confusion of fitting multiple spots the planet signature is slightly distorted, but again the sine fit reveals a similar mass and orbital period to the fake planet signature injected into the model.

Table 2. A description of the size and position of the spots placed on the variable spot model G5V star.

<table>
<thead>
<tr>
<th>Obs date (d)</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\phi_3$</th>
<th>$\phi_4$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
<td>-5</td>
<td>60</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.967</td>
<td>30</td>
<td>-5</td>
<td>60</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.278</td>
<td>40</td>
<td>-5</td>
<td>55</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1.6</td>
<td>1.2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.9685</td>
<td>40</td>
<td>-5</td>
<td>55</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1.6</td>
<td>1.2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.284</td>
<td>60</td>
<td>-2</td>
<td>50</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1.2</td>
<td>0.9</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.948</td>
<td>60</td>
<td>-2</td>
<td>50</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1.2</td>
<td>0.9</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.9597</td>
<td>60</td>
<td>-2</td>
<td>50</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>1</td>
<td>0.7</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.992</td>
<td>60</td>
<td>-4</td>
<td>45</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>0.75</td>
<td>0.5</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.9773</td>
<td>60</td>
<td>-4</td>
<td>45</td>
<td>-</td>
<td>-90</td>
<td>120</td>
<td>-</td>
<td>0.75</td>
<td>0.5</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55.1228</td>
<td>20</td>
<td>-5</td>
<td>40</td>
<td>10</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>1</td>
<td>1.2</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>57.1848</td>
<td>20</td>
<td>-5</td>
<td>40</td>
<td>10</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>1</td>
<td>1.2</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>58.0972</td>
<td>25</td>
<td>10</td>
<td>40</td>
<td>10</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>1</td>
<td>1.2</td>
<td>0.8</td>
<td>1.8</td>
</tr>
<tr>
<td>61.1231</td>
<td>25</td>
<td>-10</td>
<td>45</td>
<td>20</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>1.5</td>
<td>0.9</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>63.9251</td>
<td>25</td>
<td>-10</td>
<td>45</td>
<td>20</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>1.5</td>
<td>0.9</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>72.169</td>
<td>30</td>
<td>-5</td>
<td>50</td>
<td>15</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>2</td>
<td>0.4</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>74.1561</td>
<td>30</td>
<td>-5</td>
<td>50</td>
<td>15</td>
<td>-100</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>2</td>
<td>0.2</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>76.167</td>
<td>30</td>
<td>-5</td>
<td>50</td>
<td>15</td>
<td>-110</td>
<td>30</td>
<td>150</td>
<td>-25</td>
<td>2</td>
<td>0.2</td>
<td>1.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>
V. E. Moulds et al.

Figure 12. Left-hand panel: RV curve of the G5V model star with random spot coverage (\(T_{\text{spot}} = 4600\,\text{K}\)) with positions and sizes detailed in Table 2. Right-hand panel: RV curve after 18 iterations of the spot removal code.

I tried to use this information to determine the size and latitudinal position of the spot features on the star as well as for determining how long the spot feature should be present on the surface. Note the longitudinal position of the spots was chosen so that they would all appear on the surface of the star at the same time.

The RVs for this model were found to have an rms of 32.3 ms\(^{-1}\), as shown in Fig. 12 and CLEARASIL was able to reduce the stellar jitter rms by 65 per cent. The peak-to-peak amplitude was reduced from 105 \(\pm\) 6.3 ms\(^{-1}\) to 13 \(\pm\) 6.2 ms\(^{-1}\) showing an 80 per cent reduction. Again, we note a systemic velocity present in the resulting data which is once again a result of fitting multiple spots as explained in Section 4.3.

Again, we carried out the test when a 1 Jupiter mass planet on a 4d circular orbit was inserted into this model. Fig. 13 shows the original data points with the best fit sine wave to this data overlapped. Although this sine wave is close to the true planetary signal (shown in Fig. 14) it does not match well with the data points. Fig. 14 shows the RVs for this model with the input planet signal overlapped as the dashed line. The RV semi-amplitude due to the planet and evolving spots is 97.5 \(\pm\) 6.3 ms\(^{-1}\). Fitting this original ‘noisy’ data with a sine wave reveals a planet with a mass of 1.2 MJ on an orbital period of 4.01 d.

Even though the data size was small and covered a large time-span, CLEARASIL was still able to detect that the planet and spot fit was better at removing the spot from the line-profiles. The lowest \(\chi^2\) was reached at the ninth iteration and the resulting RVs had a semi-amplitude of 45 \(\pm\) 6.3 ms\(^{-1}\). Fitting these RVs with a sine wave revealed a 1.14 \(\pm\) 0.18 MJ on a 4 \(\pm\) 0.08 d orbital period, which is consistent with the planet signature initially placed into the model.

The code can still successfully detect the presence of a planet with sparse data covering a large time-span.

5 HD 49933

Having tested CLEARASIL on model data we now turn to observational data of the F5V main-sequence star, HD 49933. HD 49933 is known to be magnetically active and has a \(v\sin i\) of approximately 10 km s\(^{-1}\), making it an ideal candidate for testing the abilities of CLEARASIL. It was observed spectroscopically in 2004 for 10 nights by Mosser et al. (2005) who found the star to have large amplitude RV variations as a result of stellar activity. Later photometry taken by the CoRoT team in 2007 and 2008 (Benomar et al. 2009; Appourchaux et al. 2008) showed active regions crossing the stellar disc with a rotation period of 3.5 d.

We analysed 10 d of archival HARPS data from the observing run of Mosser et al. (2005). This data set consisted of a total of 1304 spectra which were then processed using the LSD method to produce line-profiles with a significantly higher SNR. The resulting LSD line-profiles were further binned into batches of 1 h 10 min over each night. This time was chosen because it corresponded to the star rotating by approximately 5° preventing any spot features from being smeared. This left us with a total of 45 line-profiles covering 8.5 d.

The RV results for these data are shown in Fig. 15. The left-hand plot shows that the data had a peak-to-peak amplitude of 334.2 \(\pm\) 2.1 ms\(^{-1}\) initially. After these line-profiles were processed through CLEARASIL the peak-to-peak amplitude was reduced to...
CLEARASIL does not need to know the stellar rotational period differs in the method it uses to remove spot ± to uncover approximately five Jupiter mass planets injected to accurately
0.01 d orbit. CLEARASIL analyses and removes spot features from the ± revealed a 2 on the data. They reveal CLEARASIL
CLEARASIL

The mass of a planet as a function of the planetary period when
P
CLEARASIL

recovered planets on a 3.2 d period ± still being able to reduce the stellar jitter by 80 per cent. The orbital period of the
∼ 0.02
± is not the first technique to remove stellar jitter from
CLEARASIL

of the semi-amplitude of the activity signal (Boisse
2.2 ms
was now able to improve on the spot CLEARASIL
CLEARASIL

differs greatly from these methods in
is also not hampered by the same limitations imposed
Left-hand panel: RV curve of HD 49933. Right-hand panel: RV curve after the spot removal code.

Figure 15. Left-hand panel: RV curve of HD 49933. Right-hand panel: RV curve after the spot removal code.

61 ± 2.2 ms\(^{-1}\). This is a reduction in the stellar jitter of approximately 80 per cent, comparable to the results from our simulations.

5.1 HD 49933 with fake planets

Currently, HD 49933 has not been found to host any planets as has been confirmed by our results in the previous section. In order to test how a planet would affect the removal of spots with CLEARASIL, we again injected fake planets with a range of orbital periods into the HD 49933 data. The size of the planet was set to ~5 Jupiter mass. This should be easily detectable by our code which was able to reduce the stellar jitter by 80 per cent. The orbital period of the planets was limited to less than 3 d due to the data only spanning 8.5 nights.

Fig. 16 shows the period and mass of planets found by the code in comparison to the model planets injected into the data. The error bars in this plot are the difference between the injected planet signals and the planets found when using CLEARASIL on the data. They reveal the code to be more accurate for short period planets than for longer

Figure 16. The mass of a planet as a function of the planetary period when using CLEARASIL to uncover approximately five Jupiter mass planets injected into the HD 49933 data. The error bars represent the difference between the fitted and the simulated parameters.

6 COMPARISON WITH OTHER TECHNIQUES

CLEARASIL is not the first technique to remove stellar jitter from planet signals; there are a variety of methods currently in use such as pre-whitening techniques (Queloz et al. 2009), Fourier Analysis (Hatzes et al. 2010) and Harmonic decomposition (Boisse et al. 2011). However, CLEARASIL differs greatly from these methods in the approach it takes at removing stellar jitter. Instead of analysing the RV data to identify and remove spurious RV signals due to activity, CLEARASIL analyses and removes spot features from the actual line-profiles themselves. This makes it a unique technique that complements these other methods.

Although CLEARASIL differs in the method it uses to remove spot features, it is still able to produce results comparable to these other techniques. Boisse et al. (2011) showed that by using harmonic decomposition to fit three sinewaves at the rotational period of the star and its first two harmonics \(P_{\text{rot}}/2\) and \(P_{\text{rot}}/3\), stellar jitter could be reduced by up to 90 per cent. This is similar to the results of our work where removal of spot features from line-profiles was also able to reduce the stellar jitter by 90 per cent in a number of cases.

CLEARASIL is also not hampered by the same limitations imposed when analysing RV points alone. One constraint required for successful removal of stellar jitter from RV data is the accurate measurement of the rotational period of the star. This results in these methods requiring numerous data points covering more than one rotational period of the star. Unlike methods which analyse the RV data, CLEARASIL does not need to know the stellar rotational period and therefore does not require data covering a long time-span. This was shown to be the case in the variable spot model where data are unevenly sampled over 76 d and spot evolution is allowed with CLEARASIL still being able to reduce the stellar jitter by 80 per cent.

RV fitting techniques also struggle to reveal planets with orbital periods that are close to the stellar rotational period. However, as described in Section 4 CLEARASIL recovered planets on a 3.2 d period which is the same as the 3.2 d stellar rotation period.

There is a further limitation of the harmonic decomposition technique in that it only works for planets with an RV semi-amplitude that exceeds \(\frac{1}{4}\) of the semi-amplitude of the activity signal (Boisse
et al. 2011). In our one-spot model (Section 4.2) CLEARASIL was actually able to recover a planetary signal that was \( \frac{1}{5} \) of the size of the stellar jitter.

Although our technique is not hindered by the same requirements as the aforementioned stellar jitter removal methods, it does have some limitations. In order for the spot to be resolved in the line-profiles we need the star to have a \( v \sin i \) higher than 10 km s\(^{-1}\). Below this level the \( v \sin i \) becomes lower than the spectrograph resolution and so any change in shape in the stellar lines by spots only results in them being shifted after convolution with the instrumental point spread function (as shown by Desort et al. 2007). Fitting sinewaves to RV data does not depend on the instrumental point spread function (as shown by Desort et al. 2007). However this current paper shows that this technique does work and can obtain results with only one major requirement and that is the \( v \sin i \) of the star needs to be within 10 to 50 km s\(^{-1}\) in order to be able to resolve the spot in the stellar line-profiles and accurately measure the RVs. CLEARASIL does not require any prior knowledge about the star or a high time series of evenly sampled data, thus making it complementary to other stellar jitter removal techniques such as harmonic decomposition which do have these limitations but can operate for stars with low \( v \sin i \).

This technique is useful for RV follow-up of transit surveys and for RV surveys of moderately rotating active stars looking for Jupiter mass planets. It is also beneficial for looking at younger stars which tend to be faster rotating and therefore more active. This could help improve our knowledge of planetary characteristics over time so we understand better how planets are formed and how they evolve.

### ACKNOWLEDGEMENTS

Victoria Moulds would like to thank the Northern Ireland Department for Employment and Learning for a PhD studentship. This work is based on data obtained from the ESO Science Archive Facility. This research was achieved using the POLLUX database (http://pollux.graal.univ-montp2.fr) operated at LUPM (Université Montpellier II – CNRS, France) with the support of the PNPS and INSU.

### REFERENCES


This paper has been typeset from a TeX/LaTeX file prepared by the author.