KIC 8462852 DID LIKELY NOT FADE DURING THE LAST 100 YEARS

Michael Hippke
Luiter Straße 21b, 47506 Neukirchen-Vluyn, Germany

Daniel Angerhausen
NASA Postdoctoral Program Fellow, NASA Goddard Space Flight Center, Exoplanets & Stellar Astrophysics Laboratory, Code 667, Greenbelt, MD 20771

ABSTRACT

A recent analysis found a "completely unprecedented" dimming of 0.165 ± 0.013 magnitudes per century in the F3 main sequence star KIC8462852. This star is interesting, as it shows episodes of day-long dips with up to 20% dimming of unknown origin. We re-analyze the same Harvard archival Johnson B photometry and find serious issues in the data processing techniques used for KIC8462852. These errors include a hard cut of all data points with magnitudes dimmer than B∼12.5, leading to a skewed distribution and incorrect error estimates of the trend. A cross-check of other stars in the Kepler field of view shows similar data quality issues with a strong dependency on quality-cuts, leading to arbitrary results. We conclude that the Harvard plates photometry is limited to an accuracy of ~ 0.1mag per century (1889–1990), which is confirmed by other studies and which does not allow for a discovery of a dimming trend at the same level. The most likely explanation for the century-long trend of KIC8462852 is thus a data artifact, in the form of a structural break, and it is probably not of astrophysical origin.

1. INTRODUCTION

The F3 main-sequence star KIC8462852 shows a unique series of brightness dips up to 20% (Boyajian et al. 2015), theorized to originate from a family of large comets (Bodman & Quillen 2015), or signs of a Dyson sphere (Wright et al. 2016). Subsequent analysis found no narrow-band radio signals (Harp et al. 2015) and no periodic pulsed optical signals (Schuetz et al. 2015; Abeysekara et al. 2016). The infrared flux is equally unremarkable (Lisse et al. 2015; Marengo et al. 2015; Thompson et al. 2015). KIC8462852 is currently one of the most discussed systems, and we have to use all available information to solve this mystery. One excellent idea by Schaefer (2016) was to check the archives for long-term trends in this star. He found a "highly significant and highly confident secular dimming at an average rate of 0.165±0.013 magnitudes per century", which is described as "completely unprecedented for any F-type main sequence star".

Our goal was to reproduce the results in Schaefer (2016) and to verify (or falsify) this extraordinary claim. In the following, we will perform an in-depth analysis of the dataset at hand, and cross-check the results for other stars.

2. REVISION OF THIS PAPER

After the release of the first version of this research paper (Hippke & Angerhausen 2016), we received a lot of feedback from the community. We appreciate this very much and update the paper according to the questions and concerns. An important element in the discovery history of KIC8462852 was the involvement of the "Zooniverse citizen science network" (Fischer et al. 2012) and the "Planet Hunter volunteers" (Schwamb et al. 2012) who searched for signals of transiting planets by "harnessing the human eye's unique ability for pattern recognition." (Boyajian et al. 2015). Community involvement and real time discussion have proven very useful, which was why we released our paper to arXiv.org before acceptance by a peer-reviewed journal. We appreciate that Bradley E. Schaefer decided to do the same with his paper, and we thank him and the whole community for valuable feedback on our first version.

Major issues identified include:

a) The selection of comparison stars. Some of the 28 stars used show variability on short (days–years) timescales, making them less useful for a comparison.

b) Application (or avoidance) of quality cuts with the Harvard DASCH (Digital Access to a Sky Century @ Harvard) data, e.g. the use of red- and yellow-sensitive plates, and/or flagged data.

We agree that these issues must be discussed carefully. In this update, and building on our earlier analysis (Hippke & Angerhausen 2016) more carefully, we will present an in-depth analysis of the DASCH photometry for KIC8462852. We use our analysis of this star as an example for the myriad of possible quality cuts and arbitrary decisions in the processing. Our conclusion is that the analysis alone casts serious doubt on the idea of a linear trend for the star. Afterwards, we will discuss the possibilities of comparing other stars as a benchmark.

3. IN-DEPTH DATA ANALYSIS OF KIC8462852

3.1. *By-eye measurements*

The Harvard DASCH photometry was taken on glass plates with an emulsion most sensitive in the blue, building the base for the later Johnson B magnitude system (Johnson & Morgan 1953). Schaefer (2016) generated two datasets. The first comes from his manual (by-eye) brightness estimates using a method described in Schaefer (1981) with a microscope, and was done for 131 data points (these procedures are described in section §2.2 in Schaefer (2016)).

These "by-eye" measurements have neither been publicly released nor been shown in the study. Also, the

hippke@ifda.eu
daniel.angerhausen@nasa.gov

selection criteria of the 131 data points are not well defined, as only those were used "which I judged to be able to return a confident and accurate magnitude". For example, it is unclear whether all 1581 plates got visually inspected, and only 131 passed this criteria; or if there was some additional (random?) selection criteria involved.

This ambiguity is unfortunate, because Schaefer (2016) claims that the manual result "yields a slope of $+0.310 \pm 0.029$ magnitudes per century", which is described as "formally different from the slope that I get from DASCH". In fact, this trend has almost twice (1.88×, 5σ confidence of differing) the slope of the scanned data. It would be very useful to compare these data points to the scanned photometry, and we urge the author to publicly release these data.

3.2. *Issues in retrieval of scanned photometry*

The second dataset in Schaefer (2016) is from automatic, scanned data, and contains "1581 plates covering the area of KIC8462852". From the latter, the author performs a quality selection and keeps 1232 data points (§2.1), which he bins into 5-year segments. In the following, we will try to reproduce his Figure 1 and his Table 1.

The DASCH data are available online[1]. One enters the coordinates of our object (20 06 15.450 +44 27 24.75) and selects one of three possible calibration bandpasses. As Schaefer (2016) did not qualify which calibration(s) were used, we will check each of the possibilities and compare the differing results.

We start by choosing the "Kepler Input Catalog" which "gives comparable accuracy for the field of the Kepler satellite". The website then produces a new window with a successful identification of our target star, labeled as "K8462852". We can check the identified coordinates with the ones we enter to make sure it is indeed the same object. Following this link presents a graph of the lightcurve, plus a choice of 44 possible data cuts. Following the methodology in Schaefer (2016), we choose to `Show: All points`, which removes all cuts. Then, we can `Hide: Yellow or Red Plates` and download the data by selecting `Show Lightcurve Summary Data (Plotted points only)`. After the header, the file contains 2468 data lines. When we remove empty flux values, 1234 data points remain. This number is close to the average of available digitized plates per sky region (some regions have more plates, some have less). It is useful to open the resulting datafile in a spreadsheet program to apply the other three quality cuts as performed in §2.1 in Schaefer (2016):

a) We identify and remove all values with quality flags indicator "AFLAGS"> 9000. There are 28 available flags[2], so that the criteria removes 20 and keeps 8 flag types.

   We searched the literature for best-practice reductions of DASCH data and found nothing that indicates that this method is "best" (see also our section 3.3). Since there is no justification for the method and its parameters in Schaefer (2016), we have to assume that this is rather arbitrary. Furthermore, the "AFLAGS" are a bit-array, and we can only assume that the value "9000" is the decimal representation. In order to continue our analysis, we ignore this issue and apply the cut for "AFLAGS"> 9000. Afterwards, 976 values remain.

b) As a next step, we follow the instructions to remove all data values with one-sigma error bars > 0.33mag. Again, no justification for the chosen limit of > 0.33mag is given. Afterwards, 708 values remain.

c) Finally, we remove all data values with the flux being within 0.2 magnitudes of the quoted plate limit. Again, this choice is arbitrary. Afterwards, 699 values remain.

We can now compare our result to Schaefer (2016): "With these selections, I have 1232 magnitudes from DASCH". We, however, have only 699 values left. Might this difference come from the different calibration that DASCH offers? The website[3] states: "We are currently using the AAVSO Photometry All-Sky Survey (APASS) Release 8 Catalog, the Kepler Input Catalog (KIC), and the GSC2.3.2 Catalog for photometry calibration. The APASS calibration gives the best photometric accuracy over the entire sky. The KIC calibration gives comparable accuracy for the field of the Kepler satellite. Finally, the GSC2.3.2 catalog provides magnitudes for objects outside the 9 to 15 magnitude range of APASS."

It is not clear from Schaefer (2016) which of the 3 calibrations was used. Therefore, we explore the possibility that the different calibrations offered by DASCH can be responsible for this discrepancy:

a) When using the KIC calibration, we get a file with 2468 (all data lines), 1234 (valid flux values), 976 (after removing "AFLAGS"> 9000), 708 (after removing all data values with one-sigma error bars > 0.33mag), 699 values (after removing all data values with the flux being within 0.2 magnitudes of the quoted plate limit). The average magnitude over the 699 values is B=12.06.

b) For the APASS calibration, we get a file with 3168, 1830, 1054, 978, 937 values. The average magnitude over the 937 values is B=12.29.

c) For the GSC calibration, we get a file with 2219, 1809, 993, 694, 679 values. The average magnitude over the 679 values is B=12.26.

From these numbers, one would probably prefer the APASS data as it holds the largest number of calibrated data values. Still, less than Schaefer (2016) could get (1232) with a total average magnitude of 12.37 (from Table 1). Therefore, we urge the author to explain the data retrieval and -cleaning process step-by-step so that it can

[1] http://dasch.rc.fas.harvard.edu/lightcurve.php, retrieved on 01-Feb 2016

[2] http://dasch.rc.fas.harvard.edu/database.php, retrieved on 01-Feb 2016

[3] http://dasch.rc.fas.harvard.edu/lightcurve.php, retrieved on 01-Feb 2016

be reproduced by the community. Until this happens, we will rely on the publicly available data (937 values) and emphasize that our following data analysis will likely be close, but not identical, to a similar analysis with the data used by Schaefer (2016). Science is based on the principle of reproducibility, and we have not managed to reproduce Figure 1 (and Table 1) in Schaefer (2016), despite following the published method and the publicly available data.

3.3. *Different approaches on data cleansing*

As mentioned in the previous section, data cleansing involves arbitrary choices. For instance, Schaefer (2016) removes all red- and yellow sensitive plates, but assures the reader in his section 2.1 that this is not required: "Critically, the removal or extension **of any or all of these cuts** does not significantly change the slope of the light curves for KIC8462852, its check stars, **or any constant star.**" (our emphasis). On the other hand, the same author described the practice of including these plates as a "beginner's blunder"[4].

The first statement, that the removal of all cuts does not change the slope of any constant star, is incorrect. Without cuts, most stars (64% of our sample of 28 stars, which have been selected from the KIC catalog with temperatures and $\log_g$ in the range of F-dwarfs) show significant trends. One example is sufficient as a falsification: KIC7180968 is a F-star with a 17σ linear trend (or structural break) in the DASCH data.

The second statement, that the use of red and yellow plates is a "beginner's blunder", is also incorrect. A literature review shows that all publications accessible for us *do* use these data (e.g., Laycock et al. (2010); Tang et al. (2013a,b); Liska & Skarka (2015)). Indeed, the DASCH team itself does this in their publications, and simply mentions that the plates "are mostly blue sensitive" (Tang et al. 2011). It is preferable to *keep* all available data (that are not amongst a few multi-σ outliers), and propagate their large(r) error bars accordingly. Indeed, these values are also classified as "Johnson B magnitude" data and have been calibrated by the DASCH team. Simply deleting arbitrarily chosen values is thus unwise.

Plate selection aside, other quality cuts are done equally arbitrary in the literature. In Tang et al. (2013b), for example, the authors analyzed 997 Kepler stars and included all plates (also the red and yellow ones), but defined their own series of quality cuts. These include blended images, "measurements within 0.75 mag of the limiting magnitude" (in contrast to Schaefer, who set a limit of 0.2mag), "images within the outer border of the plates whose width is 10%" (which was accepted by Schaefer), and more. Unfortunately, these criteria are also not precisely defined, e.g. the rejection of "Stars with strong correlation between magnitude measurements and plate limiting magnitudes", so that the results are also not reproducible. In the following, we will analyze the quality cuts as used in Schaefer (2016) and determine their effects.

3.4. *Issues of the quality cuts*

It can be seen in Figure 1 that the data cleansing performed by Schaefer (2016) causes a truncation of the data for all values with magnitudes dimmer than B=12.45. This is likely not a conscious decision, but a side-effect of the specific cleansing criteria. The largest influence on the truncation is caused by the rejection of all "AFLAGS"$>$ 9000, leading to many more dimmer than brighter values being removed.

Obviously, the truncated distribution looses normality (Figure 2). As is expected, tests for normality (e.g., Shapiro & Wilk (1965)) find a skewed distribution with very high significance.

As is well known, error estimates in linear least-squares regressions are only valid for Gaussian distributions. With the truncated data, where the distribution is not normal, error estimates from such regressions are incorrect. Of course, this problem cannot be healed by binning as is done by Schaefer (2016). Consequently, the error estimates from the linear regression in Schaefer (2016) are useless. It is also very unfortunate that Schaefer (2016) did only show his binned data, and not the individual data points. The truncation and the resulting problems would have been immediately clear to every reader.

Another issue is that the truncation removes more values in earlier years, where the scatter is much larger due to inferior technology. Figure 1 (right panel) shows that there has never been a post-cleaned data value dimmer than B=12.45, irrespective of the putative dimming trend. If the dimming trend were real, we would expect the dimmest accepted value to increase (decrease in luminosity) over time. The second problem is that the larger scatter during earlier times can only extend towards brighter values (as dimmer values are clipped), and these scattered brighter values occur more frequently during earlier times. Consequently, an artificial trend is introduced into these data, which must be quantified.

3.5. *Linear regressions*

We can test the original, and the "cleaned" dataset for the magnitude of the truncation effect on the linear regression.

a) For the original dataset, we can fit a linear regression and use normal error estimates. Without binning, we get a slope of $+0.12\pm0.02$mag per century.

b) After cleaning of the dataset (i.e., for the truncated data), we try fitting a linear regression and get $+0.07 \pm 0.013$mag per century. As mentioned in the previous section, these errors are incorrect due to non-normality.

c) To estimate errors in the cleaned data, we can use a truncated regression analysis (e.g., Cong (1999)) and get $+0.11 \pm 0.02$mag per century. Coincidentally, these error bars are consistent with an MCMC approach by Geert Barentsen[5], although the binned data were used for the MCMC.

d) Finally, we can test the effect of 5-year bins as used in Schaefer (2016). With these bins, we get $+0.1\pm 0.02$mag per century.

[4] http://www.centauri-dreams.org/?p=34933

[5] https://github.com/barentsen/did-tabbys-star-fade

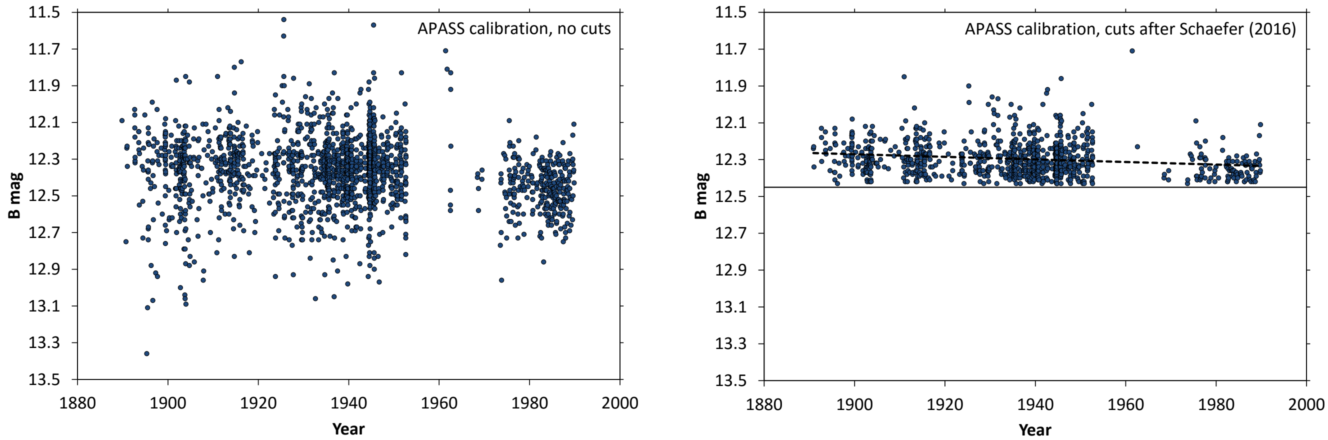


Fig. 1.— Harvard DASCH photometry in APASS calibration. Left: All data values, raw and unbinned. Right: After application of cuts by Schaefer (2016) but without bins. The dashed line is a least-squares linear trend. The straight horizontal line is the cut-off at B=12.45mag.

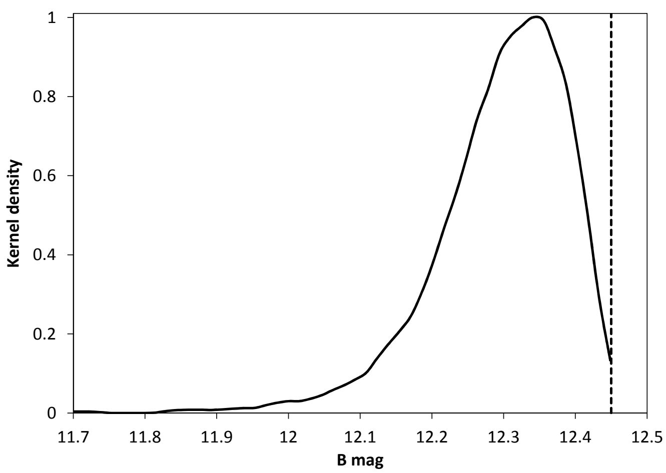


Fig. 2.— Kernel density estimate for the post data cleansing distribution shows truncated data, loosing normality.

TABLE 1
Telescopes in DASCH data for KIC8462852
(APASS, all plates)

Series	Telescope	From	To	Plates
ac	1.5-inch Cooke Lenses	1899.3	1952.7	613
rh	3-inch Ross Fecker	1928.3	1962.6	410
dnb	Damons North Blue	1962.5	1989.9	300
i	8-inch Draper Doublet	1889.8	1936.8	233
mc	16-inch Metcalf Doublet	1910.5	1951.6	79
bm	3-inch Ross	1934.2	1940.8	54
ay	2.6-inch Zeiss-Tessar	1924.3	1927.9	44
md	4-inch Cooke Lens	1911.2	1940.5	36
ca	2.5 inch Cooke Lens	1935.4	1936.7	15
ir	8-inch Ross Lundin	1935.2	1962.5	14
a	24-inch Bruce Doublet	1894.3	1906.7	10
mb	4-inch Cooke	1929.5	1929.8	10
ax	3 inch Ross-Tessar Lens	1923.2	1923.8	6
am	1-inch, 1.5-inch Cooke Lenses	1903.5	1904.7	2
me	1.5-inch Cooke Lenses	1911.2	1911.3	2
ma	12-inch Metcalf Doublet	1907.4	1907.4	1
mf	10-inch Metcalf Triplet	1917.6	1917.6	1

Dates represent the first and last observation for KIC8462852 of the respective telescope.

Before discussing the results, it must be noted again that these data are based on the publicly available APASS calibration, and differ from the ones used by Schaefer (2016). That being said, all regressions present significant slopes. However, for the truncated dataset, errors are underestimated by 53% if non-normality is ignored. Clearly, however, the apparent trend cannot be (completely) explained this way. Yet, there are more issues to be found, as will be discussed in the next section.

3.6. *Structural break: Underlying reason*

In the first version of our paper, we have hypothesized a structural break at the "Menzel Gap", a time of missing data in the 1960s. Community feedback has questioned us on the potential underlaying cause for such a break. One would have assumed that the whole time series is a single, perfectly calibrated data stream.

Quite the opposite is true. We have examined the source of these data, and find that a total of 17 telescopes have contributed measurements. Interestingly, as can be seen in Table 1 and in Figure 3, left panel, 16 of 17 telescopes were active between 1889 and 1962. For the time after 1962, all data come from only one telescope, the "Damons North Blue", which is a 4.2cm lens camera. We therefore hypothesize that the underlying root cause of a structural break would be the use of different technology (camera, lens, coatings, plates, emulsions; geography, light-pullution, airmass...) after 1962. Such differences might have canceled out from mixing 16 different telescopes in the time before 1962. Also, there is a considerable difference in limiting magnitude ("Damons North Blue": average 14.18, all others on average 13.57). This definitely affects the quality cuts. It should also have caused dimmer measurements to make it through the cut-off at 12.45mag during later years, which is not the case.

The only overlap of telescopes between the two potential segments occurs in the year 1962. Unfortunately, only 3 data points from the "3-inch Ross Fecker" and 4 data points from "Damons North Blue" overlap, all with large scatter. Formally, we get 12.22 ± 0.17 for the "Damons North Blue" and 11.91 ± 0.13 for the "3-inch Ross Fecker", so that the 1σ error bars nearly overlap. With such few data points, we can suspect, but not resolve, the issue of a potential structural break. We can, however, employ regular statistical tests to check the confidence

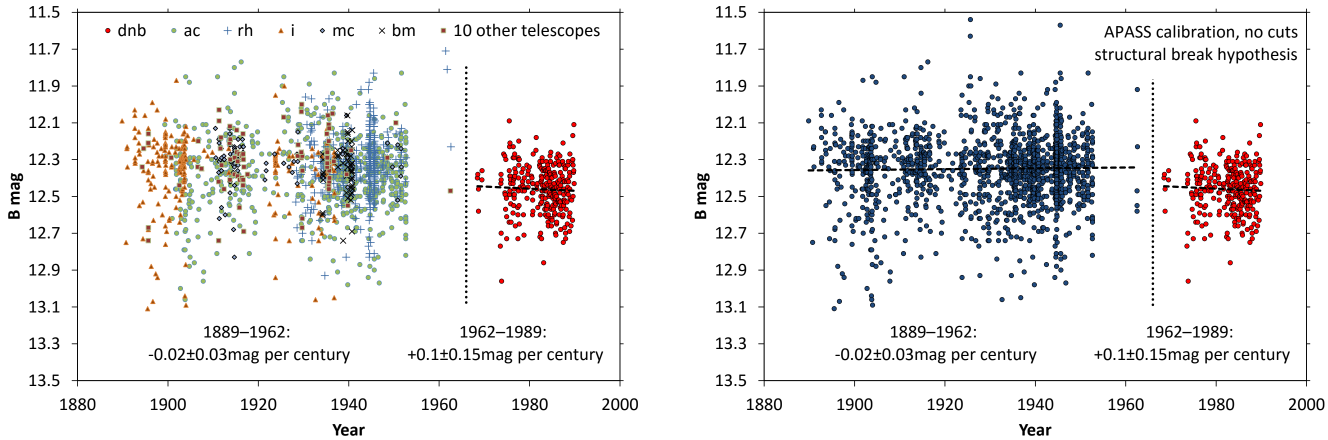


Fig. 3.— Hypothesis of a structural break. Left: The data before 1962 comes from 16 different telescopes, while the data after 1962 (red symbols) comes from only one telescope and shows an offset. Right: Linear regressions for both segments separately indicate constant luminosity within the errors. We hypothesize that the structural break is due to a different technology used after 1962 in "dnb" data, e.g. due to a different emulsion.

(if any) of such a structural break.

3.7. *Structural break: Statistical results*

It is possible that trends in the data are not slow drifts, but rather structural breaks due to e.g. abrupt changes in technology, or calibration. We have checked this hypothesis using the test by Chow (1960), which splits the data in two segments and compares the compatibility of linear regressions in both.

For KIC8462852, the Chow test prefers the year 1962 for a split, coincident with the "Menzel gap", a time of missing data. When splitting the data this way, a structural break is significant at 12σ confidence, and removes any linear trends in both parts (p=0.78 for the first part, p=0.31 for the second part). This is also visually evident from Figure 3. What is more, the (insignificant) linear trend in the first segment has a negative (brightening) slope.

3.8. *Partial data analysis*

A question related to the Chow test is about the observer's perspective. Assume that we would live at an earlier time, and only part of the dataset were available. It should give us a consistent result, albeit with larger error bars. Using this method, we can ask the question: Would an observer in the year X also have found a linear trend with the (back then) available data? Or, the other way around, what data are required to produce the trend?

To test this, we have started with the complete dataset, and then repeatedly deleted the newest data value, until the dimming becomes insignificant.

Our result is that one needs the data from 1889 through 1976.42 (at the 5% significance level), or 1889 through 1978.84 (at the 1% level). The dataset ends in 1989.89. In other words, for a researcher with only the data from 1889 to 1976 (or 1978) at hand, the star would have appeared constant within the errors. This is in stark contrast to the trend as postulated in Schaefer (2016), which should be significant even for a third of the dataset. Therefore, we argue that it is more likely that a sudden jump in apparent luminosity occurred, than a linear trend. Whether this change is of astrophysical nature is unclear. To test this, a comparison of other stars in the dataset needs to be made, and/or measurements from independent datasets need to be checked.

4. COMPARING OTHER STARS AND DATA SOURCES

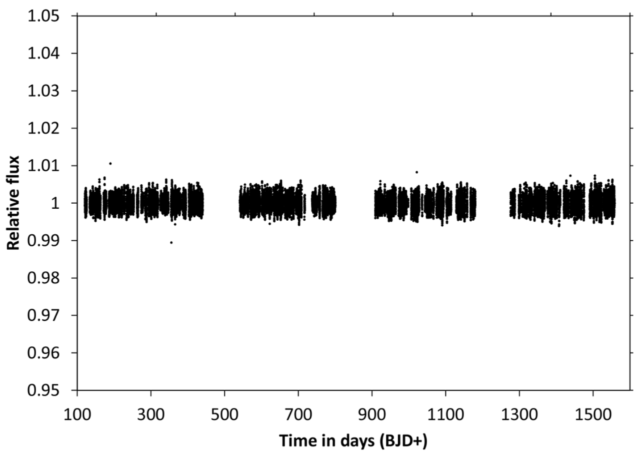
4.1. *Calibration choices*

As discussed in section 3.3, the data is offered in 3 different calibrations by DASCH. This causes another issue when judging trends and structural breaks: Which calibration is to be preferred? Not only are the flux values and error estimates different for each calibration, they also come with very different flags. One such example is KIC7180968, which qualifies as an F-star due to its temperature of 6693K, and as a dwarf according to its $\log_g = 3.887 > 3.5$, as listed in the Kepler Input catalog (see Table 2). We note, however, that SIMBAD lists the star as "F5 III", making it a giant, without any reference given. In any case, the crucial point for us is that it is a stable (not a variable) star. We cross-checked its Kepler photometry (Figure 4, left panel) and find constant luminosity within $< 0.5\%$ rms. Afterwards, we pulled the DASCH data in all calibrations, each excluding the red and yellow plates.

KIC7180968 is much brighter (B=8.7, K_P=8.4) than KIC8462852 (B=12.3, K_P=11.9). This increases the signal-to-noise ratio, making results more significant and thus clearer. As a disadvantage, we have to carefully check that the photometry is still accurate. The median light curve root-mean-square of the Harvard plates is measured to be minimal for Kp=8–9, and only marginally worse for Kp=7–8 (Tang et al. (2013a), their Figure 8, top panel). The useful range in brightness is Kp=5–15, thus saturation is not an issue. Furthermore, the DASCH pipeline adds an "AFLAGS"$>$ 9000 if the "Object is too bright for accurate magnitudes". Finally, saturation in the core of stars on photographic plates is common, but magnitudes can still be extracted correctly for our example[6].

After having established that the photometry for KIC7180968 is usable, we can pull the data from DASCH:

[6] B. Schaefer 2016, priv. comm.: "your choice (...) does not affect your analysis.", with reference to Schaefer (1981).



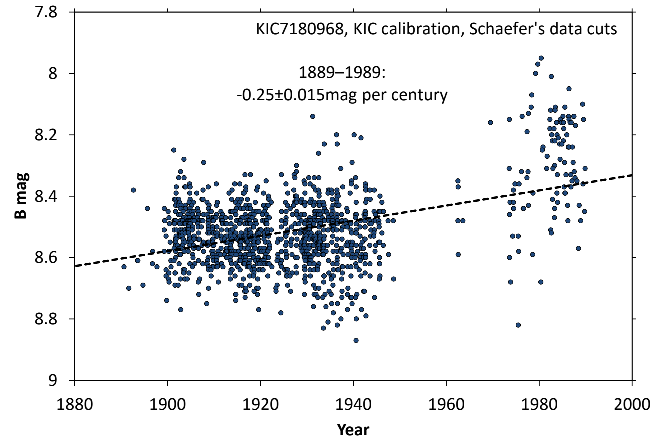


Fig. 4.— Left: Kepler data for KIC7180968 shows constant luminosity within < 0.5% rms. Right: DASCH KIC calibrated data for KIC7180968 with data cleansing as performed by Schaefer. The trend is highly significant, although a structural trend is statistically much preferred.

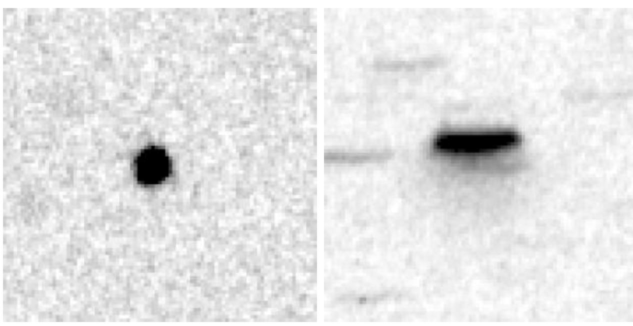


Fig. 5.— Two plates of good (left, point #604) and bad (right, flagged point #2442) data for KIC7180968 in the GSC calibration.

a) In the APASS calibration, there are 3300 data values, of which none have "AFLAGS"< 9000.

b) In the KIC calibration, there are 1971 data values, of which 1452 have "AFLAGS"< 9000.

c) In the GSC calibration, there are 3612 data values, of which 1458 have "AFLAGS"< 9000.

Now, one might argue that the APASS calibration would be superior, and all data values are to be discarded due to defects. Thus, no trend would be found, supporting the argument in Schaefer (2016) that no constant stars with trends exist (except "Tabby's"). This argument is invalid: First, comparisons should not be made on a per-star basis, by always selecting the one calibration (if any) that holds the fewest systematics. Instead, one should select one calibration beforehand. Second, all calibrations offered by DASCH should be considered as valid datasets (or invalidated and removed). We have manually checked a few dozen actual plates for KIC7180968. Two examples are shown in Figure 5. It is evident that the "good" flagged plate is indeed of good, usable quality, while the other is of bad quality. In the example shown, the sky-tracking seemed to have malfunctioned. In most cases, we did visually agree that the flags indicated good (or bad) data quality. This means, then, that the DASCH data for KIC7180968 (at least in the KIC and GSC calibrations) also holds a large number of good data values. Consequently, the KIC and GSC calibrations represent valid datasets.

TABLE 2
Properties of comparison stars

Property	KIC7180968	KIC6366512	KIC8462852
Spectral type	F5	F	F3
K_P(mag)	8.363	11.563	11.912
B(mag)	8.7	11.844	12.262
V(mag)	8.33	11.641	11.705
$\log_g$	3.9	4.0	4.0
T_{eff}	6693	6793	6750

Two comparison stars and "Tabby's star" (last column)

We can now apply the data cleansing as described in Schaefer (2016). We used the KIC calibration (in fact, the result for GSC is virtually identical). The result is shown in Figure 4 and shows a highly significant trend (or structural break). Clearly, there are datasets that pass all criteria in Schaefer (2016) and still have trends.

4.2. *Comparing very similar stars*

As the example used in section 4.1 was much brighter than our star in question, we set to also compare the most similar stars. We used the Kepler Input Catalog and sorted it by Kepler magnitude $11.5 < K_P < 12.2$, as well as gravity $3.8 < \log_g < 4.2$ and temperature $6700 < T_{eff} < 6900$ to find the nearest neighbour with these criteria. The first hit, KIC8814972, showed constant brigthness within the errors. The next best candidate, KIC6366512, presents a peculiarity that we analyze in the following. The properties of this star is shown in Table 2.

First of all, we checked the Kepler photometry to verify that the star is not variable within the 4.25yrs of Kepler observation. The result is virtually identical to KIC7180968 (Figure 4, left panel), which is why we do not show another figure. We find constant luminosity within < 0.2% rms.

Again, we pulled the DASCH data in the KIC calibration and applied the usual (arbitrary) data cuts. The resulting 832 datapoints are shown in Figure 6 (left panel) and present a brightening trend of -0.45 ± 0.03mag per century (or a structural break at very high significance). We stress that this trend persists, and the figure remains essentially unchanged, if *all* flagged data are rejected.

Also, we checked the APASS calibration and found yet *other* stars to exhibit similar trends. One example is

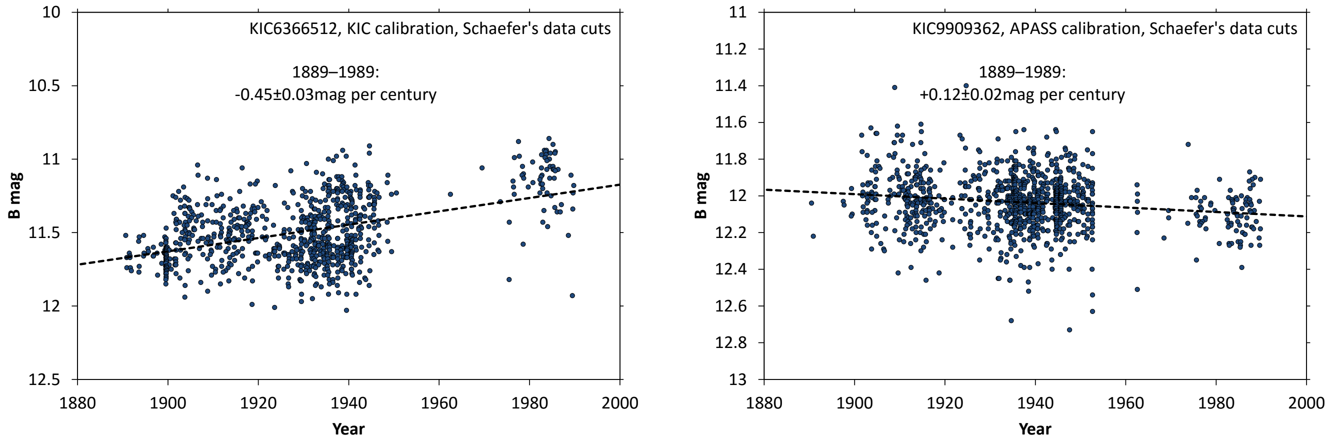


Fig. 6.— DASCH KIC calibrated data for KIC6366512 (left) and APASS calibrated data for KIC9909362 (right) with data cleansing as performed by Schaefer. The trends are highly significant, although structural breaks are statistically preferred.

KIC9909362 (Figure 6, right panel), which is a constant F-star with B= 12.2mag. We performed the same data cleansing and find a dimming trend with slope 0.12 ± 0.02mag per century.

We encourage every critical reader to blindly select a few (e.g., 10) constant stars with B$\sim 11-12$ and examine their DASCH photometry. This simple exercise makes it immediately evident that the data at hand is not always constant at the 0.1–0.2mag level.

4.3. *Comparing the check stars*

The dimming trend announced by Schaefer (2016) was compared to check stars. The author "used the same procedures and selections to produce DASCH light curves for five nearby stars with similar magnitudes" and found "that check stars have constant light curves to a level of 0.03 mag over a full hundred years." The paper reveals, however, only the identities of two of the check stars. We re-examined these.

For TYC 3162-879-1 (KIC8462775), we pulled the data from DASCH for the KIC and the APASS calibration. We find no trend, and no structural break, within the errors of 0.02mag per century.

For TYC 3162-1001-1 (KIC8398290) we tried the same. The KIC calibration, uncleaned and unbinned, results in a brightening trend of -0.07±0.01mag per century. After data cleansing as described in Schaefer (2016), we get 1002 data points and a formal slope of -0.05±0.01mag per century. This is slightly larger than the 0.03 mag found by Schaefer, and formally highly significant. We will discuss in section 4.5 why we believe that all these formally significant trends are overshadowed by long-term systematics.

Also, the slope of TYC 3162-1001-1 is ∼50% of the one we find for KIC8462852 (+0.1±0.02mag per century, section 3.5). As we have not managed to reproduce the results in Schaefer (2016) exactly (see section 3.2), we can only guess where this difference comes from. Part of it might be attributed to binning, which can change slopes and significances. Another part might be explained by the use of other calibration types. Indeed, when we repeat our analysis with the APASS calibration, we get 1772 data points and a slope of +0.03±0.01mag per century. While this might be considered a negligible slope, it is still highly significant. What is more, we also find a structural break in both calibrations. The data before, and after 1962 has a different mean magnitude. We can quantify this difference with a t-test and get significances of 2.6σ (KIC calibration, p-value 0.01) and 2.1σ (APASS calibration, p-value 0.03), respectively. Again, the check stars revealed in Schaefer (2016) do indeed fluctuate less than KIC8398290, but one of the two also shows evidence for systematics in the data.

4.4. *Benchmarking other DASCH stars*

First of all, it must be noted that Schaefer (2016) made the claim that the apparent dimming is "completely unprecedented". As is well known, there is no main-sequence star known to dim by 20% over a century. Therefore, it should in principle be the original author's task to provide proof that no instrumental trends in the DASCH data exist ("Extraordinary claims require extraordinary evidence."[7]). This could be done by comparing hundreds, or thousands of stars with the exact same methodology. This is not provided in Schaefer (2016).

Then, there is the problem of human bias that *does* affect the selection of acceptable criteria for a comparison. Assume that some researcher analyzed photometry of KIC8462852, and found a slope of 0.165mag per century. One could now define criteria for quality cuts, so that the trend in KIC8462852 persists, but trends in most other stars vanish. Clearly, quality cuts (if any!) must be defined completely independently. Then, a large number (thousands) of constant stars must be processed consistently and automatically. Afterwards, the probability distribution of slopes can be used to assess the significance of a slope of 0.165mag per century. A similar approach is used for the removal of instrumental systematics from the Kepler light curves, dubbed "Cotrending Basis Vectors" (Smith et al. 2012).

4.5. *Long-term accuracy of DASCH photometry*

It must be noted that the digitization of the Harvard Astronomical Plate Collection is an extraordinarily important project. The long-term lightcurves are and will

[7] Pierre-Simon Laplace, Théorie analytique des probabilités, 1812. The phrase was popularized by Carl Sagan (1934-1996).

be invaluable for the comparison during ongoing missions such as Kepler K2, and upcoming spacecrafts like TESS, CHEOPS and PLATO, as their long time baseline reference gives leverage in time like no other database. Great effort has been made by the DASCH team to digitize and calibrate the glass plates. We have the highest possible respect for this massive work. The volume, and the quality achieved is nothing short of impressive. As of 02-February 2016, 136,949 plates have been scanned; and 8,540,081,000 magnitudes have been measured. This work is unprecedented and of the greatest importance for the astrophysical community.

The long-term photometric calibration is described by the DASCH team to have an accuracy of $\sim \pm 0.1$mag per century[8]. A best case example is shown on the website, together with the description: "For quality control purposes we are also interested in stars that do **NOT** vary. Such constant-brightness stars enable sensitive determination of various systematic effects and provide a completely independent measure of uncertainties. At left is the lightcurve of such a star demonstrating about +/- 0.1mag photometry over 600 plates, that span 100 years and 19 different plate-series." (their emphasis). A very similar example, with the same accuracy over a century, is given in the original DASCH calibration paper (Laycock et al. (2010), their Figure 15 and caption).

This judgment on the level of long-term accuracy is similar to the one found by Tang et al. (2013a) (0.1mag) and Tang et al. (2013b) who checked 997 Kepler planet host stars and find that "Our typical photometric uncertainty is 0.1–0.15 mag". In their Figure 3, they present the uncertainty as a function of magnitude, indicating an rms of 0.2mag at 2σ for K_P=12mag. In their Figure 1, a decade-long bump is apparent for the Kepler planet host star KIC8191672, although of course it is unclear whether this trend is of instrumental, or astrophysical nature.

Lastly, there is a paper by the DASCH core team, analyzing "KU Cyg, a 5 year accretion event in 1900" (Tang et al. 2011). They state that, for the star KU Cyg, "There seems to be a slight trend of 0.1–0.2mag brightening from 1910 to 1990; however, given our systematic uncertainty over 100 years of ∼0.1 mag (S. Tang et al. 2011, in preparation), it is not convincing."

A trend on the 0.1mag level is therefore not extraordinary, and can be attributed to within the normal fluctuation of post-calibrated, "good" Harvard plate data.

4.6. *Cross-checking SuperWASP data*

As noted by Boyajian et al. (2015), KIC8462852 was observed by SuperWASP for 3 seasons. The first season shows a 0.2mag offset for KIC8462852, which is also seen for the check stars used in Schaefer (2016). Therefore, we discard these initial 22 (of 5351) data values and examine only seasons 2 and 3. As can be seen in Figure 7, KIC8462852 as well as the two check stars show constant luminosity within the errors. Due to the large number of data values, we can determine precise brightness values for the average of both seasons separately. For KIC8462852, we get 12.65701 ± 0.00054mag for the first season, and 12.65663 ± 0.00026mag for the second season, which is a brightening of -0.00038 ± 0.00054mag. In other words, constant luminosity within the errors. The same is true for the two check stars.

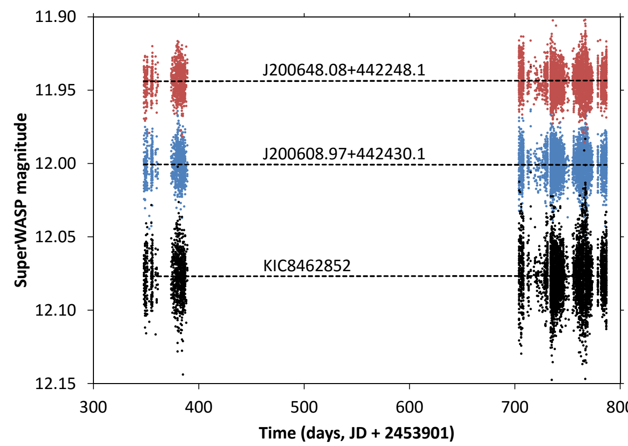


Fig. 7.— Time versus flux for SuperWASP data. Red and blue are the two check stars used by Schaefer (2016), black is KIC8462852 shifted by 0.5mag for visibility. Dashed lines are linear regressions; all with insignificant slopes. The first 22 (of 5351) data values have been removed due to a 0.2mag offset.

However, if a linear dimming trend was present, we would expect a luminosity decrease. As the two seasons are separated by an average of 377 days, luminosity would have to have decreased by 0.00170mag. Assuming constant measured luminosity (and neglecting the insignificant brightening), these results are in contradiction by 3.1σ. In other words, a linear dimming trend should have shown up in the SuperWASP data, but is not detected.

Of course, one could argue that KIC8462852 was dimming from 1889–1990, and suddenly stopped on 2007 May 29 with the begin of the SuperWASP observations. However, "Ockhams Razor tells us that it is very unlikely that one star will suffer two different mechanisms that are unique to that star and that both are only" (Schaefer 2016) present during times coincident with different observation schedules on planet Earth.

4.7. *Cross-checking Sonneberg plates*

The second largest plate archive in the world, after Harvard, is located at Sonneberg Observatory, Germany (Bräuer & Fuhrmann 1992). It was continuously active from the 1930s until today, in a very homogeneous manner, using the same optics and very similar plate scales, sizes and emulsions for many decades. The archive contains $>$ 275000 plates covering the entire northern and equatorial sky (down to declination $-33°$), without any major gaps. Since 1993, the complete archive was digitized Kroll & Neugebauer (1993); Vogt et al. (2004). For KIC8462852, ∼4000 plates exist as scanned bitmaps (∼2000 in B, ∼1800 in V, and other sensitivities). We plan to analyze these data over the course of the next months.

5. CONCLUSION

We re-analyzed time-series photometry from the Harvard plates and find a photometric sensitivity limit of ~ 0.1mag per century, which is an extraordinary achievement for a historical archive like this, and confirms the

[8] http://dasch.rc.fas.harvard.edu/photometry.php

number given in other DASCH studies (e.g., Laycock et al. (2010); Tang et al. (2011, 2013a,b)). However we therefore have to conclude that this is not good enough to derive any trend on similar levels. Assuming that no long-term dimming is present, the puzzling day-long dips in KIC8462852 might indeed be the result of a family of large comets (Bodman & Quillen 2015).

Acknowledgments. We thank Bradley E. Schaefer for his feedback on possible issues with quality cuts and plate selection. We also thank Prabal Saxena for pointing out that some comparison stars are unsuitable, and Rene Hudec and Peter Kroll for advice on photographic plates.

We express our gratitude to the DASCH project which does invaluable work for the community.

REFERENCES

Abeysekara, A. U., Archambault, S., Archer, A., et al. 2016, ApJ in press, arXiv:1602.00987
Bodman, E. H. L.; Quillen, A. 2015, arXiv:1511.08821
Boyajian, T. S., LaCourse, D. M., Rappaport, S. A., et al. 2015, MNRAS in press, arXiv:1509.03622
Bräuer, H.-J., Fuhrmann, B. 1992, Die Sterne, 68, 19
Chow, G. C. 1960, Econometrica 28, 3, 591605
Cong, R. 1999, Stata Technical Bulletin, 52, 4752
Fischer D. A. et al. 2012, MNRAS, 419, 2900
Harp, G. R., Richards, J.; Shostak, S., et al. 2015, arXiv:1511.01606
Hippke, M., Angerhausen, D. 2016, arXiv:1601.07314v1
Johnson, H. L., Morgan, W. W. 1953, ApJ, 117
Kroll, P., Neugebauer, P. 1993, A&A, 273, 341
Laycock, S., Tang, S., Grindlay, J., et al. 2010, arXiv:0811.2005
Laycock, S., Tang, S., Grindlay, J., et al. 2010, AJ, 140, 4, 1062-1077
Liska, J., Skarka, M. 2015, Open European Journal on Variable stars, 173, 1
Lisse, C. M., Sitko, M. L., Marengo, M. ApJ, 815, 2, L27, 4
Marengo, M., Hulsebus, A., Willis, S. 2015, ApJ, 814, 1, L15
Schaefer, B. E. 1981, PASP, 93, 253-262
Schaefer, B. E. 2016, arXiv:1601.03256
Schuetz, M., Vakoch, D. A., Shostak, S. et al. 2015, arXiv:1512.02388
Schwamb M. E. et al. 2012, ApJ, 754, 129
Shapiro, S. S., Wilk, M.B. 1965, Biometrika, 52, 591-611
Smith, J. C., Stumpe, M. C., Van Cleve, J. E. 2012, PASP, 124, 919, 1000-1014
Tang, S., Grindlay, J., Los, E., et al. 2011, ApJ, 738, 1, 7
Tang, S., Grindlay, J., Los, E., et al. 2013a, PASP, 125, 929
Tang, S., Sasselov, D., Grindlay, J. 2013b, PASP, 125, 929
Thompson, M. A., Scicluna, P., Kemper, F. et al. 2015, arXiv:1512.03693
Vogt, N., Kroll, P., Splittgerber, E. 2004, A&A, 428, 925-934
Wright, J. T., Cartier, K. M. S., Zhao, M., et al. ApJ, 816, 1, 17