Modeling Ideas: Transit Modeling for Kepler Light Curves

Paul Baines

June 21, 2013

Disclaimer: These are just preliminary ideas. There are likely to be typos etc. throughout the document. Use at your own risk. ☺

1 Modeling Simple Aperture Photometry Kepler Light Curves

Consider a random process Y defined as a function of time and depending on unknown parameter(s) θ i.e., $Y(t|\theta)$. For simplicity let $\theta = (\eta, \psi, \xi)$ and $\mathcal{T}$ be the interval on which the process is defined. We assume that:

$$Y_{sap}(t) = m_{sap}(t|\phi, \eta, \xi) + \epsilon_{sap}(t|\psi), \qquad \forall\, t \in \mathcal{T} \tag{1}$$

$$m_{sap}(t|\phi, \eta, \xi) = (\mu(t|\phi) + a(t|\xi)) + (1 - \tilde{q}(t|\eta)), \tag{2}$$

where $\epsilon(t|\psi)$ is a mean-zero random process, $\mu(t|\phi)$ is an overall mean corresponding to low-frequency image artifacts (i.e., the objects to be removed by detrending), $a(t|\xi)$ is a random process designed to capture any remaining non-transit-related signals (probably astrophysical) remaining after detrending, and $\tilde{q}(t|\eta)$ is the reduction in the expected intensity due to the transit i.e., $\tilde{q}(t|\eta) \equiv 0$ if there are no transits. In the event of a transit we observe a version of $m(t|\xi)$ 'dampened' by an additive amount $\tilde{q}(t|\eta)$. Typically, for detrended data the component $\mu(t|\phi)$ is divided out to give:

$$Y_d(t) = \frac{Y_{sap}(t)}{\mu(t|\phi)} = m_d(t|\phi, \eta, \xi) + \epsilon_d(t|\phi, \psi), \qquad \forall\, t \in \mathcal{T} \tag{3}$$

$$m_d(t|\phi, \eta, \xi) = \left(1 + \frac{a(t|\xi)}{\mu(t|\phi)}\right) + \frac{1 - \tilde{q}(t|\eta)}{\mu(t|\phi)} \equiv (1 + f(t|\phi, \xi)) + (1 - q(t|\phi, \eta)),$$

$$\epsilon_d(t|\phi, \psi) \equiv \frac{\epsilon_{sap}(t|\psi)}{\mu(t|\phi)}, \qquad f(t|\phi, \xi) = \frac{a(t|\xi)}{\mu(t|\phi)}, \quad 1 - q(t|\phi, \eta) = \frac{1 - \tilde{q}(t|\eta)}{\mu(t|\phi)}. \tag{4}$$

In this detrended setting the interpretation of q switches to a multiplicative reduction in the signal, and $f(t|\phi, \xi)$ models the astrophysical processes (and other artefacts not removed by detrending).

2 Modeling Detrended Kepler Light Curves

For simplicity we now focus on fitting detrended light curves of the form shown in (3) and drop the subscripts for notational brevity. Our model for this section will be:

$$Y(t) = (1 + f(t|\xi)) + (1 - q(t|\eta)) + \epsilon(t), \qquad \forall\, t \in \mathcal{T}. \tag{5}$$

There is a trade-off in the complexity we place in f and in ϵ, but for our purposes we assume a mean-zero Gaussian process for ϵ:

$$\epsilon(t)|\psi \sim \mathrm{GP}\left(0, \Sigma_\epsilon(\psi)\right), \tag{6}$$

where $\Sigma_\epsilon(t_i, t_j|\psi) = K(|t_i - t_j|; \psi)$ i.e., stationary covariance. Although the form in (6) is fairly rich, in practice we will usually assume either independent noise with constant variance, or independent noise with different variances for each quarter to reflect CCD changes i.e.,

$$\Sigma_\epsilon(t_i, t_j|\psi) = \sigma^2_{q(t_i)}\delta_{t_i t_j}$$

where $q(t_i)$ is the index corresponding to the quarter in which time point i falls. Although we only observe a noisy version of $(1+f)(1-q)$, the goal is to 'separate' this underlying mean into two pieces: one representing 'transit-like variation', the other representing 'everything else'. Fortunately we are able to approximate the form of q parametrically. Let t_0 denote the time of the first transit, t_d the transit duration, α the multiplicative reduction in the signal and P the period. Then we assume:

$$q(t|\eta) = \begin{cases} 0 & \text{if } (t \bmod P) \notin [t_0, t_0 + t_d] \\ \alpha & \text{if } (t \bmod P) \in [t_0, t_0 + t_d] \end{cases},$$

where $\eta = \{t_0, t_D, \alpha, P\}$. More sophisticated models for the transit could also be used to allow for smooth transits. The remaining modeling task is to decide upon a structure for f. The complexity of f and ϵ is heavily dependent on the detrending process, and thus we would prefer to jointly model the detrending and transit signal, something we discuss in section 3. Even with aggressive detrending we still would like to have flexible models are required to capture all possible remaining instrumental and astrophysical signals. To try to achieve this, we decide to model f in the wavelet domain. For computational tractability we apply wavelet transforms to each quarter of data separately. Let y_r, f_r and q_r denote subvectors of y, f and q that correspond to quarter $r = 1, \ldots, R$.

$$Y_r(t) = (1 + f_r(t|\xi)) + (1 - q_r(t|\eta)) + \epsilon_r(t), \tag{7}$$

$$\Rightarrow \quad (y_r - 1) - (1 - q_r) = f_r + \epsilon_r,$$

$$\Rightarrow \quad (y_r - 1) - f_r = 1 - q_r + \epsilon_r, \tag{8}$$

Taking the DWT, $\mathcal{W}$, of both sides gives:

$$\tilde{y}_r(q_r) = w_r + \epsilon_r, \quad \text{and}, \quad \tilde{y}_r(f_r) = (1 - q_r) + \epsilon_r, \tag{9}$$

where $w_r = \mathcal{W}f_r$, $\tilde{y}_r(q_r) = \mathcal{W}((y_r - 1) - (1 - q_r))$ and $\tilde{y}_r(f_r) = \mathcal{W}((y_r - 1) - f_r)$. Since $\mathcal{W}\mathcal{W}^T = I$, this yields:

$$\tilde{y}_r(q_r)|w_r \sim N\left(w_r, \sigma_r^2 I\right), \quad \text{and}, \quad \tilde{y}_r(f_r)|f_r \sim N\left(1 - q_r, \sigma_r^2 I\right). \tag{10}$$

To pool information across quarters we assume a hierarchical structure for the w_r:

$$w_r|d \overset{ind}{\sim} N\left(Ad, A\Sigma_w A^T\right), \tag{11}$$

where A is a $n_q \times n_d$ matrix that allows for extra pooling within wavelet scales across time periods. If A is the identity matrix then d is an $n_q \times 1$ vector, and the full model contains $n_q + p$ parameters where n_q is the number of time points per quarter (≈ 4320) and p is the number of additional parameters. In this context, in the special case that A contains zeros and a single one per row, then Ad is a vector that contains one of the d_j's in each position, and $A\Sigma_w A^T$ is an $n_q \times n_q$ diagonal matrix with one of the d_j's on the diagonal. The parameters of the model are therefore $\theta = \{\psi, \eta, d, \Sigma_w\}$ where $\psi = \left\{\sigma_1^2, \ldots, \sigma_r^2\right\}$ and $d = \{d_1, \ldots, d_{n_d}\}$ and $\Sigma_w = \left\{\sigma_1^2, \ldots, \sigma_{n_d}^2\right\}$. With 4 transit parameters, the model contains a total of $4 + R + 2n_d$.

2.1 Computational Details

The likelihood for this version of the model is given by:

$$\begin{aligned} p(y, w|\theta) &\propto p(y|w, \eta, \psi)p(w|d, \Sigma_w), \qquad (12) \\ \Rightarrow \qquad p(y|\theta) &= \int p(y, w|\theta)dw. \end{aligned}$$

The full Bayesian version simply requires a prior on all components of θ and may be preferable,

$$\begin{aligned} p(\theta, w|y) &\propto p(y|w, \eta, \psi)p(w|d, \Sigma_w)p(\psi, \eta, d, \Sigma_w), \qquad (13) \\ \Rightarrow \qquad p(\theta|y) &= \int p(w, \theta|y)df, \end{aligned}$$

The form of (13) lends itself to a natural Gibbs sampler of the form:

$$[\psi|w, \eta, y], \quad [\eta|\sigma^2, w], \quad [w|\eta, \psi, d, \Sigma_w], \quad [d|\eta, \psi, w, \Sigma_w], \quad [\Sigma_w|d, w].$$

Combining (10) and (11) we obtain:

$$\tilde{y}_r(q_r)|\psi, \eta \sim N\left(Ad, W_i^{-1}(\Sigma_w, \sigma_r^2)\right), \qquad (14)$$

$$W_r = \left[\sigma_r^2 I + A\Sigma_w A^T\right]^{-1}, \qquad (15)$$

which allows us to sample from the conditional posterior of all elements of d directly, without conditioning on w. This is analogous to methods for sampling in normal-normal hierarchical models.

Sampling ψ: Placing a conjugate prior on σ_r^2 of the form $\sigma_r^2 \sim \text{Inv-}\chi^2(\nu_{r,0}, s_{r,0}^2)$ we can integrate out w to obtain the partially marginalized posterior.

$$\sigma_r^2|d, \eta, \Sigma_w \sim \text{Inv-}\chi^2(\nu_{r,n}, s_{r,n}^2),$$

where

$$\begin{aligned} \nu_{r,n} &= \nu_{r,0} + n_d, \\ s_{r,n}^2 &= \frac{\nu_{r,0}s_{r,0}^2 + (\tilde{y}_r(q_r) - w_r)^T(\tilde{y}_r(q_r) - w_r)}{\nu_{r,0} + n_d}. \end{aligned}$$

Sampling η: Sampling for the parameters controlling the transit cannot be done in closed form. However, from (10), we see that values of η that produce transits that match features of $\tilde{y}_r(f_r)$ will be given higher posterior probabilities. Note that this accounts for data across all quarters, while also allowing for different noise variances for each quarter of Kepler observations. Simple box-least squares (BLS) algorithms can be used to generate a proposal distribution for η, or random walk proposals can be used if the chain is in the neighborhood of non-negligible posterior density. More efficient algorithms for exploring the parameter space may be available, but we defer discussion of these until testing on real data.

Sampling w: Recall that we have:

$$\tilde{y}_r(q_r)|w_r \sim N\left(w_r, \sigma_r^2 I\right), \qquad w_r|d \overset{ind}{\sim} N\left(Ad, A\Sigma_w A^T\right),$$

so:

$$w_r|\tilde{y}_r(q_r), \sigma_r^2, \Sigma_w \sim N\left(m_{w,r}, V_{w,r}\right), \qquad (16)$$

where:

$$m_{w,r} = V_{w,r} \left[\frac{\tilde{y}_r(q_r)}{\sigma_r^2} + \left[A\Sigma_w A^T \right]^{-1} Ad \right]^{-1},$$
$$V_{w,r} = \left[\frac{1}{\sigma_r^2} I + \left[A\Sigma_w A^T \right]^{-1} \right]^{-1}.$$

It can be seen that $V_{w,r}$, so the sampling in (16) can be done very rapidly.

Sampling d: Assuming a conjugate prior for d:

$$d \sim N(m_{d,0}, V_{d,0}),$$

then, from the marginal model in (14) we get:

$$d|w, \Sigma_w \sim N(m_{d,n}, V_{d,n}),$$

where:

$$V_{d,n} = \left[A^T \left(\sum_{r=1}^{R} W_r \right) A + V_{d,0}^{-1} \right]^{-1},$$
$$m_{d,n} = V_{d,n} \left[A^T \sum_{r=1}^{R} W_r \tilde{y}_r(q_r) + V_{d,0}^{-1} m_{d,0} \right].$$

Again, it can be seen that if $V_{d,0}$ is diagonal then $V_{d,n}$ is diagonal, so the calculations can be done very rapidly.

Sampling Σ_w: Placing a conjugate prior on $\sigma_{w,j}^2$ of the form $\sigma_{w,j}^2 \sim \text{Inv-}\chi^2(\nu_{w,j,0}, s_{w,j,0}^2)$ we can obtain:

$$\sigma_{w,j}^2 \sim \text{Inv-}\chi^2(\nu_{w,j,n}, s_{w,j,n}^2),$$

where $\nu_{w,j,n}$ and $s_{w,j,n}^2$ are posterior parameters corresponding to the choice of A matrix (the notation is more involved, but these are straightforward to compute).

3 Joint Modeling of Detrending and Transit Detection

The framework presented in (17) and (3) can be used to incorporate the uncertainty in detrending in a number of ways. The simplest approach, and one that can be applied to any probabilistic detrending procedure, is to replace the detrended data Y_d with different realizations from the detrending algorithm. If this is done at each iteration, it amounts to integrating over the uncertainty in the detrending. If a small number of deterministic detrenders are considered then the detrended data to be used at each iteration can be selected with equal (or weighted probability) from the list of detrended data. This approach amounts to assuming a discrete distribution for the overall mean term $\mu(t|\phi)$ in (17), thus ϕ essentially indexes the detrending routine. Note that this approach, while simple to implement, does not fully model the detrending and transit detection elements since it ignores the presence of $\mu(t|\phi)$ in the denominator of the terms in (3).

To try to model things more fully, recall that:

$$m_{sap}(t|\phi, \eta, \xi) = (\mu(t|\phi) + a(t|\xi)) + (1 - \tilde{q}(t|\eta)),$$
$$f(t|\phi, \xi) = \frac{a(t|\xi)}{\mu(t|\phi)}, \quad 1 - q(t|\phi, \eta) = \frac{1 - \tilde{q}(t|\eta)}{\mu(t|\phi)},$$

so we can rewrite (17) as:

$$Y_{sap}(t) = \mu(t|\phi)\,[1 + f(t|\phi,\xi) + 1 - q(t|\phi,\eta)] + \epsilon_{sap}(t|\psi), \qquad \forall\, t \in \mathcal{T}. \tag{17}$$

The algorithm described in section 3 produces posterior samples from f and q, allowing us to divide out the astrophysical and transit effects, and estimate the overall instrumental trends from this 'untrended' light curve. Note that the errors in $\epsilon_{sap}(t|\psi)$ are also scaled by the divided out astrophysical signal. Fitting the original detrender to:

$$\tilde{Y}_{sap}(t) = \frac{Y_{sap}(t)}{\mu(t|\phi)\,[1 + f(t|\phi,\xi) + 1 - q(t|\phi,\eta)]}, \tag{18}$$

accounting for the modified error bars, will then produce an updated detrended light-curve that can be sampled from. By iterating this procedure we obtain samples from the joint distribution of detrended curves, astrophysical, transit and noise parameters.