Chapter 3

Characterising and simulating stellar micro-variability

3.1 Introduction

The previous chapter focussed on the problem of transit detection, specifically in the white Gaussian noise case, isolating other issues affecting any planet search project, such as confusion and non-Gaussian noise sources. Tests of the algorithms presented therein on simulated data have shown they can perform very well in white Gaussian noise (as can other algorithms developed simultaneously or near-simultaneously by other authors), reliably detecting transits and quantifying their statistical significance.

However, rather than the detection of the transits themselves, the major difficulty for ground-based searches so far has in fact been distinguishing planetary transit-like events caused by stellar systems, such as eclipsing binaries with high mass ratios, or hierarchical triple systems (due to either a physical triple system or an eclipsing binary blended with a foreground star), from true planetary transits (Brown 2003).

In order to detect terrestrial planets, it is necessary to go to space, to avoid being affected by atmospheric scintillation and to monitor the target field(s) continuously, with minimal interruptions. With improved photometric precision comes an additional noise source, which is usually insignificant at the precisions achieved by ground based observations: the intrinsic variability of the stars, due mainly to the temporal evolution and rotational modulation of structures on the stellar disk. The Sun's total irradiance (see Figure 3.1) varies on all timescales covered by the available data, with a complex, non-white power spectrum (see Figure 3.3). The amplitude of the variations can reach more than 1 % when a large spot crosses the solar disk at activity maximum, compared to transit depths of tenths to hundredths of a percent. There is significant power on timescales of a few hours, similar to the typi-

cal transit duration. Untreated, solar micro-variability would significantly reduce the detection performance of missions such as *Eddington* or *Kepler* (see Section 2.1.4), while the variability levels of more active stars are expected to also affect COROT and even ground-based transit searches in young stellar environments.

However, it is possible to separate planetary transit signal and stellar variability because the former is of relatively well known shape and contains significant power at high frequencies, while the latter is concentrated mainly at low frequencies. Already, modified transit search algorithms, designed to distinguish between the transit and brightness variations of stellar origin, have been tested on simulated data including solar variability (Defaÿ et al. 2001; Jenkins 2002). Jenkins (2002) applied a simple scaling to the solar irradiance data to evaluate the impact of increased rotation rate. Nonetheless, a more physical model, in which the different phenomena involved can be scaled independently in timescale and amplitude for a range of spectral types and ages, is needed to simulate realistic light curves for stars other than the Sun. This will allow us to optimise, evaluate and compare different algorithms, but also different design and target field options for the space missions concerned.

The present chapter is concerned with the development of such a model. The philosophy adopted in the process is the following. Intrinsic stellar variability is by no means a well-understood process. Despite recent progress in the modelling of activity-induced irradiance variations on timescales of days to weeks in the Sun (Krivova et al. 2003; Lanza et al. 2003), the extension of these physical models to other stars remains problematic, due to the scarcity of information on how the timescales, filling factors of various surface structures, and contrast ratios, depend on stellar parameters. We have therefore adopted an empirical approach, using chromospheric flux measurements as a proxy measure of activity-induced variability. This step is possible due to the fact that a correlation between the two quantities is observed in the Sun throughout its activity cycle, as well as in other stars. Similarly, empirically derived relationships were used again to relate chromospheric activity, rotation, age and colour, rather than attempting to use models which make a number of assumptions about the physical process driving these phenomena, and generally depend on parameters which require fine-tuning.

The stellar micro-variability model is developed by extrapolating the results of a detailed Fourier analysis of total solar irradiance (TSI) variations (Section 3.2) to other spectral types and stellar ages through empirical scaling laws (Section 3.3). Tests of the model are presented in Section 3.4. The results are discussed in Section 3.5

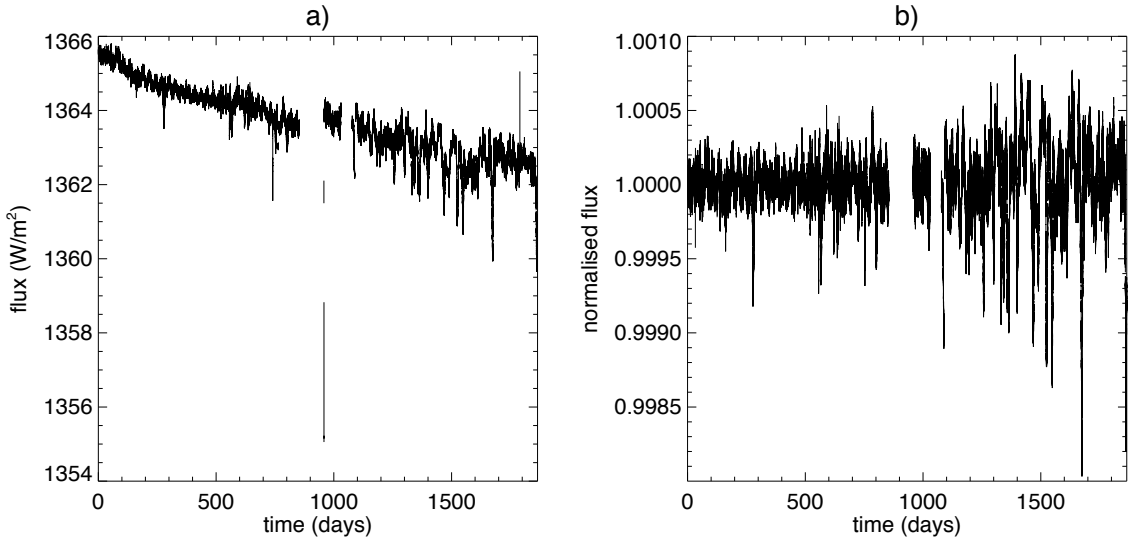

Figure 3.1: PMO6 light curve, **a)** *before and* **b)** *after the pre-processing steps described in Section 3.2.1.1. The data starts in January 1996.*

3.2 Clues from solar irradiance variations

The Sun is the only star observed with sufficient precision and frequent sampling to permit detailed micro-variability studies, thanks to the recent wealth of data collected by the SoHO spacecraft, and particularly the full disk observations obtained by VIRGO (Variability of solar IRradiance and Gravity Oscillations), the experiment for helioseismology and solar irradiance monitoring on SoHO, (Frohlich et al. 1997).

Stellar micro-variability is difficult to observe from the ground due to its very low amplitude, except for very young, active stars – which are outside the main range of interest for planet searches. There is some information available on rms night-to-night and year-to-year photometric variability of a small sample of stars monitored over many years by a few teams (Radick et al. 1998; Henry et al. 2000b). We make use of these as they present the advantage of covering a range of stellar ages, but their irregular time coverage and limited photometric precision make them unsuitable for an in-depth study, and particularly for the detailed analysis of the frequency content of the variations.

A drastic improvement in our understanding of intrinsic stellar variability across the HR diagram is expected from the very missions this work is aimed at preparing. In the relatively short term, MOST will provide valuable information for a small sample of stars, but it is not until the launch of COROT, and later *Kepler* and *Eddington*, that a wide range of stellar parameters will be covered. In the mean time, we must make use of the detailed solar data, and make reasonable assumptions to extrapolate to other stars than the Sun.

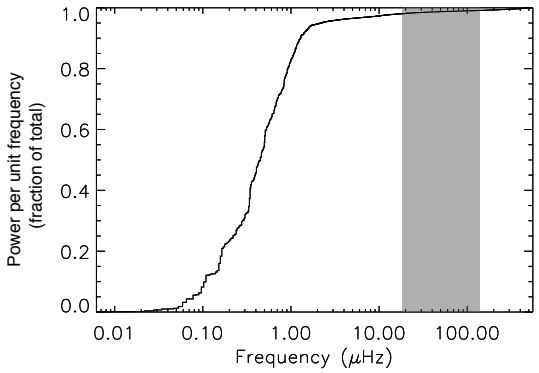

Figure 3.2: Cumulative power spectrum of the PMO6 light curve after pre-processing, as a fraction of the total power at non-zero frequencies. The grey area shows the range of frequencies corresponding to typical transit durations (2 to 15 hrs). Over 95 % of the power in the solar variations is below that range, which suggests that the two types of signal can be separated on the basis of their frequency signature.

3.2.1 SoHO/VIRGO total irradiance (PMO6) data

All SoHO/VIRGO data used in this work were kindly provided by the VIRGO team at ESTEC. The main instrument of interest was PMO6, a radiometer measuring total solar irradiance. The light-curves used here cover the period January 1996 to March 2001[1], which roughly corresponds to the rising phase of cycle 23.

3.2.1.1 Pre-processing of the data

The light curves were originally received as level 1 data, in physical units but with no correction for instrumental effects. Careful treatment was required to remove long term trends of instrumental and astrophysical origin. There was a difference of $\sim 0.24\,\%$ in the mean measured flux between the start and the end of the time series. Given that the observations roughly correspond to the interval between the minimum and the maximum of the Sun's activity cycle, one might expect to see a rise in the mean irradiance over that period. The instrumental decay may therefore be higher than the value quoted. However, the absolute value of the irradiance was of little interest for the present study, which concentrates on relative variations on time scales of weeks or less. Any long term trends in the data were therefore removed completely, regardless of whether they were of instrumental or physical origin[2]. The decay appeared non-linear and there were discontinuities and outliers in the light curves, making a simple spline fit unsuitable.

The approach that was adopted consisted of a 5 step process:

[1] Except for two interruptions roughly 1,000 days after the start of operations, corresponding to the "SoHO vacations", when the satellite was lost and then recovered.

[2] Note that fully corrected (level 2) data are now freely available from the World Radiation Centre in Davos, Switzerland for the period March 1996 to December 2002 (for data with minute sampling, more recent data is available with hourly sampling). Level 2 data is corrected for instrumental effects but contains the long-term trends due to the solar activity cycle. As only trends on timescales of 2 months or less were of interest here, the relatively crude corrections we applied to the level 1 data ourselves were sufficient.

- Visual inspection of the data was used to manually remove sections visibly affected by instrumental problems.
- Spline fits were performed on intervals chosen by visual inspection to start and end where discontinuities occurred. Each interval was divided by the corresponding fit, resulting in a normalised output light curve.
- A 5 σ cutoff was applied for outlier removal.
- The sampling, originally 1 min, was reduced to 15 min to make the size of the light curves more manageable. This was done by taking the mean of the original data points in each 15 min bin, ignoring any missing or bad data points. It is unlikely any information on timescales shorter than 15 min would significantly impact the transit detection process, as the transits of interest here generally last several hours (corresponding to orbital periods of several months or years).
- Data gaps were replaced with the baseline value of 1.0, to allow the calculation of the amplitude spectra needed for the analysis[3].

3.2.2 Modelling the 'solar background'

Table 3.1: Typical timescales for the different components of the solar background.

Component	Timescale B (s)
Active regions	1 to 3×10^7
Super-granulation	3 to 7×10^4
Meso-granulation	$\simeq 8000$
Granulation	200 to 500
Bright points	$\simeq 70$

The power spectrum of the solar irradiance variations at frequencies lower than $\simeq 8$ mHz constitutes a noise source for helioseismology, usually referred to as the 'solar background'. It is common practice to fit this background with a sum of powerlaws in order to model it accurately enough to allow the measurement of solar oscillation frequencies and amplitudes. Powerlaw models were first introduced by Harvey (1985). The most commonly used model in the literature today is that of Andersen et al. (1994), which is fairly similar: the total power spectrum is approximated by a sum of power laws, the number N of which varies between three and five depend-

[3] If d_i is a regularly sampled dataset, the dataset in which gaps have been replaced by 1.0 is $d'_i = d_i * w_i + 1 - w_i$, where w_i is window function, i.e. is 0 in the gaps and 1 elsewhere. The FT of d'_i is then, for all non zero indices k, $D'_k = D_k \circledast W_k - W_k$, where D_k and W_k are the FTs of d_i and W_i respectively.

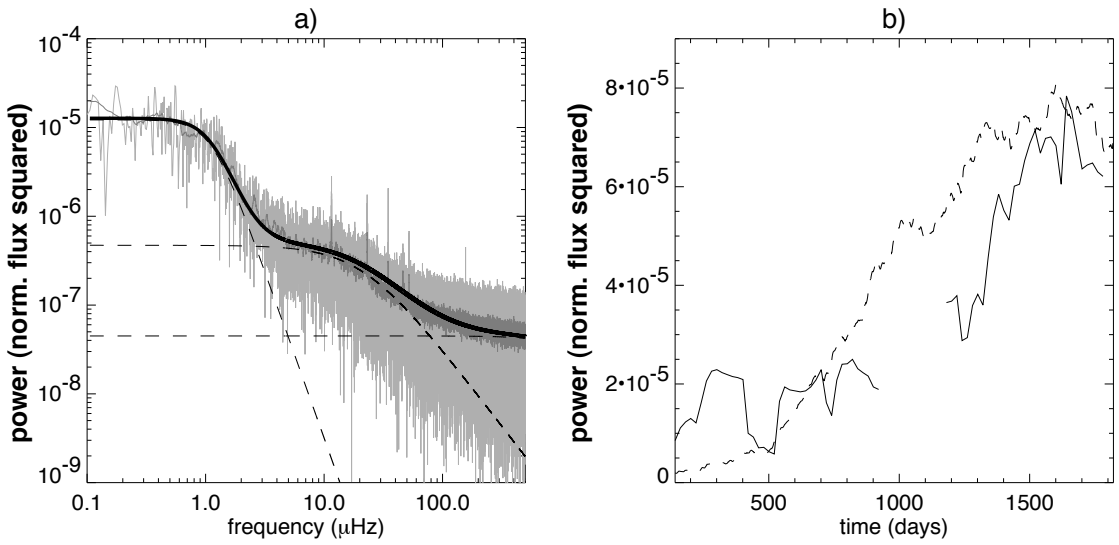

Figure 3.3: **a)** *Light grey: power spectrum of the PMO6 light curve (1996–2001). Dark grey: idem, smoothed with a boxcar filter. Thick solid line: multi-component powerlaw fit (see Section 3.2.2). Dotted lines: individual components of the fit.* **b)** *Solid line: amplitude of the low frequency component of the power spectrum (A_1, computed as described in Section 3.2.3, using L =180 days and S =20 days) versus time, from 1996 to 2001. The gap at $\sim$ 1000 days corresponds to a prolonged gap in the data. Dotted line: Chromospheric activity (BBSO Ca II K-line index) over the same period (arbitrary units), smoothed with a boxcar filter.*

ing on the frequency coverage:

$$P(\nu) = \sum_{i=1}^{N} P_i = \sum_{i=1}^{N} \frac{A_i}{1 + (B_i\,\nu)^{C_i}} \tag{3.1}$$

where ν is frequency, A_i is the amplitude of the i^{th} component, B_i is its characteristic timescale, and C_i is the slope of the power law (which was fixed to 2 in Harvey's early model). For a given component, the power remains approximately constant on timescales larger than B, and drops off for shorter timescales. Each power law corresponds to a separate class of physical phenomena, occurring on a different characteristic time scale, and corresponding to different physical structures on the surface of the Sun (see Table 3.1).

3.2.3 Evolution of the power spectrum with the activity cycle

In order to track the evolution of the solar background over the activity cycle, sums of power-laws – as given by Equation (3.1) – were fitted to the power spectrum of a section of data of duration L (typically 6 months). Such a fit is illustrated on the power spectrum of the entire dataset in the left-hand panel of Figure 3.3. The op-

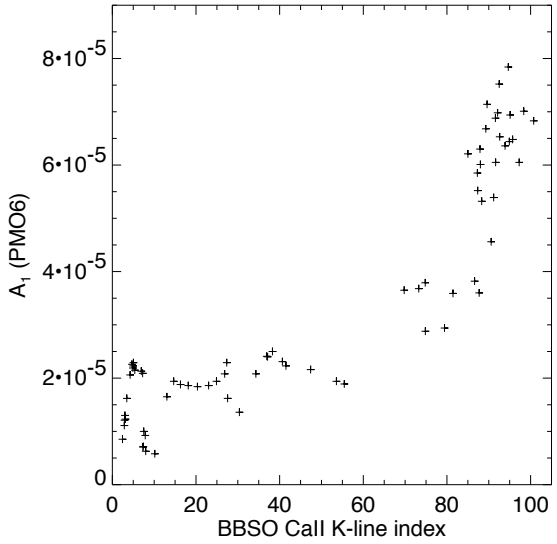

Figure 3.4: Amplitude A_1 of the low frequency component of the solar power spectrum (computed as described in Section 3.2.3, using L =180 days and S =20 days), versus the BBSO Ca II K-line index (arbitrary units) over the period 1996 to 2001.

eration is then repeated for a section shifted by a small interval S from the previous one (typically 20 days), and so on. Thus the evolution of each component can be tracked throughout the rise from solar minimum (1996) to maximum (2001) by measuring changes in the parameters defining each powerlaw.

A single component fit with parameters A_1, B_1 & C_1 is made first. Additional components are then added until they no longer improve the fit, i.e. until the addition of an extra component does not reduce the χ^2 by more than 10^{-2}. The fit to the first section is used as the initial guess for the fit to the next section, and so forth. This method allows us to track the emergence of components corresponding to different types of surface structures throughout the solar cycle, as well as monitor variations in amplitude, timescale and slope for each component.

3.2.3.1 Results

The algorithm described above was run on the PMO6 data with L = 180 days and S = 20 days, and three components were found to provide the best fit in all cases. These components have stable timescales and slope, varying in amplitude only. The physical processes giving rise to each component are thus of a permanent nature. A number of points of interest emerge from the results. The first component, with $\tau \simeq 1.3 \times 10^5$ s (active regions) shows an increasing trend in amplitude which is well correlated with the Ca II K-line index, an indicator of chromospheric activity. This is illustrated in the right-hand panel of Figure 3.3. The slope of the powerlaw is 3.8 (in good agreement with Andersen et al. 1998).

The observed correlation, which is further illustrated by the scatter plot in Figure 3.4, comes as no surprise. The passage of individual active regions across the disks of the Sun and other stars monitored by the Mt Wilson HK Project can be clearly seen in plots of the activity index S (from which R'_{HK} is derived) versus time[4]. On the other hand, the effect of the same type of event on the solar irradiance has been studied with a number of instruments, most recently VIRGO/LOI and PMO6 (Domingo et al. 1998). Recent models including contributions from faculae and sunspots of tunable size and number reproduce the PMO6 light curve to a high degree of precision (Krivova et al. 2003; Lanza et al. 2003).

However, observing and characterising a correlation throughout the Sun's activity cycle, between a chromospheric activity indicator which can be measured from the ground for a large number of stars, and total irradiance variations, whose amplitudes are so small they are only observable by dedicated high-precision photometric space missions, goes one step further. Most importantly for planetary transit searches, it implies that chromospheric activity indicators such as R'_{HK} can be used as a *proxy* to predict weeks timescale variability levels for a wide range of stars.

The amplitude of the second component also increases, but is not correlated to the Ca II index. This component corresponds to timescales corresponding to super- and meso- granulation, for which no detailed models are available to date. Our understanding of this kind of phenomenon is expected to improve dramatically when the results from space-based experiments designed for precision time-series photometry will become available.

3.2.3.2 Implications

The correlation between A_1 and the Ca II K-line index, although not extremely tight, is a clear indication that chromospheric activity indicators contain information about the variability level of the Sun on timescales longer than a few days. To establish more solidly a scaling law between photometric variability and chromospheric activity, we must use a wider stellar sample to constrain the relation over the entire expected range of activity levels (see Section 3.3.1.3).

Little useful information has been extracted from the solar data on what determines the parameters of the solar background other than A_1. The few clues available from other sources will be presented in Section 3.3.2.

[4]See the Mt Wilson HK project homepage, `http://www.mtwilson.edu/Science/HK_Project/`.

3.3 Empirical scaling to other stars

The main aim of the present model is the simulation of realistic light curves of stars more active than the Sun, if possible as a function of stellar parameters such as spectral type and age, via observables such as $B-V$ colour and rotational period $P_{\rm rot}$. The multi-component power-law model used in Section 3.2 to fit the solar background power spectrum now forms the basis of the simulation of enhanced variability light curves.

To simulate a light curve for a given star, the first task is to generate a power spectrum using this power-law model. When applying the component-by-component procedure described in Section 3.2.3 to fit solar power spectra, the optimal number of components was found to be three. We therefore use a sum of three power-laws to generate the stellar power spectra. The highest frequency component, a superposition of granulation, oscillations and white noise, has a characteristic timescale which is shorter than the typical sampling time for planetary transit searches, and thus can be replaced by a constant value(i.e. a random noise component). There are therefore a total of 7 parameters to adjust for each simulated power spectrum: three for each resolved power-law plus one constant.

The inputs of the model are the spectral type and age of the star. Starting from these theoretical quantities given, how are the power-law parameters deduced? Most of the information available concerns the amplitude of the first power-law, A_1. The scaling of this parameter is described in detail in Section 3.3.1, while that of the other parameters of the model, whose treatment is much simpler, is discussed in Section 3.3.2.

3.3.1 The amplitude of the active regions component

We have established in Section 3.2 that there is a correlation between A_1 and the Ca II K-line indicator of chromospheric activity in the Sun. We will see in this section that such a correlation holds for a wide stellar sample. On the other hand, there is a well known scaling between rotation period, colour and chromospheric activity (Noyes et al. 1984). Provided one can estimate the rotation period of the star (Section 3.3.1.1), the activity level can be computed (Section 3.3.1.2), and from this one obtains A_1 (Section 3.3.1.3).

3.3.1.1 The rotation period-colour-age relation

As detailed in Section 3.3.1.2, it is possible to deduce the expected activity level for a star of known mass (i.e. colour) and rotation period. However, the number of stars

with known rotation periods is relatively small. In the context of the present work, it would thus be useful to be able to predict the rotation period for a given stellar mass and age.

Observational constraints on rotation rates for stars of known mass and age come from two sources: star forming regions and young open clusters, where one can measure photometric rotation periods or rotational line broadening ($v \sin i$); and the Sun itself. There is little else, as rotational measurements are hard to perform for all the quiet, slowly rotating intermediate age and old stars other than the Sun, except for relatively nearby field stars, for which little reliable age information is available. The status of observational evidence and theoretical modelling in this domain is outlined by Krishnamurthi et al. (1997), Bouvier et al. (1997) and Stassun & Terndrup (2003). Here we briefly sketch the current paradigm to set the context of the present work.

The observed initial spread in rotation velocities (from measurements of T-Tauri stars, with a concentration around 10–30 $\mathrm{km\,s^{-1}}$ but a number of fast rotators, up to 100's of $\mathrm{km\,s^{-1}}$) is attributed to the competing effects of spin-up (due to the star's contraction and accretion of angular momentum from the disk) and slowing-down mechanisms such as disk-locking (Königl 1991; Bouvier et al. 1997). This spread is observed to diminish with age (by the age of the Hyades, only some M-dwarfs still exhibit fast rotation, Prosser et al. 1995), leading to a dependency of rotation on mass only. Following this homogenisation, one observes constant spin-down in a given mass range. For example, Skumanich (1972), comparing rotational velocities Sun-like (G) stars in the Pleïades, the Hyades and for the Sun, found them to decay as the square root of age, implying that the rotation period of of main sequence Sun-like stars increases as $t^{1/2}$ where t is the age on the main sequence. Both homogenisation and power-law spin-down can be explained by the loss of angular momentum through a magnetised wind (Schatzman 1962; Weber & Davis 1967), a mechanism which is more effective in faster rotators.

Angular momentum evolution models have vastly improved recently, but they still rely on the careful tuning of a number of parameters, especially for young and low-mass stars. We have therefore chosen to use empirically derived scaling laws and to restrict ourselves to the range of ages (older than the Hyades) and spectral types (mid-F to mid-K) where a unique colour-age-rotation relation can be established. In this range the aforementioned parameters become less relevant and the models reproduce the observations fairly robustly.

A relationship between $B-V$ colour (i.e. mass) and rotation at a given age was empirically derived from photometric rotation period measurements in the Hyades (Radick et al. 1987, 1995). Only rotation periods were used, rather than $v \sin i$ mea-

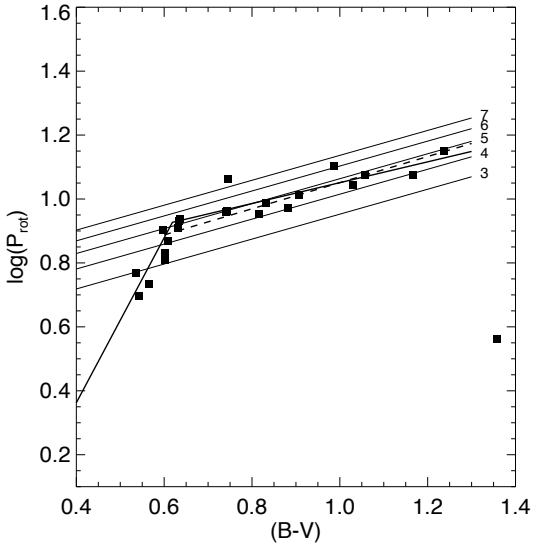

Figure 3.5: Plot of rotation period versus $B-V$ colour for Hyades stars. Data from Radick et al. (1987, 1995). The thin parallel black lines correspond to 'rotational isochrones' from Kawaler (1989) with t as indicated next to each line (in units of 10^8 yr). The dashed line is a linear fit to the data with $0.6 \leq B-V < 1.3$. The thick solid line is a composite of two linear fits, one for $B-V < 0.62$ and one for $0.62 \leq B-V < 1.3$.

surements, to avoid introducing the extra uncertainty of assigning random inclinations to the stars and having to assume theoretical radii to convert $v \sin i$ to a period. This relationship is valid for the range $0.45 \leq B-V \leq 1.3$. For stars bluer than $B-V = 0.45$, rotation rates saturate, but this is outside the range of spectral types of interest for the present work. Redder than $B-V = 1.3$, a significant spread is still observed in the rotation periods. The remaining range is divided into two zones, each following a linear trend. Redder than $B-V = 0.62$, the slope of the relation is quite close to the theoretical relation obtained by Kawaler (1989). We also estimated the age of the Hyades from the zero-point of the linear fit as done by Kawaler (1989), but incorporating the improved data from Radick et al. (1995). The age obtained in this manner is 634 Myr, consistent with recent determinations by independent methods: 655 Myr (Cayrel de Strobel 1990), 600 Myr (Torres et al. 1997), and 625 Myr (Perryman et al. 1998), thereby confirming the quality of the fit. The slope for stars bluer than $B-V = 0.62$ is much steeper, presumably due to the thinner convective envelopes of the stars in this range.

This can then be combined with the $t^{1/2}$ spin-down law into a rotation-colour-age relation:

$$\log(P_{\rm rot}) - 0.5 \log\left(\frac{t}{625\ {\rm Myr}}\right) = \left\{ \begin{array}{ll} -0.669 + 2.580\,(B-V), & 0.45 \leq B-V < 0.62 \\ 0.725 + 0.326\,(B-V), & 0.62 \leq B-V < 1.30 \end{array} \right\} \quad (3.2)$$

A comment on the adopted value of 0.5 for n_t, the index in the spin-down law

is appropriate. It has recently been suggested that a value of 0.6 might be more appropriate (Guinan & Ribas 2002, on the basis of an updated sample of Sun-like stars with some new age determinations). However, the original value of 0.5 was kept for the present work. The change would not affect the predicted rotation rates significantly, and the errors on this new value of n_t (which depends, for example, on age determinations from isochrone fitting) are larger than the difference. We have therefore kept the lower value, as it leads, if in error, to overestimated rotation rates, hence more variability and on faster timescales, and eventually conservative estimates of transit detection rates.

3.3.1.2 The activity-rotation period-colour relation

The next step consists in estimating from the colour and rotation period the expected chromospheric activity level of the star. For this purpose, the scaling law first derived by Noyes (1983) and Noyes et al. (1984) is used. It relates the mean Ca II index $\langle R'_{HK} \rangle$ to the inverse of the Rossby number R_o, and can thus be understood in terms of the interplay between convection, rotation and the star's dynamo:

$$-\log R_o = 0.324 - 0.400\, y + 0.283\, y^2 - 1.325\, y^3 \tag{3.3}$$

where $y = \log \langle R'_{HK} \rangle_5$ and $\langle R'_{HK} \rangle_5 = \langle R'_{HK} \rangle \times 10^5$. R_o is related to the rotation period P_{rot} and $B - V$ colour as follows:

$$R_o = \tau_c / P_{rot} \tag{3.4}$$

where P_{rot} is expressed in days, and the following (empirically derived) relation for τ_c, the convective overturn time, is used:

$$\log(\tau_c) = \begin{cases} 1.361 - 0.166\, x + 0.025\, x^2 - 5.323\, x^3, & x \geq 0 \\ 1.361 - 0.140\, x, & x < 0 \end{cases} \tag{3.5}$$

where $x = 1 - (B - V)$. Equation (3.3) was inverted (using an interpolation between tabulated values) to allow us to deduce the chromospheric activity index from the rotation period and $B - V$ colour of a star. For details of how the above relations, which are simply stated here, were obtained, the reader is referred to Noyes et al. (1984).

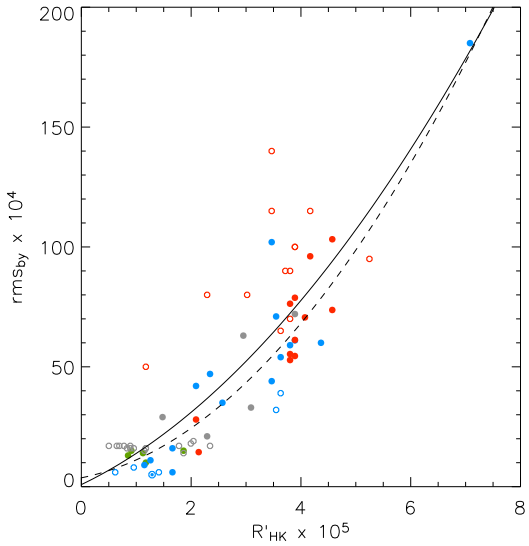

Figure 3.6: Photometric variability $[\mathrm{rms}_{by}]_4$ *versus chromospheric activity* $\langle R'_{HK}\rangle_5$ *for 72 field and Hyades stars. The colour coding refers to the source of the photometric variability data. Green: Henry et al. (2000b). Blue: Radick et al. (1998) (hollow: <10 observations per season). Grey: Lockwood et al. (1997) (hollow:* $\mathrm{rms}_{by} \leq 2$ *mmag, upper limits only). Red: Hyades stars (filled: Radick et al. 1995, hollow: averages of* rms_b *&* rms_y *from Lockwood et al. 1984 & Radick et al. 1987). Solid line: 2nd order polynomial fit to the data. Dashed line: relation used in Aigrain et al. (2004), based on 29 stars only.*

3.3.1.3 Active regions variability and chromospheric activity

Having noted that the 'active regions' component of the solar activity spectrum appears directly correlated to chromospheric activity (see Section 3.2.3), we turn to the small but valuable datasets containing both photometric variability measurements and activity indexes for a variety of stars. The data available on any one star is of course much less precise and less reliable than the solar data, but the dataset as a whole spans a much wider range of activity levels.

We use a sample of stars for which both $\langle R'_{HK}\rangle$ and photometric variability measurements are available in the literature. In Aigrain et al. (2004), we restricted ourselves to 29 stars whose variability levels and chromospheric activity indices were published in the same papers, namely Radick et al. (1998); Henry et al. (2000b). These data were then used to derive a quantitative relationship between $\langle R'_{HK}\rangle_4$ and the night-to-night rms variability in Strömgren b and y ($[\mathrm{rms}_{by}]_4 = \mathrm{rms}\{(b+y)/2\} \times 10^4$, in magnitude units).

Since then, a wider literature search has revealed a number of additional stars for which photometric variability and chromospheric activity information were available, but in separate sources: additional photometric variability data were taken from Lockwood et al. (1984); Radick et al. (1987, 1995); Lockwood et al. (1997), while activity data, when not found in the above sources, were taken from Barry et al. (1987); Duncan et al. (1991); Garcia-Lopez et al. (1993); Baliunas et al. (1995). The full sample, illustrated in Figure 3.6, now contains 72 stars.

At the same time as including the additional data in the calibration, additional care was taken to ensure consistency between datasets from different sources and

to minimise systematic errors.

Variability measurements

- Sparse time sampling: Some of the sources used, for example Radick et al. (1998), were mainly designed to study long term variations (similar to the Sun's activity cycle) and their observations were rather sparse, sometimes counting as few as 5 or 6 points per observing season. This could lead to an underestimate of, or in any case a larger uncertainty in, the short-term (night-to-night within 1 season) rms values, which were those used in Aigrain et al. (2004). In order to limit the impact of sparse time sampling, the average of night-to-night rms values from different observing seasons was used, including only seasons with more than ten observations whenever the number of observations per seasons was given. Lockwood et al. (1997) and Radick et al. (1998) gave only the total number of observations and the number of seasons. For those, stars for which the ratio of the two was less than 10 were flagged, to see if these appeared systematically less variable than stars observed more frequently.
- Exclusions: As was done in Aigrain et al. (2004), HD 95735, which was monitored by Henry et al. (2000b), was excluded from the analysis as it lies outside the range of $B - V$ colours considered in the model.
- Sensitivity limits: In the case of data from Lockwood et al. (1997), it should be noted that rms values $\leq$ 17 mmag are at the detection limit for variability and should be considered as upper limits only.
- Filters: Lockwood et al. (1984) and Radick et al. (1987) published *b* & *y* rms values separately. In later papers they noted the great similarity between the *b* & *y* light curves and averaged the two before computing rms values to improve precision. In the present work, when including data from Lockwood et al. (1984) and Radick et al. (1987), the average of the *b* and *y* rms values was used. These stars were also flagged, to highlight the fact that the rms values were not obtained in a totally consistent way with the others.
- Multiple appearances of a given star: When a given star was present in more than one source, the latest source with more than 10 observations per season was used.

Chromospheric activity measurements It is appropriate to mention here some precautions that were taken in the compilation of chromospheric activity index values. Two formulae for the temperature correction factor are commonly used (Middelkoop 1992; Rutten 1984), and published $R'_{\rm HK}$ values have been derived with both.

To ensure consistency, whenever possible S, the uncorrected quantity from which R'_{HK} is derived, was collected from the literature, and the correction factors of Noyes et al. (1984) were applied in all cases. They are somewhat intermediate between Middelkoop (1992) and Rutten (1984) and also include the photospheric correction. Several groups outside Mount Wilson Observatory have developed instrumental setups to measure the flux in the Ca II H & K lines that differ slightly from those of Mount Wilson, or setups to measure flux in other lines that are chromospheric activity indicators. In most cases, they derived conversions from their systems to S or R'_{HK}. These indexes were used for stars where no Mt Wilson system data was available, but care was taken to retain consistency in the temperature and photospheric correction factors used.

As illustrated in Figure 3.6, a 2nd order polynomial provides a good fit to the relationship between $\left[\mathrm{rms}_{by}\right]_4$ and $\left\langle R'_{HK}\right\rangle_5$. The equation of the fit is:

$$\left[\mathrm{rms}_{by}\right]_4 = 0.87 + 10.96\ \left\langle R'_{HK}\right\rangle_5 + 2.06\ \left\langle R'_{HK}\right\rangle_5^2 \tag{3.6}$$

The hollow symbols on Figure 3.6 indicate that, for one reason or another, the rms values are expected to be less reliable. As suspected, stars from Radick et al. (1998) with fewer than 10 observations per season appear on the lower envelope of the rms versus activity level trend (hollow blue circles). Also, some low activity Lockwood et al. (1997) stars that could be considered anomalously variable are in fact only upper limits (hollow grey circles). Finally, stars from Lockwood et al. (1984); Radick et al. (1987), for which the rms values were obtained by averaging single filter values, appear systematically high (hollow red circles). This could be explained if the individual rms values contain a significant contribution from measurement errors, as well as intrinsic variability. The rms of the average of the two light curves (the desired quantity) would then be systematically lower than the average of the rms of the two light curves. However, including or excluding these points does not not affect the fit significantly, so they have been kept in Equation 3.6. The older fit, based on 29 stars only, is shown as the dashed line in Figure 3.6. There has been very little change over the range of chromospheric activity levels of interest, which bodes well for the robustness of the relation.

Note that one star, EK Dra, is significantly more active and variable than the rest of the sample – because it is younger. Including it in the fit thus gives it a disproportionate weight, but we have included it because it broadens the range of $\left\langle R'_{HK}\right\rangle$ covered. Furthermore, previous studies have shown (Messina & Guinan 2002) that it fits tightly on relationships between activity and stellar parameters derived from samples of solar analogues of various ages, suggesting its behaviour is representative of the mechanisms driving activity in general – and thus, in our reasoning, variability on

the timescales under consideration.

To be usable for our purposes, Equation (3.6) must be completed by a relationship between $\left[\mathrm{rms}_{by}\right]_4$ and A_1. The desired value of A_1 corresponds to white light relative flux variations, not *b* and *y* magnitude variations. To obtain the rms of the relative flux variations one must multiply the rms of the magnitude variations by a constant factor of $2.5/\ln(10) = 1.08$. This must then be converted to a white light flux rms value. This requires the definition of a reference point in the solar cycle, at which to compare the variability levels in the two bandpasses. Radick et al. (1998) quote a single $\log\left\langle R'_{\mathrm{HK}}\right\rangle$ value of -4.89 for the Sun, which is an average of measurements performed over many years. This defines a reference solar activity level, corresponding roughly to a third of the way into the rising phase of cycle 23. The average night-to-night variability as measured in *b* and *y* by Radick et al. (1998) is $\mathrm{rms}_{by} = 5 \times 10^{-4}$ mag. The corresponding white light variability level, measured from a 6 month long section of PMO6 data downgraded to 1 day sampling, centred on the date for which the measured activity level was equal to the reference level defined above, is $\mathrm{rms}_{\mathrm{white}} \approx 1.8 \times 10^{-4}$ mag.

This allows us to convert from rms_{by} in magnitudes to $\mathrm{rms}_{\mathrm{white}}$ in relative flux. Assuming a straightforward proportionality relationship the conversion factor is 2.78. This conversion introduces a significant error in the overall conversion, as we observed that the dependency of $\mathrm{rms}_{\mathrm{white}}$ on R'_{HK} (in the Sun) is slightly different in shape to that of rms_{by} on $\left\langle R'_{\mathrm{HK}}\right\rangle$ (in the stellar sample), so a proportionality factor is inaccurate. However, until other stellar photometric time series that are sufficiently regular to perform the fitting process described in Section3.2.3 are available, it is the best we can do. Finally one must convert from $\mathrm{rms}_{\mathrm{white}}$ to A_1. As expected, a linear relationship between these quantities as computed for the Sun is observed, yielding an overall conversion between $\left[\mathrm{rms}_{\mathrm{by}}\right]_4$ and A_1:

$$A_1 \times 10^5 = -0.24 + 0.66 \ \left[\mathrm{rms}_{by}\right]_4 \tag{3.7}$$

Equation (3.6) thus becomes:

$$A_1 \times 10^5 = 0.33 + 7.23 \ \left\langle R'_{\mathrm{HK}}\right\rangle_5 + 1.36 \ \left\langle R'_{\mathrm{HK}}\right\rangle_5^2 \tag{3.8}$$

As space-based time-series photometric missions come online, we will be able to calibrate the relations above with more and more stars. Amongst these missions which have already provided useful data or will start doing so very soon are MOST (Micro-variability and Oscillations of Stars, Walker et al. 2003), a small Canadian mission which saw first light on 20 July 2003, and the OMC (Optical Monitor Camera, Giménez et al. 1999) on board ESA's new γ-ray observatory INTEGRAL. In the longer

term, COROT will also provide a wealth of information on stellar micro-variability.

These new data will allow us to calibrate the relations above with a wider variety of stars. In particular, chromospheric activity and rotational period measurements are available for a relatively large number of bright late type stars (Henry et al. 1996; Baliunas et al. 1996; Radick et al. 1998; Henry et al. 2000b; Tinney et al. 2002; Paulson et al. 2002). If any of these are observed by the missions listed above, yielding variability measurements, they will be incorporated in the present model.

3.3.2 Other parameters of the model

The previous sections were concerned with providing an estimate of one of the model's parameters, A_1, given a star's age and colour. However, there are a total of 7 parameters to adjust. We have so little information on the super- and meso-granulation component in stars other than the Sun that we have chosen, for now, to leave it unchanged in the simulations, using the solar values.

3.3.2.1 The third component: timescales of minutes or less

The third (highest frequency) component observed in the Sun is a superposition of variability on timescales of a few minutes, which is thought to be related to granulation, and higher frequency effects such as oscillations and photon noise. The distinction between these effects is not resolved at the time sampling used for the present study. There is very little information at the present time on equivalent phenomena in other stars than the Sun (see below). Solar values were therefore used in all simulated light curves for this component. Given the low amplitude of this component, and the fact that it corresponds to timescales significantly shorter than the duration of planetary transits, it should not affect transit detection significantly.

Granulation can be traced by studying asymmetries in line bisectors, leading to the possible exploration of this phenomenon across the HR diagram (see for example Gray & Nagel 1989). Trampedach et al. (1998), who modelled the granulation signal for the Sun, α Cen A and Procyon, obtained very similar power and velocity spectra for the three stars, despite the fact that convection is much more intense on Procyon. Although very preliminary, these results suggest that the granulation power may change only slowly with stellar parameters, thus supporting the use of the solar values in the present model.

3.3.2.2 The second component: hours timescale

Again, solar values were used for all simulated light curves for this component. Work is underway to identify the types of surface structures giving rise to the hours-timescale variability in the Sun (Fligge et al. 2000), and the launch of the COROT mission will provide a dataset ideally suited to improving our understanding of this type of variability. Keeping the super- and/or meso-granulation component identical to the solar case is certainly an oversimplification, and it is the area where most effort will be focused in the future, as it is highly relevant to transit detection, being on timescales similar to transits.

The possibility of measuring the power spectrum of stochastic luminosity variations in stars other than the Sun on timescales of minutes to hours, using existing data from the star-tracker camera of NASA's WIRE (Wide Field Infrared Explorer), is also under investigation[5]. However, difficulties associated with non-Gaussian noise and frequent data gaps both short and long, have impeded such a measurement so far (see Section 3.4.3.4).

3.3.2.3 Timescale of the (first) active regions component

Two parameters remain for the active regions component. We have kept the slope of the power law, C_1, unchanged from the solar case. Changing it slightly does not seem to affect the appearance of the light curve significantly. More crucial is the timescale B_1.

If the active regions component of micro-variability is the result of the rotational modulation of active regions, we expect B_1 to be directly related to the period. However, the value obtained for the Sun is $B_1 = 8.5 \times 10^6$ s, i.e. 9.84 days, compared to a rotational period of ≈ 26 days. This suggests that the timescale is not (or not exclusively) dominated by rotational modulation of active regions, but by the emergence and disappearance of the structures composing the active regions. Individual active regions evolve on a timescale of weeks to months in the Sun (Radick et al. 1998), but spots and faculae evolve faster: observed sunspot lifetimes range roughly between 10 days and 2 months (Hiremath 2002). We therefore deduce that sunspot evolution is the process which determines B_1 in the Sun.

One can in fact interpret the observed timescale purely in terms of rotational modulation: the transit of a given spot across the solar disk would last half the rotation period only, so that, given the shape of a sunspot transit signature, pure rotational modulation could contribute to signal on timescales as short as a third of the rotation

[5]This data has already been used to perform asteroseismology on a number of bright stars (Schou & Buzasi 2001) following the failure of the main instrument shortly after launch.

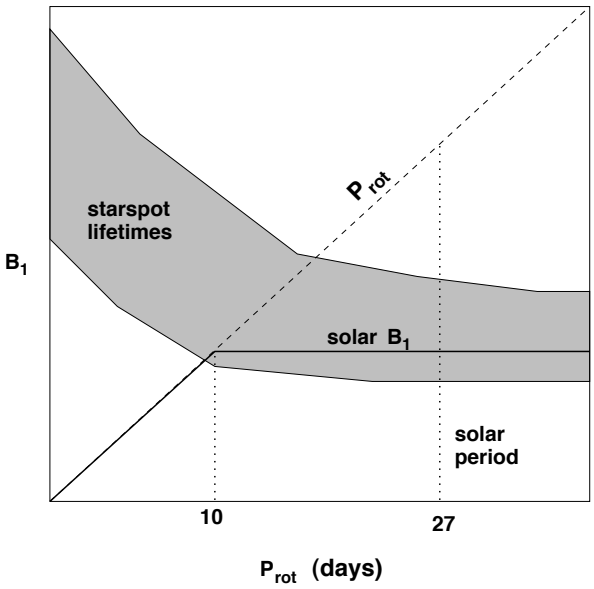

Figure 3.7: Schematic illustration of the expected dependence of the active regions component's characteristic timescale B_1 on rotational period.

period, approximately equal to the measured B_1. This serves only to highlight the need for direct observations of this timescale in other stars.

Whatever the dominating factor in the Sun, the phenomenon with the shortest timescale (in the relevant range) is expected to determine B_1 for any given star. The present version of the model assumes that, in the Sun, this phenomenon is sunspot evolution. As spot lifetimes are, if anything, longer in faster rotators (Barnes et al. 1998; Soon et al. 1999), rotational modulation should take over below a certain period. As a rough estimate we have placed the boundary between the two regimes at $P_{\rm rot}$ = 10 days (see Figure 3.7). If on the other hand the important phenomenon is rotation, but a given rotation period gives rise to a timescale $B_1 \simeq P_{\rm rot}/3$, then this rotationally dominated regime could dominate up to rotation periods of at least ~ 30 days. Such a modification would be trivial to implement should upcoming stellar data with sufficiently long time baselines, such as those from MOST and COROT, warrant it.

This completes, within the obvious limits of the assumptions used, the requirements for the simulation of white light stellar light curves with micro-variability, within the range of applicability of the scaling laws used: $0.45 \leq B-V \leq 1.3$, $t \geq t_{\rm Hyades}$, and the star must still be on the main sequence. The stellar parameters required are age (or rotation period) and $B-V$ colour (or spectral type). It is also possible to supply $R'_{\rm HK}$, or A_1 and B_1 directly. Once an artificial power spectrum is generated, phases drawn at random from a uniform distribution are applied before applying a reverse Fourier transform to return to the time domain.

Due to the use of randomly chosen phases, the shape of the variations does

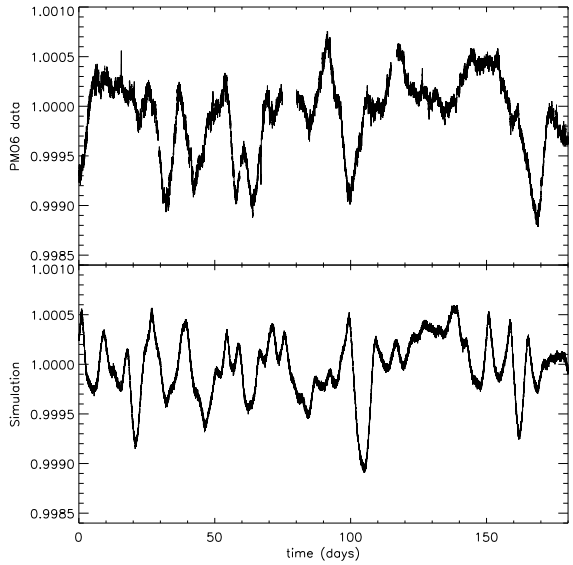

Figure 3.8: Comparison of a portion of PMO6 data starting 1300 days after the start of the full light curve (top panel) with a simulated light curve generated using the Sun's observed rotation period (25.4 days) and chromospheric activity index ($R'_{HK} = -4.89$, Radick et al. 1998). Both light curves have 15 min sampling, last 180 days and are normalised to a mean flux of 1.0.

not closely resemble the observed solar variability. It may be possible to characterise the sequence of phases characteristic of a given type of activity-related event, such as the crossing of the stellar disk by a star-spot, faculae or active region. This information could then conceivably be included, with appropriate scaling, in the model, in order to make the shape of the variations more realistic. However, how to do this in practice is not immediately obvious, and will be investigated in the future. As the model stands, the simulated light curves can be used to estimate quantities such as amplitude, timescale, the distribution of residuals from a mean level, but no conclusions should be drawn from the shape of individual variations.

3.4 Testing the model

3.4.1 Mimicking the Sun

A first check is to compare the predicted observables for the Sun to the measured values. The measured period and $\log \langle R'_{HK} \rangle$ are 25.4 days and -4.89 (Radick et al. 1998), in good agreement with the predicted values of 23.4 days and -4.84. The measured rms_{by} is $\simeq 5$ in units of 10^{-4} mag. The value measured from a simulated light curve with 1 day sampling lasting 6 months, when allowing for the conversion between $\mathrm{rms}_{\mathrm{white}}$ and rms_{by} is 6.11. This slight over-prediction is attributable to the Sun's slightly slow rotation and low activity for its age and type, and to the fact that it is slightly under-variable for its activity level (it falls slightly below the fit on Figure 3.6). Figure 3.8 compares a portion of the PMO6 light curve taken from a relatively high

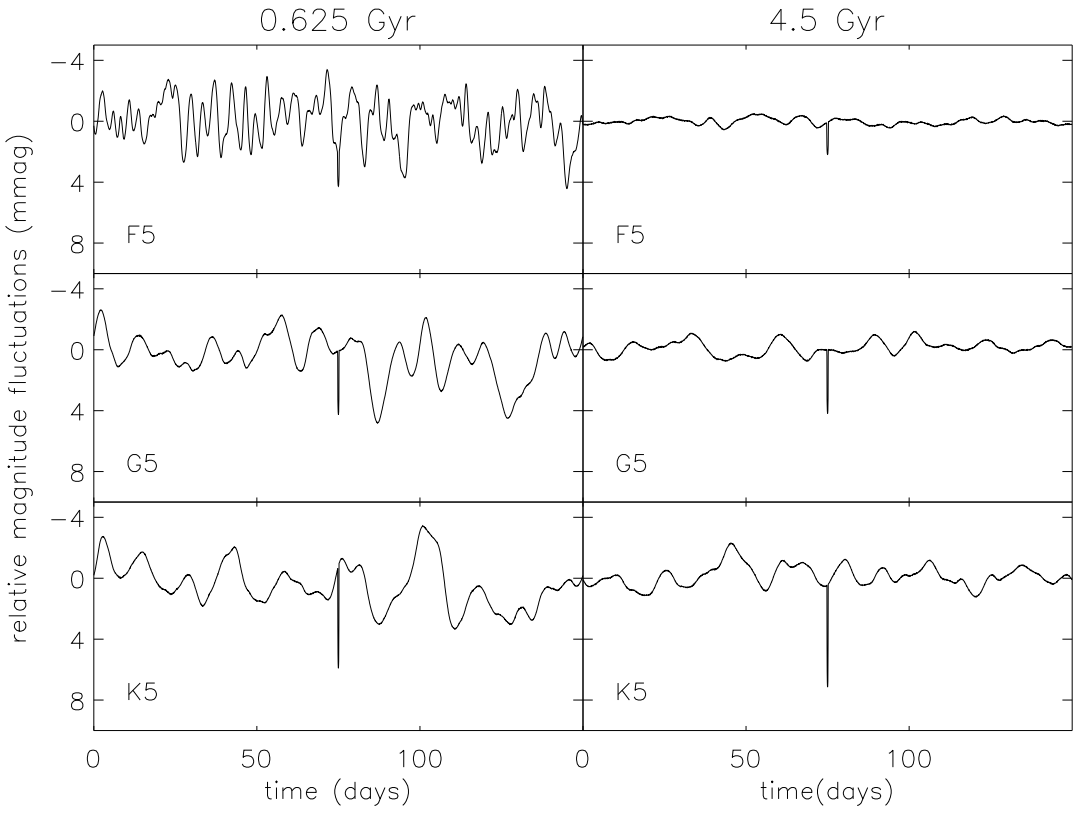

Figure 3.9: Examples of simulated light curves containing micro-variability for Hyades age (left column) and solar age (right column) stars with spectral types F5 (top row), G5 (middle row) and K5 (bottom row). A single transit by a 0.5 $R_{\rm Jup}$ planet has been added to each light curve 75 days after the start (a transit by an Earth-sized planet would be $\approx$ 25 times smaller). The light curves have 1 hr sampling and last 150 days.

activity part of the solar cycle with a light curve simulated using the observed rotation period and activity index of the Sun. The amplitude and typical timescales of the variations are well matched.

3.4.2 Trends with age and mass

A set of six light curves have been simulated, corresponding to three spectral types (F5, G5 and K5) and two ages (625 Myr and 4.5 Gyr). Examination of the light curves and the various parameters computed during the modelling process can reveal any immediate discrepancies. The light curves are shown in Figure 3.9 and the parameters in Table 3.2. They follow the expected trends, variability decreasing with age and $B-V$ and increasing with $P_{\rm rot}$, so that at solar age age the least active star is the F star, while the most active is the K star, despite its long rotation period (due to the dependence of activity on colour).

Table 3.2: Parameters of the simulated light curves.

Age Gyr	SpT	$B-V$	P_{rot} days	$\log\left(R'_{HK}\right)$	A_1 $\times 10^5$	B_1 days	rms_{white} $\times 10^4$
0.625	F5	0.44	2.9	-4.64	19.27	2.89	22.0
0.625	G5	0.68	8.8	-4.44	38.16	8.80	46.5
0.625	K5	1.15	12.6	-4.42	41.40	9.84	69.5
4.5	F5	0.44	7.8	-5.23	4.66	7.87	13.3
4.5	G5	0.68	23.7	-4.80	11.95	9.84	13.9
4.5	K5	1.15	33.8	-4.67	17.53	9.84	14.4

3.4.3 Behaviour at high activity

Comparing the light curves shown in the left column of Figure 3.9 which have amplitudes of $\sim 0.5\,\%$, to published V-band amplitude measurements for Hyades stars (Messina et al. 2001, 2003), which are of the order of $\sim 3\,\%$, immediately highlights a discrepancy. This discrepancy could be explained by a number of factors:

3.4.3.1 Bandpass

Part of the difference is readily explained by the fact that the model produces white light flux variations, while the amplitudes reported by Messina et al. were V-band magnitude variations. As previously mentioned, white light flux variations in the Sun are observed to be ≈ 2.78 times smaller that b & y magnitude variations, and a similar effect is expected with V. To estimate the amplitude of such an effect requires the comparison of simultaneous light curves in V and either b & y or white light. S. Messina (priv. comm.) did this comparison and found near-identical variability levels both in amplitude and in rms. The bandpass effect could therefore account for no more than a factor of ~ 2.8, which is not enough to explain the observed discrepancy.

3.4.3.2 Underestimated rms values at high activity

As outlined in Section 3.3.1.3, the activity-variability relation remains ill-constrained at high activity levels, with only one data point with $\left[R'_{HK}\right]_5 > 5.5$. Although many stars have been used to calibrate the relation over the range of activity levels typical of the Hyades – $1.6 < \left[R'_{HK}\right]_5 < 5$ – there is a lot of scatter over that range and different sources of data show different trends, indicating that systematics are still present. However, given that the inclusion or exclusion of such "dubious" datasets in the calibration hardly alters the scaling law, this is unlikely to be a significant factor.

3.4.3.3 Amplitudes versus rms

Simulated and observed light curves for the Sun were compared in Section 3.4.1, and both rms values and amplitudes are consistent. For younger stars, the calibration of the scaling law used should ensure that the rms of the simulated light curves approximately agrees with observations (see previous paragraph). However, if the amplitude scales differently from the rms as the activity level increases, the model could produce sensible rms values but underestimate the amplitude for active stars. This is possible if the statistical nature of the variability changes from relatively stochastic (dominated by the emergence, evolution and disappearance of spots, as we think is the case in the Sun) to close to sinusoidal (dominated by rotational modulation of small numbers of large, persistent active regions, as is generally thought to be the case in young active stars). In this case, a new component, more concentrated in period space around the stellar rotation period, could be added to the simulated power spectra.

3.4.3.4 WIRE time series data of α Centauri

The WIRE[6] or Wide field InfraRed Explorer is a NASA satellite designed to perform sky surveys in the IR, whose original goals could not be fulfilled due to loss of detector coolant early in the mission. However, the rest of the satellite is in good condition and its 5 cm aperture star-tracker telescope has been used successfully for asteroseismic studies of a number of stars, including α Cen (Schou & Buzasi 2001). α Cen is a particularly interesting target to search for low-frequency variability such as is observed in the Sun, as it is a well studied binary whose primary has a spectral type and age close to that of the Sun. Recently, Kjeldsen et al. (1999) measured excess power in ground-based radial velocity observations of it the range 600 to 3000 mHz, to which they fit a powerlaw with a slope of -1.46, and which they interpret as granulation.

An attempt was made to use the WIRE time series, which lasts 50 days with approximately 40 % duty cycle (observations are taken during just under half of each 102 min orbit) to perform an analysis similar to that which was done for the solar PMO6 data, fitting a multi-component broken powerlaw model to the power spectrum to obtain further constraints on the various parameters of our micro-variability model. However, the complex window function could induce many features in the power spectrum. To check for these, a Fisher randomisation test was carried out: a 'scrambled' dataset was constructed by drawing samples in random order from the original dataset, keeping the same window function. There were no noticeable differences between the power spectra of the scrambled and unscrambled datasets.

[6] http://www.ipac.caltech.edu/wire/

This implies that any features in the power spectrum of the unscrambled dataset are in fact due to noise and the effects of the time sampling.

As α Cen was the WIRE target with the most complete light curve to date, this suggests that little useful information relating to stellar micro-variability on the timescales of interest for transit searches is likely to be extracted from WIRE data. The first measurements of this type of variability in other stars than the Sun (excluding active stars) are expected to become available in the near future thanks to the MOST (Micro-variability and Oscillations of Stars) satellite.

3.5 Discussion

A model to generate artificial light curves containing intrinsic variability on timescales from hours to weeks for stars between mid-F and late-K spectral type and older than 0.625 Myr has been presented. This model relies on the observed correlation between the weeks timescale power contained in total solar irradiance variations as measured by VIRGO/PMO6 and the Ca II K-line index of chromospheric activity. Except for the most active cases, the resulting light curves appear consistent with currently available data on variability levels in clusters and with solar data. Further testing and fine-tuning requires high sampling, long duration space-based stellar time-series photometry and will be carried out as such data become available.

The simulated light curves can be used to test the effectiveness of pre-transit search variability filters (see Chapter 4) and to estimate the impact of micro-variability on exo-planet search missions such as COROT, *Eddington* and *Kepler* (see Chapters 5 and 6).

All the simulated light curves produced so far, as well as the routines used to generate them, have been made available to the exo-planet community through the web page: `www.ast.cam.ac.uk/~suz/simlc`. Light curves with specific parameters can be generated on request. The model has been included in light curve simulation tools developed for the COROT mission.

It is important to stress that the present approach, which is highly simplified and empirically based (and which will be referred to hereafter as 'the simulator', for clarity), differs from, but complements, more detailed theoretical models of the complex physical mechanisms that give rise to the observed micro-variability of the Sun (here after 'detailed models'). There has been recent progress in the latter area: for example, Seleznyov et al. (2003) successfully model the full power spectrum of the VIRGO/PMO6 observations by combining the magnetic activity model of Krivova et al. (2003) (which relies on identifying active regions in resolved solar disk images and modelling their photometric signature) and a simple granulation model. Scaling

this approach to other stars is non-trivial, but there will be much to learn from detailed models about how the physical mechanisms at work vary from star to star when upcoming high precision stellar photometric time series data become available. On the other hand, our simulator has already been put to use to simulate light curves for a wide range of spectral types and ages, thereby providing insights relevant to the design of transit search missions and associated data analysis techniques. However, its simplicity naturally limits its information content. Its applications rest more on the statistical side: provided the overall trends are correct, it can be used, for example, to place planetary transit detection limits for a given type of parent star.

Future improvements of the simulator will be two-fold. Of course, direct fitting of the power spectra of other stars than the Sun, from MOST or COROT, will provide direct constraints. In parallel, regions of parameter space not yet covered by these missions can be explored using detailed theoretical models, to unveil the dependence of the power spectrum on the parameters of the models. If the dependence of those on stellar parameters is known, or if reasonable assumptions to that effect can be made, the simulator can be adjusted to reproduce similar trends in the power spectra it generates.

Another step currently under consideration is the extension of the simulator to produce light curves in specific spectral band passes. This would be very useful in the context of COROT (and potentially *Eddington*) whose design includes the use of colour information, both for target selection and the preparation of colour-discriminant light curve analysis tools.