Astroconformer: Inferring Surface Gravity of Stars from Stellar Light Curves with Transformer

Jiashu Pan [* 1]   Yuan-Sen Ting [* 2 3]   Jie Yu [4]

Abstract

We introduce Astroconformer, a Transformer-based model to analyze stellar light curves from the *Kepler* mission. We demonstrate that Astrconformer can robustly infer the stellar surface gravity as a supervised task. Importantly, as Transformer captures long-range information in the time series, it outperforms the state-of-the-art data-driven method in the field, and the critical role of self-attention is proved through ablation experiments. Furthermore, the attention map from Astroconformer exemplifies the long-range correlation information learned by the model, leading to a more interpretable deep learning approach for asteroseismology. Besides data from *Kepler*, we also show that the method can generalize to sparse cadence light curves from the Rubin Observatory, paving the way for the new era of asteroseismology, harnessing information from long-cadence ground-based observations.

1. Introduction

If single-epoch data sets drove significant breakthroughs over the last decade, the coming years would undoubtedly herald the golden era of multi-epoch data. The highly successful *Kepler* space satellite (Koch et al., 2010) has yielded time-series data for hundreds of thousands of stars. And superseding it, the TESS mission (Ricker et al., 2014) further *Kepler*'s ambition by monitoring stars across the entire sky. And complementing these space satellites are the myriad of ground-based facilities. Notably, the US flagship Rubin

---
[*]Equal contribution [1]School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China [2]Research School of Astronomy & Astrophysics, Australian National University, Cotter Rd., Weston, ACT 2611, Australia [3]School of Computing, Australian National University, Acton, ACT 2601, Australia [4]Max Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, 37077 Gttingen, Germany. Correspondence to: Jiashu Pan <jspan@smail.nju.edu.cn>, Yuan-Sen Ting <yuansent.ting@anu.edu.au>.

observatory (Ivezić et al., 2019) will start taking data in 2023, whose goal is to scan the sky every few days. As ground-based telescopes generally have a much larger aperture, Rubin will observe much fainter stars than *Kepler* and TESS, thus collecting an unprecedented sample of stellar light curves from the Milky Way's outer halo.

Stellar light curves are invaluable: as the brightnesses of stars vary due to physical processes, such as their oscillations and granulation, the light curves can reveal fundamental properties of stars, including their masses, radii and intrinsic brightnesses, a technique known as asteroseismology. Since the masses and intrinsic brightnesses of stars can then be used to infer precise ages and the distances of them, asteroseismology has been playing a critical role in our quest to unravel the history of the Milky Way.

Despite its central role, the tools to analyze the light curves remain rudimentary to date. For a large part, stellar light curves are often assumed to be a Gaussian process and then characterized by the power spectrum to investigate oscillations and granulation (Toutain & Appourchaux, 1994; Benomar et al., 2009; Gruberbauer et al., 2009). However, as solar-like oscillations are stochastically excited and intrinsically damped (Chaplin & Miglio, 2013), the excitation and damping process of solar-like oscillations is a realization of a Levi process, which contain subtle non-Gaussian signals. Additionally, the long-range granulation, i.e., the convection on stellar surfaces, is also primarily non-Gaussian signals.

Asteroseismology is further limited by the lack of the short-cadence observation to detect high-frequency modes of less evolved stars. In light of this limitation, the community has since seen the potential of applying machine learning to capture lower-frequency information. For example, Hon et al. (2018) applies convolutional neural networks (CNNs), and Sayeed et al. (2021) applies a k-nearest neighbors method to study asteroseismic data. However, such methods come with their limitations. The k-nearest neighbors method fails to make robust inference in the less-populated sample space. The convolution layers of CNNs restrict them to only a local perceptive field. As time-series data often have more long-range correlated information excited at different times, such information is better captured with the attention-based mechanism.
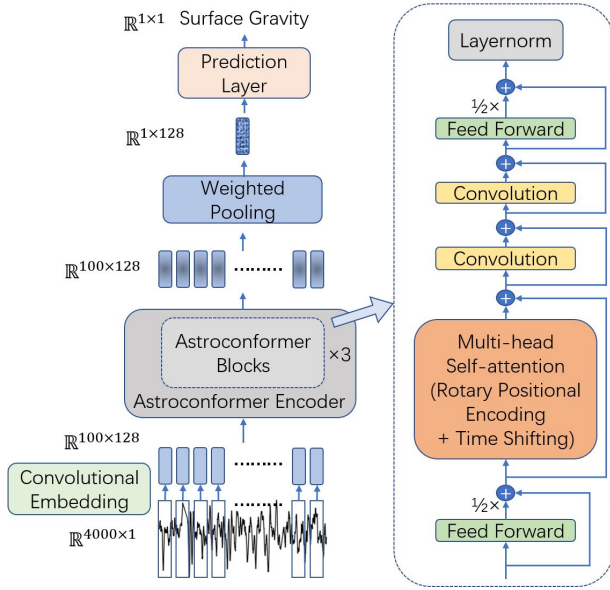
Figure 1. Architecture of Astroconformer. Astrconformer ingests stellar light curves and infer the surface gravity of stars. Our final model comprises three Astroconformer blocks. Each Astroconformer block consists of the convolution and multi-head self-attention layers to capture both the local and long-range information, respectively, going beyond the more explored CNNs method in asteroseismology.

In this study, we will explore the utility of Transformer (Vaswani et al., 2017) to capture long-range information. As a proof of concept, we will apply the method to infer the stellar surface gravity ($\log g$) from stellar light curves.

2. Astroconformer

This study adopts a variant of the Transformer framework known as the Conformer. Conformer was first proposed in the context of speech recognition (Gulati et al., 2020). At its core, Conformer extracts both long-range and local information simultaneously by deploying the self-attention and convolution modules in each "Conformer block." By interleaving convolutions between Multi-head Self Attention (MHSA) module and the Feed Forward module, Conformer aims to learn the global interaction with MHSA while capturing local features with convolutions. We posit that the Conformer framework is well suited for asteroseismic light surveys because, like speeches, stellar light curves also have correlations on varying timescales.

Fig 1 illustrates our model, which we dub the name Astroconformer. Our method differs from the native implementation in the following aspects. (1) We adopt simply a convolution layer to carry out preliminary feature extraction

and tokenize light curves instead of transforming signals into spectrograms as implemented in Gulati et al. (2020). This is motivated by the fact that the sampling rate of light curves (mins) is much lower than that of speech signals. (2) We include an extra convolution layer in every Astroconformer block. (3) We apply time-shifting technique (Bo, 2021) and rotary positional encoding (Su et al., 2021) in MHSA module. (4) After the Astroconformer blocks, we assume weighted pooling across different tokens to adaptively combine and yield the final embedding (Zhu et al., 2018).

We will benchmark Astroconformer against the state-of-the-art machine learning method in asteroseismology. In particular, we will compare with THE SWAN (Sayeed et al., 2021), a local linear regression method based on k-nearest neighbors. Noting that others also have explored various CNNs to perform a wide range of tasks in asteroseismology (e.g. Hon et al., 2018; Dhanpal et al., 2022), we will also demonstrate the importance of self-attention by comparing with CNNs through ablation experiments.

3. Experiments

3.1. *Kepler*'s Data and Mock Rubin Observatory Data

We adopt an identical subset of *Kepler* long-cadence light curves as THE SWAN, which facilitates a direct comparison. The data consists of 14003 one-quarter light curves with a median noise of ~ 100 parts per million (ppm). Our training labels are asteroseismic $\log g$ spanning 0.2-4.3 dex with a typical precision of ~ 0.01 dex (Mathur et al., 2017; Yu et al., 2018). All light curves are normalized by applying a high-pass filter, as was done in Sayeed et al. (2021), to remove any low-frequency variations due to the *Kepler* spacecraft.

In addition to the *Kepler* data, we will demonstrate that Astroconformer can be generalized to deal with ground-based observations with a much sparser cadence. In particular, the Rubin Observatory will scan the sky for ten years with a typical cadence of about four days, with an expected noise of 4600 ppm (Ivezić et al., 2019). To simulate observations from Rubin, we combine the *Kepler* light curves of the same stars from all *Kepler*'s observed quarters, spanning typically three years in duration. We subsequently degrade the *Kepler* light curves to mimic the noise profile and cadence (assumed to be fixed) of that of the Rubin Observatory.

3.2. Performance Comparisons and Ablation Studies

We first evaluate the performance with our *Kepler* training data to compare Astroconformer with the state-of-the-art method. Then the critical role of self-attention is showed by ablation studies. The ablation studies are tailored to be a fair comparison with approximately the same number of learnable parameters for all models. Evaluations are done by calculating the root mean square (RMS) errors from the
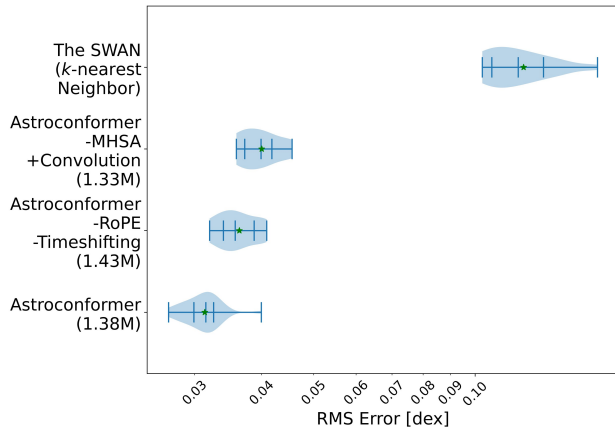
Figure 2. Long-range information harnessed by Astroconformer leads to a more precise inference of the surface gravity of stars. The violin plot shows the RMS errors from the 10-fold cross-validation from the *Kepler* training data and their quantiles. Astroconformer outperforms THE SWAN (*k*-nearest neighbors) by threefold and the mere convolution model by about 30%.

10-fold cross-validation of the training data. The results are summarized in Fig 2.

First of all, and perhaps unsurprisingly, deep learning models outperform THE SWAN by a large margin ($\sim$3 folds), with or without MHSA. Our cross-validation studies with *Kepler* data demonstrate that even with only 10^4 samples in the training data, deep-learning models learn more than just "linearly interpolating" the data set, as was done in THE SWAN. More importantly, Astroconformer shows that, even with the same training set and the number of learnable parameters, transformer-based model captures long-range information that is critical in characterizing stellar light curves. As shown in Fig 2, removing MHSA from Astroconformer and substituting it with convolution layers leads to the degradation of performance by about 30% in RMS errors.

On top of that, our ablation experiments also show that, with the still limited training set from *Kepler*, the inductive bias introduced by the bag of tricks in this study (rotary positional encoding and time shifting) has helped the Astroconformer attains better RMS errors ($\sim$ 20%).

3.3. What has Astroconformer Learned?

A key advantage of Transformer models, compared to some other deep learning models, is that visualizing the attention can often shed light on what the Transformer models have learned (e.g. Vig, 2019; Hoover et al., 2019). In Fig 3, we consider a specific *Kepler* light curve and visualize its attention in the first MHSA module.
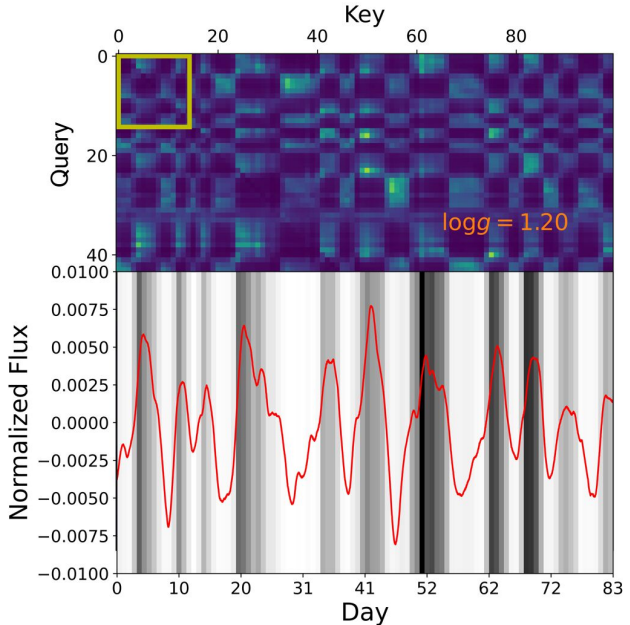

Figure 3. Attentions as learned by Astroconformer. *Top:* Shown is the attention map evaluated for a particular stellar light curve from *Kepler* (KIC 4252170). The yellow square shows the characteristic stellar oscillation period of this star. *Bottom:* The light curve with background illustrates the attention value from the first row of the attention map. The attention map shows that Astroconformer extracts not only the characteristic Gaussian oscillations but also more subtle non-Gaussian information within the oscillations.

The attention map illustrates a plethora of information captured by Astroconformer. (1) The attention map presents a periodical block pattern whose period is close to the characteristic oscillation period of the star, signifying the model learns about stellar oscillations. This is further demonstrated in the bottom panel. (2) Importantly, the off-diagonal block shows that our model has understood pixels at varying time stamps are strongly correlated regardless of their location, essentially capturing the long-range information (e.g., granulation) that CNNs miss. (3) Furthermore, Astroconformer also harnesses information at a time scale smaller than the characteristic oscillation, extracting other local "non-Gaussian" information, presumably from local dynamos that power the stellar oscillations. It is through extracting this combination of Gaussian and local non-Gaussian information and the advantage of capturing long-range correlation through self-attention that explains the superb performance Astroconformer compared to CNNs (see Fig 2).

3.4. Leveraging Ground-Based Light Curves

The US-led flagship Rubin Observatory will provide an unprecedented number of stellar light curves from the faintest
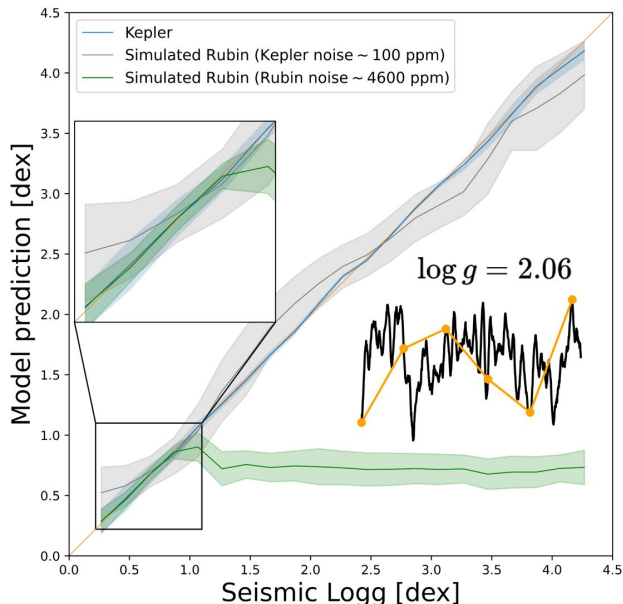
Figure 4. Prospects for inferring surface gravity from Rubin observations. Shown are the 1σ error of the $\log g$ inferences on the *Kepler* data (blue) and the simulated data for Rubin (gray = 100 ppm, green = 4600 ppm). Analyzing the much higher cadence of *Kepler*'s data (light curve in black) with Astroconformer leads to an RMS error of 0.027 dex. Although the simulated Rubin Observatory data (light curve in orange) has a cadence of four days, Astroconformer achieves an RMS error of about 0.22 dex with *Kepler*-like noise for all stellar types. With a Rubin-like noise, Astroconformer attains a precision of 0.07 dex for stars with $\log g < 1$.

stars (with magnitude > 26). However, given the ultra sparse cadence (~ 4 days) from ground-based observations, to our best knowledge, no one has thus far attempted to infer asteroseismic properties of stars through such light curves. That is because only stars with $\log g < 1$ have oscillation periods longer than $\mathcal{O}(10)$ days. Added to the challenge is the limitation of the power spectrum to determine fundamental stellar properties from uneven, sparse light curves; power spectra have a low-frequency resolution and are subject to the observed cadences that could vary over stars for Rubin.

Without uniform sampling assumption of CNNs, transformer also has the potential to extract information beyond stellar oscillations from unevenly spaced time series (Tipirneni & Reddy, 2021), leading to the possibility of studying the light curves from ground-based missions like Rubin in a meaningful way. To make a preliminary validation of the vision, we attempted Astroconformer for the mock Rubin data which is assumed to be regularly sampled. Fig 4 shows that in the limit with the same noise as *Kepler* (100 ppm), even with a cadence of four days, Astroconformer

can infer $\log g$ of stars to about 0.22 dex. Since the typical oscillation period of stars for $\log g = 1$ is ten days (near the Nyquist frequency) and even longer for higher $\log g$, extracting only stellar oscillation information would fail. Although here we assume an idealistic noise level, the result demonstrates further that Astroconformer indeed extracts information beyond just the stellar oscillations and can harness subtle non-Gaussian and long-range information even at a cadence of four days.

Nonetheless, with a more realistic noise of 4600 ppm for Rubin, the noise significantly "Gaussianizes" the signals. The Rubin noise reduces the ability of Astroconformer to most stars with an oscillation period shorter than the Rubin cadence ($\log g > 1$, with a period shorter than 10 days). In this case, we can only infer $\log g$ for stars with $\log g < 1$ to 0.07 dex. Despite that, we emphasize that a 0.07 dex precision for stars with $\log g < 1$ approaches the $\log g$ precision typically attained only by spectroscopy (Zhao et al., 2012). As the surface gravity of giant stars is tightly correlated with the stars' intrinsic brightness, Astroconformer will help refine our measurement of distances from these evolved M-giants, which constitute the bulk of the stellar sample in the Milky Way's outer halo measured by Rubin.

Other considerations also call for further optimism. In this study, we degrade the three-year *Kepler* data to mimic the Rubin Observatory data, but Rubin's 10-year baseline is 2.5 times longer. In addition, Rubin will observe six different photometric bands. The multi-band time series data will contain other subtle information beyond what is in our simplistic mock with degraded monochromatic *Kepler*'s data.

4. Concluding Remarks

Astronomy today is a fundamentally different field than it was just a decade ago. The change has been powered by our ability to gather orders of magnitude larger data sets from ever more powerful instruments. But despite this rapid growth in data collection, well-labelled training data remains scarce in astronomy. With the limited training set, implementing neural networks with the relevant inductive bias is critical. As most information from astronomical observations, such as spectra and time series, is long-range by nature, we argue that the local receptive field of the more widely adopted CNNs approach is somewhat limiting.

Our study showed that Astroconformer leverages long-range information and outperforms CNNs in asteroseismology. Since long-range information is prevalent in most astronomical observations, apart from asteroseismology, the same architecture can easily be adapted and might also advance other domains in astronomy. Furthermore, the true power of Transformer lies in unsupervised learning, learning from all the unlabelled data, beyond only those with supervised

labels, a subject which we will explore next. Just like the revolution Transformer has sprung across all studies in computer vision and natural language processing, Transformer-based models might prove important to take Astronomy x Machine Learning to the next level.

References

Benomar, O., Appourchaux, T., and Baudin, F. The solar-like oscillations of hd49933: A bayesian approach. *http://dx.doi.org/10.1051/0004-6361/200911657*, 506, 10 2009. doi: 10.1051/0004-6361/200911657.

Bo, P. Blinkdl/rwkv-lm: 0.01, August 2021. URL https://doi.org/10.5281/zenodo.5196577.

Chaplin, W. J. and Miglio, A. Asteroseismology of Solar-Type and Red-Giant Stars. ARA&A, 51(1):353–392, August 2013. doi: 10.1146/annurev-astro-082812-140938.

Dhanpal, S., Benomar, O., Hanasoge, S., Kundu, A., Dhuri, D., Das, D., and Kaul, B. Measuring frequency and period separations in red-giant stars using machine learning. *The Astrophysical Journal*, 928(2):188, apr 2022. doi: 10.3847/1538-4357/ac5247. URL https://doi.org/10.3847%2F1538-4357%2Fac5247.

Gruberbauer, M., Kallinger, T., Weiss, W. W., and Guenther, D. B. On the detection of lorentzian profiles in a power spectrum: a bayesian approach using ignorance priors. A&A, 506(2):1043–1053, aug 2009. doi: 10.1051/0004-6361/200811203. URL https://doi.org/10.1051%2F0004-6361%2F200811203.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., and Pang, R. Conformer: Convolution-augmented transformer for speech recognition, 2020. URL https://arxiv.org/abs/2005.08100.

Hon, M., Stello, D., and Yu, J. Deep learning classification in asteroseismology using an improved neural network: results on 15 000 kepler red giants and applications to k2 and TESS data. *Monthly Notices of the Royal Astronomical Society*, 476(3):3233–3244, feb 2018. doi: 10.1093/mnras/sty483. URL https://doi.org/10.1093%2Fmnras%2Fsty483.

Hoover, B., Strobelt, H., and Gehrmann, S. exbert: A visual analysis tool to explore learned representations in transformers models, 2019. URL https://arxiv.org/abs/1910.05276.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Gilmore, D. K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby, S. H., Jain, B., Jammes, F., Jee, M. J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Jones, R. L., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Krughoff, K. S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Langton, J. B., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J., Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., Mills, D. J., Miraval, C., Moeyens, J., Moolekamp, F. E., Monet, D. G., Moniez, M., Monkewitz, S., Montgomery, C., Morrison, C. B., Mueller, F., Muller, G. P., Muñoz Arancibia, F., Neill, D. R., Newbry, S. P., Nief, J.-Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S. S., Olsen, K., O'Mullane, W., Ortiz, S., Osier, S., Owen, R. E., Pain, R., Palecek, P. E., Parejko, J. K., Parsons, J. B., Pease, N. M., Peterson, J. M., Peterson, J. R., Petravick, D. L., Libby Petrick, M. E., Petry, C. E., Pierfed-

erici, F., Pietrowicz, S., Pike, R., Pinto, P. A., Plante, R., Plate, S., Plutchak, J. P., Price, P. A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A. P., Regnault, N., Reil, K. A., Reiss, D. J., Reuter, M. A., Ridgway, S. T., Riot, V. J., Ritz, S., Robinson, S., Roby, W., Roodman, A., Rosing, W., Roucelle, C., Rumore, M. R., Russo, S., Saha, A., Sassolas, B., Schalk, T. L., Schellart, P., Schindler, R. H., Schmidt, S., Schneider, D. P., Schneider, M. D., Schoening, W., Schumacher, G., Schwamb, M. E., Sebag, J., Selvy, B., Sembroski, G. H., Seppala, L. G., Serio, A., Serrano, E., Shaw, R. A., Shipsey, I., Sick, J., Silvestri, N., Slater, C. T., Smith, J. A., Smith, R. C., Sobhani, S., Soldahl, C., Storrie-Lombardi, L., Stover, E., Strauss, M. A., Street, R. A., Stubbs, C. W., Sullivan, I. S., Sweeney, D., Swinbank, J. D., Szalay, A., Takacs, P., Tether, S. A., Thaler, J. J., Thayer, J. G., Thomas, S., Thornton, A. J., Thukral, V., Tice, J., Trilling, D. E., Turri, M., Van Berg, R., Vanden Berk, D., Vetter, K., Virieux, F., Vucina, T., Wahl, W., Walkowicz, L., Walsh, B., Walter, C. W., Wang, D. L., Wang, S.-Y., Warner, M., Wiecha, O., Willman, B., Winters, S. E., Wittman, D., Wolff, S. C., Wood-Vasey, W. M., Wu, X., Xin, B., Yoachim, P., and Zhan, H. LSST: From Science Drivers to Reference Design and Anticipated Data Products. ApJ, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

Koch, D. G. et al. Kepler Mission Design, Realized Photometric Performance, and Early Science. ApJ, 713(2):L79–L86, April 2010. doi: 10.1088/2041-8205/713/2/L79.

Mathur, S. et al. Revised stellar properties of kepler targets for the q1-17 (DR25) transit detection run. *The Astrophysical Journal Supplement Series*, 229(2):30, mar 2017. doi: 10.3847/1538-4365/229/2/30. URL https://doi.org/10.3847/1538-4365/229/2/30.

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., and Villasenor, J. Transiting Exoplanet Survey Satellite (TESS). In Oschmann, Jacobus M., J., Clampin, M., Fazio, G. G., and MacEwen, H. A. (eds.), *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, volume 9143 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 914320, August 2014. doi: 10.1117/12.2063489.

Sayeed, M. et al. The swan: Data-driven inference of stellar surface gravities for cool stars from photometric light curves. *The Astronomical Journal*, 161(4): 170, Mar 2021. ISSN 1538-3881. doi: 10.3847/1538-3881/abdf4c. URL http://dx.doi.org/10.3847/1538-3881/abdf4c.

Su, J., Lu, Y., Pan, S., Wen, B., and Liu, Y. Roformer: Enhanced transformer with rotary position embedding, 2021. URL https://arxiv.org/abs/2104.09864.

Tipirneni, S. and Reddy, C. K. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series, 2021. URL https://arxiv.org/abs/2107.14293.

Toutain, T. and Appourchaux, T. Maximum likelihood estimators: an application to the estimation of the precision of helioseismic measurements. A&A, 289:649–658, September 1994.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need, 2017. URL https://arxiv.org/abs/1706.03762.

Vig, J. A multiscale visualization of attention in the transformer model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 37–42, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-3007. URL https://www.aclweb.org/anthology/P19-3007.

Yu, J. et al. Asteroseismology of 16,000 kepler red giants: Global oscillation parameters, masses, and radii. *The Astrophysical Journal Supplement Series*, 236(2): 42, Jun 2018. ISSN 1538-4365. doi: 10.3847/1538-4365/aaaf74. URL http://dx.doi.org/10.3847/1538-4365/aaaf74.

Zhao, G., Zhao, Y., Chu, Y., Jing, Y., and Deng, L. Lamost spectral survey, 2012. URL https://arxiv.org/abs/1206.3569.

Zhu, Y. et al. Self-Attentive Speaker Embeddings for Text-Independent Speaker Verification. In *Proc. Interspeech 2018*, pp. 3573–3577, 2018. doi: 10.21437/Interspeech.2018-1158.