Earths within Reach: Evaluation of Strategies for Mitigating Solar Variability using 3.5 years of NEID Sun-as-a-Star Observations

Eric B. Ford [1,2,3,4] Chad F. Bender [5] Cullen H. Blake [6] Arvind F. Gupta [7] Shubham Kanodia [8] Andrea S.J. Lin [1,2] Sarah E. Logsdon [7] Jacob K. Luhn [9] Suvrath Mahadevan [1,2,*] Michael L. Palumbo III [1,2] Ryan C. Terrien [10] Jason T. Wright [1,2,11] Jinglin Zhao [12] Samuel Halverson [13] Emily Hunting [7] Paul Robertson [9,†] Arpita Roy [14] and Guðmundur Stefánsson [15]

[1]*Department of Astronomy and Astrophysics, Penn State University, 525 Davey Laboratory, 251 Pollock Road, University Park, PA, 16802, USA*

[2]*Center for Exoplanets and Habitable Worlds, Penn State University, 525 Davey Laboratory, 251 Pollock Road, University Park, PA, 16802, USA*

[3]*Institute for Computational and Data Sciences, Penn State, University Park, PA, 16802, USA*

[4]*Center for Astrostatistics, Penn State University, 525 Davey Laboratory, 251 Pollock Road, University Park, PA, 16802, USA*

[5]*Steward Observatory, University of Arizona, 933 N. Cherry Ave, Tucson, AZ 85721, USA*

[6]*Department of Physics and Astronomy, University of Pennsylvania, 209 S 33rd St, Philadelphia, PA 19104, USA*

[7]*U.S. National Science Foundation National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85719, USA*

[8]*Earth and Planets Laboratory, Carnegie Science, 5241 Broad Branch Road, NW, Washington, DC 20015, USA*

[9]*Department of Physics & Astronomy, The University of California, Irvine, Irvine, CA 92697, USA*

[10]*Carleton College, One North College St., , MN 55057, USA*

[11]*Penn State Extraterrestrial Intelligence Center, Penn State University, 525 Davey Laboratory, 251 Pollock Road, University Park, PA, 16802, USA*

[12]*DTU Space, National Space Institute, Technical University of Denmark, Elektrovej 328, DK-2800 Kgs. Lyngby, Denmark*

[13]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109*

[14]*Astrophysics & Space Institute, Schmidt Sciences, New York, NY 10011, USA*

[15]*Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

ABSTRACT

We present the results of Sun-as-a-star observations by the NEID Solar Telescope at WIYN Observatory, spanning January 1, 2021 through June 30, 2024. We identify 117,060 observations which are unlikely to be significantly affected by weather, hardware or major calibration issues. We describe several high-level data products being made available to the community to aid in the interpretation and inter comparisons of NEID solar observations. Solar observations demonstrate excellent performance of NEID, including radial velocity (RV) accuracy and long-term stability of better than $\simeq 0.37\,\mathrm{m\,s^{-1}}$ over $\simeq 3.5$ years, even though NEID was not originally designed or optimized for daytime observations of the Sun. Currently, intrinsic stellar variability is the primary barrier to detecting Earth-analog planets for most nearby, Sun-like stars. We present a comparison of the effectiveness of several methods proposed to mitigate the effects of solar variability on the Sun's estimated radial RV. We find that the Scalpels algorithm performs particularly well and substantially reduces the root mean square (RMS) solar RV from over $2\,\mathrm{m\,s^{-1}}$ to $0.277\ \mathrm{m\,s^{-1}}$. Even when training on a subset of days with NEID solar observations and testing on a held-out sample, the RMS of cleaned RV is 0.340–$0.422\ \mathrm{m\,s^{-1}}$. This is significantly better than previous attempts at removing solar variability and suggests that the current generation of EPRV instruments are technically capable of detecting Earth-mass planets orbiting a solar twin if provided with sufficient observing time allocations ($\sim 10^3$ nights of observations).

1. INTRODUCTION

1.1. *NEID Solar Observations*

NEID is a highly-stabilized, high-resolution spectrograph designed to conduct extremely precise radial velocity (EPRV) measurements during night-time observations from the WIYN 3.5m observatory[1] (Schwab et al. 2016). An auxiliary NEID Solar Telescope allows for routine daytime observations of the Sun (Lin et al. 2022). While the spectrograph and detector are shared between night-time and solar observations, there are some important differences in the light path and instrument stability between daytime solar and night-time stellar observations (e.g., longer fiber from the solar telescope, finite angular size of the sun, different optical path and calibration, no atmospheric dispersion corrector, pointing errors are driven by the solar tracker and not the WIYN 3.5m telescope) that we will discuss in §4.

Most aspects of NEID, the official data reduction pipeline (DRP)[2], and the detailed error budget (Halverson et al. 2016) apply to both night-time and solar observations[3]. The typical formal measurement uncertainty reported by the NEID DRP for each NEID solar spectrum is 0.285 m s^{-1}, not significantly different than for a bright nighttime star.

However, unlike nighttime star observations, it is practical to obtain hundreds of solar observations every clear day, providing the opportunity to bin observations to obtain very high signal-to-noise, $\sim 2,000$ which may be useful for mitigating the impact of stellar line shape changes (Davis et al. 2017; Palumbo et al. 2024). If there were no contamination from stellar variability or other sources of systematic "noise" (e.g., telluric variability, instrumental variability), then the information content of one hour of NEID solar observations could achieve a precision of ~ 0.05m s^{-1}and daily averaged NEID solar RVs could achieve a precision of ~ 0.02m s^{-1}.

In practice, the root mean square deviation (RMS) of daily-averaged RVs observed by NEID (and other solar telescopes) is multiple m s^{-1}. The specific value depends on the dates where observations were taken, as well as the the quality of the instrument calibration and data reduction methods. The large number of solar observations (and favorable observing cadence) could prove valuable for evaluating strategies for mitigating the impact of stellar variability on EPRV surveys (see §1) and for empirically constraining some terms of the instrumental error budget and evaluating the effectiveness of various methods proposed for mitigating them.

Several planet-hunting spectrographs are collecting solar observations (Dumusque et al. 2015; Lin et al. 2022; Llama et al. 2024; Luque 2022; Rubenzahl et al. 2023; Cabona et al. 2024). NEID solar observations offer a unique combination of features, including high spectral resolution (R$\sim$120,000), an ultra-stabilized spectrograph enclosure (Robertson et al. 2019; Stefansson et al. 2016), an unusually large wavelength grasp ($\leq$380–$\geq$930nm), rapid observing cadence (55s exposures with gaps for readout lasting 28s (38s) starting (before) September 17, 2021), a dataset that includes both low and high levels of solar magnetic activity, and data that become public promptly after processing. Thus, we envision the NEID solar observations as providing a valuable community resource for quantifying the impact of intrinsic stellar variability on the sensitivity of future EPRV surveys to detect and/or characterize low-mass planets. This paper presents NEID solar observations collected during January 2021—June 2024 and provides high-level data products that will be useful for future analyses.

1.2. *Evaluating Stellar Variability Mitigation Strategies*

Thanks to the dramatic improvements in spectrographs and calibration systems, stellar variability is often the dominant source of "noise" limiting the effective detection sensitivity of EPRV surveys (Crass et al. 2021). Most methods for measuring "the radial velocity of a star" assume that the true stellar spectrum remains constant in the star's center-of-mass frame (Hara & Ford 2023). From this perspective, line-shape changes in the stellar spectrum can be seen as an additional source of astrophysical "noise" that contaminates measured RVs. Several previous studies have

---

attempted to clean the observed RVs of this additional noise, based on other changes in individual spectra (e.g., classical spectroscopic activity indicators, data-driven variability indicators) and/or the temporal structure in measured RVs. While several approaches appear to help, the community has not yet identified a single strategy that can be counted on to enable the detection of an Earth mass-planet orbiting in the Habitable Zone of a Sun-like star.

One challenge in evaluating stellar variability mitigation strategies is the results from each method inevitably depend on many additional details of the data and analysis, such as which instrument was used, the pipeline (and version) used for extracting spectra from raw observations, the dates of the observations (due to constantly changing extent and distribution of magnetically active regions), how the researchers decide which observations should be included/excluded in the analysis, and the choice of which stellar features to include in the analysis. Therefore, it is often not straightforward to compare the effectiveness of two proposed strategies based on the results published in two separate studies.

To address this and related challenges, the EPRV community has organized multiple data challenges to help researchers compare different methods to search for planetary signals (Dumusque 2016; Dumusque et al. 2017), to estimate the evidence for planetary detections (Nelson et al. 2020), and to reduce the impact of stellar variability on RV measurements (Zhao et al. 2020, 2022b, 2023). Dumusque (2016) provided data challenge participants with just a few potential indicators, while Zhao et al. (2020) provided extracted spectra and let participants apply a wider variety of strategies to mitigate stellar variability. For the latter study, eleven research teams submitted their attempts to clean the measured RVs of stellar variability using 22 methods. While many of the methods succeeded in reducing the RMS of cleaned RVs, the different methods disagreed on the magnitude (and sometimes the sign) of the correction that was needed at each time (Zhao et al. 2022b). The differences were significant and demonstrated that further research is needed to determine which methods (if any) are powerful, robust, and reliable for detecting low-mass planets. While the EXPRES Stellar Signals Project (ESSP) Data Challenge was valuable, the interpretation of the results was complicated by the use of real observations of stars for which the true number, masses, and orbits of planets are unknown (Zhao et al. 2022b). Not having access to the ground truth limited how precisely the data challenge could evaluate and compare the various mitigation strategies.

Solar observations provide a unique opportunity to evaluate strategies for mitigating RV "noise" from stellar variability, since the relative motion between the Sun and observatory is known precisely. In this study, we apply multiple strategies for mitigating stellar variability and compare their effectiveness using a common NEID solar dataset.

This paper presents results of analyzing solar observations and describes various data release products to ease the analysis of NEID solar observations for other researchers. It also presents a comparison of several wavelength-domain methods for characterizing the instantaneous level of stellar variability and its impact on RV measurements. The manuscript is organized as follows. In §2, we describe a series filters applied to select high-quality observations. In §3, we describe our methods for analyzing extracted spectra to calculate cross correlation functions (CCFs), radial velocities (RVs), and CCF bisectors. In §4, we describe the high-level data products that we provide and present results from their first analysis. In §5, we evaluate and contrast the ability of several wavelength-domain strategies for mitigating solar variability. In §6, we discussion the implications of our findings and opportunities for future improvements.

2. FILTERING OF OBSERVATIONS

Before performing a scientific analysis of NEID solar observations, one needs to filter out observations which were corrupted by weather, known hardware issues, and significant wavelength calibration issues. We describe a series of filters that are useful in constructing a clean dataset for subsequent analysis. We start with essential filters and work towards less common, smaller, and/or less well understood effects.

2.1. *Essential filters*

Here, we describe several filters that we consider essential and recommend all subsequent analyses of NEID solar observations perform.

2.1.1. *Wavelength calibration applied*

Since EPRV measurements are differential, it is important to only compare RVs based on data reduced using the same major and minor versions of the NEID DRP. This paper presents results based on NEID DRP v1.3. Since solar observations were paused at the end of June 2024, no solar observations were affected by issues necessitating the bug fixes included in DRP v1.3.1.
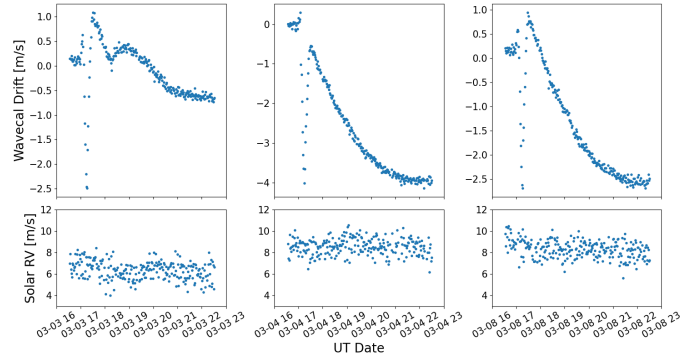
Figure 1. Top: Measured instrumental drift over the course of solar observations for each of three days. Bottom: Measured solar RVs after accounting for measured instrumental drift. The NEID DRP does an excellent job of removing the short-term transients in wavelength calibration following the daily refill of the liquid nitrogen dewar. The RMS RV from a smoothed version of the instrument drift (after the initial large transient) is 0.052–0.055 m s^{-1}. However, an observed trend in solar RVs over the course of each day is described in §4.1. The solar RVs are potentially subject to additional systematic effects not probed by the etalon in the calibration fiber. These effects could include pointing, coupling-related behaviors with the solar feed or sci-cal fiber drift, and differential extinction.

Since we aim to focus our attention on high-quality observations, we filter for observations with a wavelength calibration based on a combination of the laser frequency comb (LFC) in the red and Thorium-Argon lamp in the blue and that use the `dailymodel0` drift function. These are indicated by the FITS header keywords WAVECAL and DRIFTFUN of the FITS files provided by the NEID DRP (level 0, 1 and 2). `dailymodel0` is the default drift model in DRP v1.1-v1.3 which compares Fabry-Perot (FP) etalon measurements which are taken simultaneously with the solar exposures to the daily wavelength solution derived from the standard NEID dawn/dusk calibration sequences. [4] Despite efforts to stabilize NEID's environment, the instrument does experience a significant RV drift during the daylight hours due to the daily filling of the liquid nitrogen dewar (see Figure 1). Most of the drift is removed as part of NEID's DRP using the `dailymodel0` drift function that combines information from calibration observation taken at the beginning (end) and end (beginning) of each night (day) of observing. The NEID wavelength calibration procedure and wavelength calibration model was designed to maximize accuracy of night-time observations on stars. Unfortunately, that means that there are large, non-linear changes in the wavelength calibration during solar observations. In §4.1 we show that there is room for further improvements in the daily drift model for solar observations.

2.1.2. *Hardware issues*

On three occasions, and likely due to inclement weather, the NEID solar telescope cable bundle became dislodged from its optimal wrap path. During these windows, NEID solar observations were adversely affected as a result of tension on the cable bundle causing very poor tracking. The NEID fiber scrambling system minimizes the effects of pointing errors (Kanodia et al. 2023) and an oversized input port in the integrating sphere further suppresses the impact of pointing errors on NEID solar observations (Lin et al. 2022). However, the pointing errors due to the cable bundle issue were so extreme that they still resulted in a substantial RV signal, likely due to a truncation of the stellar disk. Therefore, we exclude data taken during the intervals October 2-27, 2021, February 5-27, 2023, and March 18-31, 2023. Initially, these were attributed to bad weather events, but after multiple occurrences, the root cause of the problem was found and the cable bundle was more securely attached to the roof in April 2023. We do not anticipate additional data gaps due to this issue.

2.1.3. *Contreras fire*

The Contreras Fire led to a pause in NEID solar operations, with the last day of pre-fire solar observations being June 13, 2022. The spectrograph was warmed up in order to mitigate severe risks to the hardware stemming from uncontrolled thermal shocks. Once the fire was put out and facilities recovered, NEID resumed both solar and nighttime operations in low precision mode on November 12, 2022, while the instrument was still cooling down and settling to

---

[4] See the NEID Data Reduction Pipeline documentation at https://neid.ipac.caltech.edu/docs/NEID-DRP/algorithms.html# wavelength-calibration-sources for details.

its new state. Some high-precision night time observations resumed on November 23, 2022, and the NEID Guaranteed Time Observations program resumed December 1, 2022. While one could likely recover additional observations during the cool down period with custom attention, we will exclude data taken after the fire and prior to December 1, 2023. After applying these filters (plus excluding solar eclipse days §2.2.5, we are left with 192,373 "available" solar observations, $\simeq 84\%$ of all NEID solar data collected.

We define data taken during January 1, 2021 – June 13, 2022 as "run 1", and during December 1, 2022 – June 30, 2024 as "run 2". Starting July 1, 2024 solar observations were paused while the etalon simultaneous calibration source was unavailable for monitoring instrumental drift.

2.1.4. *Solar telescope on target*

Solar observations are scripted and automated to run from 16:30 to 22:30 UT. We found that the last solar observations of a day were sometimes impacted (e.g., solar telescope starts slewing to stow position before final exposure completes, potentially a partial obscuration or specular reflection). Therefore, we restrict our analysis to observations which start between 16:30:00 and 22:12:00 UT each day. It is likely that about a dozen additional clean observations could be recovered on many clear days, but additional effort would be required. This filter resulted in losing 5.5% of available observations.

2.1.5. *Weather*

Next, we apply a series of filters designed to exclude observations potentially contaminated by poor observing conditions. The NEID solar telescope takes observations even when the Sun is partially obscured by clouds. Since the equatorial limbs of the sun are rotating at nearly 2 km s^{-1}, even small variations in transmittance as a function of position on the disk can cause a substantial spurious RV. NEID has a dedicated exposure meter that picks off light coming through the solar fiber and reads out once a second. The NEID solar tracker also has an independent pyrheliometer that records the instantaneous total solar irradiance during observations in the level 0 files from the NEID DRP. The pyrheliometer has a substantially larger field of view ($\sim$5 degrees) than the telescope assembly and is more sensitive to infrared radiation that is susceptible to absorption by water vapor.

We require a minimum flux (averaged over the exposure) from both the pyrheliometer (pyrflux_mean $\geq 10^{2.95}$ W/m^2) and exposure meter (expmeter_mean $\geq 10^5$ ADU). Fluctuations in either the exposure meter or pyrheliometer reading during an exposure suggest transparency variations in the vicinity of the Sun could affect observations. Therefore, we compare the mean (expmeter_mean) and root mean square (RMS) deviation from the mean of the exposure meter during each exposure (expmeter_rms) and require that expmeter_rms $\leq 0.003\times$expmeter_mean. Similarly, we compute the mean and RMS deviation from the mean duration each exposure (pyrflux_rms) and require pyrflux_rms $\leq 0.0035\times$ pyrflux_mean. While the precise values could be tweaked, we have found these thresholds to be useful for a preliminary NEID data release based on January 2021–June 2022[5] and found no reason to change them. The weather filters resulted in losing an additional 33% of available observations.

2.2. *Additional filters*

Here, we describe several filters that we found were valuable for the current analysis. However, we recognize that other researchers may want to adapt different threshold values depending on their specific science goals.

2.2.1. *Pointing/tracking/fiber coupling*

The minimum exposure meter flux provides some check on the pointing. However, we found a more powerful filter to be based on the ratio of the flux from the exposure meter and the pyrheliometer. We note that this is the only filter which we have updated since the original recommendations developed based on run 1. Previously, we filtered for expmeter_mean / ADU $\geq 150\times$ pyrflux_mean /(W/m^2) in order to minimize risk of any undetected pointing errors. However, the system throughput decreased between run 1 and run 2 by nearly 14% (as measured by the exposure meter). Therefore, we have replaced this filter with one that requires that the ratio expmeter_mean/pyrflux_mean exceed 95% of the median value for that ratio within each day. This filter resulted in losing less than 0.2% of available observations.

---

[5] 10.5281/zenodo.7857521

2.2.2. *Airmass*

Differential atmospheric extinction can induce a spurious RV signal. For this analyses, we restrict our attention to observations with an airmass less than 2.5. This filter resulted in losing less than 1% of available observations. For some science cases, it may be useful to limit observations to those taken at lower airmass, so as to limit the impact of by differential extinction.

2.2.3. *Binning & Minimum number of observations*

In order to reduce the RV impact of solar oscillations, we will focus on analyzing binned observations. For the analyses presented below, we will work primarily with binned observations, requiring five consecutive exposures (within the same day) to pass the above criteria. This corresponds to $5 \times$ 55s integration time and a wall clock time of 7.12 minutes (for early observations) or 6.45 minutes (starting September 16, 2021). Working from binned observations has the additional benefit of reducing the risk of weather impacting RVs, since points passing thresholds by chance are unlikely to be consecutive. Our analysis of interday RV and CCF variations will focus on days with at least 15 complete binned observations to further reduce the impact of oscillations, reduce the chance of observations during poor atmospheric conditions, and ensure a high signal-to-noise ratio.

Depending on one's science goals, other analyses might choose to bin data differently (e.g., characterizing pulsations), so we provide a table with each unbinned observation as well as a table binning 10 consecutive exposures. Since we have not inspected unbinned observations for data quality issues, many users will likely prefer to work from the binned observations. Similarly, we provide data for all days with at least one binned observations, but have not performed a detailed inspection of data quality for days with less than 15 binned observations.

2.2.4. *Wavelength calibration issues*

In §3 we will describe how we calculate RVs from each order of the spectrum (with sufficient lines, signal-to-noise, and accurate wavelength calibration). By comparing the order RVs within each observation, we can recognize days where there is an unusually large dispersion of derived RVs across orders, most likely due to calibration issues (e.g., LFC flux, pipeline configuration). Most of these outliers can be tracked to a small number of orders near the transition between orders with and without LFC calibration. This has been traced to sporadic episodes of low flux from the LFC in the bluest orders where the NEID DRP uses the LFC for wavelength calibration. While this primarily affects echelle orders 117 & 118, thanks to the large number of solar observations, we can also recognize it sometimes affecting echelle orders 113-118. While we provide tables with the RVs extracted from these orders, we exclude these orders for computing CCFs and RVs for subsequent analysis in this paper. With adequate investment, a future version of the NEID DRP could likely recover additional RV information from these orders.

2.2.5. *Solar eclipses*

NEID observed partial solar eclipses on October 14, 2023 and April 8, 2024. These provide a unique opportunity for obtaining spatially resolved information with the NEID spectrograph and will be the subject of a future publication (Gonzales et al., in prep). We exclude these days from plots and summary statistics.

2.2.6. *Outcome of filtering*

We computed CCFs and RVs for 129,889 NEID solar observations. After applying the above filters, we are left with 117062 NEID solar observations and 20438 binned observations that include all five observations. There are 698 days with at least one complete binned observation and 544 days with at least 15 binned RVs to be used for subsequent analysis.

3. ANALYSIS METHODS

3.1. *Penn State Research Pipeline*

In order to help evaluate various strategies for analyzing EPRV observations, we have developed the Penn State Research Pipeline (PSRP). The PSRP complements and does not replace existing instrument-specific data reduction pipelines, as it starts from order-extracted spectra. This way, PSRP makes use of instrument-specific knowledge that instrument teams have incorporated into their data reduction pipelines. PSRP adds value by providing multiple alternatives for computing CCFs, estimating RVs, and applying data-driven methods for mitigating stellar variability. For the official NEID DRP, reliability and stability of the pipeline are essential. In contrast, PSRP prioritizes flexibility,

computational speed, and a modular and composable design, so as to allow researchers to rapidly try out multiple algorithms and parameter values.

The PSRP is based on a combination of Julia packages that are part of the `RvSpectML` organization. `RvSpectML` provides many options for processing of spectral time series. Here we provide an overview of the methods used for the analysis presented in this paper.

3.1.1. Ingesting FITS files

`EchelleInstruments.jl` provides a common interface for reading individual order extracted spectra and providing information specific to each instrument. We perform a blaze normalization based on the blaze function estimated for each observation by the NEID DRP. Next, we perform continuum normalization that closely follows the Rassine algorithm (Cretignier et al. 2020). The anchor points are computed from the summed spectrum over one clear day and then reused for continuum normalizing all other days.

3.1.2. CCFs

`EchelleCCFs.jl` computes the cross-correlation function between each observed spectrum and a mask. Each mask is specified by list of entries (each including the order, lines, and weight) and the mask shape. We present analyses based on two line lists. One is based on the ESPRESSO G2 mask, so as to facilitate comparisons to previous work. Since the ESPRESSO mask is limited to 380-786nm, we also compute a custom line list to use as much of NEID's wavelength grasp as practical based on `RvLineList.jl`. In both cases, stellar lines can be rejected due to being near telluric absorption lines based on the predicted slope in atmospheric transmittance (as a function of velocity) as computed by TAPAS (Bertaux et al. 2014).

For both line lists, our baseline analysis presented here uses a top-hat mask with width of 620.953 m s^{-1}and the CCF was evaluated at velocities within 12.5 km s^{-1}of a reference velocity roughly centered on the solar RV. In addition to the CCFs for each observation, we also compute daily averaged CCFs that are particularly useful for studying rotationally-linked stellar variability. While we tabulate daily averaged CCFs for days with only a few observations, we recommend focusing on days with at least 15 binned observations and follow this recommendation in our analysis below.

3.2. Radial Velocities

`RvSpectML.jl` provides multiple algorithms for estimating RVs from CCFs. For this study, RVs are measured separately for each order by fitting a Gaussian line profile to each order CCF. For each order, we compute RVs from the individual observations as well as daily averaged RVs. For intraday RV analyses (§4.2 and earlier), order weights are proportional to the inverse mean (over days with at least 15 binned observations passing the filter above) of the mean squared deviation of the binned order RVs from that day's mean RV. For comparing RVs across days, we compute an RV offset for each order and each observing run based on the daily means. The run offsets are subtracted from each order RV and then we add back the weighted mean RV offset for each observation's run based on the mean of a set of reliably LFC-calibrated orders (physical orders 98-112). Once RV offsets have been computed for each (run, order) pair, a single RV is estimated for each observation by taking a weighted average of order RVs. At this stage, updated order weights are computed proportional to the inverse of the mean squared deviation of the daily binned order RVs from their mean (over days with at least 15 binned observations). Orders with known wavelength calibration issues (see §2.2.4) and large measurement uncertainties are assigned zero weight.

3.3. CCF bisectors

We compute a weighted sum (over orders) of daily binned order CCFs, using weights proportional to the inverse of the mean squared deviation of the daily binned order RVs from their mean (over days). Orders with known wavelength calibration issues (see §2.2.4) are assigned zero weight. To compute the velocity at each of several target CCF depths (on both sides of the minimum), we interpolate using a Matern-$\frac{5}{2}$ Gaussian process conditioned on the observed CCF. We average and subtract the values from the left and right side of the CCF to obtain the CCF bisector and CCF velocity span at each target depth. We estimate RVs separately from the top and bottom portion of the daily average CCF. We compute a bisector inverse slope (BIS), here defined as the difference of the RV measured from the top of the CCF (RV_{top}, using fractional depths of 10–40%) and the RV measured from the bottom of the CCF (RV_{bottom} using fractional depths of 55–95%). We compute the bisector curvature as $(v_{top} - v_{mid}) - (v_{mid} - v_{bottom})$. For estimating

the curvature, the top is defined as the mean of the bisectors between depths of 30 and 50%, the middle is based on depths from 50 and 75%, and the bottom is based on depths from 75 and 90%. In each case, bisector is measured relative to the best-fit CCF normalization at depths spaced by 0.01. The depths used for measuring the BIS and CCF curvature have been adjusted slightly from their original definitions in the literature based on the observed shape of CCF and robustness of measurements at shallow CCF depths.

3.4. *Stellar activity indicators*

The NEID DRP provides measurements of several classical activity indicators (e.g., Ca II H&K and Hα) in FITS headers. Our tables merely propagate these values to enable their convenient analysis. Figure 8 shows time series for daily averages of four indicators that appear likely to be the most useful for characterizing rotationally-modulated variations and long-term activity variations.

3.5. *Additional solar-specific corrections*

Data quality summary statistics: We compute several additional quantities that can be useful when analyzing Sun-as-a-star observations. Summary statistics based on the exposure meter and pyrheliometer were described in §2.1.5 & 2.2.1. The NEID DRP already provides the Doppler boost necessary to transform observations into a (fictitious) frame such that the true center-of-mass solar velocity would be a constant in FITS headers SSBZNNN, where NNN specifies the echelle order (Wright & Kanodia 2020). We compute differential extinction and apparent rotation velocity using `NeidSolarScripts.jl`.

Differential extinction: Next, we estimate the apparent Doppler shift due to differential extinction, Δv_diff_ext or Δv_r, following notation of Collier Cameron et al. (2019). Δv_r can be expressed as the product of the extinction coefficient, k, and a complex function of geometry and a limb darkening coefficient (set to 0.6). Since NEID is able to identify measurements affected by transparency variations using data from the exposure meter and pyrheliometer individually (as opposed to in a statistical sense), we do not need to employ a mixture model like that of Collier Cameron et al. (2019) for estimating k. Instead, we simply fit a linear model to the signal-to-noise in echelle order 98 ($619.5 - -631.4$ nm) (which has similar wavelengths to HARPS-N "order 60" used by Collier Cameron et al. 2019) versus airmass. Modeling differential extinction as a single, grey shift is an approximation to be improved in a future version of the PSRP. Improving this approximation could be particularly important for making use of NEID's far red and NIR orders.

One complication is that there are significant changes in the signal-to-noise due to imperfect coupling of the solar fiber to the science fiber. The solar observations are scripted to periodically delay the start of exposure, so the port fiber can realign and recenter the solar fiber on the NEID science fiber, leading to occasional jumps in the SNR (Logsdon et al. 2022). We use the timestamps of all NEID solar observations within a day (even those unusable due to weather or pointing) to identify when the port fiber aligner makes adjustments. We group observations taken with gaps between consecutive exposure start times of less than 100s as belonging to a single "segment". When fitting for each day's parameters, we allow each segment to have its own SNR offset, but require all segments to share a common extinction coefficient.

There can also be changes in the throughout near solar noon, when the solar telescope flips to position itself for afternoon observations. Therefore, we exclude observations starting within 15 minutes of solar noon when fitting a differential extinction coefficient.

While a simple linear fit works well on days with plenty of data taken in clear weather, it is not practical to make a good measurement for some days with few clear observations. Therefore, we apply Bayesian inference, adopting priors that guide the model parameters to reasonable values on days with few observations. We adopt diffuse Normal priors for each segment intercept with zero mean and unit variance. We adopt a log Normal prior for the extinction coefficient with $\mu = \ln(0.1) + 0.5^2$ and $\sigma = 0.5$. This results in $k \simeq 0.1$ on days with few usable observations, likely leading to underestimating the magnitude of Δv_r on such days. For the jitter parameter, σ_j we adopt a log Normal prior distribution with $\mu = \ln(0.05) + 2^2$ and $\sigma = 2$. We use Markov chain Monte Carlo (MCMC) with the No U-Turn Sampler (NUTS) to obtain a posterior sample, the mean, median, and standard deviation of the marginal posterior for k, and credible intervals.

Given the potential for future work to provide updated differential extinction coefficients, we tabulate corrections Δv_r based on formalism of Collier Cameron et al. (2019) and a constant extinction coefficient of $k_\circ = 0.157482$ (the value of a preliminary fit derived from a subset of the NEID solar dataset, done as part of a preliminary analysis).

When applying differential extinction corrections for a given day, we first rescale the tabulated Δv_r by the ratio $k_{\rm med}/k_{\rm o}$, where $k_{\rm med}$ is the median of the marginal posterior for k for that day based on MCMC.

Apparent rotation velocity: The apparent solar angular rotation rate varies due to Earth's rotation and eccentric orbit. These affect the measured line widths and propagates to affect the CCFs. We calculate the difference between the squares of the full width half maximum (FWHM) of the CCF as observed and the FWHM of the CCF if it were observed in a sidereal frame ($\Delta {\rm fwhm}^2 = F_{\rm obs}^2 - F_{\rm sid}^2$) following conventions of Collier Cameron et al. (2019) We adopt the same scaling factor $\gamma = 1.04$ estimated by Collier Cameron et al. (2019). Note that this field is in km s^{-1} instead of m s^{-1}. Since flux is preserved, the depths of lines and CCF contrast are also affected. We calculate a sidereal CCF contrast, $C_{\rm sid} = C_{\rm obs}\frac{F_{\rm obs}}{F_{\rm sid}}$.

For each daily averaged CCF, we fit a Gaussian CCF profile allowing for an unknown normalization, location, width, and depth. The observed FWMH is estimated based on the best-fit parameters assuming a Gaussian CCF shape. The observed contrast is estimated as the product of the best-fit normalization and depth. Results are shown in the left panel of Figure 11. We find that the product CCF area, $A = C_{sid} \times F_{\rm obs}$ is nearly constant, so A and $C_{\rm sid}$ are not shown. The RMS of daily-averaged CCF areas relative to the mean CCF area is 0.00056.

3.6. *Data products*

We provide CSV files with tables of: (1) individual observations (`rvs.csv`), (2) binned RV observations (with all constituent observations satisfying the filters above) (`rvs_binned_5.csv` & `rvs_binned_10.csv`), and (3) daily averaged RV observations (`rvs_daily.csv`). Each table includes both a combined RV based on our chosen order weights and its formal uncertainty and order RVs (e.g., rv_NNN_gauss) and uncertainties (σrv_NNN_gauss), where NNN refers to the physical echelle order (equal to NNN=174-i, where i is the 1-based array index of orders in the NEID DRP FITS files). Unless otherwise specified, we report combined RVs using the order weights recorded in the (`order_weights_interday`) column of the `order_weights.csv` file.

The table of individual observations includes the name of the input L2 FITS file, several fields propagated from the FITS headers, the exposure meter and pyrheliometer data necessary to apply the filters described in §2.1.5 & §2.2.1, potentially using alternative thresholds, and estimated solar-specific corrections (§3.5).

The table of daily averaged RVs also includes statistics of that day's observations, including the number of binned observations that satisfy the above filters, the best-fit linear slope to that day's observations, the RMS deviation of that day's RVs from either a constant or linear model, and the best-fit linear slope of order RVs versus order index. We also provide a FITS file with daily averaged CCFs (`ccfs_daily.fits`).

Finally, we provide several additional files with ancillary data to facilitate reanalyses of the data that would like to replace one step of the pipeline, while preserving other steps. These include: (1) daily extinction model parameters (`extinction_daily.csv`), (2) classical spectroscopic activity indicators (`activity.csv`) computed by NEID DRP 1.3 pipeline and copied from FITS headers for convenience and their daily averages (`activity_daily.csv`), and (3) a list of input FITS filenames with their dates and bin they were assigned to, so others can easily reproduce our exact binning of data (`filenames.csv`), and (4) a list of observing "runs", their start and stop dates, RV offset adopted based on reliable LFC-calibrated orders, and offsets for each order included in our analysis (`run_offsets.csv`). The data release is archived on Zenodo (Ford 2024).

4. RESULTS

We performed a visual inspection of the binned order RVs within each day with at least 15 binned observations to search for evidence of any additional data quality issues. In order to evaluate results for orders that extend beyond the ESPRESSO mask, we show results based on our custom line list. We compare the mean daily RV using the PSRP to those from the NEID DRP v1.3 in Figure 2. For orders covered by both masks, results were very similar with the ESPRESSO G2 mask often appearing to perform slightly better. We leave an exploration and tuning of parameters with RvLineList for future research.

4.1. *Daily trend*

One concerning feature is an apparent slope in RVs versus time of day, as shown in Fig. 3. We find a median (over all days with sufficient data) slope of $\simeq -0.24$m s^{-1} hr^{-1}. The implied change in velocity over the course of a day of solar observations (over 1m s^{-1}) is significantly larger than the corrections expected based on differential extinction ($\simeq 0.06$m s^{-1}). Further, the steady downward trend is distinct from what is expected based on differential extinction,
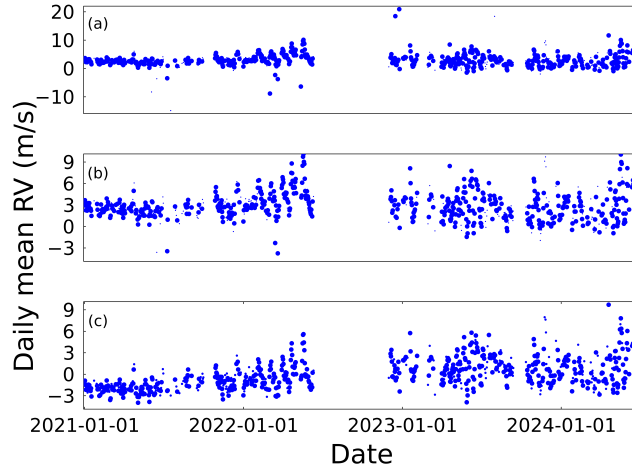
Figure 2. Top and Middle: Mean daily averaged RVs from the NEID DRP v1.3. The two panels differ only in their y-axis scale. Bottom: Mean daily RVs computed for this study using the Penn State Research Pipeline. The RMS of daily average RVs is 2.11 m s^{-1}. The formal photon noise measurement uncertainties are negligible. Both timeseries were filtered as described in §2. In both panels, the small points are days with less than 15 binned observations which are more affected by solar oscillations, have lower effective signal-to-noise, and are more susceptible to including an observation contaminated by atmospheric fluctuations.
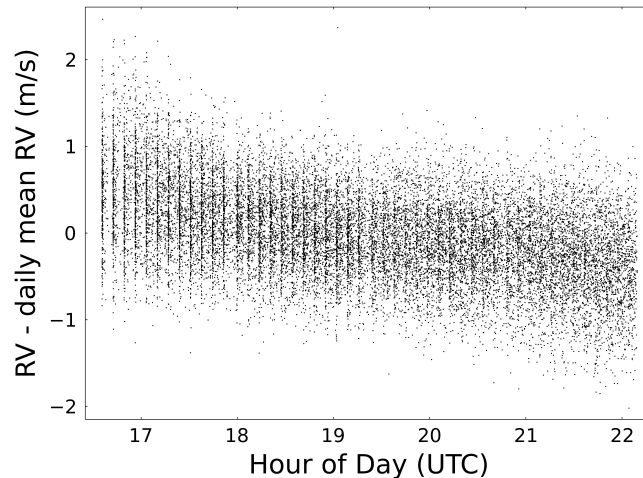

Figure 3. We show all binned NEID RVs from the PSRP as a function of the hour of day (UTC). The slope suggests that the NEID wavelength calibration model has room for further improvement.

since that depends on airmass and would lead to a reversal in the trend each solar noon. These observations suggest that the trend is due to an imperfection in the daily wavelength drift model. Next, we look for trends in the daily RV slope versus time with either the echelle order and/or time of year to gain insights into the cause and develop an empirical model for correcting RVs of this signal.

4.1.1. *Order*

The day-to-day dispersion in best-fit RV slope varies with echelle order, as shown in Fig. 4 (left). After excluding several orders in the red with especially large dispersion of daily slopes (presumably due to variable absorption from water lines), the typical dispersion in daily slopes is $\simeq 0.156$m s^{-1} hr^{-1}. We fit a linear model to the daily slope versus echelle order number (using physical orders 98:101, 108:111, 119:120, 123:124, 126:143, and 145:150) and find a slope of -0.179m s^{-1} hr^{-1} order^{-1}.
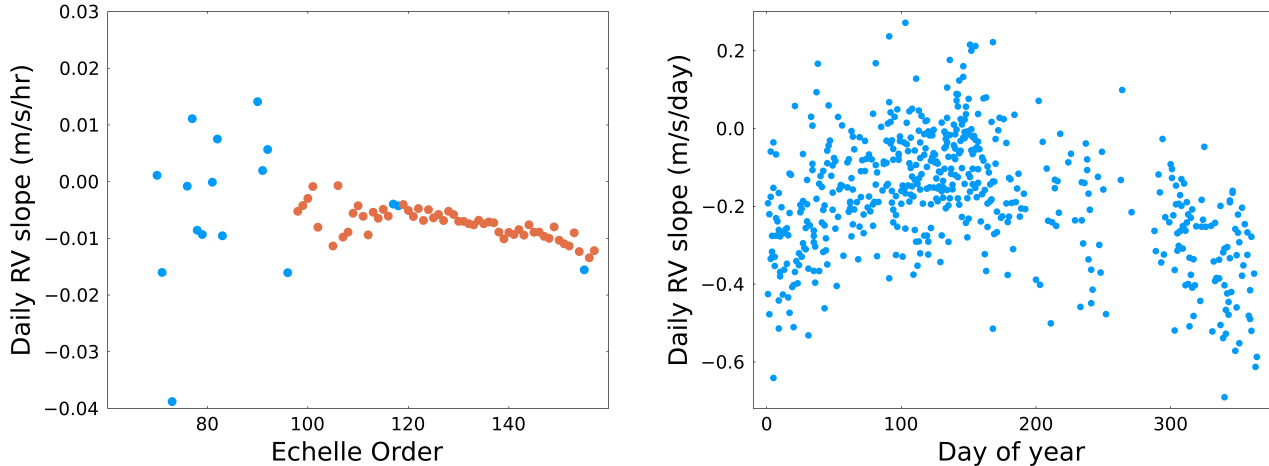
Figure 4. Left: Mean daily RV slope as a function of echelle order. Points in red were used for fitting the trends with RV slope on order and day of year. Right: Mean daily RV slope as a function of the day of the year.

4.1.2. *Time of year*

Next, we show the daily RV slopes versus time of year in Fig. 4. Here we have computed an average slope over the orders included in the fit in §4.1.1. There appears to be an annual pattern. When fitting a sinusoidal model with (with argument $2\pi \times$ (day of year)/365), we measure an amplitudes of 0.0817m s^{-1} hr^{-1} and -0.0883m s^{-1} hr^{-1} for the sine and cosine components, or 0.116m s^{-1} hr^{-1} for their quadrature sum.

Tables of RVs include the measured RV (`rv` and a "corrected" RV (`rv_corrected`) that subtracts off an estimate of the spurious RV due to these effects. We adopt a linear model for the spurious RV within each day, where the slope versus time that depends linearly on both the echelle order and the sine and cosine of the annual phase.

4.2. *Intraday RMS RV*

We compute the intraday RMS RV deviation from the above linear model for each echelle order, and report the mean over all days with at least 20 binned observations (see Figure 5, left). These are used to set the order weights for examining the intraday RMS RV. Figure 6 shows the distribution of the RMS deviation of the binned RV from both a constant (before correcting for differential extinction; top) and from one of two linear trends (after correcting for differential extinction). The middle panel subtracts off the linear trend estimated based on §4.1. The bottom panel subtracts off the best-fit linear trend for each day. While fitting for a linear trend within each day is intended to mitigate presumed wavelength calibration issues during solar observations, we recognize that it could also remove some astrophysical signals. Therefore, the distribution in the lower panel is expected to underestimate the within day RMS RV achievable with NEID. In contrast, the distribution in the middle panel is expected to overestimate the within day RMS RV that could be obtained with further pipeline improvements due to imperfections in the differential extinction and wavelength drift models. The median RMS RV when binning ten observations (0.343m s^{-1}) is slightly larger than the formal measurement uncertainties (typically 0.285m s^{-1} for a single observations) due to a combination of stellar variability (primarily pulsations and granulation, but also small scale magnetic activity), and instrumental noise sources (e.g., errors in the wavelength calibration).

4.3. *Across Day RMS RV*

For analyzing variations across days, we compute daily average RVs for each day with at least 10 binned observations. Order weights are based on the inverse variance of daily averaged RVs (see Figure 5 right). Orders with an RMS daily averaged RV greater than 3 m s^{-1} were assigned zero weight (and plotted with blue points). When computing the RMS RV within a run, each (order, run) pair gets its own RV offset. When computing the RMS RV over the full timespan of observations, the offset between the mean RV of each run is set based on the weighted mean of the reliable LFC orders. For computing the final daily average RVs, we update the order weights to be proportional to the inverse of RMS of daily averaged RVs (see Figure 5, right).
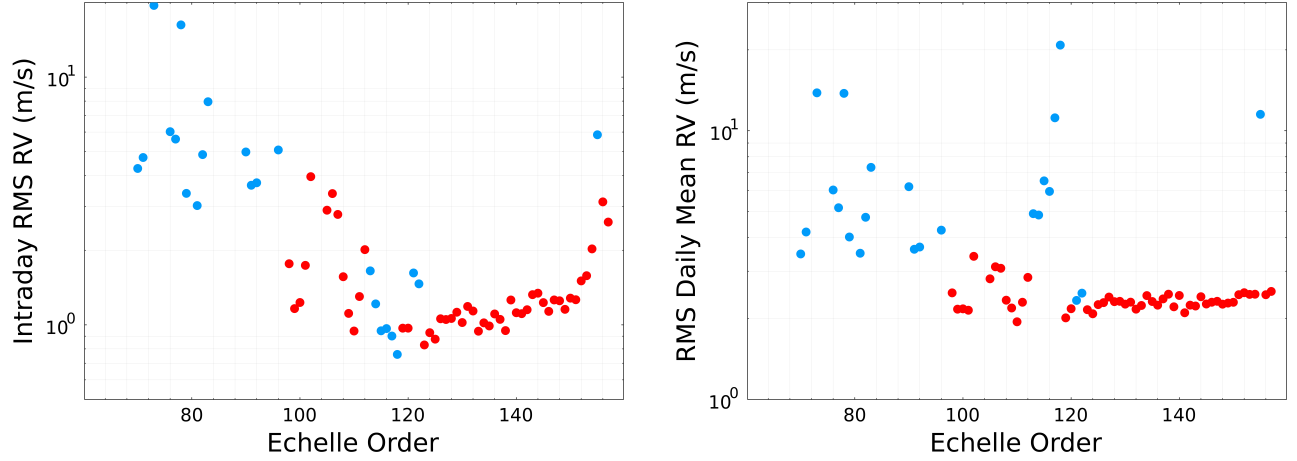
Figure 5. Left: Intraday RMS RV vs echelle order. The RMS RV within each day was used for setting initial order weights. Right: Interday RMS RV vs echelle order. These were used for setting the order weights used for comparing measurements across multiple days. The excursions near order 115-118 are due to days where the LFC flux was low. Including these orders in the computation of RVs results in some days showing up as large outliers. Setting the order weights of these to zero results in significantly reduced RMS daily RVs.
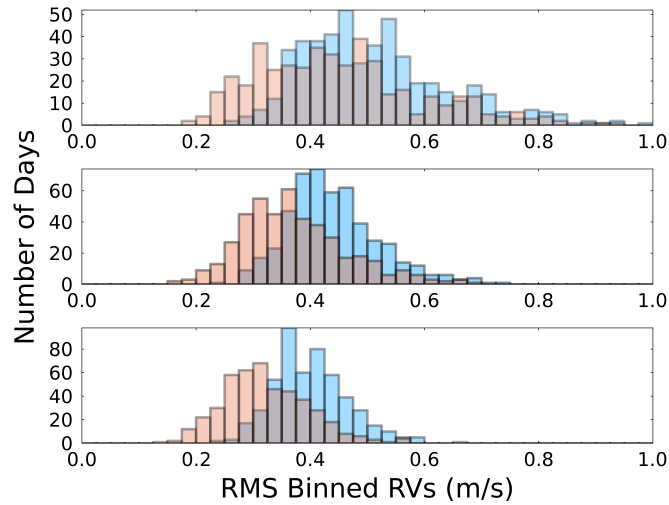
Figure 6. Histograms of RMS of binned RVs within each day. The light blue histograms are based on binning 5 consecutive observations and at least 20 binned observations within the day. The light red histograms are based on binning 10 consecutive observations and at least 10 binned observations within the day. The overlpa region appears as grey. Top: The RMS RV is measured relative to a constant mean for each day before applying corrections for differential extinction or imperfections in the wavelength drift model. Middle: The RMS RV is measured relative to a constant mean for each day using RVs that include our correction for differential extinction. This is likely to overestimate the true RMS RV due to imperfect models for differential extinction and drift in the the wavelength model. Bottom: The RMS RV is measured relative to a best-fit linear model for the RVs within each day. This may underestimate the RMS RV due to solar variability, since the linear model could absorb a linear trend due to changes in the solar spectrum over the course of the day.

4.4. *Daily Average RV*

Figure 7 shows the mean daily RV versus date. The top two panels make up run 1 (pre-fire) and the bottom two panels make up run 2 (post-fire). Small points show the daily average RV on days with less than 15 binned observations. During early 2021 (top panel), the sun was still in a phase of relatively low activity. By late 2022, we can clearly see an astrophysical signal in the RV time series on the solar rotation timescale of ∼26 days.
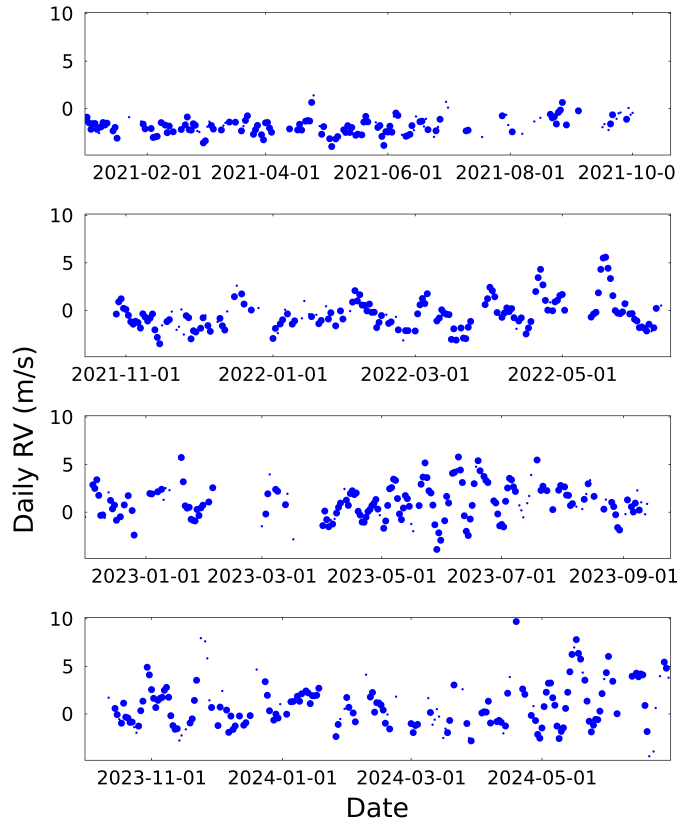
Figure 7. Mean daily RV as a function of time. We show four sections of data each with a roughly similar number of days with observations passing our quality filters, but with precise breakpoints set by large gaps in the data. Small dots indicate days with less than fifteen binned observations that passed our filters.

4.5. *Daily Average of Classical Spectroscopic Activity Indicators*

Stellar astronomers have identified several spectroscopic indicators useful for studying magnetic variability and other astrophysics. For Sun-like stars, the Ca II H & K lines are particularly effective at diagnosing both rotationally modulated magnetic activity and long-term magnetic activity cycles. Since the Ca II H & K lines are in the near UV, measuring this indicator is challenging for cooler stars and for some optical spectrographs. In these cases, other indicators such as Hα have also proved useful for recognizing stellar magnetic activity. The NEID DRP provides measurements of these as well as several other classical spectroscopic indicators (He I D3 at 587.562 nm, Na I D1/D2 at 589.592 and 588.995 nm, H-α, Ca I at 652.2795 nm, Ca II Infra-red Triplet (IRT) at 849.8018, 854.214 and 866.214nm, Na I NIR Doublet at 818.326 and 819.482, Paschen Δ, and Mn I at 539.4677 nm). We show the time series for daily averages of four selected indicators in Figure 8, Ca II H & K (Duncan et al. 1991), Ca IRT (Zechmeister et al. 2018), H-α (Gomes da Silva et al. 2011), and Mn 539 (Danilovic et al. 2016).

These confirm that the Sun was in a quiet state in early 2021 for the beginning of NEID solar observations and began ramping up in activity shortly before the Contreras fire shutdown. The Sun has been quite active since the NEID restart (for all of run 2). The classical activity indicators demonstrate that the spurious RVs measured by the NEID Solar Telescope are dominated by astrophysical effects (rather than instrumental or pipeline), reveal periods of time with prominent rotationally-modulated activity signals, and show a long-term trend during the timespan of these NEID solar observations that qualitatively resembles that seen in the measured RVs (see Figure 7). Thus, these measurements will prove helpful in interpreting results in §5 where we attempt to mitigate the effects of solar variability on measured RVs.

5. EVALUATING SOLAR VARIABILITY MITIGATION STRATEGIES

While several methods have been proposed to mitigate the impact of stellar variability, it is very difficult to validate these methods on stars other than the sun, since their true center-of-mass motion is unknown. Further, different methods can suggest significantly different "corrections" (Zhao et al. 2023). Here, we evaluate several wavelength-domain methods for mitigating stellar variability. Each involves measuring empirical correlations between the measured RV shift and one or more input "features" where both are calculated from individual spectra and using those correlations to "clean" the measured RVs of (some of the) effects of intrinsic solar variability. Some features were defined for other purposes and unlikely to be optimal for mitigating solar variability, but can have benefits of being easier to interpret. Other feature sets are derived using statistical or ML methods. We anticipate that these are likely to have greater predictive power, but may be more difficult to interpret, to explain, and to earn the trust needed to support the discovery of Earth-analogs. By comparing the results of different stellar variability mitigation methods on a common high-quality solar dataset, we find many similarities and find one method that is particularly promising for future application and further development.

5.1. *Classical Spectroscopic Activity Indicators*

Some previous studies have attempted to clean exoplanet survey RVs by detrending with classical activity indicators like Ca II H&K or Hα. We apply this technique to the daily averaged solar data, so as to provide a reference point for evaluating the effectiveness of detrending against other ML-based stellar variability indicators in subsequent sections. We apply the same filters as for RVs and compute daily averages of the values provided for each NEID observation by the NEID DRP v1.3. Just as the pre/post-fire shutdown introduced offsets in the wavelength calibration, there are also clear offsets in some of the classical activity indicators. The bulk of these offsets have been traced to the fact that NEID DRP v1.3 did not use the blaze function derived for each observation when computed activity indicators. This issue should be fixed starting in DRP v1.4. For this study, we simply add unknown coefficients for the mean of each activity indicator during each of run 1 and 2. Among classical activity indicators, detrending against Ca II H & K was the most effective at reducing the RMS of daily averaged RV from 2.11 m s^{-1} to 1.39 m s^{-1}. For instruments that can not measure Ca II H & K, H-α is a common alternative spectroscopic indicator. Detrending against H-α reduce the RMS of daily averaged RV to 1.48 m s^{-1}. Previous studies have noted that often there appears to be a lag between local extrema of the RV timeseries and spectroscopic activity indicators such as Ca II H & K (Collier Cameron et al. 2019). As this study focuses on purely wavelength-domain methods for mitigating solar variability, we leave a joint analysis that allows for temporal lags for a future study.

5.2. *Daily Red & Chromatic RVs*

One feature of NEID is its broad wavelength grasp. This was intended to allow NEID to measure classical stellar activity indicators that run from near UV/blue to far red/NIR. It also has the potential to be useful for measuring differences in apparent RV with wavelength or order. In Figure 9, we examine this wavelength dependence by dividing the order RVs into two disjoint groups that we refer to as Optical RVs (physical echelle orders $\geq$ 97 or $\lambda \leq 631.4$nm) and Red RVs (echelle orders 70-96, 632.4nm $\leq \lambda < 884.0$nm). These are daily average RVs, with order weights based on our standard order weighting. There is a strong correlation between the optical and Red RVs, and a moderate correlation between the optical RVs and the difference between the two ("chromatic RV") (see Figure 10). Attempting linear decorrelation using the chromatic RV reduces the RMS of daily mean RVs from 2.11 m s^{-1} to 2.03 m s^{-1}. The chromatic RVs appear to be tracing long-term changes more than stellar variability on the rotational timescale.

5.3. *CCF-based Mitigation Strategies*

Next, we attempt to mitigate the effects of stellar variability on the daily mean RVs by considering several proposed stellar variability indicators calculated from the CCF. We compute a mean daily CCF for 544 days with at least 15 binned observations passing our data quality filters. Below, we apply several approaches for characterizing the CCF shape and test their effectiveness for linear decorrelation with measured solar RVs.

5.3.1. *Classical CCF Shape Indicators*

Figure 11 (left) shows time series for multiple common daily CCF shape measurements, including the full width at half maximum (FWHM; measured based on fitting a Gaussian line profile to the CCF), the bisector inverse slope (BIS), and the bisector curvature (see §3.3). The CCF FWHM and BIS show little variability in 2021, but significant rotational-linked variability in subsequent years, and a long term increase which we attribute to the solar magnetic
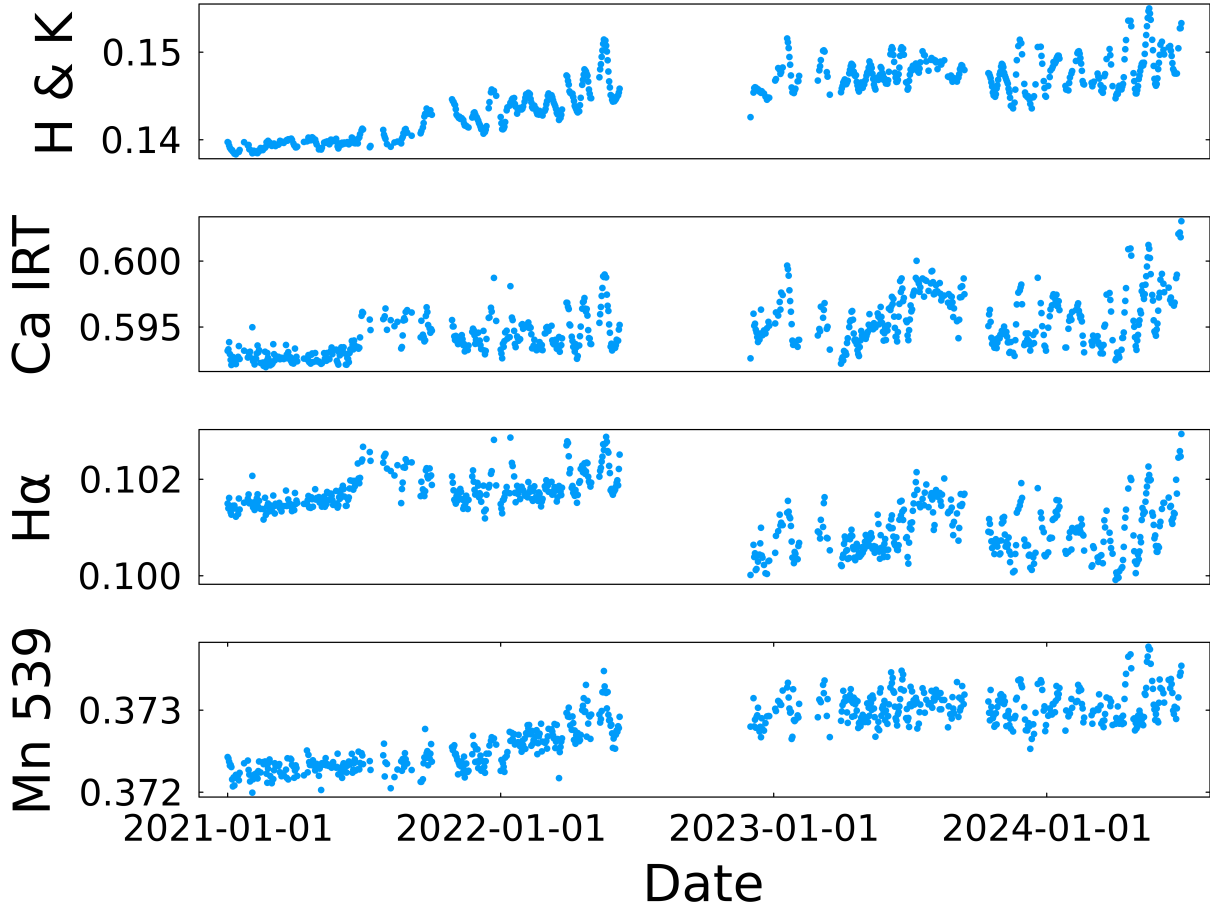
Figure 8. Daily averages of selected classical spectroscopic activity indicators versus time. Some indicators display a substantial offset between run 1 and run 2 (large gap in late 2022) due to a pipeline configuration issue.
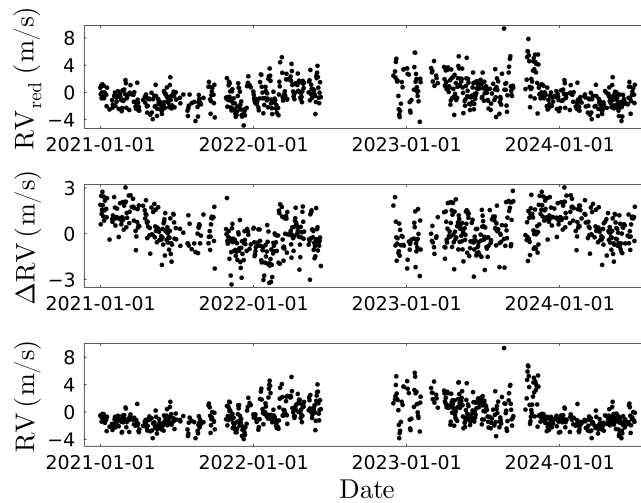
Figure 9. Top: Daily averaged Red RVs based on echelle orders 70-96 versus time. Middle: Daily averaged chromatic RVs (difference between the Red RVs and Optical RVs based on echelle orders $\geq$ 97) versus time. Bottom: Daily mean of Optical RVs after decorrelating based on the chromatic RVs.
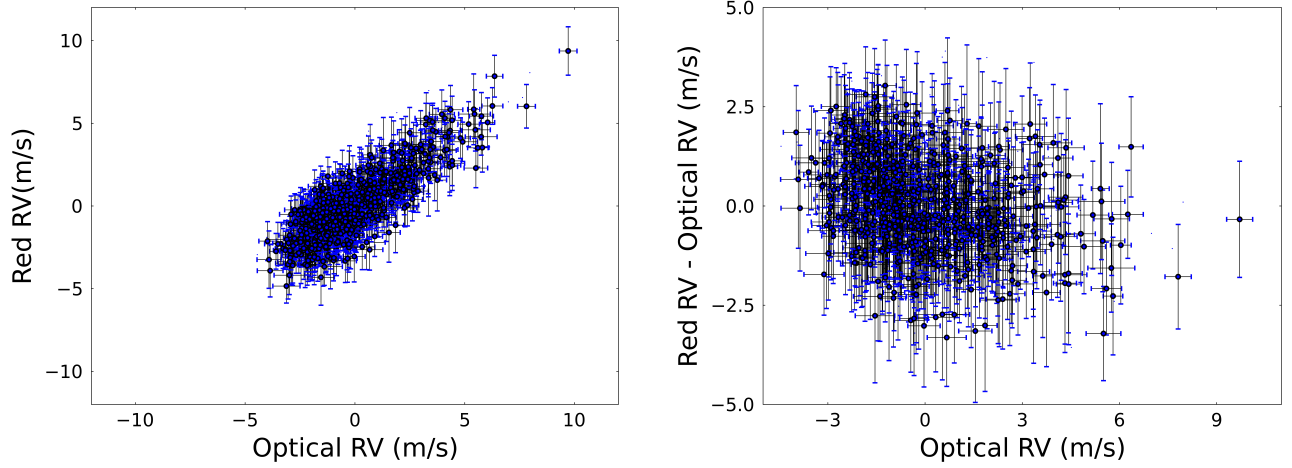
Figure 10. Left: Daily average RVs for days with at least 15 binned RVs using our baseline order weights ("optical") and weights using only "red" portions of the spectrum. Errorbars show the within-day RMS RV for based on the two sets of order weights. Right: Difference in Red RV and Optical RV versus Optical RV.
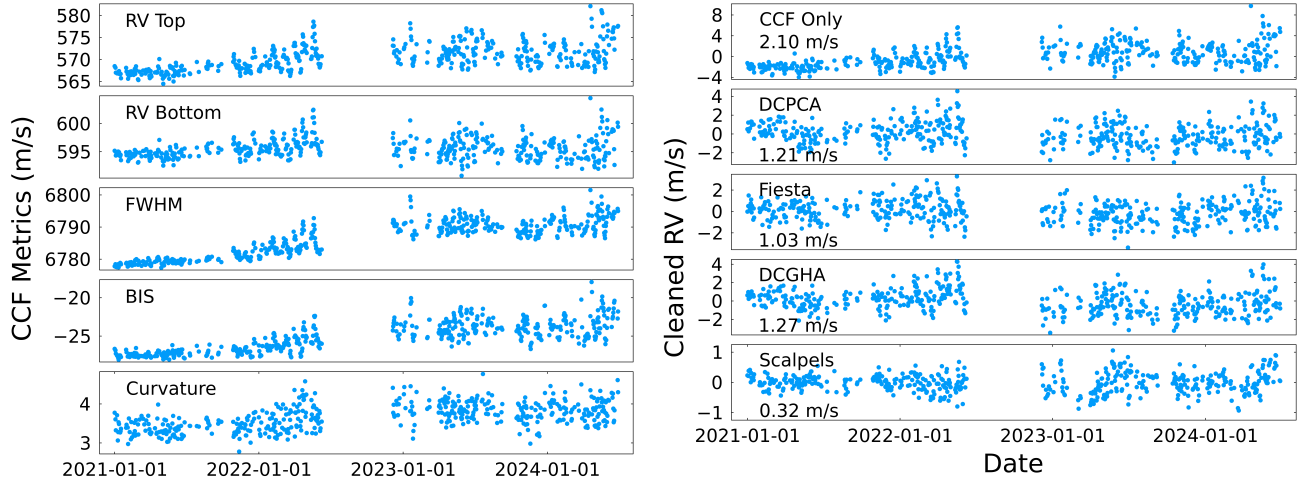
Figure 11. Left: Daily-averaged CCF-based solar variability metrics. Since the CCF area is very nearly constant, the product of the CCF FWHM and CCF contrast (ratio of depth to CCF normalization) is very nearly constant, and the CCF contrast (not shown) would appear like an inverted version of the CCF FWHM. Right: Daily-averaged RVs after detrending with different solar variability indicators.

cycle. Both are consistent with trends observed in Ca II H&K and other classical activity indicators. When we attempt to detrend using any either the classical CCF shape indicators, we find using the CCF BIS yields the lowest RMS of daily RVs, 1.66 m s^{-1}. While an improvement, this proved less effective than using Ca II H&K in §5.1.

5.3.2. *Doppler-constrained Principal Components Analysis (DCPCA)*

Next, we consider using several different features measured from daily-averaged CCFs using data-driven or machine learning (ML) approaches. First, we consider a variant of Doppler-constrained Principal Components Analysis (Jones et al. 2017; Gilbertson et al. 2020a, DCPCA). Instead of analyzing the spectrum directly with DCPCA, we apply it to the daily-averaged CCFs. We summarize the steps: (1) shift CCFs to remove large known barycentric corrections, (2) compute a template CCF, (3) subtract the template from each CCF to form a residual CCF, (4) project the residual CCF onto the derivative of the template CCFs to estimate the RVs, (5) compute the projection of the residual CCF onto the plane orthogonal to the RV component (CCF$_{\text{shape}}$), (6) perform a standard principal components analysis (PCA) of CCF$_{\text{shape}}$'s, and (7) perform linear regression on the PCA scores with the estimated RVs (after subtracting

the Doppler shifts from any known planets) to estimate model coefficients for predicting the portion of the estimate RVs that can be attributed to shape driven RVs (RV_{shape}), (8) predict the shape driven RVs (RV_{shape}) for each observation based on the PCA scores, and (9) estimate the shift-driven or "cleaned" RVs by subtracting RV_{shape} from the estimated RVs.

The results are shown in second right panel of Figure 11. DCPCA has one important algorithmic parameter, the number of PCA feature vectors and scores to retain in steps 7 & 8 when computing (RV_{shape}). The blue curve in the left panel of Figure 12 shows how the RMS of cleaned RV decreases as the number of scores and feature vectors increase. Using even a single feature vector yields a substantial improvement. There is a gradual improvement with the use of additional feature vectors 2–8. Some algorithms are susceptible to increasing out-of-sample prediction errors when the number of scores and feature vectors used increases since the model coefficients can start to increase due to overfitting. Fortunately, we find that DCPCA yields a long plateau extending to several dozen feature vectors. In such cases, a common strategy is to pick the smallest number of feature vectors that reaches the plateau. Applying DCPCA to the NEID solar daily-averaged CCFs dataset yields an RMS RV of 1.30m s^{-1}. This represents a substantial improvement over detrending RVs based on either classical spectroscopic activity indicators or traditional CCF shape metrics.

One could worry about the potential for "overfitting", or overestimating the improvement in RMS RVs when applied out-of-sample, due to using the same spectra being used in steps 6 & 7 ("training") and steps 8 & 9 ("testing"). To address this concern, we applied a form of cross-validation, where we train and test using disjoint subsets of daily-averaged CCFs. Rather than drawing random subsets of our data, we consider two physically motivated partitionings. Our primary cross validation test is based on separating the data taken as part of run 1 and run 2, i.e., before and after the Contreras fire. Between the two runs the instrument was warmed up to its long-term safe mode and cooled back down, and we need to allow for RV offset between runs in any analysis. There is a multi-month gap between the observations, so there is negligible risk of information about the distribution of magnetic active regions during the training set providing useful information for the test set. The results are shown in the blue curve in the right panel of Figure 12. The solid (dotted) line corresponds to training of run 2 (1) and testing on run 1 (2). Since there are clear differences in the level of solar activity between the two runs, the two train/test pairs are expected to yield different results. For these reasons, our train/test scenario represents a much steeper challenge than a more typical application of cross validation. When using five DCPCA scores, we find out-of-sample RMS's of the cleaned daily RVs of 1.34 and 1.32 m s^{-1}, only slightly larger than when we trained and tested on the full dataset. Using ten DCPCA scores, the out-of-sample RMS's of the cleaned daily RVs improve to 1.31 and 1.21 m s^{-1}.

We also performed an analogous cross validation scheme separating the data into four sections (shown in separate panels of Figure 7), where we train using three segments and test using the one held-out segment. The performance was typically comparable or improved relative to our reference cross validation using runs 1 & 2. This is not unexpected, since using four segments has the advantage of using a larger amount of training data (and to a lesser extent the effect of including a few training observations near in time to a few testing observations).

5.3.3. *FourIEr phase SpecTrum Analysis (FIESTA)*

Next, we apply the FIESTA algorithm (Zhao et al. 2022a) to the daily NEID solar CCFs. Here, we summarize the steps: (1) shift CCFs to remove large known barycentric corrections, (2) normalize each CCF by the "CCF continuum", (3) transform the normalized CCFs by negating them and adding one (i.e., 1−CCF), (4) compute coefficients for discrete Fourier transform on the transformed CCFs, (5) estimate the RVs from each Fourier mode (RV_k's) using the Fourier phase information, (6) subtract the weighted mean RV from the Fourier modes to obtain the shift-invariant ΔRV_k's, (7) perform linear regression on the ΔRV_k's with the estimated RVs (after subtracting the Doppler shifts from any known planets) to estimate model coefficients for predicting the portion of the estimate RVs that can be attributed to shape driven RVs (RV_{shape}), (8) predict the shape driven RVs (RV_{shape}) for each observation based on the ΔRV_k's, (9) estimated the shift-driven or "cleaned" RVs by subtracting RV_{shape} from the estimated RVs.

The results are shown in the third panel on the right of Figure 11 and with the red curves in Figure 12. We find that FIESTA yields an RMS of daily cleaned RV of 1.03m s^{-1} when using 6 feature vectors. Like DCPCA, FIESTA also yields a broad plateau, but reaches this plateau with just four feature vectors. (We caution that the number of feature vectors that FIESTA needs depends on the velocity span of the input CCFs. We adopted a velocity span of 25 km s^{-1}, as we found that concentrated the useful information into a small number of scores and feature vectors.) When we perform cross validation using runs 1 and 2, we find that RMS of FIESTA 's out-of-sample RV predictions for run 2

(1.21m s^{-1}) are significantly worse than for run 1 (1.06m s^{-1}). These are somewhat smaller than the analogous from analysis applying DCPCA to the daily averaged CCFs.

5.3.4. *Doppler-constrained Gauss-Hermite Analysis (DCGHA)*

Next, we apply Doppler-constrained Gauss-Hermite Analysis (DCGHA) which combines ideas from several previous studies, particularly Jones et al. (2017); Holzer et al. (2021); Gilbertson et al. (2020a). Here, we summarize the steps: (1) shift CCFs to remove large known barycentric corrections, (2) compute a template CCF and estimate its mean and width by fitting a Gaussian line profile, (3) subtract the template from each CCF to form a residual CCF, (4) project the residual CCF onto the derivative of the template CCFs to estimate the RVs, (5) compute coefficients for projecting the residual CCF onto Gauss-Hermite basis functions using the mean and width estimated in step #2, (6) perform linear regression on the Gauss-Hermite coefficients with the estimated RVs (after subtracting the Doppler shifts from any known planets) to estimate model coefficients for predicting the portion of the estimated RVs that can be attributed to shape driven RVs (RV$_{\mathrm{shape}}$), (8) predict the shape driven RVs (RV$_{\mathrm{shape}}$) for each observation based on the Gauss-Hermite coefficients, (9) estimate the shift-driven or "cleaned" RVs by subtracting RV$_{\mathrm{shape}}$ from the estimated RVs.

The results are shown in the fourth panel on the right of 11 and with the green curves in Figure 12. There is little benefit to including the first, second and third Gauss-Hermite coefficients unless the fourth Gaussian-Hermite is also included. (The number of feature vectors used is one greater than the order due to the 0th order coefficient which is quite useful even by itself.) When using 5 feature vectors, we find that the Doppler-constrained Gauss-Hermite analysis yields an RMS daily cleaned RV of 1.29 m s^{-1}, very similar to that of DCPCA once both use at least 6 feature vectors. This is not unexpected. Since both are linear function of the CCF, they should yield nearly identical results if the maximum number of feature vectors were used by both algorithms. The two algorithms are taking different approaches to choosing how to parameterize and reconstruct the projection of the CCFs onto the plane orthogonal to a Doppler shift. Once both have retained most of the useful information, their predictions for RV$_{\mathrm{shape}}$ are very similar.

When we perform cross validation using runs 1 and 2, we find that RMS of DCGHA's out-of-sample RV predictions for runs 1 (1.15 m s^{-1}) and 2 (1.11 m s^{-1}) are comparable to the results from DCPCA, once sufficient feature vectors are included. Interestingly, the RMSs for the two test sets are more similar to each other than when using DCPCA or FIESTA. Based on comparing the predictive performance from analyzing the full data set and subsets for cross validation purposes, the DCGHA appears to perform better with smaller datasets than DCPCA or FIESTA.

5.3.5. *Self-Correlation Analysis of Line Profiles for Extracting Low-amplitude Shifts (Scalpels)*

Finally, we apply Self-Correlation Analysis of Line Profiles for Extracting Low-amplitude Shifts (Scalpels; Collier Cameron et al. 2021) to the NEID daily solar CCFs. Here, we summarize the steps: (1) shift CCFs to remove large known barycentric corrections, (2) compute the discrete autocorrelation function (ACF) of each CCF, (3) perform a Singular Value Decomposition (SVD) on the autocorrelation function, (4) perform linear regression on the eigenvalues of the SVD decomposition with the estimated RVs (after subtracting the Doppler shifts from any known planets) to estimate model coefficients for predicting the portion of the estimate RVs that can be attributed to shape driven RVs (RV$_{\mathrm{shape}}$), (5) predict the shape driven RVs (RV$_{\mathrm{shape}}$) for each observation based on the the ACF of the CCF for that observation, (6) estimate the shift-driven or "cleaned" RVs by subtracting RV$_{\mathrm{shape}}$ from the estimated RVs.

The RMS of daily RVs after cleaning with 5 Scalpels features (0.981 m s^{-1}) is a substantial improvement over the CCF-based RV and significantly less than that of any of the other algorithms considered in this study. Allowing Scalpels to use ten features results in additional improvements (0.277m s^{-1}; see Figure 12 left). Scalpels performs better than the other methods considered when holding the number of feature vectors used fixed as long as at least 2 features are used. Scalpels shows a rapid improvement as the number of features used increases from 1 to 4, a stall from 5 to 7, and then a very broad plateau for 8 or more features used.

When we perform cross validation using runs 1 and 2 and less than ten features (see Figure 12 right), we find that RMS of Scalpels's out-of-sample RV predictions for run 1 (0.784 m s^{-1}) are less than those for run 2 (0.969 m s^{-1}, values for 5 feature vectors). When training on either run 1 or 2 alone, more feature vectors are needed before it reaches its plateau. When we performed cross validation scheme training on three of the four sections (shown in separate panels of Figure 7) and testing on the held-out segment, Scalpels performs better than when using less training data, acheiving a smaller RMS of out-of-sample RV predictions (0.350, 0.422, 0.340and 0.373m s^{-1}when using 10 feature vectors) and reaching its plateau reliably with fewer feature vectors. Based on comparing the predictive performance from analyzing the full data set and subsets for cross validation purposes, it appears that Scalpels's performance does
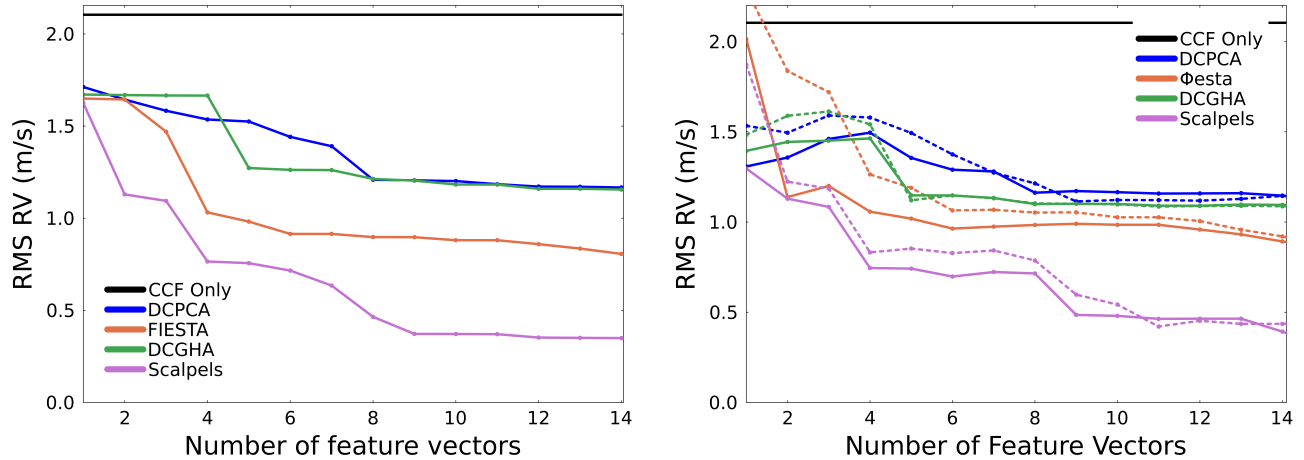
Figure 12. RMS of daily average RVs after detrending with different sets of data-driven stellar variability indicators as a function of the number of indicators used. Left: Results from training and testing each method to the full data set. Right: Results from training each method on one of two runs and testing on the other run. The dotted (solid) curve show results training on run 1 (2) and testing on run 2 (1).

degrade somewhat as the number of observations in the training sample decreases. However, once Scalpels uses at least ten features, the RMS of Scalpels's out-of-sample cleaned RVs for the two runs become comparable (0.426 m s^{-1} and 0.510 m s^{-1}). This represents a substantial improvement in the level of stellar variability mitigation compared to all the other methods considered here and to previous published analyses (for actual stellar observations as opposed to simulated data).

6. DISCUSSION

Here we discuss implications of our findings for NEID, mitigating stellar variability, and the design of future EPRV instruments and surveys.

6.1. *RMS of Daily Cleaned RVs*

The measured RMS of daily-averaged RVs before attempting to clean them of stellar variability is 2.11 m s^{-1}. This greatly exceeds the photon noise for each daily binned spectrum (varies with the number of observations included, $\simeq$0.02–0.07 m s^{-1}) and the NEID instrumental error budget (0.27 m s^{-1}). Decorrelating the daily RVs with any of the data-driven, wavelength-domain stellar variability mitigation strategies considered yielded a significant reduction in the RMS RV (resulting in an RV RMS of 0.277– 1.03 m s^{-1} depending on algorithm). Each data-driven algorithm we considered performed better than decorrelating against classical spectroscopic indicators or chromatic RV information. Of particular note, Scalpels performed particularly well, reducing the RV RMS of daily cleaned data to 0.277 m s^{-1} and achieving an out-of-sample RV prediction accuracy of 0.340– 0.422m s^{-1} when training of approximately three quarters of the data.

As the NEID solar telescope continues to collect more data, we'll be able to reassess how much of the difference between these and the 0.277 m s^{-1} when applied to the full dataset is due to the reduction in the size of the training set and how much is due to out-of-sample predictions being harder than within sample predictions.

6.2. *Effect of limiting integration time*

Given the success of Scalpels in cleaning daily averaged CCFs, one naturally wonders whether the excellent performance was only possible due to combining hundreds of spectra to obtain a very high effective signal-to-noise. Therefore, we repeated our Scalpels analysis of daily averaged spectra, but computing daily averaged CCFs that used no more than one hour of integration time each (see Figure 13). Limiting the integrating time per day to no more than one hour increased the RMS of the cleaned daily RVs by only $\simeq$ 16% relative to our baseline analysis. In cross validation tests, the increase in RMS of daily cleaned RV based on just one hour of observations per day was similar, resulting in an RMS of $\simeq$ 0.4 m s^{-1}.
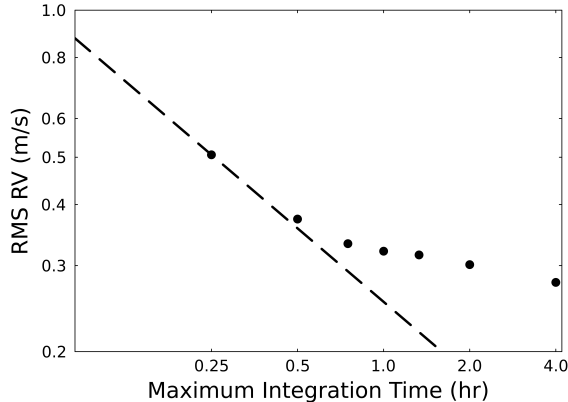
Figure 13. RMS of daily mean cleaned RV as a function of the maximum integration time allowed per day. Daily average RVs are computed for days with at least fifteen binned observations (each bin consisting of five observations). The number of binned observations included in each daily average CCF was limited, so the total integration time did not exceed the duration on the x-axis. For days with more observations than needed, binned observations with the lowest airmass were selected. Daily averaged CCFs were cleaned with Scalpels using 10 feature vectors. The dashed line shows a $-\frac{1}{2}$ power law with arbitrary normalization.

The RMS of cleaned RVs increases slowly as the maximum integration time decreases from four to $\simeq 1$ hour, and then increases more rapidly for maximum integration times of less than 45 minutes. For comparison, the typical photon noise for measuring a pure Doppler shift using one hour of integration time is $\simeq 0.045$ m s^{-1}. While Scalpels is removing the impact of stellar variability to the $\simeq 0.3$ or 0.4 m s^{-1} level, doing so requires precisely characterizing the CCF shape and thus much higher signal-to-noise to than if the solar spectrum were constant and errors were dominated by photon noise.

For Sun-like stars where magnetic activity is effectively suppressed using techniques such as Scalpels as shown here for the Sun, granulation and supergranulation are expected to become limiting factors for the accuracy of EPRV mass measurements (Luhn et al. 2023; Palumbo et al. 2024). Therefore, An additional potential factor in how well Scalpels can clean RV is the wall time of solar observations significantly exceeds the integration time. Solar oscillations, a complex beat pattern of stochastically excited p-modes, granulation, super granulation, and potentially other processes contribute to the observed solar RV variability on timescales of minutes to days (Crass et al. 2021). A portion of the improvement as a function of integration time may be due to the greater wall time more effectively averaging over these processses. For night-time observations, the duty cycle is higher than for solar observations ($\simeq 66\%$). Therefore, night-time observations are expected to be more effective in averaging over oscillations than NEID solar observations. However, once granulation and/or supergranulation become significant sources of RV contamination, it may be important to make multiple visits to a target star within a night to spread observations over a greater wall time.

6.3. *Consistency of ML-based Detrending*

Figure 14 compares the RV corrections suggested by different methods. We find that the corrections suggested by each algorithm considered are highly correlated with each other, in contrast to the findings of Zhao et al. (2022b). This could be due to a variety of factors. First, Zhao et al. (2022b) included more algorithms that took a more diverse set of approaches to cleaning RVs. For example, this study does not attempt to use any time information, while several methods presented in Zhao et al. (2022b) did. Second, since many teams performed their own prepossessing of data, the analyses in Zhao et al. (2022b) were less homogeneous than in this study. Third, this study focused solely on Solar observations and the daily binned observations are at higher signal-to-noise than the datasets from (Zhao et al. 2020). While each of these factors is likely to have contributed to the different findings, further research will be need to quantify how much is due to each factor.

6.4. *Comparison to NEID Error Budget*
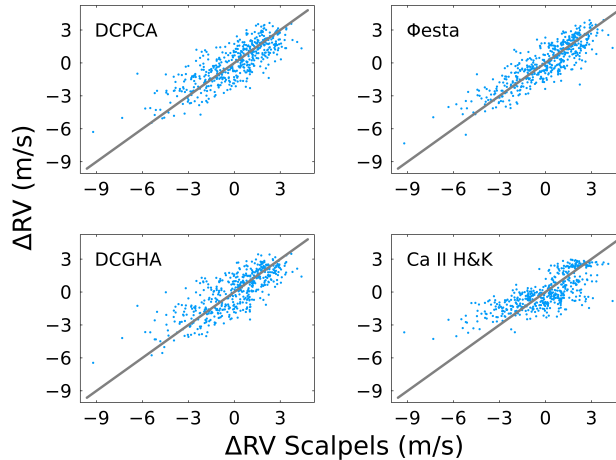
Figure 14. Comparison of the correction applied to daily-averaged RVs when detrending using different stellar variability indicators. In each panel, the x-axis is the correction applied by Scalpels. The y-axis is the correction is applied based on trending against scores from DCPCA (top left), ΔRV_k's from FIESTA (top right), scores from projecting the CCF_{shape} onto a Gauss-Hermite basis (bottom left), and the classical Ca II H&K activity indicator.

After cleaning by Scalpels, the RMS of daily averaged RVs over the NEID solar dataset (0.277 m s^{-1}) is significantly larger than the photon noise limit, but comparable to the 0.27 m s^{-1} expected based on NEID instrumental error budget (Halverson et al. 2016). While a formal error budget for the NEID solar telescope was not developed at the same level of rigour, clearly the solar telescope is also performing well, after applying the filters described in this paper. This provides empirical evidence that the NEID instrument and pipeline are approaching their design goals (at least for bright G-star targets after applying corrections for known pipeline issues described in §2 and §4.1). Further research will be needed to identify those terms in the error budget that are most significant after cleaning data of solar variability. The largest four terms in NEID's error budget are: software algorithms for data reduction, software algorithms for the calibration process, micro-telluric contamination, and sky fiber subtraction, each of which is budgeted 0.1 m s^{-1}. Our results in §4.1 suggest that the algorithms for data reduction and calibration in NEID DRP v1.3 have room for further improvement. When we performed multiple analyses with different line list that were more/less aggressive in avoiding telluric contamination, we did see small, but potentially interesting changes in the results, suggesting that further research in avoiding telluric contamination may yield additional improvements. The sky fiber subtraction is not used for solar observations, so the solar observations are not a diagnostic for that term during night time observations.

6.5. *Comparison to HARPS-N Solar RVs*

The 0.277 m s^{-1} RMS of NEID's daily averaged RVs after cleaning by Scalpels is significantly less than the 1.25 m s^{-1} obtained by applying Scalpel to HARPS-N solar observations (Collier Cameron et al. 2021). This is particularly impressive, since the sun was significantly more active during much of the NEID observations than during the HARPS-N observations analyzed in Collier Cameron et al. (2021). The explanation for this difference is unclear.

One candidate explanation is NEID's improved environmental stability. Interestingly, the application of Scalpels to HARPS-N solar observations benefited from using 13 feature vectors, more than needed to achieve a much lower RMS of daily RVs with NEID solar observations. Collier Cameron et al. (2021) linked multiple feature vectors to instrumental interventions that likely led to changes in the instrument LSF. In contrast, inspecting the Scalpels scores for NEID solar observations does not reveal any feature that can be clearly mapped to changes in the instrument before and after the Contreras fire. The smaller number of feature vectors needed here suggests that NEID's improvement environmental stability is paying off in terms of improved LSF stability and ability to attribute observed line shape variation to astrophysical variability. We interpret this as evidence that NEID's improved LSF stability may contribute to better removal of spurious RVs due to stellar variability.

Another potential explanation for the difference in results when comparing NEID and HARPS-N solar observations is NEID's use of both the pyrheliometer and exposure meter to reject observations affected by poor atmospheric

conditions. Therefore, we recommending including/adding a pyrheliometer to other solar telescopes. Yet another possibility is that it is more effective to train data-driven algorithms on datasets that include significant solar variability, and most previous studies using HARPS-N solar data have focused on a range of data when the Sun was relatively inactive. This explanation could be tested by performing updated analyses of HARPS-N solar observations that include a wider range of solar activity levels.

6.6. *Prospects for EPRV detection of Earth twins*

Collier Cameron et al. (2021) applied Scalpels to synthetic a dataset based on 853 days of HARPS-N solar observations, but with Doppler signature of multiple hypothetical low-mass planets added. They found that Scalpels would allow for accurate mass measurements of low-mass planets, with RV amplitude uncertainties of $\simeq 0.066$ m s^{-1} . Scaling their results by the ratio of the RMS of cleaned daily RVs from NEID to the RMS of cleaned daily RVs from HARPS-N would yield an expected uncertainty on the Doppler amplitude of 0.015 m s^{-1}, sufficient to obtain a "6-σ" detection for an Earth twin. Thus, it appears we are now on the precipice of having the technology to detect true Earth twins with modern EPRV spectrographs (and generous observing time allocations). §6.8.2 discusses some of the key remaining challenges.

6.7. *Summary of key findings*

Our findings show that the NEID solar telescope is performing extremely well, measuring solar RVs accurate to 0.489 m s^{-1}, once binning over p-mode pulsations. This and the measured RMS of daily mean RV observations (2.11 m s^{-1}) are comparable to recent results from other solar telescopes such as those operating at HARPS-N, EXPRES, and HARPS (Zhao et al. 2023). We find that postprocessing the NEID daily averaged CCFs with one of several data-driven techniques can yield significant reduction in the accuracy of daily averaged solar RVs. The Scalpels algorithm stands out for its excellent performance in mitigating solar variability, reaching an RMS of daily averaged spectra of just 0.277 m s^{-1}. Even a strenuous cross validation test using only half of the available data for training yielded out-of-sample RMS of daily RV predictions of 0.426– 0.510 m s^{-1}. Limiting the integration time per day to 1 hour resulted in only a modest increase in the RMS of RVs after cleaning by Scalpels. Based on scaling the results from a previous study that jointly fit Scalpels with planetary scaling, it appears that detecting and characterizing the mass of an Earth-twin may now be within reach of existing EPRV instruments, if provided sufficient observing time.

This motivates further effort devoted to refining pipelines and data analysis techniques in hopes of reducing the amount of observing time required and to allow EPRV facilities to search as many nearby stars as practical. To support further research and future comparisons to other stellar variability mitigation strategies, we provide a large public data set that has been consistently processed and cleaned of known instrumental effects. By applying algorithms to the exact same observations, the results of subsequent analyses can be directly compared to the results presented here.

6.8. *Future research opportunities*

While our results are very encouraging for the prospects of NEID and other EPRV spectrographs to discover and characterize rocky planets in the habitable zone of Sun-like stars, there are still several areas ripe for further research.

6.8.1. *Calibration and pipeline improvements*

First, we find opportunities for further improving the NEID wavelength calibration and drift model, both within each day and across days. It seems likely that improvements in the drift model could lead to Scalpels and other methods becoming even more effective at removing spurious RVs due to variability. This finding demonstrates that one of the powerful benefits of making solar observations is the ability to stress test an instrument and its data reduction pipeline, so as to motivate future improvements in data process, operations, and/or hardware. Similar analyses of solar observations from other EPRV instruments would likely be helpful in characterizing their performance, prioritizing issues to be improved, and enabling more precise cross-instrument comparisons.

6.8.2. *Extending to night-time observations of other stars*

Second, our analysis was restricted to solar observations that offer several advantages over night time observations. It is practical to obtain hundreds of solar observations every clear day with a favorable observing cadence since solar observations do not impinge on nighttime observing programs. While each individual exposure is not particularly high

signal-to-noise, combining information from several hours of observations effectively provides access to unusually high signal-to-noise data. Additionally, combining solar data taken over hours helps to suppress stellar variability due to oscillations and granulation. Additional research is needed to understand how the performance of algorithms such as Scalpels degrades as the number of observations, signal-to-noise, and integration time observations degrades. A preliminary exploration indicates the performance is only mildly degraded ($\simeq$ 16%) if the total integration time per day is reduced to 1 hour. Further research will be needed to explore the interaction of integration time and wall time (for marginalizing over granulation), as well as how these vary with stellar type.

Third, night time observations have some additional sources of noise or systematics that do not affect solar observations significantly. For example, solar observations due not need to worry about removing background contamination via a sky fiber. As another example, the throw of barycentric correction for solar observations is ~ 1km s^{-1}, while other targets can have up to $\simeq 30$km s^{-1} throw. Here we used a line list which was cleaned to avoid lines that would suffer significant contamination even if the barycentric corrections reached a 30 km s^{-1}. However, further research will be needed to characterize the performance of stellar variability mitigation algorithms in the presence of both stellar and telluric variability.

Finally, additional research is needed to understand how these algorithms perform on other stars and to what extent solar data can be used for training and testing stellar variability mitigation techniques for other stellar types. This is particularly important for EPRV to play a supporting role in discovery and mass characterization of potentially Earth-like planets for future large direct imaging missions (e.g., Habitable Worlds Observatory).

6.8.3. *Combining wavelength & time-domain information*

Stellar mitigation strategies can be divided into three broad categories: (1) wavelength-domain strategies that use changes in each stellar spectrum individually to distinguish true Doppler shifts from stellar variability, (2) time-domain strategies that use the known shape true planetary signals to separate true Doppler shifts from stellar variability, and (3) hybrid strategies that analyze spectroscopic timeseries jointly, so as to combine wavelength and time-domain information. In §5, we explored several wavelength-domain strategies that make use of classical spectroscopic indicators, the CCF bisectors, or machine learning approaches using input features computed from the CCF. In this study, we have considered mitigating the impacts of solar variability on derived RVs using only the instantaneous spectrum, rather than incorporating temporal information. Wavelength-domain approaches has the significant benefit of improving the interpretability and explainability of results relative to hybrid strategies. We recommend future studies to consider time-domain and hybrid strategies. In particular, previous studies have found that some stellar activity indicators appear to have a lag, i.e., reach their local maximum a few days before/after the spurious RV reaches a local extrema. Complementary techniques such as Gaussian process regression using a latent process and its derivatives (e.g., Gilbertson et al. 2020b), a Gaussian process with a non-standard kernel that explicitly incorporates a lag, or a convolutional neural network may be capable of making better use of the stellar variability indicators considered here. We defer making use of temporal information for future studies.

Facilities: WIYN (NEID)

Software: EchelleCCFs.jl v0.2.9 (https://github.com/RvSpectML/EchelleInstruments.jl), EchelleInstruments.jl v0.2.11 (https://github.com/RvSpectML/EchelleInstruments.jl), NeidArchive.jl v0.1.3 (https://github.com/RvSpectML/NeidSolarScripts.jl), NeidSolarScripts.jl v0.1.6 (https://github.com/RvSpectML/NeidSolarScripts.jl), RvLineList.jl (https://github.com/RvSpectML/RvLineList.jl), RvSpectML.jl v0.2.8 (https://github.com/RvSpectML/RvSpectML.jl), RvSpectMLBase.jl v0.2.3 (https://github.com/RvSpectML/RvSpectMLBase.jl), SunAsAStar.jl v0.1.2 (https://github.com/RvSpectML/SunAsAStar.jl)

REFERENCES

Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46, doi: 10.1051/0004-6361/201322383

Cabona, L., Claudi, R., Ghedina, A., et al. 2024, in EAS2024, 837

Collier Cameron, A., Mortier, A., Phillips, D., et al. 2019, MNRAS, 487, 1082, doi: 10.1093/mnras/stz1215

Collier Cameron, A., Ford, E. B., Shahaf, S., et al. 2021, MNRAS, 505, 1699, doi: 10.1093/mnras/stab1323

Crass, J., Gaudi, B. S., Leifer, S., et al. 2021, arXiv e-prints, arXiv:2107.14291, doi: 10.48550/arXiv.2107.14291

Cretignier, M., Francfort, J., Dumusque, X., Allart, R., & Pepe, F. 2020, A&A, 640, A42, doi: 10.1051/0004-6361/202037722

Danilovic, S., Solanki, S. K., Livingston, W., Krivova, N., & Vince, I. 2016, A&A, 587, A33, doi: 10.1051/0004-6361/201527039

Davis, A. B., Cisewski, J., Dumusque, X., Fischer, D. A., & Ford, E. B. 2017, ApJ, 846, 59, doi: 10.3847/1538-4357/aa8303

Dumusque, X. 2016, A&A, 593, A5, doi: 10.1051/0004-6361/201628672

Dumusque, X., Glenday, A., Phillips, D. F., et al. 2015, ApJL, 814, L21, doi: 10.1088/2041-8205/814/2/L21

Dumusque, X., Borsa, F., Damasso, M., et al. 2017, A&A, 598, A133, doi: 10.1051/0004-6361/201628671

Duncan, D. K., Vaughan, A. H., Wilson, O. C., et al. 1991, ApJS, 76, 383, doi: 10.1086/191572

Ford, E. B. 2024, High-level Data Products for NEID Solar Observations (January 1, 2021 - June 30, 2024), 0.1.0, Zenodo, doi: 10.5281/zenodo.13363762

Gilbertson, C., Ford, E. B., & Dumusque, X. 2020a, Research Notes of the American Astronomical Society, 4, 59, doi: 10.3847/2515-5172/ab8d44

Gilbertson, C., Ford, E. B., Jones, D. E., & Stenning, D. C. 2020b, ApJ, 905, 155, doi: 10.3847/1538-4357/abc627

Gomes da Silva, J., Santos, N. C., Bonfils, X., et al. 2011, A&A, 534, A30, doi: 10.1051/0004-6361/201116971

Halverson, S., Terrien, R., Mahadevan, S., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99086P, doi: 10.1117/12.2232761

Hara, N. C., & Ford, E. B. 2023, Annual Review of Statistics and Its Application, 10, 623, doi: 10.1146/annurev-statistics-033021-012225

Holzer, P. H., Cisewski-Kehe, J., Zhao, L., et al. 2021, AJ, 161, 272, doi: 10.3847/1538-3881/abf5e0

Jones, D. E., Stenning, D. C., Ford, E. B., et al. 2017, arXiv e-prints, arXiv:1711.01318, doi: 10.48550/arXiv.1711.01318

Kanodia, S., Lin, A. S. J., Lubar, E., et al. 2023, arXiv e-prints, arXiv:2307.12403, doi: 10.48550/arXiv.2307.12403

Lin, A. S. J., Monson, A., Mahadevan, S., et al. 2022, AJ, 163, 184, doi: 10.3847/1538-3881/ac5622

Llama, J., Zhao, L. L., Brewer, J. M., et al. 2024, arXiv e-prints, arXiv:2407.07967, doi: 10.48550/arXiv.2407.07967

Logsdon, S. E., Wolf, M. J., Li, D., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 121844N, doi: 10.1117/12.2629004

Luhn, J. K., Ford, E. B., Guo, Z., et al. 2023, AJ, 165, 98, doi: 10.3847/1538-3881/acad08

Luque, R. 2022, in European Planetary Science Congress, EPSC2022–90, doi: 10.5194/epsc2022-90

Nelson, B. E., Ford, E. B., Buchner, J., et al. 2020, AJ, 159, 73, doi: 10.3847/1538-3881/ab5190

Palumbo, M. L., Ford, E. B., Gonzalez, E. B., et al. 2024, AJ, 168, 46, doi: 10.3847/1538-3881/ad4c6d

Robertson, P., Anderson, T., Stefansson, G., et al. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 015003, doi: 10.1117/1.JATIS.5.1.015003

Rubenzahl, R. A., Halverson, S., Walawender, J., et al. 2023, PASP, 135, 125002, doi: 10.1088/1538-3873/ad0b30

Schwab, C., Rakich, A., Gong, Q., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99087H, doi: 10.1117/12.2234411

Stefansson, G., Hearty, F., Robertson, P., et al. 2016, ApJ, 833, 175, doi: 10.3847/1538-4357/833/2/175

Wright, J. T., & Kanodia, S. 2020, PSJ, 1, 38, doi: 10.3847/PSJ/ababa4

Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, A&A, 609, A12, doi: 10.1051/0004-6361/201731483

Zhao, J., Ford, E. B., & Tinney, C. G. 2022a, ApJ, 935, 75, doi: 10.3847/1538-4357/ac77ec

Zhao, L., Fischer, D. A., Ford, E. B., et al. 2020, Research Notes of the American Astronomical Society, 4, 156, doi: 10.3847/2515-5172/abb8d0

Zhao, L. L., Fischer, D. A., Ford, E. B., et al. 2022b, AJ, 163, 171, doi: 10.3847/1538-3881/ac5176

Zhao, L. L., Dumusque, X., Ford, E. B., et al. 2023, AJ, 166, 173, doi: 10.3847/1538-3881/acf83e