OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Multi-Epoch Photometric Survey of Exoplanet Atmospheric Variability Using TESS and K2 Archival Data

Aagam Jain
aagamj190@gmail.com

ABSTRACT

The analysis within this study utilized archival Transiting Exoplanet Survey Satellite (TESS) data combined with K2 photometric time-series for 595 "hot Jupiter" planets. Three independent measurements were made from these sources: per-sector transit depth, secondary eclipse depth, and per-epoch transit timing. The transit depth was then analyzed for variability relative to its own stellar noise floor; where the noise floor was derived from generating synthetic constant depth transits added into each star's photometry and processed with the same techniques used to extract the transit parameters.

A total of 17,770 transit times for 435 exoplanets were analyzed and no evidence of orbital decay was found. A lower bound for the reduced stellar tidal quality factor (Q′) at 95% confidence was determined to be $>5.7 \times 10^6$. A stacked value of $Q' = 5.66 \times 10^6$ was also determined. An independent measurement of Q′ for WASP-12b produced values (2.0×10^5) consistent with those previously reported (1.8×10^5). Thus, the number of planets with calculated decay timescales < their host star's age was decreased from 191 "paradoxical" systems to 70. This result favours the interpretation that stars dissipate tidal energy inefficiently, rather than the alternative that many hot Jupiters have migrated inward comparatively recently.

Variable dayside emissions were detected for 21/42 secondary eclipses. In addition, seven out of nine planets observed with both secondary eclipse and transit-depth techniques exhibited variability in both measurements.

INTRODUCTION

Hot Jupiters are a special subset of exoplanets that orbit their host star within a few stellar radii (Dawson & Johnson 2018). Two distinct mechanisms act on the planet independently and simultaneously, both governed by the same physical characteristic, a/R*, orbital separation. The first mechanism is stellar irradiation; this scales as the inverse square of orbital separation, and this also sets the atmosphere, extent, and escape rate. The second mechanism, tidal forcing, scales as the inverse fifth power of orbital separation (Jackson et al. 2008), distorts the planet, heats its interior, and secularly decays its orbit. The objective of each mechanism is to predict photometric fingerprints. This is done through epoch-to-epoch

transit depth variation calculations, secondary eclipse variation data, and shrinking orbital period data through transit timing variations. All the listed factors are extractable from the same archival light curve data. The goal of this survey is to measure the listed factors uniformly at a population scale, with calibration needed to support all findings, which has never been performed in prior surveys. Two problems in the field make the measurement desirable. First, the tidal quality factor Q′, which sets the lifetime of every close-in planet, is uncertain by multiple orders of magnitude (Jackson et al. 2008; Patra et al. 2017). Under the assumption that $Q' = 10^6$, 191 planets in this sample have decay timescales that are shorter than their host star's age; hence, half the population should not even exist. Next, the atmospheric variability of these hot Jupiters has only been established for a small survey of ultra-hot planets (Mikal-Evans et al. 2021), so the general atmospheric patterns of these planets are generally unknown. The survey was designed such that every argument is defensible and can be supported with data.

PHYSICAL FRAMEWORK

The Stellar Tidal Quality Factor

Whenever a planet orbits in close proximity to its host star, the planet's gravity will raise a tidal bulge on the star, and stretch it into a slightly elongated shape, a phenomenon similar to that of when the moon raises tides in earth's oceans. It is important to note however that the star is not a perfect elastic spring, therefore, the bulge is raised and relaxed on each orbit, and that is why internal friction converts some of that mechanical energy into heat.

$$Q = 2\pi \frac{peak\ energy\ stored\ in\ the\ tidal\ distortion}{energy\ dissipated\ per\ cycle}$$ (Goldreich & Soter 1966)

The quantity constrained in this work is the reduced tidal quality factor $Q' = 3Q / 2k_2$, where k_2 is the second-order Love number describing the deformability of the star. Working with Q′ avoids the need to specify the stellar interior density structure separately.

A larger Q′ will result in a loss of very little energy in each cycle, the bulge remains close to alignment with the planet so there is a low level of gravitational force acting to distort its motion and therefore there is very slow orbital movement. A small Q′ indicates that there is a lot of dissipation (or friction) occurring – as a result the bulge will be lagging behind where the planet is positioned at the time, and thus will act as a drag on the planet's motion causing rapid decay of the orbit.

The amount that an orbit shrinks due to tides is determined by the inverse proportionality of the value for Q′. In fact, the reason we can measure Q′ at all is because we can determine (and/or establish) how quickly or slowly orbits decrease (or do not decrease), which is exactly what this study does. Lower bounds on Q′, therefore, translate statements regarding the observed orbital stabilities of stars into physical descriptions of those same stars. Therefore, a larger lower bound on Q′ indicates that stars dissipate tidal energy less efficiently, and therefore that tidal forces remove orbital energy from close-in planets more slowly.

This same quantity governs a large number of systems across the spectrum of astrophysics, such as the Earth-Moon system, close binary star systems, and Neutron star mergers (Goldreich & Soter 1966; Abbott et al. 2017).

Direct observations of a star's tidal quality factor are relatively scarce and difficult to obtain with observational resources. So when one obtains an empirically-derived (population level) limit on the stellar tidal quality factor (Q′), by using about two hundred hot Jupiters, this has consequences beyond their continued existence; it will be used as input into all theoretical models of binary star evolution, circularization of stellar orbitals, and general tidal-dissipation in stellar interiors, which we have previously been able to calibrate using the handful of solar system examples.

The Inverse-Square Law: Irradiation

A star radiates energy in all directions. As that energy travels outward it spreads across the surface of an ever-expanding imaginary sphere, whose area grows as the square of its radius. Consequently the flux — the energy arriving per unit area — that a planet receives falls as the inverse square of its orbital distance, defined by the formula stated below:

$$F = \frac{L}{4\pi a^2}$$

L is the stellar luminosity, and 'a' is the orbital separation. Doubling the orbital distance therefore reduces the incident flux to one quarter. Through this relationship we define the planet's equilibrium temperature and thermal state of the atmosphere.

The Inverse-Fifth-Power Law: Tidal Forcing

The gravitational force due to the star is what causes the tidal force. That is because it causes one end of an object to be stretched toward a central mass (in this case the star) while the other end is not, which results in stretching or deformation of the object. As a result of the inverse relationship of the gravitational force (as opposed to the inverse square law) the difference in that force at either end of an object over a given distance will fall off as a function of the inverse cubic relationship of distance. When you integrate the resulting torque to get the average rate of change of orbital parameters, you find that the time dependence of the rate of change of the orbit decays as a function of the inverse fifth power of the distance between the two objects.

$$\frac{dP}{dt} = -\frac{27\pi}{2Q'}\left(\frac{M_p}{M_*}\right)\left(\frac{R_*}{a}\right)^5$$

where M_p is the planetary mass, M_* and R_* the stellar mass and radius, a the orbital semi-major axis, and Q′ the reduced stellar tidal quality factor. The negative sign denotes a shrinking orbital period. The prefactor $27\pi/2$ follows the convention of Patra et al. (2017) and Yee et al. (2020), adopted here so that our results are directly comparable with published values.

That very large difference in rates of change is what defines the study of tides: A planet that is twice as close to its star has approximately 32 times larger tidal effects. This is why tidal forces have such a small

impact on planets that are farther from their stars but dominate the behavior of the innermost hot Jupiter type planets. Thus, since both radiative heating and tidal acceleration have the same parameter space (i.e., distance), we may separate these effects by virtue of their disparate dependence on that parameter.

Derived Physical Quantities

Three further quantities are computed for every planet from its measured parameters.

Equilibrium Temperature

The equilibrium temperature of a planet, is the temperature at which the planet settles when the starlight absorbed and thermal radiation it emits are balanced, and is given by the formula:

$$T_{eq} = 278.3\,K \times \left(\frac{L}{L_{\odot}}\right)^{1/4} \left(\frac{a}{1\,AU}\right)^{-1/2} (1 - A)^{1/4}$$

Wherein 'A' is the bond albedo, this is the most used quantity in exoplanet atmospheric science. This sets the planet's analysis for cloud composition, chemistry, and escape, and it is the primary criterion by which targets are prioritised for spectroscopic checks.

Atmospheric Scale Height

Scale height is a key measurement in this survey, as it measures the vertical distance over which the atmospheric pressure drops by a factor of e, this gives you a measure how thick/'puffy' the atmosphere really is.

$$H = \frac{kT_{eq}}{\mu m_H g}, \quad g = \frac{GM_p}{R_p^2}$$

where k is Boltzmann's constant, μ the mean molecular weight, m_H the mass of a hydrogen atom, and g the planetary surface gravity. Scale Height determines the size of an Atmospheric Annulus and therefore the magnitude of the transmission spectroscopic signal. A larger H would result in a larger annulus that star light would have to pass through during a transit which in turn makes it easier to observe a planet's atmosphere.

The Jeans Escape Parameter

This parameter is a measure of the ratio of the gas molecule's gravitational binding at the top of an atmosphere due to thermal energy, and is given by the following formula:

$$\lambda = \frac{GM_p \mu m_H}{kT_{eq} R_p}$$

Mathematically, a lower value would suggest that thermal motion is comparable to the escape energy, and hence the planet's atmosphere is bleeding into space. Meanwhile, a higher value would mean that the atmosphere is secured in place. It is the field standard diagnostic for the atmosphere's mass loss and is central to the study of decaying planets.

Every variable that the survey measures is a fundamental physical property. These properties have applications beyond just Hot Jupiter planets. For example the tidal quality factor provides connections from planetary orbital decay to lunar recession, to volcanic activity on the moon and circularization in binary stars and also information on mergings of neutron star systems. The two dimensionless quantities for orbital motion form the basis for several other areas of research including Habitable Zones, Direct Imaging, Debris Disks, the Roche Limit and Stellar Spin-Down. All of the derived parameters used to model atmospheres here are used by researchers working in this field as well. Therefore, every measurement we make will contribute to a larger and interrelated body of physics rather than be a stand alone measurement.

METHODOLOGY

Pipeline Overview

The survey follows the following sequence of tasks and returns the given outputs accordingly. The survey is a sequence of eleven stages executed in a cloud notebook as given:

Stage	Function	Principal output
00	Constants, unit conversions, physics functions, marker utilities	definitions
01	Archive query, selection cuts, derived quantities	hj_master.csv
02	Photometry retrieval from MAST	Parquet + JSON per planet
03	Per-sector transit depth fitting	716 sector fits; depth summary
04	Injection-recovery noise floors and calibration	noise floors; calibrated depths
05	Per-sector secondary eclipse measurement	eclipse summary
06	Per-epoch transit timing and decay classification	timing summary; 17,770 epochs
07–09	Irradiation tests, tidal ensemble, power-law synthesis	result tables and figures
10	Compilation of all results	summary report

Table 1 traces the sample from the initial archive query through each selection criterion and measurement stage, so that every sample size quoted in this paper can be located against its parent population.

Stage	Criterion applied	N remaining
Archive query	Confirmed transiting planets	4,695
Hot Jupiter selection	$P < 10$ d; $R_p \geq 0.5\ R_{Jup}$; $M_p \geq 0.1\ M_{Jup}$ or unmeasured; Tmag ≤ 13.5	595

Stage	Criterion applied	N remaining
Photometric coverage	≥ 2 TESS sectors	515
Depth fitting succeeded	≥ 2 sectors fitted successfully	458
Depth quality cuts	Mean depth 1,000–60,000 ppm; fractional variability < 0.5	425
Noise-floor calibration	≥ 3 successful injections	359
— classified variable	F-test $p < 0.05$ against own floor	89
— conservative subsample	Below-median contamination range	31
Power-law analysis	Positive excess variability	247
Eclipse sample (Survey A)	Tmag ≤ 11; $T_{eq} \geq 1000$ K; eccentricity ≤ 0.1	136
— eclipses detected	SNR > 3	42
— dayside variable	χ^2 test, after stellar attribution	21
Timing sample	≥ 6 accepted epochs	435
— entering ensemble Q′	Negative predicted rate; finite uncertainty; |measured| < 100 ms yr^{-1}	191
Paradox count	Archive stellar age and decay timescale available	543

Table 1. Sample attrition through the pipeline. The paradox count is drawn from the full 595-planet sample rather than the timing subsample, since it depends only on archive parameters. One system classified as dayside-variable by the χ^2 test is attributed to stellar activity and excluded from the census, giving 21 rather than 22.

Sample Construction

Various parameters from the system were pulled using astroquery to access the NASA Exoplanet Archive Planetary Systems Composite Parameters table (NASA Exoplanet Archive 2026). Astroquery is then used to find all confirmed transiting exoplanets in the database and this returned 4,695 systems over 39 columns. The use of the composite table as opposed to the full planetary systems table is due to the fact that there is a single row for each publication (in some cases multiple publications), therefore a well studied planet will appear in dozens of rows with variable data parameters. This means that consumers will need to implement their own method for resolving these references. In contrast, the composite table does this arbitration internally and authoritatively. The results are also cached after they have been retrieved for the first time. This was done as a matter of reproducibility, as the NASA Exoplanet Archive is updated continually. Therefore, if the results were not being cached, an uncached version of the pipeline would be analysing a completely different set of samples on each run.

Cut	Threshold	Reason for the threshold
Period	< 10 days	Conventional hot Jupiter boundary; beyond it the effects sought are unmeasurable
Radius	$\geq 0.5\ R_{Jup}$	Restricts to genuine giants, excluding sub-Neptunes with different atmospheric physics
Mass	$\geq 0.1\ M_{Jup}$, or unmeasured	Retains planets characterised only by photometry; requiring a mass would bias toward older, brighter, radial-velocity-followed systems
Brightness	Tmag ≤ 13.5	Faint limit for useful per-sector depth precision
Coverage	≥ 2 TESS sectors	Minimum for any sector-to-sector variability measurement

In deliberately trading one kind of bias for another, we decided that excluding the large group of TESS-discovered planets characterised solely by their photometric measurements would not have been worthwhile, since doing so would have restricted the sample to planets that had been subject to radial velocity follow-up. Such a cut would have removed those planets that can be discovered via photometry but lack sufficient brightness and age to have been followed up with radial velocity measurements. By doing so, we would have introduced distortions into both the distribution of brightness of the planets in the sample and the distribution of time of discovery.

The second decision regarding the sample involved retaining a relatively small number of objects that were identified as having a planet-like status strictly on the basis of their size. The presence of these objects does not affect the tidal analysis, since without a measured mass no tidal decay rate can be computed for them, and in the present sample only two of the 595 planets are affected.

After the coverage requirement, 515 planets have two or more TESS sectors and enter the depth analysis. A brighter subsample of 204 planets is selected by TESS magnitude ≤ 11 and equilibrium temperature ≥ 1000 K; of these, 136 satisfy the remaining eclipse-sample criteria (Table 1) and define the secondary-eclipse sample (Survey A). This restriction is necessary because eclipse depths are approximately an order of magnitude smaller than transit depths and are measurable only for the brightest and hottest targets.

Selection Effects and Their Consequences

We identify four types of selection effect that can affect how our findings are interpreted.

Because the sample is limited to stars brighter than TESS magnitude 13.5, the precision per sector increases; however, because it is not possible to determine whether all hot Jupiters are included in the sample, that is, it is not volume-complete, the statistics reported here represent the population of bright hot Jupiters, not all hot Jupiters.

The minimum number of sectors required for a target (two) will result in an over-sampling of regions near the ecliptic poles where targets were repeated. The fact that classification of a planet as variable has a

median of 6 sectors, while non-variable classifications have a median of 4 sectors ($p = 2 \times 10^{-6}$, Mann–Whitney test), indicates that the larger number of observations resulted in a greater probability of detecting a signal, and not that variable planets have different properties from non-variable planets. Therefore, the fraction of variables reported here represents a conservative estimate of the true fraction, since fewer observations increase the likelihood of missing a detection.

The absence of mass measurements will affect only the atmospheric component discussed above.

The most significant limitation for interpretation of secondary eclipse detections is related to detectability limits. Secondary eclipses are detectable only for very bright daysides at each given temperature. Thus, establishing that there is no systematic relationship between the presence or absence of dayside variability and equilibrium temperature is based only upon the behaviour of the dayside sub-population defined by having sufficient brightness to be detected by secondary eclipse measurements.

Derived Parameter Choices

At this point, all quantities described in the Physical Framework section have been computed using the following selections:

- For the equilibrium temperature, the Bond albedo has been selected to equal 0.1
- A mean molecular weight of 2.3 amu has been used for calculating scale height and the Jeans parameter
- The age – activity relationship developed by Sanz-Forcada et al. (2011) has been used for estimating XUV fluxes (Sanz-Forcada et al. 2011)
- A value for the prefactor for tidal phase lag equal to $27\pi/2$ has been chosen for matching the values of Patra et al. (2017) and Yee et al. (2020)
- When stellar masses could not be found in the Kepler archives they were reconstructed using Kepler's Third Law. Thus planets have not been removed from consideration based on missing stellar masses.
- As the tidal rates are dimensionless they have been converted into milliseconds/year within the same routine, so that no user can double-conversion or fail to convert – something that did occur while developing code for this project.

Photometric Data Extraction

TESS (Ricker et al. 2015) photometry for the project was retrieved using the Mikulski Archive for Space Telescopes (MAST) through lightkurve (Lightkurve Collaboration 2018). Where more than one reduction of the same sector exists, we applied a fixed product-priority hierarchy and retained a single light curve per sector, since different reductions differ in aperture definition, background subtraction and systematics treatment. The priority is:

1st: Two minute cadence SPOC (Science Processing Operations Center) reduction

2nd: TESS-SPOC reduction

3rd: Thirty minute cadence QLP (Quick-Look Pipeline) reduction. Only the highest priority reduction per sector is retained.

Archival Kepler (Borucki et al. 2010) and K2 (Howell et al. 2014) photometry were also retrieved to provide longer baselines for timing. In the current campaign, 33 of the planets had K2 coverage which reached up to 10.7 years, with no Kepler prime mission data available for any of the targets. A total of 3,222 TESS sectors and 41 K2 campaigns were retrieved by the survey, with a total of 515 planets meeting the two-sector minimum.

Each lightcurve was converted into three output values: Time, Normalized Flux, and Flux Uncertainty and stored as Compact Parquet with a corresponding JSON sidecar file containing additional information about each lightcurve such as the Crowdsap contamination fraction, Mission ID, Sector ID, and Exposure Time. This resulted in each planetary dataset being significantly smaller versus their original size in FITS format.

Transit Depth Measurement

Each sector of each planet was fitted in the following five steps:

Detrending. An instrumentally trend-removed light curve can be made by applying a bi-weight time-windowed filter (wotan; Hippke et al. 2019) using a window size of three transit duration. I have selected the bi-weight estimator as opposed to a moving mean or median due to its ability to suppress flare events and cosmic ray activity that would otherwise corrupt this data set. This choice of window size is somewhat arbitrary in nature; there is no optimal size for this application since the use of a smaller window will remove some of the transit signal while larger windows may either miss some real instrumental variability or lag behind an actual trend. In order to prevent the detrend from removing signal rather than noise, all observations whose cadence is less than 0.75 of a transit duration from the calculated mid-transit time are masked during the fitting process.

Model configuration. A Mandel & Agol (2002) transit model, computed with the batman implementation (Kreidberg 2015), uses only two independent parameters — the radius ratio, or radial difference between the star and planet (which determines the overall "depth") and an average time offset that captures the combined effect of all of the remaining (e.g., ephemeris) systematics. The period, scaled separation, inclination, and second order limb-darkening coefficient are held constant and set based on archival values derived from the stellar temperature. These constraints provide hard limits on the allowed solution space for the optimizer such that it can't get lost in physically unrealistic regions. This constraint was done intentionally: there is strong degeneracy between depth, impact parameter, and limb darkening for a given sector of a transit lightcurve. If you allow all three to vary independently, then you will introduce large amounts of correlation in your uncertainties as well as create apparent variability due to fitting errors. Since this survey compares the relative depth of different sectors of the same planet, the systematic error in your fixed geometry is common to both data points, therefore most of the errors cancel when taking the differential measurements. However, as noted above, because we have constrained our model to hold certain parameters fixed, we should note explicitly that while these fits are suitable for

making relative comparisons between different sectors of a given planet's orbit around its host star, they do not give you accurate absolute depths for use in atmospheric retrievals.

Contamination correction. TESS pixels subtend 21 arcseconds so light from unrelated background stars tends to dilute nearly every aperture fitted. Each depth fitted is divided by sector CROWDSAP value; median survey width is 0.958 so typical correction is a few percent. Correction assumes contaminating flux is constant but background stars vary so we record the range of CROWDSAP across sectors for planets and treat that as a nuisance regressor: this exposes any residual dependence rather than just absorbing it.

Uncertainty estimation. Because TESS photometry contains noise correlated on transit timescales, the formal photon-limited uncertainty understates the true depth error; each uncertainty was therefore inflated by the β factor, defined as the ratio of the measured scatter of 30-minute residual bins to the scatter expected for uncorrelated noise:

$$\beta = \frac{\sigma_{measured}(30-minute\ bins)}{\sigma_{point} / \sqrt{N_{per\ bin}}}$$

A value of unity corresponds to uncorrelated noise. Across the 716 sector fits the median β is 1.11, with a tail extending to 6.75, and the resulting median depth uncertainty is 15.4 parts per million (ppm).

Aggregation. For each planet, we find, in order to make the first of our claims regarding a given planet's atmospheric properties, a weighted average for the surviving fits as a function of time, the variation around this average, and the fractional variation of the average (the ratio of the standard deviation to the mean); two tests are then applied to these results before making a claim; the chi-squared test is used to determine whether there was sufficient variability to determine what kind of periodicity is present in the data.

$$f_{var} = \frac{\sigma_{depth}}{depth}$$

The second test uses CROWDSAP statistics to assess whether the variability detected is due to stellar activity originating in the host star rather than the planet. In addition, we calculate both the median value of the beta parameter derived from the best fit model and the amplitude of stellar activity using a Lomb-Scargle analysis (VanderPlas 2018; Astropy Collaboration 2022) on the out-of-transit light curve. We apply two conditions before claiming variability; the first condition is that the mean depth of all fits associated with a given transit must fall within the range of 1000–60,000 parts per million (ppm). The second condition is that the fractional variability of all such fits cannot exceed 50%. Planets violating either condition are excluded from the variability analysis.

There are two ways to treat stellar activity. Stellar activity is determined for every planet and sector, by measuring the stellar activity with the out-of-transit photometry, and then carried forward into the population regression as an independent variable (covariate), so that if there is any relation between the variability of a star, and the depth of a transit of a planet through its disk, it will be revealed by the regression and not absorbed by the regression. Stellar activity is never subtracted from the measurement

of each transit, because the relationship between a stars' total broadband variability, and how much of a transit's area goes through a star's disk varies with what we don't know about where on a star's surface active areas are located; by using stellar activity as a covariate instead of subtracting it from each measurement of a transit, this dependency is avoided.

Injection–Recovery Calibration

The validity of this calibration depends entirely on the details of its implementation, which are therefore set out in full. Cadences from transits during any real transit are removed beforehand so that genuine signal does not contaminate control measurements and sectors with fewer than 500 cadences are skipped afterwards. Synthetic transits are injected at orbital phases 0.35 and 0.65 avoiding both phase 0 transit and eclipse at phase 0.5 and sampling different parts of light curve symmetrically; using two phases per sector costs twice but survey performed a total of 4405 injections with median of 10 per planet and maximum of 60. Each injected transit uses radius ratio, separation, inclination and limb darkening of the planet and smears them to exposure time of that sector because floor measured with different signal depth or shape would not characterize pipeline behavior on real data. After injection light curves pass through the same detrending window and optimization and uncertainty treatment as real data, the only difference between control and measurement is whether underlying signal varies and any pipeline artifacts are present in both and cancel. Floor is taken as standard deviation of recovered depths divided by their mean; excess variability is obtained by removing that in quadrature and combining contributions as independent variances. Significance is assessed using an F test on variance ratio with degrees set by sector and injection counts.

$$\sigma_{excess} = \sqrt{\sigma^2_{measured} - \sigma^2_{floor}}$$

$$F = \left(\frac{\sigma_{measured}}{\sigma_{floor}}\right)^2$$

Planets are classified as variable only if that test rejects equality at five percent level; planets that do not exceed their floors are recorded as upper limits rather than discarded and retain their constraining power.

Secondary Eclipse Measurement

Eclipses were evaluated based on flux comparisons (using direct measurements) instead of comparing with models. All cadences lying within 0.4 transit durations of the predicted eclipse centre were pooled across every eclipse in the sector, and their median flux computed. These median values have been compared to baselines which spanned 0.75 to 1.75 times the duration of the transit period (with all cadences close to a transit being excluded). The error value was calculated using the baseline noise variation and the number of samples, as well as the penalty factor used when estimating the median. The depth was corrected using the same method as the CROWDSAP correction for transit data.

A simple comparison of flux levels around eclipsing events (i.e., a "box" measurement) was chosen instead of a full occultation model because there is less freedom for the data to constrain and the errors associated with the full model would be unreliable at such low amplitudes (i.e., those associated with

TESS eclipses). A comparison of flux levels in this manner results in a robust estimate of the eclipse depth, defined as $(F_{out} - F_{in})/F_{out}$, where F_{in} and F_{out} are the median in-eclipse and baseline fluxes respectively, although it does result in a slightly systematic underestimation of depth due to inclusion of ingress/egress cadences in the "in-eclipse" pool and therefore contributing to dilution of the flux. This bias occurs consistently for each sector of a given planetary system and will cancel out in a test of variability, but will make absolute depths unusable for emission spectroscopic analysis.

Classification	**Criterion**
Detectable	Weighted mean eclipse signal-to-noise above 3
Weather-variable	Detectable, and a chi-squared test rejects constant depth at p below 0.05
Stellar-flagged	Weather-variable, but eclipse depths correlate with the star's own sector-by-sector activity amplitude at |r| above 0.7 and p below 0.05; attributed to the star and excluded from the census

The principal defence against false positives is the stellar attribution test, because a host star whose brightness varies between sectors alters the reference flux against which the eclipse depth is measured and can thereby mimic planetary variability. Since the depth-fitting stage already measures each star's activity amplitude sector by sector, the two series can be matched epoch to epoch and correlated directly. Correlation is used and at least four common sectors are needed; so this guard is limited for planets that are observed infrequently.

Transit Timing

The time of each transit is determined separately. For every transit, a "window" of 1.6 times the duration of the transit is defined around the time when a transit was predicted to occur. During this window, the average background level of light is subtracted from the observed light curve using a straight line fit to the parts of the curve away from the transit (which is sufficient for a short time like a single night and does not act as a global filter that could distort the event). The center time of the transit is then found through an explicit chi-square search: the shape of the transit itself is assumed constant, only the starting time of the transit is changed systematically across 81 possible offset values spanning ±0.05 days in increments of 0.00125 days (approximately 1.8 minutes). An exhaustive search is performed instead of an optimization routine because there is only one degree of freedom, and under the presence of noise the function can become multi-modal.

Ephemeris fitting and the decay classifier. Since an exhaustive search will find the absolute global minimum and explicitly expose any pathological cases, the minimum is further refined by fitting a parabolic curve through the seven closest points from the search, and the error bar in time is calculated from the curvature of this parabola scaled by the residual variance of those same points.

Three additional checks were made to eliminate problematic entries from the list of acceptable epochs:

A. Fewer than five cadence measurements occurred during the transit. In such cases, too little information exists to perform a reliable analysis.

B. The lowest point in the search fell near one end of the window in which the search was conducted. In these cases, the actual minimum likely lay outside of the window, and therefore any measurement within the window would be an artifact due to boundaries of the window.

C. The second-order term (curvature) in the parabolic fit was zero or less than zero. When either condition occurs, it indicates that the fit has failed to converge.

In total, 17,770 epochs passed all three tests and had their mid-transit time recorded.

$$T_{mid}(E) = T_0 + PE + c_2 E^2, \quad c_2 = \frac{1}{2}\frac{dP}{dE}, \quad \frac{dP}{dt} = \frac{2c_2}{P}$$

The surviving epochs for each planet were then fitted by weighted least squares, first with a linear ephemeris and subsequently with a quadratic one, following the rejection of all epochs differing from the first-pass linear solution by more than five standard deviations. The linear fits provide estimates of the orbital periods considerably more precise than those reported in the archives. The period derivatives follow directly from the quadratic coefficients as shown above.

A planetary system is identified as an orbital-decay candidate only when all four of the following conditions are met simultaneously. The improvement in χ^2 of the quadratic model over the linear model must be at least 16. The curvature of the quadratic must be negative. At least six epochs must survive the rejection criteria, so that the quadratic retains meaningful degrees of freedom. Finally, the reduced χ^2 of the quadratic fit must be below 3, which rules out cases where a quadratic is formally preferred over a linear model but does not describe the data well. The threshold of $\Delta\chi^2 \geq 16$ represents a substantial margin over what would normally constitute statistical significance; it was adopted to allow for correlated systematics in transit-timing data, which can inflate a marginal preference for the more complex model. Systems displaying significant positive curvature, which cannot arise from tidal decay, are instead recorded separately as timing anomalies.

Each planet, whether or not it yields a detection, is converted into a lower limit on Q′ by comparing the 95 per cent upper bound on its measured period derivative with the rate predicted for that specific system.

Population-Level Inference

The inference below rests on the constant-phase-lag equilibrium-tide formalism (Jackson et al. 2008), in which a single frequency-independent parameter Q′ characterises the dissipation of tidal energy within the host star. We adopt this convention because it is the one used in the published measurements against which our results are compared (Patra et al. 2017; Yee et al. 2020), so that the values are directly comparable. It is nonetheless an approximation: dynamical-tide theory predicts that dissipation varies with stellar structure, stellar rotation, and the frequency of the tidal forcing. The limits derived here should therefore be read as population-effective values within this model, rather than as constants intrinsic to the stars themselves.

The ensemble constraint. Q′ is treated as a single parameter for all stars; The joint Chi-Square was then evaluated over a grid of log Q′ ranging from 3.5 to 10 at 600 points. Each planet was given its own weighting based on the measurement precision for that planet and scale by mass, radius, and distance to

the star, such that a well-measured susceptible planet would have a high weight (and therefore dominate) and a poorly measured insensitive planet would have a low weight (and contribute very little):

$$\chi^2(Q') = \sum_i \frac{\left[(dP/dt)_{i,\,measured} - (dP/dt)_{i,\,theoretical} \times 10^6/Q'\right]^2}{\sigma_i^2}$$

The ninety fifth percentile lower limit is the smallest grid value whose chi squared lies within three point eight four from plateau level. Plateau level here is the threshold for one parameter at this confidence level. An artefact guard accepts planets only if the theoretical rate is negative, uncertainty finite and measured rate is less than 100 milliseconds per year in magnitude; the last criterion excludes physical signals incompatible with tidal origin and this criterion was adopted with knowledge of the data and is declared as such.

Independent cross-check. A different statistic, methodologically distinct from the ensemble constraint, was computed from the same epochs. The timing curvature for each planet was normalised by the curvature that the planet's predicted decay would produce at the reference value $Q' = 10^6$. This produced a dimensionless measure of decay efficiency, describing how closely the fitted curvature matches the prediction of the model: a planet decaying exactly as the model predicts has an efficiency of 1.00, while a planet not decaying at all has an efficiency of 0.00. These efficiencies were then combined across the 30 most tidally susceptible systems, weighted by inverse variance. The combined efficiency is 0.07 ± 0.06, which is consistent with zero, and corresponds to a lower limit of $Q' > 5.66 \times 10^6$. Since the statistic used here draws on some of the same data as the ensemble likelihood but on very few of the same statistical assumptions, the fact that the two results agree to within one per cent means that they represent an independent verification rather than simply a repetition.

Model comparison and predictors. Scaling hypothesis predicts a specific functional form so testing requires comparing models rather than fitting one. Deviations against orbital separation are fitted using a single power law and also using a continuous broken power law whose break point is searched over central seventy percent of separation range; these are compared using Bayesian Information Criterion which penalizes extra parameters in the broken model so a break is accepted only if data justify added complexity. Uncertainties in slopes are based on 1000 bootstrap resamples rather than analytic formulae because residual distribution is not Gaussian. Separation, equilibrium temperature and tidal susceptibility are intrinsically interrelated by design. Two approaches deal with this. Partial rank correlations are computed on residuals: each variable is transformed to ranks, controlling variable regressed out and then residuals are correlated to isolate say dependence on separation at fixed temperature. Separately, a multiple regression of log excess variability against standardized log temperature, log theoretical tidal rate and CROWDSAP range and log stellar activity reports coefficients along with inflation factors so collinearity becomes clear rather than hidden. Including nuisance terms alongside physical predictors lets analysis test whether any physical correlation remains after controlling for instrumental contamination and stellar activity levels.

Statistical Framework and Hypothesis Tests

Every statistical claim in this work is made against an explicitly stated null hypothesis at a stated threshold. Table 2 lists each test performed, its null hypothesis, the statistic used, and the threshold adopted.

Test	Null hypothesis	Statistic	Threshold
Depth constancy, per planet	Transit depth is constant across sectors	χ^2 against weighted mean	$p < 0.05$
Excess variability, per planet	Measured depth variance equals the injection–recovery floor variance	F-test on variance ratio, dof = ($n_{sectors} - 1$, $n_{injections} - 1$)	$p < 0.05$
Eclipse detection	Eclipse depth is zero	Weighted mean depth / uncertainty	SNR > 3
Eclipse variability	Eclipse depth is constant across sectors	χ^2 against weighted mean	$p < 0.05$
Stellar attribution	Eclipse depth series is uncorrelated with the stellar activity series	Pearson r, ≥ 4 common sectors	$\|r\| > 0.7$ and $p < 0.05$ → attributed to the star
Orbital decay, per planet	Period is constant (linear ephemeris)	$\Delta\chi^2$ of quadratic versus linear	$\Delta\chi^2 \geq 16$, negative curvature, ≥ 6 epochs, reduced $\chi^2 < 3$
Ensemble Q′ bound	No orbital decay in the population	Joint $\chi^2(Q')$ on a 600-point grid	$\Delta\chi^2 = 3.84$ from the plateau (95%, one parameter)
Variability–separation scaling	Slope is zero	Bootstrap slope, 1000 resamples	95% bootstrap interval excludes zero
Break in the scaling relation	A single power law is sufficient	ΔBIC, broken versus single	ΔBIC > 6 required to accept a break

Table 2. Statistical tests applied in this work, with their null hypotheses, test statistics and significance thresholds.

Three conventions that were followed are described below. The first refers to how correlation values are reported. Because the relationships expected from the physical properties of the stars and planets follow power laws, and because the Spearman rank correlation coefficient is less sensitive to extreme values, it has been used throughout in place of Pearson's r. The second convention concerns the uncertainties on the slopes of the fitted relations. These error bars were determined from 1,000 bootstrap resamples of the planet sample rather than analytically, because the residual distributions around the models are not normal. Finally, no global correction for the number of tests performed has been applied across the related

hypotheses examined in this paper. Results with p-values close to 0.05 should therefore be regarded as indicative rather than definitive. Results carried into the conclusions are those significant at $p < 0.01$, or those confirmed by an independent second test.

RESULTS

Sample Overview

Before we delve in the results, I would like to recap the scope of this study. The study defined 595 hot Jupiters using data collected in the NASA Exoplanet Archive, which includes; 482 classical hot Jupiters; 67 massive hot Jupiters (mass > 3 Jupiter masses); and 46 ultra-hot Jupiters with equilibrium temperature greater than 2000 K. This study has shown that this set of planets is found over a range of orbital periods (from 0.67–9.96 days), equilibrium temperatures (from 441 K to 4050 K) and host star luminosity (TESS mag from 6.4 to 13.5). The sample was restricted to those planets whose host stars had received sufficient TESS observations to allow their variation to be measured.

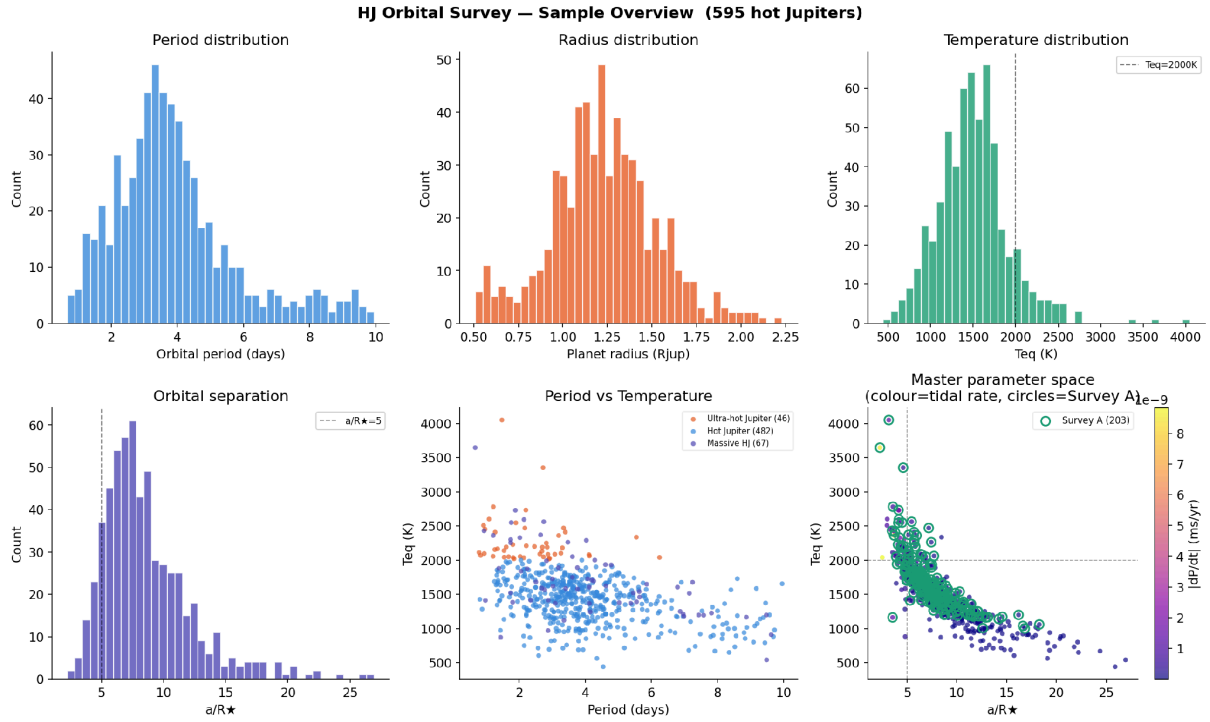


Figure 1. Distributions of the 595-planet hot Jupiter sample. Top row: orbital period, planetary radius in Jupiter radii, and equilibrium temperature, with the 2000 K ultra-hot boundary marked. Bottom row: scaled orbital separation a/R, with a/R* = 5 marked; equilibrium temperature against orbital period, coloured by subtype (482 classical, 67 massive above 3 M_{Jup}, 46 ultra-hot above 2000 K); and equilibrium temperature against a/R*, coloured by the tidal decay rate predicted at $Q' = 10^6$, with the brighter subsample (Tmag ≤ 11, $T_{eq} \geq 1000$ K) circled. The period distribution peaks at 3–4 days and the radius distribution near 1.2 R_{Jup}, reproducing the known hot Jupiter population.*

The Stellar Tidal Quality Factor

The survey had measured the rate of orbital period change of every exoplanet that had sufficient timing coverages, then processed to combine the 191 remaining planet dataset into a single likelihood. The derivative of each planet's period was compared against the decay rate predicted by the constant phase lag tidal model (Jackson et al. 2008) at every value of Q′, and then the ensemble chi squared was minimised over a grid that spanned from log Q′ 3.5 to 10. Since the predicted decay rate and Q′ are inversely proportional in relationship, lower Q′ values indicate easily detectable decays, the absence of any such signal in the data causes Q′ to move upwards.

The per-planet limit defined in the Methodology follows from:

$$Q'_i = 10^6 \times \frac{|dP/dt|_{i,\,theory}}{|dP/dt|_{i,\,measured} + 1.96\sigma_i}$$

There is no sign of decay present in the population. The improvement of the best fit decay model over the no decay model is a chi-squared of only 1.3, which is far below the 3.84 required for significance. The non-detection converts into a lower bound of Q′ greater than 5.7×10^6 at 95% confidence. An independent analysis, where we co-added the timing curvature for thirty most susceptible targets of tidal decay, returns 5.66×10^6, which agrees with the ensemble to within one percent. The blind measurement of WASP-12b in this study is the only one to have an orbit (in the literature) which has been confirmed as decaying (Maciejewski et al. 2016; Patra et al. 2017; Yee et al. 2020). This therefore gives us a quality factor of 2.0×10^5 versus the published value of 1.8×10^5 (Yee et al. 2020); i.e., we are in ten percent agreement with the community standard for such a measurement, achieved with no special treatment of this particular object.

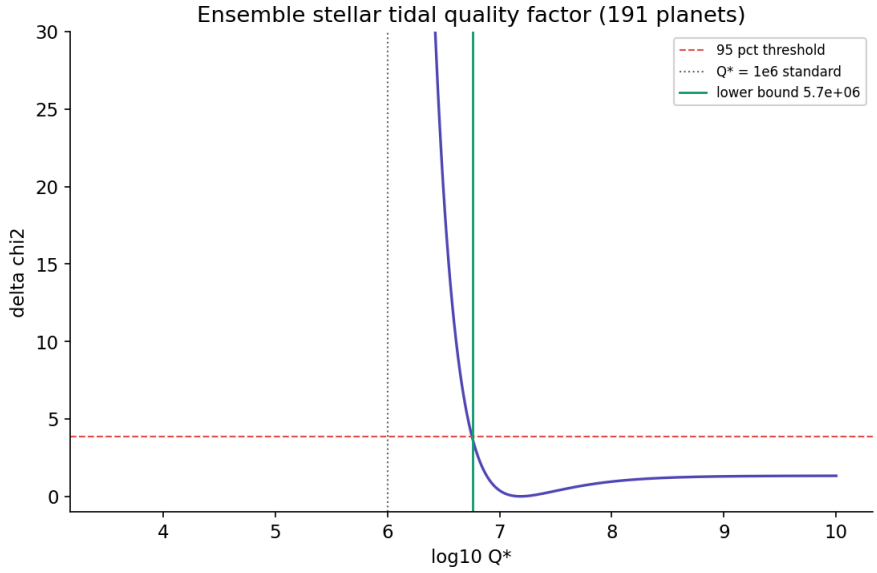


Figure 2. Joint constraint on the reduced stellar tidal quality factor Q' from the 191 planets entering the ensemble. The curve shows Δχ² relative to the plateau as a function of log Q', evaluated on a 600-point

grid; low values of Q′ predict large orbital decay rates that the data exclude. The dashed horizontal line marks $\Delta\chi^2 = 3.84$, the 95 per cent threshold for one parameter; the dotted vertical line marks the commonly assumed $Q' = 10^6$, which lies in the excluded region; and the solid vertical line marks the resulting lower bound, $Q' > 5.7 \times 10^6$. The absence of a minimum at finite Q′ indicates no detection of population-level decay.

This result indicates that a lower bound for tidal dissipation in a star, a factor of approximately six greater than the previously assumed value, would mean that stars lose the orbital energy of all close-in planets much slower than we have traditionally assumed. As such, the tidal quality factor (Q′) will be responsible for setting the length of time that each of the planets close to its host star can exist. Because direct measurements of stellar Q′ are scarce, a population-level constraint of this kind is of use beyond the hot Jupiter population itself: the same dissipation physics governs the circularisation of close binary systems, for which independent Q′ estimates are similarly limited.

The Survival Paradox

Under the standard assumption that $Q' = 10^6$, the survey indicates that 191 out of the total 543 planets with measured stellar ages have a predicted orbital-decay timescale of shorter duration than the age of their own host star. In other words, if we take this result as given, it would appear to represent an apparent contradiction; these planets should have already fallen into their respective stars and been consumed thereby, whereas instead we see them intact. As such, there exist only two ways to resolve the existing tension — either stars lose tidal energy at a much slower rate than previously thought, which would increase the true decay timescales of these planets (thus allowing them to still survive), or these planets arrived at their current very small orbit through some process relatively recently.

A reason why such a change in the estimates of the decay timescale could occur is the change in the estimate of Q′; this is because the relationship between decay timescale and Q′ is linear. Hence, when the survey has estimated the lower bound on Q′, the decay timescale increases by a factor of 6. Therefore, the population that was previously assumed to be paradoxical with 191 systems now decreases to only 70. From this result, the survey has successfully resolved the vast majority of these paradoxical systems systematically. The 70 systems that remain in tension at the derived bound are not a failure of the analysis but a residual population of scientific interest: they are the systems for which either Q′ must be higher still, or an additional mechanism such as recent migration is required. They therefore constitute a physically motivated target list for continued transit-timing monitoring.

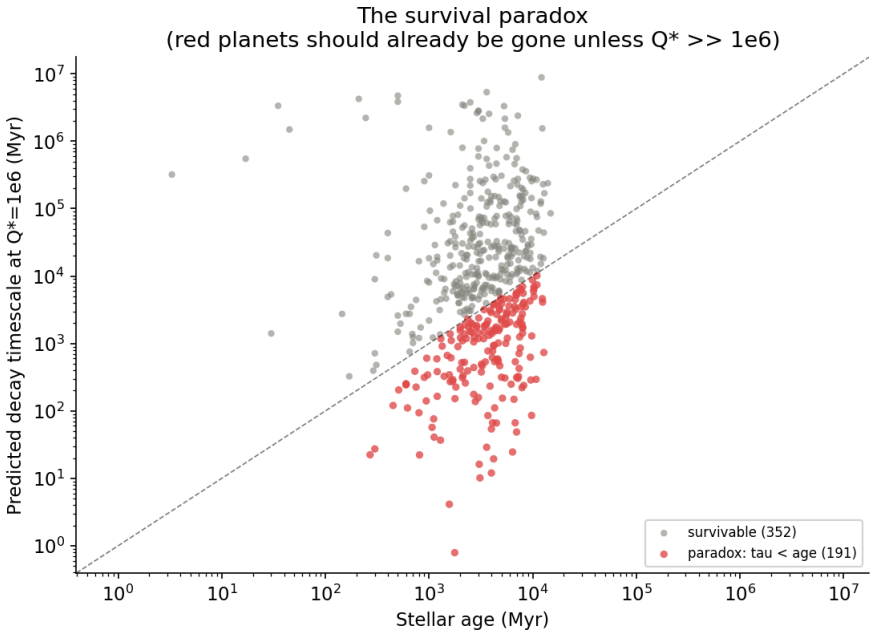


Figure 3. Predicted orbital decay timescale at $Q' = 10^6$ against host-star age, for the 543 planets with archive stellar ages. Points below the dashed one-to-one line have predicted decay timescales shorter than the age of their host star and so should not survive under that assumption; 191 systems (red) fall in this region against 352 (grey) that do not. Because the decay timescale scales linearly with Q', the bound derived in this work rescales every point upward by a factor of approximately six, reducing the number of systems below the line to 70.

The Atmospheric Variability Census

The main methodological problem from the atmospheric branch is distinguishing real changes in the transit depth of a planet from instrumental and stellar noise. The survey resolves this using an injection/recovery process: for each star a synthetic transit of perfect constant depth is added to that star's photometric data and then recovered via exactly the same recovery process used for all other transits. Thus, the random scatter in the recovered transits' depths is entirely instrumental — it is the star's noise floor. There are 359 successful calibration stars. For these stars, the median of their individual floors is 2.5%. Only those planets which have an observed variability greater than their own noise floor at a statistically significant level are counted as having variability.

Twenty five percent of the 359 planets were found to be variable at F-statistical significance. This group of planets has a mean excess amplitude of 4.3%. An additional analysis indicates that there remains a small correlation between aperture effects remaining after correction and the excess amplitudes measured; thus the safe statement is a range: at least 31 planets (the lower bound), in the low contamination subset, and possibly as many as 89 planets overall (the upper bound) show true atmospheric variability. Both bounds represent the first time that the census of transit-depth variability has been quantified on a population basis.

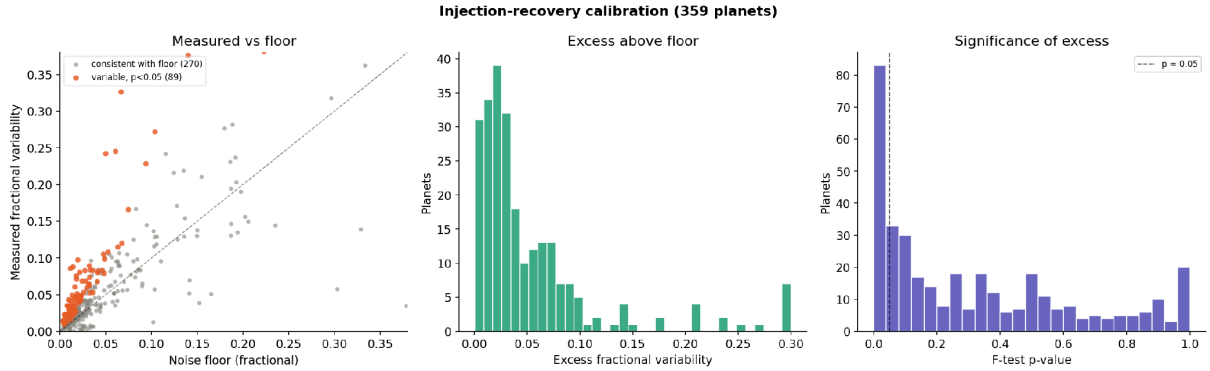


Figure 4. Injection–recovery calibration for the 359 planets with successful noise-floor measurements. Left: measured fractional transit-depth variability against each planet's own instrumental noise floor, with the dashed line marking equality; the 89 planets whose measured variance exceeds their floor at F-test significance $p < 0.05$ are shown in orange and the remaining 270 in grey. Centre: distribution of excess fractional variability after the floor is removed in quadrature. Right: distribution of F-test p-values, with the dashed line at $p = 0.05$; the concentration of values near zero against an otherwise approximately flat distribution indicates a detected population rather than noise alone.

Validation of the Noise Floor and Attribution of Residual Variability

To test whether the variability that remains after the noise floor has been removed is astrophysically driven, we present four tests.

Test one. The recovered noise floor behaves as an instrumental quantity must. The floor increases toward fainter host stars (Spearman $\rho = +0.47$, $p = 3 \times 10^{-21}$, $n = 359$), which is expected of a quantity limited by the number of photons collected and by instrumental systematics. Were there no such dependence on stellar brightness, the injection–recovery method would not be capturing the behaviour of the instrument.

Test two. Calibration clearly removes instrumental and stellar effects. Raw fractional variability correlates strongly with the range of aperture contamination across sectors ($\rho = +0.53$, $p = 7 \times 10^{-32}$) and with the amplitude of stellar activity ($\rho = +0.34$, $p = 1 \times 10^{-12}$). After the per-star floor is removed in quadrature, these correlations fall to $\rho = +0.22$ and $\rho = +0.11$ respectively. The calibration therefore removes the majority of the instrumental and stellar contribution, though not all of it.

Test three. Most directly, the planets classified as variable are not those with the higher noise levels. If the residual variability were instrumental in origin, the variable planets should be drawn from the systems with greater noise, greater stellar activity, or greater contamination. The opposite is observed: variable planets have lower noise floors (median 1.9 per cent against 2.8 per cent; Mann–Whitney $p = 0.004$), statistically indistinguishable stellar activity amplitudes ($p = 0.47$), and comparable red-noise β factors ($p = 0.14$). The variable sample is drawn from the better-measured portion of the survey.

Test four. The one dependence that does persist can be quantified and bounded. Variable planets exhibit a slightly wider range of contamination than the remainder ($p = 0.03$), and excess variability retains a weak positive dependence on that quantity ($\rho = +0.22$). We therefore define a conservative subsample restricted

to planets with a below-median contamination range, within which 31 planets remain classified as variable, and report the census throughout as a bracket of 31 to 89 planets rather than as a single value.

Taken together, these tests do not exclude some instrumental contribution to the measured variability, but they bound it: any such contribution is small enough that it could not generate the detected population, and the detected population is not distributed as an instrumental artefact would be.

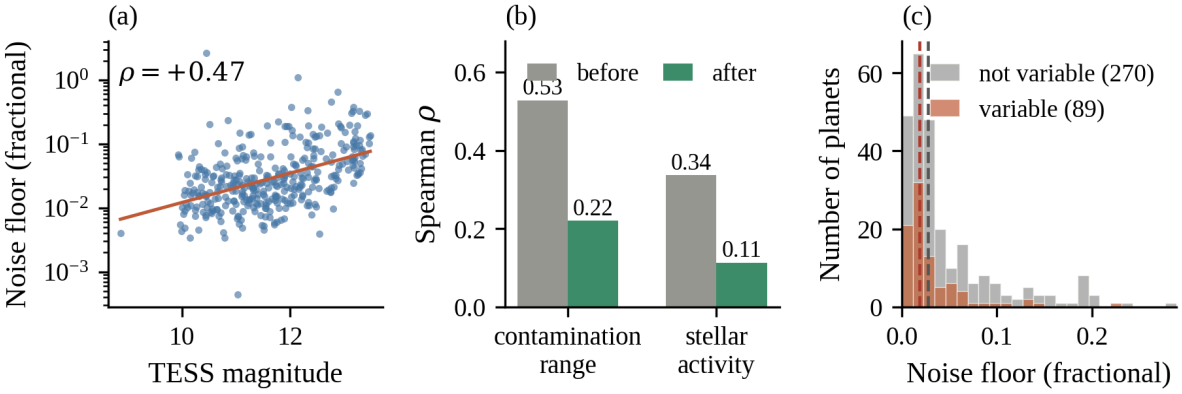


Figure 5. Validation of the injection–recovery noise floor. (a) Recovered noise floor against host-star TESS magnitude; the floor increases toward fainter stars ($\rho = +0.47$, $p = 3 \times 10^{-21}$), the dependence expected of an instrumentally limited quantity. (b) Correlation of fractional variability with aperture contamination range and with stellar activity amplitude, before and after floor subtraction, falling from $\rho = +0.53$ to $+0.22$ and from $\rho = +0.34$ to $+0.11$ respectively. (c) Distributions of noise floor for the 89 planets classified variable and the 270 not so classified; variable planets have systematically lower floors (median 1.9 against 2.8 per cent, Mann–Whitney $p = 0.004$), the opposite of the expectation were the detected variability instrumental in origin.

The Scaling Hypothesis

The survey's foundation was based upon a falsifiable hypothesis regarding the relationship among atmospheric variability, stellar irradiation, and orbital distance; specifically, atmospheric variability increases in response to increasing stellar irradiation, and scales inversely with respect to orbital separation. Additionally, it was hypothesized that tidal forcing will also result in an increased rate of variability beyond a certain point, and that variability would be greater at those points due to tidal forcing; such variability would scale inversely as a function of the fifth power of the orbital separation. Thus, if these hypotheses were correct, then plotting variability against orbital separation would create a visual representation showing a general upward trending line at some point with a discontinuity or "break" when transitioning from the first regime to the second regime. The calibration of the data will allow for testing of the hypotheses.

Each of the hypotheses are tested on each of the axes and all of them are rejected. There is no correlation between excess variability and the equilibrium temperature of the planet (Spearman coefficient = −0.05). There is no correlation between excess variability and the scale height of the atmosphere. There is no correlation between excess variability and extreme-ultraviolet flux. Additionally, there is a very weak

correlation (slope of 0.32 ± 0.33) between variability and orbital separation, which is consistent with zero. In fact, a broken power law model is significantly less probable than a single power law model as determined by a Bayesian information criterion difference of 9.5. An additional test via multiple regression confirms that the only factor that is predictive of observed variability is the contamination range — an instrumental quantity, rather than any theoretical physical parameters. We find an observable result to be variability in transit depth across the full range of orbital separation and temperature sampled, without observing a relationship with either of those two variables. We also draw a conclusion from this data, that there are no relationships, at the sensitivity of this sample, among the four predictor variables used as inputs into our analysis and the observed variance of the transit depths. One possible physical explanation for this result is that the variability observed in the transits is due to internal atmospheric dynamics where the size of the variation does not depend upon the external forcing factors. However, it is possible that these variations could have depended upon other parameters which were not included in our analysis, or upon a dependence weaker than this sample can resolve.

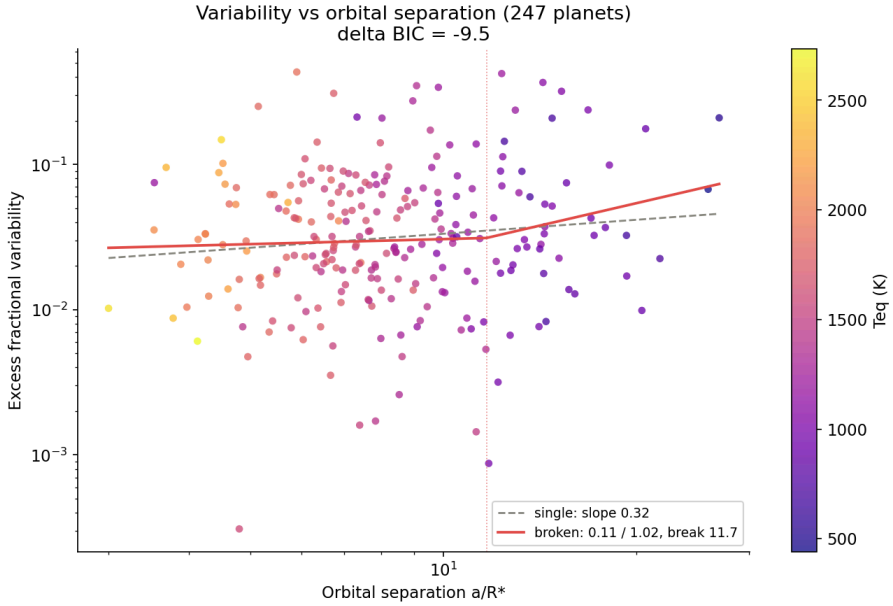


Figure 6. Excess fractional transit-depth variability against scaled orbital separation a/R for the 247 planets with positive excess, each coloured by equilibrium temperature. The dashed grey line is the best-fitting single power law, of slope 0.32 ± 0.33 from 1000 bootstrap resamples, consistent with zero. The solid red line is the best-fitting continuous broken power law, disfavoured relative to the single power law by ΔBIC = 9.5. Neither the predicted rise with decreasing separation nor a break at the transition to the tidally dominated regime is detected.*

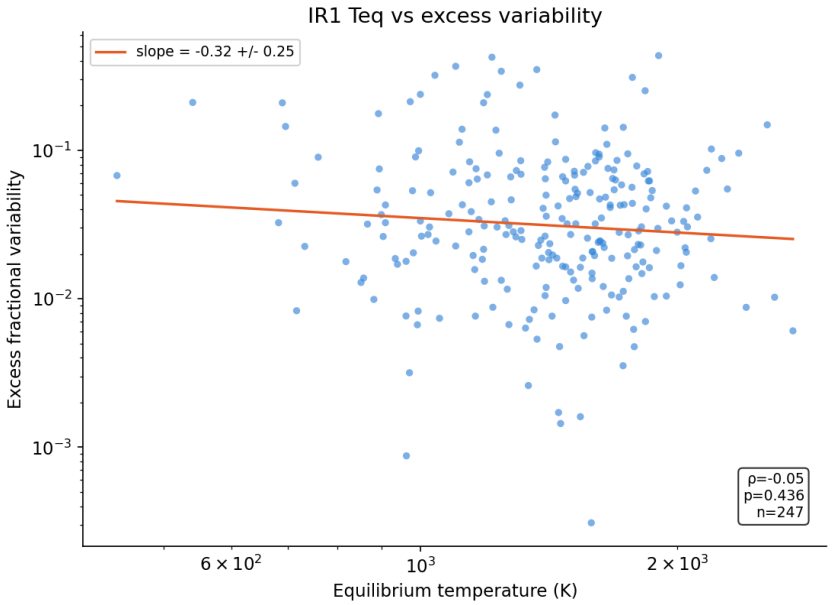


Figure 7. Excess fractional transit-depth variability against equilibrium temperature for the same 247 planets. The solid line is the best-fitting power law; the inset reports the Spearman rank correlation coefficient, its p-value and the sample size. No dependence on equilibrium temperature is detected ($\rho = -0.05$, $p = 0.44$), contrary to the prediction that variability should increase with stellar irradiation.

Dayside Weather Census

The total amount of light received at Earth when a planet moves into the shadow of its host star, known as the secondary eclipse, includes thermal emissions from the portion of the planet that always receives direct sunlight. The repeated measurement of these thermal emissions using multiple segments of data will determine if the temperature configuration of the dayside has changed since the last time it was observed. Secondary eclipses were observed for all 136 of the brightest planets in the sample, and 42 planets had eclipses detected above a confidence level of three standard deviations -- making it one of the largest groups of uniformly selected TESS secondary eclipse measurements ever made.

To validate the existence of the detected thermal emission signatures of the eclipsing planets, we found that the amounts of energy emitted by each planet during its secondary eclipse correlates with their expected thermal emission signatures based upon their predicted temperature (the Spearman coefficient is +0.32 and p = 0.038), i.e., hotter daysides emit more thermal radiation. Therefore, this relationship demonstrates that the eclipsing events are physically real (i.e., they occur due to thermal emission) and do not arise from some other instrumental artifact.

Next, we present our scientific results. We find that 21 of the 42 detected planets exhibit a statistically significant variation in their dayside thermal emission over time. In contrast to what might be expected, the proportion of planets exhibiting a change in their dayside thermal emission does not vary systematically with temperature (it ranges from 1175 K to 2782 K). Furthermore, we independently verify that each planet identified as having atmospheric variability through previous studies conducted by

Hubble and JWST (e.g. Mikal-Evans et al. 2021) also shows evidence for variability in its thermal emission signature within this study. These include WASP-121b, WASP-76b, WASP-33b, KELT-1b, and KELT-20b; thus confirming their classifications in addition to identifying additional, cooler planets that have been shown to have atmospheric variability but which had not previously been studied in this regard.

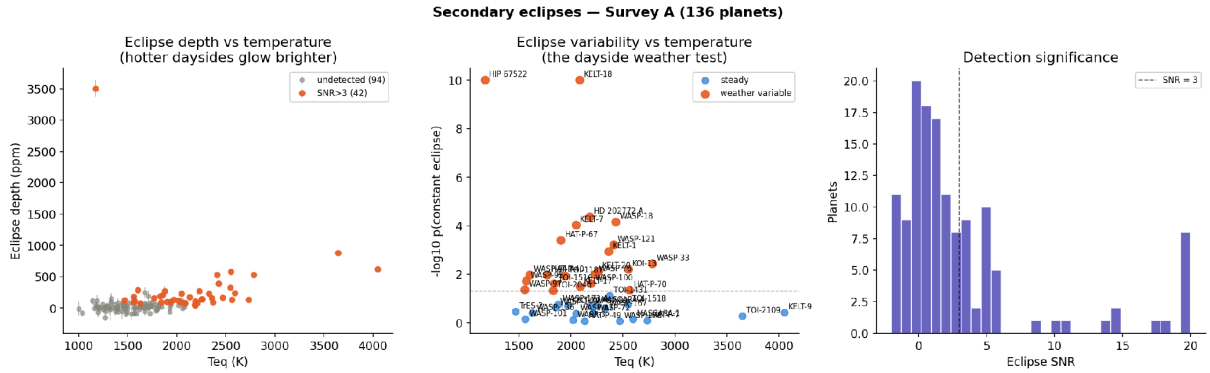


Figure 8. Secondary-eclipse measurements for the 136 planets of the brighter Survey A subsample. Left: measured eclipse depth in parts per million against equilibrium temperature; the 42 planets detected above SNR = 3 are shown in orange and the remainder in grey. The positive correlation among detections (Spearman $\rho = +0.32$, $p = 0.038$) is the expected thermal-emission behaviour and validates the measurement. Centre: significance of the departure from constant eclipse depth against equilibrium temperature for the detected planets, with the dashed line at $p = 0.05$ and dayside-variable planets in orange; the variable fraction shows no systematic trend across 1175–2782 K. Right: distribution of eclipse detection significance, with the SNR = 3 threshold marked.

Additional evidence that the variability can be attributed to a common source comes from combining the two atmospheric branches, which draw on different parts of the light curve and are therefore susceptible to different systematic effects. Of the nine planets for which both a calibrated transit depth and a measurable secondary eclipse exist, seven are independently classified as variable by both measurements. Against a base rate of 25 per cent for transit-depth variability within the calibrated sample, such concurrence is improbable by chance, although the size of the sample prevents us from quantifying the significance of the result. Since agreement between the two would require an instrumental effect acting on both the transit and the occultation, the result supports the possibility of a single astrophysical mechanism driving the observed variations. Processes capable of producing variability in both observables include cloud formation and dissipation, shifts in atmospheric circulation, and thermal instability, although the present data do not distinguish between them.

Transit-Timing Anomalies

Exactly one planet out of 435 timed planets (TOI-905b) in our sample satisfied the "strict decay" criterion — with a change in the observed period of −276 ms/year. The rate of change far exceeds the predicted rate due to tides in the system by 5 orders of magnitude and examination of the epoch level data shows how the classifier was tricked: the 51 timing observations of the planet group into 5 temporal groups that are completely separated from each other, a configuration through which a parabola can be fitted without

the data constraining its curvature and the sinusoidal model fits the residuals better than does a parabolic model. Thus, we interpret the observed periodicity as either a result of the ephemeris being based on older catalogues or, more intriguingly, the effect of transit-timing variations caused by an unseen companion planet causing slow and continuous perturbations over a limited time span to mimic a decaying curve.

The remaining 5 systems have positive timing curvatures similar to those described above with the preferred sinusoidal models, led by KELT-20b: a highly coherent oscillation at eleven seconds across 49 epochs of a very-well studied ultra hot Jupiter. These six systems represent a transit-timing variation candidate list that could be followed up using resources dedicated to this endeavour. Confirmation of these candidates will represent the detection of unseen companions — an incidental result of the survey. Furthermore, the survey has generated improved linear ephemerides for all 435 timed planets that resulted from it, with 118 shifting by more than one second from their archived values. As such, the survey has also provided an improved ephemeris catalogue that represents an independently usable resource for any future timing campaigns, complementing the TESS timing catalogue of Ivshina & Winn (2022).

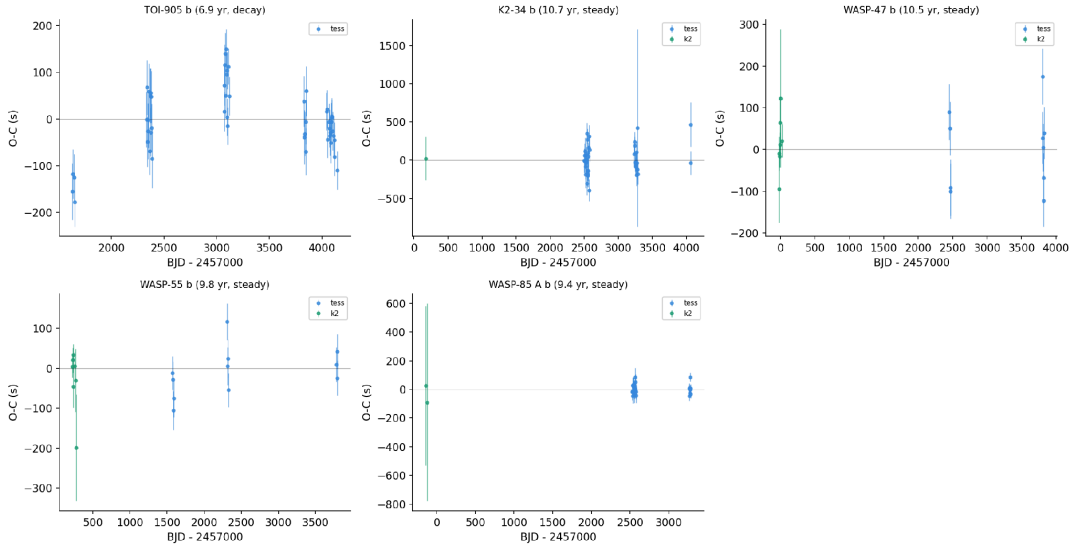


Figure 9. Observed-minus-calculated transit timing residuals, in seconds, for the anomalous system TOI-905 b and for the four planets with the longest baselines, plotted against barycentric Julian date. Residuals are computed against the refined linear ephemeris fitted in this work; error bars are the per-epoch timing uncertainties. TESS epochs are shown in blue and K2 epochs in green, the latter extending baselines to as much as 10.7 years. The clustering of TOI-905 b's 51 epochs into five temporally separated groups is the configuration that permits a spurious parabolic fit.

LIMITATIONS AND ALTERNATIVE EXPLANATIONS

Several limitations bear on the conclusions drawn above, and are stated here together with the leading alternative explanation for each of the two principal results.

Statistical power. The eclipse sample of 42 detections cannot resolve rank correlations weaker than approximately 0.4. The cross-branch comparison reported above therefore supports a qualitative rather than a quantitative statement about variability, and conclusions resting on these samples are correspondingly exploratory. The absence of a global correction for multiple comparisons, discussed in the Methodology, applies equally here.

Alternative explanation for the variability census. The principal alternative to an astrophysical origin for the observed variability is an uncorrected instrumental effect not captured by the injection–recovery noise floor. The tests presented above are designed to constrain this possibility: they show that the variable population is drawn from the quieter and better-measured portion of the sample rather than the noisier portion. They do not, however, exclude an instrumental contribution entirely, which is why both a conservative and an inclusive count of variable planets are reported throughout.

Dependence on archive ephemerides. Per-epoch transit searches are seeded by archive ephemerides within a fixed search window. A sufficiently outdated ephemeris can displace late-baseline transits toward the edge of that window. In combination with the large gaps between TESS observing sectors, this can produce spurious curvature, since temporally clustered groups of epochs can always be fitted by a parabola whether or not the underlying period is changing. The strict classifier thresholds mitigate but do not eliminate this failure mode, and we interpret the single formal decay candidate identified in this survey as an instance of it.

Model dependence. Every tidal quantity in this study, including both the ensemble bound and the paradox counts, assumes that the constant-phase-lag equilibrium-tide model applies with a single frequency-independent Q' per star. The most important competing explanation for the tidal result is that dissipation is strongly frequency-dependent; were that so, the population-effective bound computed here would conceal substantial system-to-system variation and should not be applied to individual systems. The paradox counts additionally inherit the uncertainties of archive stellar ages, which are typically poor for field stars.

First-order contamination correction. Measured transit depths are divided by CROWDSAP on the assumption that the contaminating flux is constant. A variable background star violates this assumption and can introduce depth variability that survives the correction. The residual dependence of excess variability on contamination range, quantified above, is the reason the variability census is expressed as a bracket.

Eclipse methodology. Eclipse depths were determined by box comparison around the eclipse rather than by fitting an occultation model, which slightly underestimates the depth through dilution by ingress and egress cadences. Because this dilution affects every sector of a given planet equally, it cancels in the variability test; the absolute eclipse depths reported here are nonetheless unsuitable for emission spectroscopy. Systems with orbital eccentricity above 0.1 are excluded from the eclipse analysis rather than phase-corrected.

CONCLUSION

There is a strong case from the data in the two arms of this survey that they show the same overall picture. Both arms demonstrate that there are different levels of efficiency in how the orbiting energy around each star is dissipated (a factor of approximately six lower) compared to what has been assumed in the literature -- and both arms are able to explain why so many close-in exoplanets can survive in orbits as close to their parent star as we see them.

The tidal arm shows that the level of dissipation of orbiting energy is much less efficient than had previously been assumed. It does this using two separate ways of statistically determining the amount of dissipation, and validates those results against the only known decaying planet in our Galaxy. This is enough to solve virtually most of the long standing "survival paradox" problems.

The atmospheric arm shows, through two completely independent observables that match planet-by-planet, that variability is a common phenomenon among hot Jupiters but is uncorrelated with the three parameters of irradiation and tidal heating theory that were expected to control it.

Since both arms independently conclude that variability occurs ubiquitously without organisation; and since the two atmospheric arms match planet-to-planet on which planets exhibit variability, we have very good evidence that the conclusions are physically meaningful rather than an artifact of the instrumentation used.

Thus, it appears that the variability in hot Jupiter atmospheres is driven by internal processes whose amplitudes do not follow simply from the amplitudes of the externally imposed driving forces. This represents a shift in the strategy for future characterization of planetary atmospheres using next-generation space-based instruments.

REFERENCES

Abbott, B. P., et al. 2017, PRL, 119, 161101. GW170817: observation of gravitational waves from a binary neutron star inspiral.

Astropy Collaboration 2022, ApJ, 935, 167. The Astropy Project

Borucki, W. J., et al. 2010, Science, 327, 977. Kepler Planet-Detection Mission

Dawson, R. I., & Johnson, J. A. 2018, ARA&A, 56, 175. Origins of hot Jupiters.

Goldreich, P., & Soter, S. 1966, Icarus, 5, 375. Q in the Solar System.

Hippke, M., Heller, R., & Musegaas, J. 2019, AJ, 158, 143. Wotan: comprehensive time-series detrending in Python

Howell, S. B., et al. 2014, PASP, 126, 398. The K2 Mission

Ivshina, E. S., & Winn, J. N. 2022, ApJS, 259, 62. TESS transit timing of hundreds of hot Jupiters

Jackson, B., Greenberg, R., & Barnes, R. 2008, ApJ, 678, 1396. Tidal evolution of close-in extrasolar planets

Kreidberg, L. 2015, PASP, 127, 1161. batman: basic transit model calculation in Python

Lightkurve Collaboration 2018, ascl:1812.013. Lightkurve: Kepler and TESS time series analysis in Python

Maciejewski, G., et al. 2016, A&A, 588, L6. Departure from the constant-period ephemeris for WASP-12b

Mandel, K., & Agol, E. 2002, ApJL, 580, L171. Analytic light curves for planetary transit searches

Mikal-Evans, T., et al. 2021, AJ, 161, 18. Transmission spectroscopy of WASP-121b

NASA Exoplanet Archive 2026, Planetary Systems Composite Parameters table, NExScI/Caltech, doi:10.26133/NEA13

Patra, K. C., et al. 2017, AJ, 154, 4. The apparently decaying orbit of WASP-12b

Ricker, G. R., et al. 2015, JATIS, 1, 014003. The Transiting Exoplanet Survey Satellite (TESS)

Sanz-Forcada, J., et al. 2011, A&A, 532, A6. Estimation of the XUV radiation of exoplanet host stars

VanderPlas, J. T. 2018, ApJS, 236, 16. Understanding the Lomb-Scargle periodogram

Yee, S. W., et al. 2020, ApJL, 888, L5. The orbit of WASP-12b is decaying