THE VARIABLE SKY OF DEEP SYNOPTIC SURVEYS

Stephen T. Ridgway1, Thomas Matheson1, Kenneth J. Mighell1, Knut A. Olsen1, and Steve B. Howell2

1 National Optical Astronomy Observatory, Tucson, AZ 85725, USA; ridgway@noao.edu
2 NASA Ames Research Center, P.O. Box 1, M/S 244-30, Moffett Field, CA 94035, USA

Received 2013 October 20; accepted 2014 September 9; published 2014 November 5

ABSTRACT

The discovery of variable and transient sources is an essential product of synoptic surveys. The alert stream will require filtering for personalized criteria—a process managed by a functionality commonly described as a Broker. In order to understand quantitatively the magnitude of the alert generation and Broker tasks, we have undertaken an analysis of the most numerous types of variable targets in the sky—Galactic stars, quasi-stellar objects (QSOs), active galactic nuclei (AGNs), and asteroids. It is found that the Large Synoptic Survey Telescope (LSST) will be capable of discovering \( \sim 10^5 \) high latitude \((|b| > 20^\circ)\) variable stars per night at the beginning of the survey. (The corresponding number for \(|b| < 20^\circ\) is orders of magnitude larger, but subject to caveats concerning extinction and crowding.) However, the number of new discoveries may well drop below 100 per night within less than one year. The same analysis applied to GAIA clarifies the complementarity of the GAIA and LSST surveys. Discovery of AGNs and QSOs is each predicted to begin at \( \sim 3000 \) per night and decrease by 50 times over four years. Supernovae are expected at \( \sim 1100 \) per night, and after several survey years will dominate the new variable discovery rate. LSST asteroid discoveries will start at \( > 10^2 \) per night, and if orbital determination has a 50% success rate per epoch, they will drop below 1000 per night within two years.

Key words: galaxies: active – minor planets, asteroids: general – quasars: general – stars: general – surveys – techniques: miscellaneous

Online-only material: active – minor planets, asteroids: general – quasars: general – stars: general – surveys – techniques: miscellaneous

1. INTRODUCTION

The Large Synoptic Survey Telescope project (LSST; Ivezic et al. 2008) is being designed to carry out a 10 yr survey of the available southern sky, with \( \sim 800 \) visits per field and an expected single visit depth \( r \sim 24.5 \) (5\( \sigma \)). The survey will detect unprecedented numbers of variable/transient targets, including variable stars, active galactic nuclei (AGNs), supernovae (SNe), and variable and main belt asteroids (MBAs). The LSST Project is committed to identifying variable targets and rapidly publishing information packets on them with a time lag not longer than 60 s. The total number of these detections and the rate at which they occur will have a significant impact on the processing power required to support the LSST pipeline and feed subsequent follow-up resources. The LSST Science Requirements Document, v5.2.3 (henceforth SRD; LSST Science Collaboration 2011) specifies a minimum alert rate capability of \( 10^3 \) per visit, and a stretch-goal rate of \( 10^5 \), with visits on a \( \geq 40 \) s cadence. It has not been clear how these specifications compare to likely peak variable detection rates.

While the production of alerts completes the LSST near-real-time response to variability detection, it signals just the beginning of the scientific response. From the alert stream, it will be necessary to apply filters, some simple and some complex, to select the targets for particular programs. For those requiring rapid follow-up, the filtering process must respond on a \( \sim \) minute timescale to avoid degrading the timeliness of the LSST alert stream. This filtering process is sometimes referred to as a Broker (Bloom et al. 2012)—e.g., a SkyAlert3 event stream. We understand a Broker to refer to a process that applies one or more selection criteria, possibly including correlation with the target history if any, other LSST data including nearby objects, other surveys, and very likely including among its services a probabilistic classification of many or all targets (Richards et al. 2012). Naively, the required Broker processing power should scale with the LSST alert production rate, though we will question this below.

There have been estimates of alert rates in the past, but they have not been detailed or documented and have not always been consistent. The objective of this study is to organize basic data and a rationale needed to determine order of magnitude (goal 50%) estimates for discovery rates of variable targets in deep synoptic surveys. In most of the paper, LSST will be taken as a model, guide, and example for observing parameters. Variable Galactic stars will be examined in most detail, with briefer discussion of AGNs, quasi-stellar objects (QSOs), SNe, and MBAs. There are of course other synoptic surveys. In Section 10.2, the methodology will be applied to GAIA as a second example.

2. VARIABLE STARS

The most numerous variable/transient targets in the LSST catalog will be stars, active galaxies, and asteroids. Simple estimates suggest that any one of these could dominate actual detections, hence all should be considered. Each of these requires a different approach.

Initially, we address the number of Galactic variable stars that will be detected and generate alerts in the LSST survey. It is known that at the level of a few percent variation, a few percent of stars are variable (Howell 2008). This suggests very large variable object counts. However, even if more exactly formulated, such a single data point does not suffice to determine the actual detection rates within an order of magnitude.

There have been several studies for prediction of variable star numbers in surveys, e.g., Mauder & Høg (1987) and Eyer & Cuypers (2000), which typically have taken a bottom-up approach of considering each major variable type.

---

3 http://www.skyalert.org/
in turn, reviewing the status of information on the occurrence of these types, and extrapolating to the number that will be detected in a future survey. These suffer from uncertain completeness. Estimating the probability of variability can be problematic. Most large surveys give some attention to identifying variable targets. However, in order to estimate probability, it is necessary to know the non-variable count for each stellar type, which involves determining the properties of the non-variable objects. Furthermore, we need data that goes deep and wide enough to produce good statistical characterization of the targets that will be the most numerous variable stars in the deep surveys. Existing surveys provide large and important databases of variability. An example is the OGLE-II catalog of 200,000 candidate variables (Woźniak et al. 2002). Published studies of such data sets have typically been focused especially toward detecting specific types of periodic variables rather than a high-level description of the incidence of variability.

Two elements are needed to estimate the detection rates: a (possibly simulated) catalog of the objects and a statistical basis for the probability of variability of each type of object as a function of variability amplitude (which we will call a variability probability distribution function, or VPDF) for at least the most abundant variable types. Our approach to predicting variable star detections combines knowledge of the variability of stars by spectral type, with a Galactic model. As will be explained below, this requires photometry suitable for estimating effective temperatures, \( T_{\text{eff}} \) (and to a lesser extent, luminosity), in combination with well-populated photometric time series. Thus, for example, we considered the overlap of the Sloan Digital Sky Survey (Gunn et al. 2006) for multicolor photometry and the Catalina Sky Survey (Drake et al. 2009) for time series. This remains a promising avenue.

We have chosen to work with data from the \textit{Kepler} mission. These data have obvious strengths and conveniences owing to the high target count, the photometric quality, and the pre-survey characterization effort. They have some minor drawbacks discussed below. Our approach is to characterize the stars primarily by temperature as determined in the \textit{Kepler} target pre-selection studies, and we find simple statistics of variability from their photometric time series. In order to apply the \textit{Kepler}-derived statistics to a region of the sky, we generate star catalogs for a grid of Galactic coordinates with the Besançon online Galaxy model. These catalogs provide complete descriptions of each simulated star.

3. THE BESANÇON GALAXY MODEL

The Besançon Galaxy model is described by Robin et al. (2003), and a web-based interface for its use is available.\(^4\) For an input Galactic coordinate pair (or range), the model computes a simulation of the star population in the designated sky area. The simulation is realized as a catalog of stars, with both fundamental parameters (age, mass, luminosity, effective temperature) as well as predicted observed parameters (brightness in selected filters). We use only the \( T_{\text{eff}} \) and magnitudes. The Galaxy model includes values for typical interstellar extinction, and should be directly comparable to observed stellar populations, although in the Galactic plane, where extinction is complex and clumpy, it is represented by a smooth model. (An improved representation of extinction for the Besançon model is under development; A. Robin 2013, private communication.)

\(^4\) http://model.obs-Besancon.fr

4. THE KEPLER VARIABLE STAR DATA SETS

We have created a Besançon model for the \textit{Kepler} field in order to explore whether or not the model population distribution is consistent with the color distribution of the \textit{Kepler} target sample (not required, but reassuring), and also as a guide to the luminosity content of the \textit{Kepler} target set. The model was computed for an area of 20\(^2\), covering Galactic coordinates \( b = 75.3\)–77.3 and \( l = 8.5\)–18.5, and thus representing a fair sample of the \textit{Kepler} field. With a faint limit imposed by \( R = 18 \) (\textit{UBV} system), the catalog returned 209,742 stars. This synthesis was done after the 2013 March 6 update in the Besançon late M dwarf model.

Both the Besançon catalog and the \textit{Kepler} catalog include \( T_{\text{eff}} \) values for each object, and these are used as a primary basis for associating \textit{Kepler} variability statistics with Besançon catalog stars. We have adopted the spectral type ranges and \( T_{\text{eff}} \) boundary values from spectral type–\( T_{\text{eff}} \) tables of Cox (2000), and listed here in Table 1. Figure 1 shows the distribution of the field contents by spectral type for the 210,000 Besançon model stars sampling the \textit{Kepler} field.

Figure 1. Distribution of main-sequence stars \( R \leqslant 18 \) in the Besançon model for the \textit{Kepler} field (solid line), and of all stars in Kepler Quarter 13 (dashed line). \textit{Kepler} target selection was designed to enhance the proportion of solar-type stars, as shown by the peak near G0.

(A color version of this figure is available in the online journal.)

For our purposes, it is important to have good statistics for variability near the level of survey photometry calibration limits, which for LSST is to be \( \pm 5 \) mmag (SRD design specification). This quality of photometry is not common in large ground-based surveys, but has been reported by Pan-STARRS (Schlafly et al. 2013). For a data source, we have chosen to go with the \textit{Kepler} mission which monitors 170,000 stars brighter than \( R \approx 16 \), with a photometric precision \( \approx 80 \) ppm. The \textit{Kepler} targets are intentionally drawn from a large range of main-sequence types. Giants are also represented, though not proportionately.

\textit{Kepler} benefits from a pre-survey study (Batalha et al. 2010) that provided measures of \( T_{\text{eff}} \) and surface gravity, \( \log(g) \), and the former are useful as they allow us to associate the \textit{Kepler} stars to our Galaxy model stars. \textit{Kepler} has undergone a careful selection process associated with maximizing its planet detection performance. This is only relevant to us if it tends to bias the variability statistics for any stellar type. We will return to this point later, but it proves to be a small effect.
The Astrophysical Journal, 796:53 (19pp), 2014 November 20

Ridgway et al.

Table 1

Spectral Type Groups Used for Analysis of Variability

| Spectral Type Range | $T_{\text{max}}^a$ | log $g$ cut | $K_p$ | $a$   | $b$   | $c$   | $d$   | Min Valid (mmag) | Max Valid (mmag) |
|---------------------|-------------------|------------|-------|-------|-------|-------|-------|------------------|------------------|
| M5                  | 3170              | ...        | 13.0  | ...   | ...   | 0.004 | 1.35  | 1                | 1000             |
| M2–M5               | 3520              | <3.5       | 13.0  | ...   | ...   | 0.06  | 2.10  | 1                | 1000             |
| M0–M2               | 3840              | >3.5       | 13.0  | ...   | ...   | 0.12  | 1.30  | 1                | 100              |
| K5–M0               | 4410              | <3.5       | 13.0  | ...   | ...   | 0.15  | 1.80  | 1                | 100              |
| K2–K5               | 4830              | <3.75      | 13.5  | −0.50 | −1.10 | ...   | ...   | 1                | 200              |
| M0–K2               | 5150              | >3.75      | 13.5  | ...   | ...   | 0.40  | 1.50  | 1                | 200              |
| G8–K0               | 5310              | <4.0       | 14.0  | −0.75 | −1.20 | ...   | ...   | 0.55             | 200              |
| G5–G8               | 5560              | >4.0       | 14.0  | ...   | ...   | ...   | ...   | 0.55             | 200              |
| G2–G5               | 5790              | <4.3       | 14.0  | −0.65 | −0.90 | ...   | ...   | 0.55             | 200              |
| G0–G2               | 5940              | >4.3       | 14.0  | −0.75 | −0.95 | ...   | ...   | 0.55             | 200              |
| F8–G0               | 6250              | <4.3       | 14.0  | −0.75 | −0.85 | ...   | ...   | 0.55             | 200              |
| F5–F8               | 6650              | >4.3       | 14.0  | −0.75 | −1.20 | ...   | ...   | 0.55             | 200              |
| F2–F5               | 7000              | <4.25      | 14.0  | −0.70 | −0.80 | ...   | ...   | 0.55             | 200              |
| F0–F2               | 7300              | >4.25      | 14.0  | −0.75 | −1.00 | ...   | ...   | 0.55             | 200              |
| A5–F0               | 8180              | <3.7       | 15.0  | ...   | ...   | 0.90  | 1.2   | 0.1              | 150              |
| A2–A5               | 9000              | >3.7       | 15.0  | ...   | ...   | 0.90  | 1.0   | 0.1              | 150              |
| B8–A2               | 11400             | <0.75      | 15.0  | ...   | ...   | 0.55  | −0.4 | 0.1              | 7                |
| B5–B8               | 15200             | >0.5       | 15.0  | ...   | ...   | 0.15  | 1.8   | 0.1              | 10               |

Notes. Spectral type ranges are defined by the $T_{\text{max}}$ limits noted for each range. The log $g$ cut is an empirical criterion, with no independent significance, based on published Kepler log $g$, for selecting a clean set of main sequence stars. The Kepler magnitude $K_p$ is a cut in brightness that was found convenient in eliminating higher luminosity stars from the main sequence set. The $a$, $b$, $c$, $d$ parameters are for the functional representations described in Section 5, and the following columns give the approximate range of usefulness of the empirical functions.

$^a$ The maximum $T_{\text{eff}}$ assigned to the spectral type range

Kepler data was acquired in three-month intervals, as constrained by regular reorienting of the space craft to maintain illumination of the solar panels. Data first became publicly available with the release of Q1, the first quarter Kepler data set in 2009. Soon after, Ciardi et al. (2011) published a summary of variability statistics for the Kepler stars. This study was not completely sufficient for our purposes, and D. R. Ciardi et al. (2013, private communication) recommended that we use their results cautiously and that a new analysis of the statistics of variability was merited when a better Kepler pipeline became available. We have followed that advice.

Since that time, the Kepler team has made much progress in the photometry pipeline, as described in data release notes of MAST (2013). Although the team takes care to point out that the bulk data processing does not produce the best photometry product for individual stars, it is now very effective at achieving good data quality across the breadth of the survey. We have undertaken a new study of stellar variability with Kepler Quarter 13 (Q13) as described in Kepler Release 13 Notes (Barclay & Christiansen 2012), based on the SOC Pipeline 8.0.

Each star data file contains a number of quality flags, and deleting all object data sets with any flag setting other than nominal, as well as a few non-standard FITS files, leaves a total of 155,370 stars, with the so-called long cadence sampling of ~30 minutes, over a total interval of 93.4 days, with two gaps of less than 1 day each, as described in the Kepler Data Handbook (Data Analysis Working Group 2012).

Using the Kepler-supplied values of $T_{\text{eff}}$, the Kepler stars are shown in Figure 1, for comparison with the Besançon model. The comparison shows the good correspondence in distributions between the Kepler target list and the simulated Galaxy, and reflects some of the intended selection effects in the Kepler catalog, to enhance coverage of Sun-like stars, cooler dwarfs, and hot stars.

The majority of Galactic stars detected in any deep survey will be dwarfs. The Kepler sample is likewise dominated by dwarfs, with an estimated $\geq5\%$ contribution of giants. In our application, luminosity is important primarily in avoiding contamination of the dwarf statistics with giants. Following Mann et al. (2012), we have used apparent magnitude as a secondary indicator of luminosity. By examining log $g$–$M_k$ distributions, several cleaned dwarf data sets were generated for each spectral group, with differing log $g$ and $M_k$ cuts. This

http://archive.stsci.edu/kepler/
established and confirmed consistent characterization for clean luminosity data sets. The isolation of higher-luminosity groups was less secure, especially for spectral types earlier than G. Therefore, we accept a lower degree of confidence in our evaluation of giant variability and statistics. This is of little importance to this study, but should be noted if the results are used for other purposes, and we do report results for luminosity dependence where the data justifies it (all spectral types except A and B).

Figure 2 shows an example of the kind of variability distribution found and also one of the cases of clearly defined luminosity dependence. The M0–M2 stars are separated into two groups, log \( g \) greater than and less than 3.5. There are more than 2000 stars in this spectral group, and the luminosity dependence is distinct.

5. REPRESENTING THE VARIABILITY FRACTION

Figure 2 for M0–M2 stars is typical of all the VPDFs in several respects. These empirical VPDFs predominantly have a power-law relationship to the rms variability amplitude in the range \( \sim 1–100 \) mmag. It shows a turnover for very low amplitude variability essentially due to running out of stars that are less variable. It also shows strong, spectral-group-dependent, systematic deviations from a simple analytical description. Therefore, we have not tried to work with a functional representation.

For an application in counting detectable variables, it is convenient to make use of a cumulative VPDF, from high to low variability, giving the probability that a target will be more variable than a given rms amplitude. Figure 3 presents the cumulative distributions for Figure 2. Not surprisingly, the integrated distribution shows less scatter than the original data, but this is more a numerical effect than a reduction in noise, as the data in Figure 2 has excellent signal-to-noise ratio (S/N). It appears that the cumulative VPDFs might have a simple analytical representation, and indeed a rather good fit is obtained with a function of the form \( f(\sigma) = 1/(1 + c\sigma^d) \), as shown in the figure. In some cases, a function \( f(\sigma) = a + \sigma^b \) is more appropriate. In Table 1, we give the parameters of empirical fits. However, it is not clear that the fitted functions are even a better description than the actual data, so we have chosen to work with the numerical cumulative VPDFs. Most of the cumulative VPDFs approach unity at or near 1 mmag rms. This study does not closely examine the behavior much below this level. The cumulative VPDFs for all spectral type groups are shown in the Appendix in Figures 18–26.

It is reasonable to wonder why the cumulative VPDFs tend to relatively uniform distributions since the underlying physics undoubtedly involves different processes at different amplitudes. The VPDFs are possibly revealing a log-normal distribution (Limpert et al. 2001) that can commonly characterize statistics of quantities that are predominantly small but cannot be negative (in this case, variability amplitude).

6. HOW REPRESENTATIVE OF THE SKY IS THE KEPLER FIELD?

The Kepler field, at galactic coordinates (76.32, +13.5), samples regions near but mostly above the plane and at approximately the solar distance from the galactic center. It is reasonable to wonder whether or not this sample population, to \( R \sim 18 \), will be representative of a deeper survey and in other directions. The Everett et al. (2002) and Tonry et al. (2005) surveys of variability studied fields at \((167, +12)\) and \((99, +60)\), to \( V \sim 19.5 \), and showed a similar result for field stars, though based on a much smaller sample and omitting the cooler part of the population. The VPDF-equivalent representation of their data appears in Figure 4. The sense of the difference is that the previous work did not include a Galactic mix of cool dwarfs, which are among the more variable stars, so it is understandable that the Besançon mix shows somewhat more variability. The Groot et al. (2003) Faint Sky Variability Survey studied 78 fields at mid- and high galactic latitudes over a large range of longitudes, to \( V \sim 24 \) mag. In the Huber et al. (2006) analysis of this data set, data related to the VPDF are presented in Figures 13 and 14 of that paper. This information is in terms of peak-to-peak variation detected with \( \sim 13 \) epochs, which we convert to an rms estimate by dividing by three. The inferred cumulative mean VPDFs for magnitudes 18.15–19.25, in their regions 01 and 19, are shown in Figure 4 as open symbols. The paper reports on fainter stars, for which variability fraction is estimated...
For a single visit to a target, what is the probability, 
\( f_\sigma \), estimate could be computed directly from the 
Kepler data, but it would be a great simplification if the 
target is variable. This could be Monte Carlo modeled with 
photometric measurements will be required to determine that 
are qualitatively and to some extent quantitatively supportive. 

This range should reasonably span the continuously variable 
governed by Gaussian statistics, the probability will be 
\( \frac{1}{\sqrt{2\pi}} \) as described in Section 5.

The fit in the range 1–100 mmag (dotted line) has \( a = 2.2 \) as described in 
Section 4. The dashed line shows an analogous result from Everett et al. (2002), 
also representing a mixed population, but based on fewer stars and a restricted 
range of spectral types. The open squares represent a cumulative VPDF derived 
from Huber et al. (2006) for their group 01 and the open triangles similarly for 
their group 19 (further discussion in our Section 6).

(A color version of this figure is available in the online journal.)

7. THE RMS REPRESENTATION OF 
VARIABILITY AMPLITUDE

In Section 9.5, we will be interested in estimating how many 
photometric measurements will be required to determine that 
a target is variable. This could be Monte Carlo modeled with 
the Kepler data, but it would be a great simplification if the 
estimate could be computed directly from the \( \sigma_{\text{var}} \) distribution associated with a \( T_{\text{eff}} \) (and where appropriate, a luminosity). 
For a single visit to a target, what is the probability, \( f_\sigma \), that it will be undergoing a brightness excursion (positive or negative) greater than the \( \sigma_{\text{var}} \)? For stochastic variability governed by Gaussian statistics, the probability will be 
\( f_\sigma \sim 0.317 \). For sinusoidal variability, the probability is \( f_\sigma = 0.5 \). 
This range should reasonably span the continuously variable stars. Transient variable types such as eclipsing binaries and 
flare stars can have low values of \( f_\sigma \).

In order to evaluate the utility of the rms value in characterizing the Kepler stars, we examined each photometric time 
series for both the rms value and for the amount of time spent with excursion amplitudes up to 100 \( \sigma_{\text{var}} \). Counting all Q13 observations for the set of Kepler stars, we found \( f_\sigma = 0.314 \), showing that Kepler variability statistics are at least somewhat noise-like (perhaps dominated by the more numerous stars with low-amplitude variations).

The statistics for the measurements of all stars are shown in 
Table 2. This shows that the actual distribution of brightness 

| Excursion Range (\( \sigma_{\text{var}} \)) | Observed | Gaussian | Excess |
|--------------------------------------|---------|---------|-------|
| Fraction | Fraction | Fraction |       |
| 0.0–0.2 | 0.15914 | 0.15852 | 0.00062 |
| 0.2–0.4 | 0.15259 | 0.15232 | 0.00027 |
| 0.4–0.6 | 0.14087 | 0.14066 | 0.00021 |
| 0.6–0.8 | 0.12558 | 0.12478 | 0.00080 |
| 0.8–1.0 | 0.10761 | 0.10640 | 0.00121 |
| 1.0–2.0 | 0.27168 | 0.27182 | 0.00014 |
| 2.0–4.0 | 0.04180 | 0.04544 | 0.00364 |
| 4.0–inf | 0.00070 | 0.00006 | 0.00063 |

Notes. Each measured excursion in flux is converted to units of \( \sigma_{\text{var}} \) for the particular star, and the measures are binned as shown in units of \( \sigma_{\text{var}} \), showing a small excess of high amplitude excursions. Based on 620 M measurements of 155,178 stars.

systematically lower with fainter magnitudes due to a natural 
sensitivity cutoff in variable detection.

This comparison between surveys is somewhat problematic, 
as the comparison studies have different cadences and magni-
tude ranges. Future surveys will probably detect significant pop-
ulations dependencies, but we conclude that these other studies do 
not offer a preferred or distinctly different characterization, and 
are qualitatively and to some extent quantitatively supportive.

(A color version of this figure is available in the online journal.)
points were contributed by one or more of three major brightness variation types.

1. Most Kepler time series have small numbers of (usually low-amplitude) single-point glitches.

2. A significant fraction of time series have small discontinuities at one or more of the breaks where the spacecraft was rotated and repointed.

3. In some time series, the slow instrumental drifts characteristic of Kepler appear to have been corrected inexactily, leaving one or more spurious features.

These “events” are important for peak-finding searches, but are generally not important for the kind of statistical measurements considered here, which are dominated by the ~4000 well-behaved measurements for each star rather than by the few outliers.

The same three effects appear in the light curves with a larger number of excursions, but in this group, eclipsing binaries are increasingly common. In the examined sample with more than 50 large excursion data points, 100% were eclipsing binaries, with a range of periods and amplitudes. Not surprisingly, these spend a relatively smaller fraction of their time at peak photometric excursion. These mostly eclipsing binaries have \( f_\sigma \sim 0.07–0.1 \) for the high-exursion tail of the distribution, and, of course, for individual rare objects, the fraction could be arbitrarily small. From this sample we deduce evidence for the signature of at least ~1000 eclipsing binaries in Q13, consistent with Slawson et al. (2011), who have identified as eclipsing binaries 2165 Kepler stars, or 1.4% (integrated over multiple data releases, with overlapping but larger target lists).

In summary, there are only small and qualitatively understandable deviations from Gaussian statistics. For purposes of a high-level characterization of variability, we will adopt brightness rms and \( f_\sigma = 0.314 \) as the parameters for estimation of variability detectability.

At this point, the tools described can be used to compare the statistics derived here for Q13 with those available to Ciardi et al. (2011) in their analysis of Q1. D. R. Ciardi et al. (2013, private communication) provided us with the Q1 data file used for their paper, which included a somewhat different set of stars and additional selection criteria that we did not apply. We compared the variability distributions between the two quarters. The summed distribution functions (i.e., weighting all spectral type groups equally) are in good agreement, show no systematic difference, and in particular no variability amplitude-dependent difference above 1 mmag. The Kepler photometric catalog quality improved significantly between Q1 and Q13, without changing the basic result for frequency of variability.

8. LIMITATIONS OF THE KEPLER SAMPLE

At the faint limit of the Kepler survey, the S/N per data point is reduced, and this could potentially bias the results toward overestimating the number of variables among the faint stars. We have compared spectral type subsets grouped by brightness and confirmed that there is no strong brightness dependence at the 1–2 mmag level.

The Kepler survey target list (Batalha et al. 2010) constitutes 150,000 stars selected from 0.5 million with \( m_K < 16 \). The selection process included explicit inclusion of the full range of temperature classes, M to O, suppression of higher-luminosity stars in favor of main-sequence stars, and a consideration of the estimated observability of a low-mass planet for each potential target. These biases may be obviated by not relying on the relative counts between spectral types. Within each spectral class, it is more important to consider what selection factors may impact variability. Batalha et al. (2010) note that all then-known eclipsing binaries in the Kepler field (>600) were included in either the primary or ancillary target lists. Therefore, the Slawson et al. (2011) eclipsing binary count of 1.4% in the Kepler target list may include an enhancement from the fraction in the host population. This likely error will be accepted here.

From examination of individual Kepler light curves, there appears in some cases a possible ambiguity between slow drifts of the instrument and slow evolution of the stellar brightness—a point discussed in the Kepler Data Handbook (Data Analysis Working Group 2012). The frequency of occurrence of changes over intervals longer than \( \sim 20 \) days may be diluted.

The Kepler survey provides limited information about the frequency of stochastic transients such as flares in dwarfs or mass transfer events in cataclysmic binaries. These are discussed separately below.

9. ESTIMATING NUMBERS OF DETECTABLE VARIABLE STARS IN SYNOPTIC SURVEYS

The approach used to determine the count of detectable variables is as follows: (1) for representative areas on the sky, generate a model catalog of all stars—this catalog gives for each star an apparent magnitude and fundamental parameters; (2) step through the list of stars; (3) for the assumed survey technology, determine the fractional variation detectable for each star; (4) from the VPDF function for that spectral type, find the probability that the star is more variable than the detection limit; (5) sum the probabilities over the catalog in useful ways.

In order to narrow the range of possible survey and analysis strategies, we assume a simple survey model, consisting of a sequence of measurements, with identification of variables by differential photometry between two or among many epochs. We consider two cases of variable detection: rapid detection, as required for urgent alert generation (discussed here), and post-processing detection (Section 9.4). For clarity, the discussion is referenced to the S/N achieved in a single epoch, represented by its inverse, \( \sigma_\text{phot} \), whether derived from single or multiple images, with the added assumption that the S/N in a stacked multi-epoch image will improve as \( \sqrt{n} \).

LSST reports discuss a 5\( \sigma \) criterion for detections (Ivezic et al. 2008). The development of criteria to ensure a desired level of statistical significance is complex. In sophisticated difference imaging simulations and source detection simulations, Becker et al. (2013) predict that the 5\( \sigma \) cut would allow \( \sim 3–13 \) false positives per sensor, or \( \sim 500–2200 \) per field per visit. In light of this number, Becker et al. suggest perhaps increasing the threshold to 5.5\( \sigma \). Indeed, a stringent criterion is needed to limit contamination of precious blank-field transient alerts. However, only a tiny fraction of 5\( \sigma \) sky-photon/detection noise events will coincide with a known star position, thus a 5\( \sigma \) cut may be far more severe than necessary for identification of variables from among already cataloged stars.

This example emphasizes that the significance threshold is not fundamental to a survey. Furthermore, it may be adjustable for various purposes and differently for different data streams. In the following, we will focus on significance of detection of variability only and not extend the discussion to the significance criteria for catalog inclusion.
9.1. Photometric Detection and Calibration Noise

The brightness-dependent component of photometric noise is the information typically offered by instrument measurement calculators that determine the detection rms noise, \( \sigma_{\text{det}} \), including detector noise, sky photon noise, and possibly observing process factors.

In variability detection, we should also consider error in photometry calibration, \( \sigma_{\text{cal}} \). All-sky calibration closure better than 0.005 has been reported (Schlafly et al. 2013) and is increasingly the expectation for surveys. An all-sky calibration of this quality may only be achieved post-facto months or years after data acquisition. On the other hand, relative photometry within a single field of view is less subject to systematics, and \( \sigma_{\text{cal}} = 5 \text{mmag} \) can be a conservative limit for such differential photometry.

For the total measurement noise, we use the root of the sum of squares of the detection noise and calibration noise, \( \sigma_{\text{phot}}^2 = (\sigma_{\text{det}}^2 + \sigma_{\text{cal}}^2) \). This \( \sigma_{\text{phot}} \) represents the single-epoch measurement noise. The following discussion will refer to \( \sigma_{\text{phot}} \).

9.2. Procedure

For application to a survey of the sky, we will use the Besançon Galaxy model to compute a star catalog for representative fields. If the number of stars in a field greatly exceeds \( 10^9 \), the model will be used to generate a catalog for a representative sub-area and the results scaled to the full area.

For each star in the model catalog, the apparent brightness of the star in the selected filter will be used with the measurement model to predict the single-epoch measurement error, \( \sigma_{\text{phot}} \). The threshold criterion, \( t \) in units of \( \sigma_{\text{phot}} \), will determine the minimum level of variability detectable for that star. The Besançon temperature will be used to select the appropriate Kepler VPDF. The threshold, \( t \sigma_{\text{phot}} \), will be used to read from the cumulative VPDF the probability that the star has a detectable variability. The fractional probabilities per star will be simply summed with binning appropriate to track across the catalog variable star counts that may be sorted by spectral type and/or apparent magnitude.

9.3. Rapid Detection and Alerts

A common objective of synoptic surveys is the rapid identification of transient targets, in some cases requiring timely follow-up, hence the generation of alerts. A low probability of false positives will greatly enhance the value of the alerts.

For rapid identification, we assume simple comparison of one measurement at the current epoch with one measurement from a previous epoch. We expect that a few percent of stars will have detectable variability. Each star will be observed multiple times, so if the probability of a false positive indication for variability is \( \gamma \), then the integrated probability after \( n \) epochs will be \( \sim n \gamma \) (for small probability). Requiring that the false attribution of variability be less than 0.0001 of the number of true detected variables, a clearly severe criterion, requires \( \gamma \ll 10^{-5}/n \). For a survey duration of the order of \( 10^2 \)–\( 10^3 \) epochs, this implies a statistical threshold of \( 3.4 \text{–} 4.0 \sigma_{\text{phot}} \). Though much different from \( 5 \sigma_{\text{phot}} \) as a statistical property (2+ orders of magnitude), it is only 20% different as a variability amplitude, hence perhaps not importantly different, so in the following we will use a threshold of \( 5 \sigma_{\text{phot}} \) for rapid detection.

9.4. Variability Detection by Post-processing

For a survey with many epochs, variability will be detectable in the full data set to a lower fractional level than possible with one visit pair. We note from Table 2 that stars with rms variability \( \sigma_{\text{var}} \) will typically spend a fraction \( f_\gamma = 0.314 \) of their time at excursions exceeding amplitude \( \sigma_{\text{var}} \), and 0.683 of their time at smaller excursions. For a survey with \( n \) epochs, assume that one can optimally sort the data into bins greater than and less than \( \sigma_{\text{var}} \). Then, for a per epoch noise of \( \nu = 1/\sigma_{\text{phot}} \), the noise in the mean of the two bins will be improved by \( \sqrt{1/0.314n} \) and \( \sqrt{1/0.686n} \), respectively. The noise in the comparison of two epochs will be \( \sqrt{2}\nu \), and the noise in the comparison of \( n \) epochs binned will be \( \sim 2.15\nu/\sqrt{n} \) (for \( n \gg 2 \)). This shows that a survey of 100–1000 epochs may reach variability levels \( \sim 5 \text{–} 15 \) times lower than the alert detection process. Combining this and Section 9.3, a useful threshold range for variable detection by post-processing of 100–1000 epochs would be 1 or 0.3 \( \sigma_{\text{phot}} \), respectively. Of course the availability of other information or priors may allow further reducing the limits and further enlarging the set of detectable variables, so these thresholds probably imply lower bounds to the totals.

9.5. The Declining Rate of Detection in a Survey

It can be anticipated that the number of detectable variable stars will be large. If the survey commences with no knowledge of the particular targets, the rate of discovery of variables should start out high and then taper off as the survey continues and an increasing fraction of the variables become known variables. While this may not reduce the required alert rate, if all new variations are announced, the existence of previous detections and alerts can greatly reduce the alert generation and filtering tasks.

The rate of variable detection can be cast with reference to the probability, \( f_\sigma \), that a star when observed at one epoch will be undergoing a brightness excursion greater than the selected threshold. We will assume a simple model in which all targets have only two distinguishable states, with a frequency of occurrence of \( f_\sigma \) and \( (1 - f_\sigma) \). A variable will be detected when the star switches from one state to the other, assumed to be random. Then at the \( n \)th epoch, the number of targets that have not yet been observed to have switched states for the first time will be \( N_0[f_\sigma^n + (1 - f_\sigma)^n] \), where \( N_0 \) is the total number of detectable variables and \( n \) is the number of epochs. There are two terms here, one for beginning in the low state and switching to the high state, and vice versa—that are about equally important in this case. This is a two-process exponential decay problem, and the expression for the number of yet undetected variables can also be written \( N = N_0[f_\sigma e^{-\lambda_1 n} + (1 - f_\sigma)e^{-\lambda_2 n}] \), where \( \lambda_1 = \ln f_\sigma \) and \( \lambda_2 = \ln(1 - f_\sigma) \).

The exponential drives very rapidly toward zero, so for most types of continuous variability, the discovery rate is strongly front-loaded and decreases to “noise” within a few tens of epochs.

As used here, the term “epoch” refers to the characteristic time over which observations of the brightness of the type of variable considered decorrelate, so it is the larger of the intrinsic decorrelation time of the target and of the revisit interval. Values will be introduced below in the discussion of discovery rates for specific surveys and target types.
9.6. Discussion of False Positives

Controlling the number of false positives in a survey can have multiple motivations and criteria, and the latter can differ across the data products and can evolve with time.

In a mature survey, the rate of false detections should hopefully be relatively stable. The rate of real discoveries of variability will start out higher than pessimistic false positive rates, but after a few tens of epochs, should fall below some estimates of false positive rates (Becker et al. 2013). This will be true for virtually any threshold, for example, in the range 3–6 σphot, hence the actual value assigned is more a matter of convenience than a fundamental decision. For example, the threshold could be set low initially, as long as the ratio of false to real detections was small, or it could be set high initially to suppress both false and lower-significance real detections and reduce the volume of the alert stream. Late in the survey, one might consider setting the threshold to zero for convenience, hopefully a fair representation of expected survey products, but subject to adjustment for different assumptions.

10. Application of Besançon Models and Kepler Statistics to Future Surveys

For future surveys, the following results depend not only on hardware not yet tested on the sky, but on assumptions about survey execution and analysis. Therefore, the results should be understood as illustrative, hopefully a fair representation of expected survey products, but subject to adjustment for different assumptions.

10.1. The Large Synoptic Survey Telescope

Quantifying the LSST alert rate and Broker tasks is a major motivation for this study. The LSST (Ivezić et al. 2008) will survey the southern sky available from its site on Cerro Pachón repeatedly in 6 filters for 10 yr. The actual cadence is not yet fixed, and hence the sky coverage per night and the distribution among filters is uncertain, so we have chosen to consider the variable statistics on a per-field-visit basis, rounding the LSST field size up from the design ~9.6 to 10.0 deg^2 for convenience of scaling. The LSST photometric errors are constrained by a number of design specifications (LSST Science Collaboration 2011), but of course the delivered performance is not known and depends on observing conditions. We have used an LSST “exposure calculator,” version 4.3 (Gee 2008), that implements Equations (4) and (5) from Ivezic et al. (2008), calculating S/N for point sources including sky subtraction. The Kepler bandpass has a response centroid of 643 nm. We have used the LSST r bandpass for comparison (response centroid 620 nm). It is understood that LSST will have a saturating bright limit—we have adopted r = 15, but the total counts are not significantly dependent on this cut.

LSST is expected to reach all-sky photometry calibration of σcal ~ 5 mmag. Though it will sometimes operate under conditions that are not photometric in the classical sense of low and stable extinction, the large field and great depth of LSST imagery will allow strong constraints with differential photometry relative to nearby stars. Pending better analysis, or until it is resolved by on-sky experience, we adopt σcal = 5 mmag.

Figure 6. Number of detectable variables per field of 10 deg^2, per spectral type, for an LSST-type survey, with detection in the r band at 5σphot significance. Stars brighter than r = 15 are excluded. The ordinate follows a great circle, which for the stars starts at the Galactic center, to the south Galactic pole (at 90 degrees). Solid lines show main sequence plus G and K giants, T Tauri (TT), and white dwarf (WD). These are total TT and WD counts, not considering level of variability. The predicted dip at the Galactic equator is due to higher extinction in the plane. The text symbol MIII represents a single point for the M giants, which have a non-zero density only in the galactic plane (where extinction brings them into the r > 15 brightness range). The text AV represents a similar single point on the plane for A dwarfs. Additional material is included to suggest rates for other targets. Dashed line—estimated upper limit to the possible number of cataclysmic variables (CV; Section 12), without regard to possible detectability—the observable number is surely much lower. The dash-dot line shows the predicted distribution of detectable main belt asteroids (MBAs; Section 14.4), for which the ordinate shows ecliptic latitude. The dotted line connects four circles that show the predicted number of detectable M-dwarf flare events (amplitude greater than 0.1 mag in the u band) per visit (Hilton et al. 2011)—most of these stars will be included in the dwarf data sets at the top of the figure.

(A color version of this figure is available in the online journal.)
as the rms calibration noise for a single visit, and we will require $5\sigma_{\text{phot}}$ significance for variability detection and alert generation.

Figure 6 shows the expected numbers of detectable variable stars at the $5\sigma$ threshold, per 10 deg$^2$, as a function of sky position, along a great circle from the Galactic center to the south Galactic pole. The total counts are dominated by the dwarfs, with the M dwarfs most numerous except near the Galactic plane, where K and G dwarfs are competitive due to a combination of star column density and line-of-sight extinction. The cataclysmic variables (CVs) will be discussed in Section 12, and the MBAs in Section 14.4.

Figure 7 illustrates how the number of detectable variables depends on the brightness range considered. Each curve represents the total stars from Figure 6 sorted by 0.5 mag bins. At high latitudes, the total initially increases by $\simeq 3$ times for each magnitude of additional depth. However, at $r \simeq 20$, the trend turns around and the number of stars per half-magnitude decreases steadily, owing to the Galactic distribution of stars above the plane. Thus, the majority of detectable variables at high latitude will be brighter than $r \simeq 21$. Near the Galactic center, the peak brightness for the number of detectable variables shifts over the range 18–22 and will be direction dependent anywhere near the plane.

Figure 8 shows how the total number of detectable variables depends on the detection threshold. The $5\sigma_{\text{phot}}$ solid curve effectively sums the information in Figures 6 and 7 and shows the number of detectable variables with that criterion, which should approximate the number of stars that can alert for variability. The dashed curve for $1\sigma_{\text{phot}}$ shows a likely number when end-survey post-processing for variable detection is considered. The dotted line shows for comparison the total number of stars (variable and non-variable) detectable in one visit at $5\sigma_{\text{phot}}$, and $6 \times 10^7$ ($1\sigma_{\text{phot}}$). Figure 9 shows the integrated total detectable variables for $|b| \geq 20^\circ$, plotted as a histogram in 0.5 mag bins. The totals are modest, even for a threshold of $1\sigma_{\text{phot}}$. The distributions show a strong peak at 21 mag, respectively, confirming that the most numerous variables will be detected 2–3 mag brighter than the LSST single-visit detection threshold.

In the plane of the Galaxy, the focus will be on Galactic targets, there will be interest in detecting large numbers of variable targets of all types, and compiling complete samples may be useful. We have studied the Galactic plane with samples on a grid. An appropriately weighted sum for the Galactic plane ($|b| \leq 20^\circ$) gives counts of $6 \times 10^{10}$ and $8 \times 10^{11}$ for $5\sigma_{\text{phot}}$ and $1\sigma_{\text{phot}}$, respectively, always subject to caveats about extinction and crowding. Again, most of these variables will not be close to the LSST single visit faint detection limit.
The capability of the GAIA mission to discover variable stars has been studied with the same tools described above. The GAIA mission (Perryman et al. 2001) will deliver time-resolved photometry for large numbers of Galactic variables, with $\simeq 70$ observations per target over an interval of five years. The mission will make photometric measurements in two passbands, designated blue and red (de Bruijine et al. 2006). The performance in terms of detection error per visit (i.e., taking into account the information obtained from multiple detectors during the course of a single survey) has been characterized in terms of the GAIA $G$-magnitude, for stars in the range $G = 12–20$, as $\sigma_{\text{det}} = 10^{-3}(0.02076 z^2 + 2.7224 z + 0.004352)^{1/2}$, where $z = 10^{0.4(G-15)}$ (ESA 2014). Note that the GAIA $G$-bandpass is very similar to the Kepler bandpass, and hence may be compared to LSST $r$.

The photometry calibration error for GAIA will naturally be less than for LSST, but it is a little unclear what value to use, particularly for the rapid processing and alert generation (Jordi et al. 2011). We have experimented with calibration noise levels of 0, 1, and 5 mmag and show these in Figure 10. The GAIA variability detection limit at faint magnitudes is surely limited by photon noise rather than calibration, and the lower (and zero) calibration error assumptions simply increase the number of very bright and only slightly variable targets identified.

Quantitatively, we find that for 1 mmag calibration error, the number of detectable variables will be $3.9 \times 10^6$ at high latitude and $4.8 \times 10^9$ in the plane. The numbers with 0 mmag calibration error are larger by $\simeq 20\%$, and the numbers for a photometric noise of 5 mmag are $\simeq 35\%$ smaller, so the total range is $<2$ times. In the following, we will take as representative a GAIA photometric error of 1 mmag. The large star count at low-latitude of course depends on GAIA’s ability to deal with crowded fields, given the drift scan and data compression strategy. In this context, it is worth noting that GAIA operations have an optional Modified Scanning Law for crowded regions, that the angular resolution of GAIA is very good, and that scans will eventually be combined to form a high-resolution image of the vicinity of cataloged targets.

As with LSST, the detectability of variability for the complete survey data set will be greater than for the single-epoch determination and we estimate that by considering a threshold noise level of $1\sigma_{\text{phot}}$.

A numerical experiment shows that for a two times increase in variability amplitude, the expected number of detectable GAIA variables increases by a factor of $2.1$ ($3.1$) for latitudes $>20$. LSST will reach fainter stars, showing a satisfying performance, assuming $5$ mmag photometric error and at the $1\sigma_{\text{phot}}$ (single-scan noise) threshold. This will be compared to the LSST early survey performance, assuming $5$ mmag photometric performance, and a $5\sigma_{\text{phot}}$ (single-visit noise) threshold.

Figure 11 shows the expected number of detectable variable stars for GAIA and for LSST in the sky region of overlap. Thanks to its expected high photometric quality, the GAIA survey will detect lower variability amplitudes for stars brighter than $G = 20$. LSST will reach fainter stars, showing a satisfying level of complementarity. The integrated totals are shown in Table 3.

11. THE ALERT RATE FOR LSST AND GAIA

Both GAIA and LSST plan to provide the community with timely alerts for newly identified targets of interest. Particularly early in the survey, it will sometimes be difficult to distinguish novel from mundane targets. We expect that Galactic stars will be a major contributor to the count of variable/transient targets, and we have now assembled the information needed to predict the discovery rates for such variable stars. Our approach
is to use the tools and results described above, especially in Section 9.5. The initial total pool of targets to be discovered for LSST is the sum of the low- and high-latitude columns in Table 3. The corresponding total for GAIA is given by the GAIA-S entries in the table. The predicted discovery rates for GAIA and LSST are shown in Figure 12. This is for a required $5\sigma_{\text{phot}}$ detection in all cases, and the photometric $1\sigma_{\text{phot}}$ noise is 1 mmag for GAIA and 5 mmag for LSST.

The timescale is determined by the rate of measurement of decorrelated epochs of variability. In the case of GAIA, the ~70 visits will be obtained in a complex and precisely known pattern over approximately five years. For LSST, ~180 $r$-filter visits will occur in a yearly modulated but not yet known complex pattern over 10 yr. For both surveys, there will be close pairing in time for some or most observations, reducing by $\approx$ 2 times the number of independent epochs. With this correction, we estimate typical values of 52 days between well-separated $G$-band observations for GAIA and 42 for LSST $r$-band visits. These are long enough to exceed the variability timescale of most stars in our study, and so we use these numbers as the interval between measurement epochs.

Figure 11. GAIA as a precursor to LSST. Distribution of detectable variables per 0.5 $G$- or $r$-magnitude bin, comparing the single-epoch sensitivity for LSST, $5\sigma_{\text{phot}}$, to the GAIA performance for $1\sigma_{\text{phot}}$, approximating the performance for the end-of-mission processing of the full survey data set—hence the target counts are higher here than in Figure 10. The curves stopping at $G = 20$ (dashed lines) represent GAIA performance, and the curves extending to $r = 25$ (solid lines) represent LSST. The lower count groups (diamond symbols) correspond to the integrated high latitude sky ($|b| > 20^\circ$), and the higher count groups (squares) correspond to the integrated low latitude sky ($|b| \leq 20^\circ$). The complex structure for the Galactic plane arises due to interplay of the stellar distribution model and the extinction model. The integrated totals of the GAIA curves correspond to the -E entries in Table 3, and the integrated totals for LSST correspond to the -S entries.

(A color version of this figure is available in the online journal.)

Table 3

| Sky Region | Survey | $G \geq r = 15$–$20$ | $= 20$–$25$ | Total |
|------------|--------|-----------------|-------------|-------|
| $|b| > 20$   | GAIA-E | 2.5e+7          | ...         | 2.5e+7|
| $|b| > 20$   | GAIA-S | 3.9e+6          | ...         | 3.9e+6|
| $|b| > 20$   | LSST-S | 1.0e+7          | 1.0e+7      | 2.0e+7|
| $|b| \leq 20$| GAIA-E | 2.8e+10         | ...         | 2.8e+10|
| $|b| \leq 20$| GAIA-S | 4.8e+9          | ...         | 4.8e+9 |
| $|b| \leq 20$| LSST-S | 2.5e+10         | 3.8e+10     | 6.3e+10|

Note. E = detectable by end of survey analysis, S = detectable in a single scan or visit.

Figure 12 shows the variable star “alert problem” for GAIA and LSST, initially severe for the first year of the survey, nominally dropping to modest numbers within a few years. The timescales are similar for the two surveys. Of course, the numbers are imprecise, but to the extent that an exponential relation applies, the qualitative conclusion is inescapable—most variable stars will be identified relatively early in the surveys and need not overwhelm the search for rarer targets subsequently. (Note that GAIA and LSST report different planned policies, with LSST attempting a more comprehensive alert broadcast stream and GAIA more selective.)

Figure 12 carries an important message for the processing power required to handle alert production. In the case of LSST, the planned program is to emit an alert every time a target is observed to deviate from a reference image or value. Variable stars will be the object of repeated alerts (in many cases, alerts will be produced on virtually every observation). From the number of variable stars, this clearly implies a lot of alerts. However, after a short break-in period, the majority of stars generating alerts will already have a previously generated data package, the only additional information will be a few added data points, and derivative characterizations can be formulated in incremental algorithms for rapid updating.

In the following sections, we will gather information on detection rates for other types of variables, and these will be collected in Figure 13. This figure shows numbers only for the LSST high-latitude sky. It is not directly applicable to GAIA because the detection limits and cadence are different, but a similar comparative analysis of GAIA will lead to similar qualitative conclusions. The LSST high-latitude detection rate for stellar variables from Figure 12 appears in Figure 13 as a single line labeled “Stars.” The additional information in Figure 13 will be discussed below.

12. CATAclySMIC VARIABLES

CVs are binary stars that consist of a compact object (almost always a white dwarf) and most commonly a relatively normal low-mass, main-sequence companion. They are formed in systems initially close enough to participate in a stage of
common envelope evolution, resulting in a shrinkage of the orbital separation, and eventually mass transfer events and outbursts. CVs are somewhat problematic for variable counts as there are several CV types, and the CV phenomenon is known primarily through the longer orbital period, high-mass transfer objects, whereas the majority of the systems should be of shorter orbital period with lower-mass transfer rates (Howell et al. 1997), which are less well known.

The number of CVs reported from among the Kepler targets is small (Howell et al. 2013 and references therein), and the Kepler faint limit is not well matched to the quiescent brightness of the most numerous components of the CV population. A deep survey may well reveal disproportionately more detectable variables of this type than projected from less deep surveys (Howell et al. 2001). Modeling the CV population is complex, with several distinct stages of evolution and poorly constrained statistics.

The approach we have taken is to set a limit to the maximum possible number of CVs. We assume that all binaries in the galaxy that could evolve to CVs have already done so and remain in the active CV population indefinitely. The terms to include are the fraction of stars that are binaries and the fraction of binaries that have the right combination of orbital parameters and masses to evolve into CVs. For the fraction of star systems that are binary or multiple, an observational study of F and G dwarfs by Duquennoy & Mayor (1991) led to an estimate of 0.57 (not counting substellar companions), whereas Lada (2006) suggests that knowledge of the IMF implies a value of 0.33. For the binary fraction, we use a compromise number of 0.5. The fraction that evolves to CV has been estimated from Monte Carlo simulations of binary evolution by Howell et al. (1997) and Howell et al. (2001) as 0.0038. The Besançon model galaxy does not give explicit binary statistics, so we used the model white dwarf population and estimate that the probability that each is a member of a CV system to be no greater than (0.5 × 0.0038).

We have run a small set of Besançon models with the limits set to output all white dwarfs, regardless of apparent brightness, and applying our limit, show these as a dashed line in Figure 6. There it can be seen that the CV number density upper limit is about an order of magnitude below that of main sequence dwarfs, hence if all potential CVs were active CVs, with detectable outbursts, these would be an important contributor, but would not dominate, the variable star totals.

The large number of potential CVs suggests that CV outbursts may be an important contributor to transient object alerts. Some of these objects will be cataloged prior to flaring, though most will not have been distinguishable as CVs in the quiescent state and some will appear as blank sky transients. An estimate of the upper limit to the CV discovery rate can be derived. The outburst frequency can be estimated in the range 0.1–1 per year. The length of outbursts will typically ensure detection of sufficiently bright events when they occur, unless during the fraction of the year, estimated as 0.5, when the field is not regularly observed. A simple model gives an upper limit for the CV discovery rate as shown in Figure 13 for an assumed outburst frequency of 0.5 per year. It is important to note that this rate is several orders of magnitude higher than estimates based directly on CV observations. However, even if the fraction of CVs that gives detectable outbursts is small, it could dominate the stellar variable discovery rate after an initial survey startup period, and perhaps more importantly, could be a major contributor to the blank sky transient rate. More frequent outbursts would give a steeper decline in the discovery rate. In fact, synoptic surveys will contribute to clarifying CV evolution and statistics.

13. DWARF FLARES

M dwarf flares could be numerous because the stars are so common. It was noted above that the Kepler variability rms numbers are not very sensitive to rare outliers—the complementary fact is that the rms values do not tell us much about rare, transient events, so the Kepler analysis does not really describe flare frequency. Furthermore, flares are represented poorly in the Kepler survey, owing to the small number of cool M stars in the Kepler catalog and perhaps also their characteristic ages. What we do know is that Walkowicz et al. (2011) have studied the flare rates in Kepler K and M dwarfs, finding flaring in ~1.5% of the some 23,000 cool dwarfs observed, with the vast majority having peak flare amplitudes below the likely LSST 5σcal. Since flaring is thought to be correlated with spotting and rotation, it is likely that most flaring dwarfs will appear in the variable dwarf counts.

Kowalski et al. (2009) have studied the flare rates for M dwarfs, and Hilton et al. (2011) have applied the results to the specific question of the frequency of flaring event detections in the LSST survey. While the flare rates, particularly for large flares and in the Galactic plane, are uncertain, and the predictions depend on several assumptions, the numbers are an important guide, as they set a floor to the transient rate. The Hilton et al. (2011) predictions for the u band are included in Figure 6.

An interesting prediction is the frequency of detectable flares from stars that are below the survey limit. Hilton et al. (2011) estimate this frequency (per epoch per 10 deg² field) to range from 0.01 at high latitudes to 0.1 at low latitudes, with the
frequency in the plane higher but not estimated. This prediction is for the $u$ band, but it is included in Figure 13, using the low latitude estimate of 0.1, and assuming visits to 600 distinct fields per night. Although the discovery rate should drop with time, we have no basis for estimating a slope. Expecting that fields per night. Although the discovery rate.

14. OTHER VARIABLE AND TRANSIENT SOURCES IN DEEP SURVEYS

Galactic stars are only one component of the variable sky, and other types of targets may compete numerically, especially toward the higher latitudes. We started with the Galactic stars for good reasons: they are important in driving alert totals, and there are good sources of data for statistical descriptions of variability with considerable detail by stellar type. Beyond that, they serve as a useful model for how to approach the topic of variable detection and alert generation.

We will mention here three other important sources of variable/transient targets.

14.1. Quasi-stellar Objects

QSOs are an important subgroup of AGNs, and for our purposes are defined by the selection criteria of our primary data source, MacLeod et al. (2012). The amplitude of observed QSO variation depends on the time interval studied, with the occurrence of large amplitude changes increasing with time. The characteristic timescales, up to hundreds or thousands of days, are similar to survey durations. We have utilized the cumulative VPDFs from MacLeod et al. (2012). We describe the cumulative probability as a linear function of $\Delta m$. MacLeod et al. (2012) present a number of data sets for different observing intervals ranging from 10 to 1000 days. Data sets for different magnitude cutoffs in different bandpasses are not obviously consistent, so we have brought the different data sets into a single representation for comparison.

We describe the relation analytically as $\log p = -2\Delta m/\Delta m_{-2}(\Delta)$, where $\Delta m_{-2}$ is the value of the magnitude difference between two measurements corresponding to a cumulative probability $p$ of $\log p = -2$, and is a function of the interval between measurements. Values for $\Delta m_{-2}$ can be read from the MacLeod et al. (2012) figures for different data sets corresponding to different apparent brightnesses. Figure 14 shows our evaluation of its dependence on $\Delta$. We see an approximately linear relationship between $\Delta m_{-2}$ and $\log(\text{days})$, while there is scatter, there does not appear to be a systematic deviation from linearity over the timescales up to 10 yr for different selections of targets. We adopt the relationship $\Delta m_{-2} = 275 \times \log(\text{days})$, shown as a straight line in the figure.

The QSO luminosity function is derived from Hopkins et al. (2007), with a $K$ correction (Palanque-Delbrouille et al. 2013) and converted to $r$ magnitude using a median $g - r$ color relation with redshift. The adopted relation is shown in Figure 15.

With this information, we can integrate over the luminosity function, determining the limiting $\Sigma \sigma$ variability detection limit of LSST (single visit), the cumulative fraction of QSOs with expected variability exceeding that value, and summing the total number of detectable variable QSOs at each time step. The differential with time of that function gives the predicted discovery rate, assuming efficient detection. To illustrate this, we have used the techniques described above to evaluate the number of QSOs for which LSST could detect variability with time baselines of 50, 600, and 3600 days. These distributions are shown in Figure 15.

Because most detectable variability is at magnitudes brighter than 23, the exact trend of the LF at the faint limit is not important. While most detectably variable QSOs will be found early in a survey, the discovery rate falls off more slowly than for stars due to the increase in variability amplitude with time. The expected detection rate from this model is shown in Figure 13.

By simulation, LSST should detect variability in $8 \times 10^5$ QSOs in a 10 yr survey. If the amplitude of QSO variations is increased by 2 times, the number of variable QSOs detectable increases by 21%. If the characteristic time constant is increased by 2 times, the number decreases by 18%.

14.2. Active Galactic Nuclei

A careful prediction for discovery of AGNs should take into account a description of the broad range of variability characteristics. However, the determination of temporal structure
functions for large samples is less satisfactory than for QSOs, and hence the behavior of variability amplitude with survey duration is not well established. Instead, following the basic logic of the analysis of Galactic stars, we propose to estimate the number of galaxies, the variable fraction of galaxies, the number of variable AGNs detectable, and then comment on the temporal sampling.

A VPDF was constructed to satisfy a combination of specific survey data and general constraints. In surveys, AGNs are identified largely by variability, so the number detected is a function of the photometry and the temporal sampling, and it is not surprising that estimates of the frequency of occurrence vary. We adopt a recent report (Klesman & Sarajedini 2012) that tabulates varying that estimates of the frequency of occurrence vary. The VPDF is set to zero for variability amplitudes greater than 700 mmag, owing to lack of sufficient data and low probability. (A color version of this figure is available in the online journal.)

![Figure 16. Cumulative VPDF for galaxies, showing the data from Klesman & Sarajedini (2012) as squares. A boundary condition shown as a triangle, explained in the text, implicitly defining AGNs to have variability $\geq 10$ mmag, is the basis for extrapolating the function to smaller amplitudes. The VPDF is set to zero for variability amplitudes greater than 700 mmag, owing to lack of sufficient data and low probability. (A color version of this figure is available in the online journal.)](image)

AGNs would dominate the total variable count. However, going through the calculation with the VPDF and the Galaxy luminosity function, we find that the number of detectably variable AGNs from a pair of epochs will be $\sim 150$ per deg$^2$. The reason is that the Galaxy luminosity function is so strongly weighted toward the faint end that the variability of the overwhelming majority will be undetectable. The expected distribution with magnitude is shown in Figure 17 for 20,000 deg$^2$.

In the case of stars, it was a fair assumption that the brightness of most variables would decorrelate between observation epochs. For galaxies this is not a good assumption, so we also need information about the temporal power in Galaxy variations (e.g., Gaskell & Klimek 2003). For present purposes we simply note that the variation is more stochastic than periodic (so $f_\sigma \sim 0.317$, Section 7), and the time constant is typically months. This means that the rate of first-time detection of variability in AGNs will be lower, and the number yet to be detected will decay more slowly than was the case for stars. For a synoptic survey, the discovery total will be determined more strongly by the duration of the survey rather than by the number of observations.

For purposes of fielding an estimate, we assume stochastic variation and a time constant of six months, so that approximately one-half of the yet undetected variables will be discovered in each six-month interval (neglecting such details as accessibility through the year). Now the fraction of AGNs observed for the first time to have switched to a detectably different state will be described as in Section 9.5. As shown in Figure 13, in this model, the discovery rate will begin at $\sim 3000$ per night, and decrease by 50 times within four years. It is noted that the difference between the AGN and QSO models may not be significant, as may be seen from the similar trajectories in this figure.

![Figure 17. Expected number of AGNs per 0.5 $r$-magnitude bin whose variability can be detected by LSST at the single visit detection threshold, $5\sigma_{\text{phot}}$, integrated over 20,000 deg$^2$ LSST-accessible sky. (A color version of this figure is available in the online journal.)](image)
in detection rate would be faster. The total detected in a 10 yr survey would not change significantly in this model.

Predictions for AGN detections by GAIA are problematic because of the GAIA optimization for bright point sources, whereby host galaxies may not be identifiable or properly cataloged.

### 14.3. Supernovae

SNe are a target of LSST key science, but they can also be a source of clutter in the search for certain rarer sources (“one man’s trash is another man’s treasure” and vice versa). All of the targets discussed above have the common characteristic that they come from a limited pool, and once a member has been discovered, the number remaining to discover is reduced. This is not the case for SNe, which will be as numerous on the last day of the survey as on the first. Bailey et al. (2009) estimate an SNe Ia detection rate for LSST of $1.4 \times 10^5$ per year or 380 per night, and we adopt this result. Based on Bulbul et al. (2012), we multiply this number by three to estimate the discovery rate for SNe of all types.

### 14.4. Main Belt Asteroids in the LSST Survey

The last category of variable targets that will be considered here is the MBAs. A wide and deep survey will detect large numbers. For LSST, Ivezić et al. (2009) utilize model orbital element and magnitude distributions to estimate that $N_{\text{MBA}} = 5.5 \times 10^6$ MBAs will be detectable by LSST. For comparison with other target types, we partially reverse engineer that simulation, assuming that the asteroids are distributed uniformly in ecliptic longitude, and with an ecliptic latitude dependence from Wolfe & Zissis (1978), we model the distribution of MBAs as $F(\beta) = 385 e^{-0.14\beta}$, where $F$ is in objects per square degree and $\beta$ is the latitude in degrees. Near the ecliptic plane, the number of objects will be in the range of 3000 per LSST field, while at ecliptic latitudes greater than about 30°, the density will drop to a few tens per field. In the following, it will be assumed that the orbits of all of these targets are initially unknown. Since asteroids move rapidly, this corresponds in some respects to up to thousands of detections per field per observation. As may be seen in Figure 6, if these cannot be readily identified and tracked, then in the ecliptic at high galactic latitude, asteroids will compete with stars in target numbers.

To add the number of MBAs detected per night to Figure 13, we reduce the total number of detectable MBAs by $f_{\text{lat}} = 0.5$ to count just southern hemisphere objects. We apply a factor of $f_{\text{sky}} = 0.25$ for the fraction of the sky observed each night and a factor $f_{\text{opp}} = 0.5$ to select the fraction that are in a relatively favorable position (with respect to opposition) for detection. Therefore, the MBAs should be observed at a (very uncertain) initial rate of about $N_{\text{MBA}} f_{\text{lat}} f_{\text{sky}} f_{\text{opp}} = 280,000$ per night. Every subsequent detection on following nights will be a new “discovery” until the orbit is characterized.

An adequate analysis of MBAs requires orbital modeling, which is possible, and currently unknown information about the efficiency of orbit characterization from the LSST cadence. Without detailed justification, we simply assume that the detectable asteroids will be observed with data sets suitable for orbit determination on average once every two months and that the characterization will be successful 50% of the time. This model gives the predicted observation rate of objects whose orbit has not yet been characterized, as shown in Figure 13. (A proper simulation of the MBA discovery rate with a distribution of orbits will show that it has a long, low-level tail, due to objects that are not detectable every year.)

From Figure 13, it appears that MBA discoveries will dominate other variable target counts at the start of the survey by approximately 10 times. MBA targets are of course moving significantly, and to some extent (to be determined), this motion will enable them to be distinguished in individual exposures. However, even if only 1% of the MBAs cannot be identified by apparent motion, the unidentified MBAs will still represent a very large contribution of anonymous transients. Unless MBAs can be immediately identified as moving objects, they will dominate the variable discovery problem at the start of the survey. The schematic model suggests that a modest success rate at orbit determination will suppress the problem by, initially, about an order of magnitude per year.

### 15. SUMMARY AND DISCUSSION

This study was undertaken to determine the scale of the LSST alerts and Broker tasks with respect to the number of variable and transient targets that will be discovered. The Kepler Q13 was used to “calibrate” variability in a Besançon Galactic model and predict the number of detectable variable stars as a function of position on the sky. An upper limit to the number of CVs shows that they should be less numerous than variable dwarfs. Applying the model to LSST and GAIA predicts that each will detect $>10^7$ variable stars at high latitude and $>10^{10}$ at low latitude.

It is suggested that the alert and Broker tasks will depend mainly on the rate of discovery of new variables, as on subsequent detections there will already by a near-complete data package and analysis of the target. The temporal characteristics of the variable stars and the surveys were used to predict the variable star discovery rates for LSST and GAIA, starting at $10^4$ and $2 \times 10^6$ per night for high- and low-latitude sources, respectively, but with efficient discovery dropping by about one order of magnitude per year.

Further discussion focused on just LSST at high Galactic latitudes, where the variable star discovery rate begins at $10^4$ per night. For comparison, the rate derived for AGNs begins as $3 \times 10^3$ per night, but decays more slowly due to the longer time constant. QSOs are close to the AGN trend. Figure 13 suggests that discovery of rare events will become rapidly easier over the interval two to three years into the survey. On the other hand, attempting to push this activity sooner might be an exercise in frustration.

The expected discovery rate for MBAs begins very high, a few $10^5$ per night. In a deep survey, asteroids present an almost existential problem in the ecliptic, which must be mastered to a high level of success. They are so numerous that even a small incompleteness could threaten other, difficult, science objectives. Reducing the discovery rate depends on the success of characterizing each target—a success rate of 50% per data set or better appears satisfactory.

The SN discovery rate is predicted to average $\sim 1100$ per night with no decay and may exceed most other discovery rates a few years into the survey.

Before undertaking this study, we could not determine which of the three categories—stars, galaxies, asteroids—might dominate LSST discovery and alert tasks, and Figures 6 and 13 show why it was not obvious, as they all compete for dominance in various parts of the sky and at various times.
15.1. What Could Go Wrong and What Could Be Done Better?

This study gives estimated counts for numbers of variable stars. The source sample is large and heterogeneous, but does not represent all infrequent variable types. However, uncommon targets will represent small corrections to the total counts.

The Kepler statistics may not fairly represent low amplitude, very slow stellar variation, and its incidence is unknown. Our statistical representation of Kepler data does not capture rare transient variability.

The population dependence of the stellar variability characteristics has not been studied. In a simple exercise, it was found that increasing the variability amplitudes by 2 times, the number of detected variables in the plane increased by 3.25 times, and the number at high latitude by 2.3 times.

The predicted star counts in the Galactic plane are enormous. Some will be inaccessible owing to blending. However, variable star discovery may be less limited by confusion than one intuitively expects since the requirement to have photometrically detectable variability ensures that the variables counted here are magnitudes brighter than the survey faint limit.

The star counts in the plane strongly depend on extinction. We have used the default Besançon extinction. Increasing the extinction by 2 times (in absorption) results in a decrease in the number of detected variables by a factor of 2 times.

The upper limit for CVs may greatly exaggerate their observable numbers.

The study of AGN variability needs long duration sampling of a large number of targets to give a stronger observational basis for the VPDF and to quantitatively characterize the variability timescales.

The calculations are largely schematic and represent very simplified descriptions. The assumption of Gaussian statistics is an approximation—it could be improved on the variable side by using more elaborate descriptions of observed variability, and on the instrument side with simulations using actual detection and photometry algorithms, but ultimately probably requires on-sky experience with survey detectors as well.

If false detections are generated by systematic effects such as PSF differences rather than statistical noise and cannot be suppressed, the variable detection rates suggested here may not be achievable.

Improved dwarf flare statistics are needed, especially for deep observations, over a range of latitudes.

Asteroid discovery/characterization rates and speed of identification could be improved with existing models, Monte Carlo simulations and experience at ongoing surveys.

Finally, when the actual discovery rates for various target types have fallen by one to two orders of magnitude from the initial rates, second order effects not considered here will enter in and the discovery rates will likely change to a shallower slope or possibly plateau. Happily, this will be not just a shortcoming of this analysis, but a welcome discovery space.

We are grateful to our colleagues of the NOAO LSST Science Working Group for discussions and to participants in Oxford IAU Symposium 285 and the 4th GAIA Alerts.
Figure 20. **Left:** cumulative VPDF for Kepler stars K2–K5. The red (light) line represents stars of Kepler log $g \geq 3.75$, and the black (dark) line of $< 3.75$. **Right:** cumulative VPDF for K0–K2 stars. The red (light) line represents stars of Kepler log $g \geq 4.0$, and the black (dark) line of $< 4.0$. In both figures: the dotted lines show the empirical fits described in Table 1.

(A color version of this figure is available in the online journal.)

Figure 21. **Left:** cumulative VPDF for Kepler stars G8–K0. The red (light) line represents stars of Kepler log $g \geq 4.1$, and the black (dark) line of $< 4.1$. **Right:** cumulative VPDF for G5–G8 stars. The red (light) line represents stars of Kepler log $g \geq 4.2$, and the black (dark) line of $< 4.2$. In both figures: the dotted lines show the empirical fits described in Table 1.

(A color version of this figure is available in the online journal.)

Figure 22. **Left:** cumulative VPDF for Kepler stars G2–G5. The red (light) line represents stars of Kepler log $g \geq 4.3$, and the black (dark) line of $< 4.3$. **Right:** cumulative VPDF for G0–G2 stars. The red (light) line represents stars of Kepler log $g \geq 4.3$, and the black (dark) line of $< 4.3$. In both figures: the dotted lines show the empirical fits described in Table 1.

(A color version of this figure is available in the online journal.)

Figure 23. **Left:** cumulative VPDF for Kepler stars F8–G0. The red (light) line represents stars of Kepler log $g \geq 4.3$, and the black (dark) line of $< 4.3$. **Right:** cumulative VPDF for F5–F8 stars. The red (light) line represents stars of Kepler log $g \geq 4.25$, and the black (dark) line of $< 4.25$. In both figures: the dotted lines show the empirical fits described in Table 1.

(A color version of this figure is available in the online journal.)
Workshop for feedback and encouragement. David Ciardi provided advice and unpublished data. Zeljko Ivezic and Richard Green gave valuable advice on QSO counts, and Ian McGreer provided the derivation of the r-magnitude QSO luminosity function from published data. Patricia Burchat gave valuable insights.

NOAO is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

APPENDIX

CUMULATIVE VARIABILITY PROBABILITY DISTRIBUTION FUNCTIONS

The most useful description of the variability fraction proved to be the cumulative VPDF (Section 5) as illustrated in Figure 3 M0-M2 stars. The cumulative VPDPs are shown here for all spectral types in Figures 18-26, with empirical analytical approximations as described in Section 5 and Table 1.
REFERENCES

Bailey, S., Bernstein, J. P., Cinabro, D., et al. 2009, LSST Science Book, ed. A. Tyson, M. Strauss, & Z. Ivezic (Tucson, AZ: LSST), 383, http://www.lsst.org/lsst/scibook

Barclay, T., & Christiansen, J. (the Data Analysis Working Group). 2012, Kepler Data Release 13 Notes, KSCI-19053-001 (Moffett Field, CA: NASA)

Bathala, N. M., Borucki, W. J., Koch, D. G., et al. 2010, ApJL, 713, L109

Bulbul, E., Smith, R. K., & Loewenstein, M. 2012, ApJ, 753, 54

Ciardi, D. R., von Braun, K., Bryden, G., et al. 2011, AJ, 141, 108

Cox, A. 2000, Allen’s Astrophysical Quantities, ed. A. N. Cox (4th ed.; Washington, DC: AIP), 388

Data Analysis Working Group 2012, Kepler Data Characteristics Handbook, KSCI-19040-003 (Moffett Field, CA: NASA)

de Bruijne, J., Perryman, M., Lindegren, et al. 2006, Gaia Astrometric, Photometric, and Radial-velocity Performance Assessment Methodologies to be used by the Industrial System-level Teams, GAIA-JdB-022, http://www.rssd.esa.int/SYS/docs/ll_transfers/project=PUBDB&id=448635.pdf, http://www.rssd.esa.int/index.php?page=Science_Performance&project=GAIA

Drake, A. J., Djorgovski, S. G., Mahabal, et al. 2009, ApJ, 696, 870

Duquennoy, A., & Mayor, M. 1991, A&A, 248, 485

ESA 2014, GAIA Science Performance, http://www.cosmos.esa.int/web/gaia/science-performance

Everett, M. E., Howell, S. B., van Belle, G. T., & Ciardi, D. R. 2002, PASP, 114, 656

Eyer, L., & Cuypers, J. 2000, in ASP Conf. Ser. 203, The Impact of Large-Scale Surveys on Pulsating Star Research, ed. L. Szabados & D. Kurtz (San Francisco, CA: ASP), 71

Gaskell, C. M., & Klimek, E. S. 2003, A&AT, 22, 611

Gee, P. 2008, LSST Exposure Time Calculator (ver. 4.3.2; Davis, CA: UC Davis), http://dls.physics.ucdavis.edu:8080/etc4_3work/servlets/LsstEtc.html

Groot, P. J., Vreeswijk, P. M., Huber, M. E., et al. 2003, MNRAS, 339, 427

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731

Howell, S. B. 2008, AN, 329, 259

Howell, S. B., Everett, M. E., Seebode, S. A., et al. 2013, AJ, 145, 109

Ivezić, Z., Tesicanin, P., & Jones, L. 2009, LSST Science Book (ver. 2.0; Tucson, AZ: LSST), http://www.lsst.org/lsst/scibook, 101

Jordi, C., Gebran, M., Carrasco, J. M., et al. 2011, A&A, 523, A48

Klesman, A. J., & Surajiedini, V. L. 2012, MNRAS, 425, 1215

Mann, A. W., Gaidos, E., Lépine, S., & Hilton, E. J. 2012, ApJ, 753, 90

Mast 2013, Data Release Notes (Baltimore, MD: STScI), http://archive.stsci.edu/kepler/data_release.html

Metcalfe, N., Shanks, T., Campos, A., McCracken, H. J., & Fong, R. 2001, MNRAS, 523, 795

Palanque-Delabrouille, N., Magneville, C., Yéche, C., et al. 2013, A&A, 551, A29

Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339

Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012, ApJS, 203, 32

Slawson, R., Proș, A., Welsh, W. F., et al. 2011, AJ, 142, 160

Schlafly, E., Green, G., & Finkbeiner, D. P. 2013, BAAS, 45, 145.06

Tonry, J. L., Howell, S. B., Everett, M. E., et al. 2005, PASP, 117, 281

Walkowicz, L. M., Basri, G., Batalha, N., et al. 2011, AJ, 141, 50

Woźniak, P. R., Udalski, A., Szymański, M., et al. 2002, AcA, 52, 129