Foreword to the Focus Issue on Machine Intelligence in Astronomy and Astrophysics

Giuseppe Longo[1], Erzsébet Merényi[2], and Peter Tiňo[3]

[1] Department of Physics, University Federico II in Napoli, via cintia, Napoli I-80126, Italy; longo@na.infn.it
[2] Department of Statistics, Rice University, Houston, Texas, USA; erzsebet@rice.edu
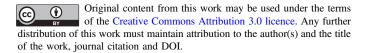[3] School of Computer Science, The University of Birmingham, Birmingham, UK; P.Tino@cs.bham.ac.uk

1. The Need for Machine Learning in Astronomy

Astronomical observations already produce vast amounts of data through a new generation of telescopes: Atacama Large Millimeter Array (ALMA), Jansky VLA, and through large surveys, e.g., the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Zwicky Transient Facility (ZTF; Bellm 2014), Pan-STARRS (Kaiser et al. 2010), VLT Survey Telescope-VST, and many others that cannot be analyzed manually. Next-generation telescopes such as the Large Synoptic Survey Telescope (LSST; Ivezic et al. 2008) and the Square Kilometer Array (SKA; Dewdney et al. 2009) are planned to become operational in this decade and the next, and they will increase the data volume by many orders of magnitude. The increased spatial, temporal, and spectral resolution afford a powerful magnifying lens on the physical processes that underlie the data, but at the same time, generate unprecedented complexity that is hard to exploit for knowledge extraction. It is therefore imperative to develop machine intelligence—machine learning (ML) in particular—suitable for processing the amount and variety of astronomical data that will be collected and which is capable of answering scientific questions based on the data (Baron 2019).

Astronomical data exhibit the usual challenges associated with big data such as immense volumes; high dimensionality; and missing or highly distorted observations. In addition, astronomical data can exhibit large continuous observational gaps, very low signal-to-noise ratios, and the need to distinguish between true missing (i.e., noncollected) data and nondetections (i.e., due to upper limits). There are strict laws of physics behind the data production that can be assimilated into ML mechanisms to improve over general off-the-shelf, state-of-the-art methods. An additional peculiarity is that these large and heterogeneous data sets (Pesenson et al. 2010) need to be simultaneously queued, merged, and mined by many independent groups of researchers, posing problems uncommon in many other application domains. In this context, it is important to mention the crucial role played by the International Virtual Observatory Alliance (IVOA; Arviset & Gaudet 2011), which aims at seamless access to astronomical data and which focuses on the standardization of data and metadata; data exchange methods; and the implementation of a registry that lists available services and what can be done with them.

Significant progress in the face of these challenges can be achieved only via the new discipline of astroinformatics (Brescia et al. 2017), a symbiosis of diverse disciplines such as ML; probabilistic modeling; astronomy and astrophysics; statistics; distributed computing; and natural computation. The importance of the task resulted in the emergence of (often large) interdisciplinary collaborations. Representing 69 authors from 15 countries and six continents, this focus issue offers a sample of recent progress enabling scientific discoveries using ML.

Machine intelligence provides capabilities for discovering intricate relations and extracting information, making accurate inference from complex multidimensional data; it also has the capability of processing large amounts of data quickly. This collection of papers focuses primarily on machine intelligence's former capability, although its latter speed considerations also surface in most works. A noteworthy point is that the majority of the ML methods in these papers are artificial neural networks (ANNs), including multi-layer perceptrons with back propagation (MLP/BP), self-organizing maps (SOMs), and deep-learning networks (convolutional and recurrent convolutional neural nets). This confirms a trend already noted by many: namely the slow adoption by the astronomical community of less well-known, but possibly more effective, ML methods than those that have traditionally been used. In the recent past, most works focused only on supervised classification/regression methods applied to a handful of problems, mainly due to the

lack of suitably large knowledge bases (i.e., template data sets) on which to train the models. Instead, the broad spectrum of applications presented in this special issue shows that ML methods are becoming widely used and that new applications arise every time a new data set is made publicly available. In what follows, we briefly summarize the papers contained in this focus issue, grouping them by the astronomical problems targeted, with the aim of emphasizing the spectrum of ML applications.

2. Astronomical Applications of ML in this Issue

Classification of astronomical sources is one of the central problems addressed in this issue. The respective approaches can be divided into two broad groups. One group involves ML methods that aim to reproduce, on large data sets of objects, predefined classification schemes specified on subjective grounds by human experts (supervised methods). The other group uses methods aimed at obtaining an unbiased classification based on the statistical properties of the objects (unsupervised clustering methods). Both approaches have their pros and cons, and both are useful in their own terms.

1. In *Deep Learning for Image Sequence Classification of Astronomical Events*, Carrasco-Davis et al. (2019) propose a novel approach to the classification of variable stars and transients (and, in general, of astronomical objects in large data streams). They bypass the traditional preprocessing of temporal sequences of images for extracting features—often light curves further reduced to parameters derived from the light curves—that allows the classification of much lower-dimensional signals than the image sequence. Instead, they train a recurrent convolutional neural network (RCNN) to extract the relevant features (latent variables) directly from the raw data, then use the latent feature vector as input to the classifier component of the network. This allows automatic, data-driven feature extraction without the need of computing difference images, making model assumptions, and corrections throughout the various steps; it also helps avoid errors that may be introduced by computing light curves. The authors also describe a method to simulate synthetic image sequences on the basis of the instrument and observation characteristics of a given experiment (the HiTS survey in their example), for the purpose of training the RCNN (which requires a large number of labeled samples). The RCNN can subsequently be fine-tuned to the distribution of data actually measured from the assumed experiment by using a small number of the real labeled samples.

A more traditional approach to classification problems, based on morphological features, is the topic of several papers.

1. *Multiband Galaxy Morphologies for CLASH: A Convolutional Neural Network Transferred from CANDELS* by Perez-Carrasco et al. (2019) uses a convolutional neural network (CNN) to classify galaxies into broad morphological classes: spheroid, disk, irregular, point source, or unclassifiable. The authors use a large image collection of over 8400 galaxies from the CLASH in 16 photometric bands, from ultraviolet to nearinfrared. A big challenge in such studies is the knowledge of the "truth" labels attached to the galaxy images and used in the processes of training and evaluating the classifiers, because the morphological labeling provided by humans may not be perfect. The authors take great care to address this issue. In particular, five experts were used to provide labels for 100 randomly selected galaxies in each of the 16 filters. This subset of galaxies with highly reliable labels is then used to evaluate the classifier performance in greater detail and to fine-tune the deep network.

2. SOMs play the key role in *Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing Maps* by Galvin et al. (2019) to facilitate automatic labeling (classification) of a large number of galaxies according to their morphological complexity. A topological feature map of RGZ (Banfield et al. 2015) images is learned, and the SOM prototypes (weight vectors) are subsequently labeled according to the plurality of the labeled samples in their receptive fields. The labeled training samples are objects that have high consensus among citizen scientists who labeled the RGZ images. The remaining citizen-scientist-labeled images are then mapped to the learned SOM prototypes to see how they align with the topology learned from the high-consensus labeled images, as well as to classify them according to their best-matching SOM prototypes. For the labeling/classification of RGZ objects, they use SOM heat maps as input to a quantile regression forest (QRF), where a heat map is the matrix of similarities between an input image and each of the SOM prototypes. The QRF then performs a probabilistic assignment into classes. One noteworthy aspect is that for the winner selection in SOM learning a modified Euclidean distance is utilized, in the framework of the PINK package, which incorporates affine transformations. This accounts for rotated versions of source images, providing rotational invariance in a more elegant way than the customary generation and storage of transformed images.

3. In *Radio Galaxy Zoo: Unsupervised Clustering of Convolutionally Auto-encoded Radio-astronomical Images* by Ralph et al. (2019) also use SOMs to cluster radio astronomy images from the RGZ. The paper is part of the ongoing preparation for the future SKA radio surveys and is focused on demonstrating that a very much needed increase in computational efficiency (speed) can be achieved via intelligent compression of the data fed to the SOM. In their work, they use compressed versions of the raw images (vectors of extracted latent variables) obtained by a (supervised) CNN autoencoder, which proves capable of preserving the relevant information for an accurate representation of the manifold structure by the SOM. A very interesting

aspect of this paper is that SOM knowledge is also viewed as a way to transform a discrete (hard) classification task into a continuous regression problem, which could open the way to a more physical classification of the radio morphologies. This aspect is subject of future works by the same group.

4. In a similar line of work, *Unsupervised Classification of Galaxies. I. Independent Component Analysis Feature Selection*, Chattopadhyay et al. (2019) tackle the problem of categorizing galaxies using an unsupervised approach—K-means clustering —for over 300,000 galaxies from the Value Added Galaxy Catalog (VAGC), and utilizing ICA feature selection prior to clustering to reduce the dimensionality of the data. Importantly, the authors use 49 measurable physical attributes including photometry, spectroscopy, morphology, chemical composition, and kinematics, but no non-observable features. Their results are very encouraging, as they succeed in identifying clusters of objects with different Sérsic index, active or non-active, and with different star formation history, etc., thus paving the way toward a physically based classification scheme. We also note that unsupervised approaches do not suffer from the label biases always present in attempts to reproduce galaxy classification using supervised methods, and are therefore more likely to be able to capture—discover—physical phenomena (such as the presence of an active galactic nucleus (AGN)) that are often missed by traditional morphology.

Often, in huge collections of data, one would like to discover unexpected phenomena that are not aligned with the general trends in the data. This is known as "novelty" or "anomaly" detection.

1. In *Identifying Complex Sources in Large Astronomical Data Sets Using a Coarse-grained Complexity Measure* Segal et al. (2019) explore the possibility of finding such phenomena using a detector based on a complexity measure applied to data items. In this context, data items with unusual complexity values will stand out. In particular, the authors introduce a measure on image data, termed apparent complexity, that is based on a combination of information theoretic considerations (entropy) and a smoothing function to filter out random elements in the images. This measure can be used to segment and identify interesting observations in huge data repositories by quantifying their morphological complexity. This method is demonstrated on the Australia Telescope Large Area Survey (ATLAS) to distinguish between images of galaxies with simple and complex morphologies.

The evaluation of photometric redshifts is among the most common applications of ML methods in astrophysics. The number of applications is expected to drastically increase in the coming years, when new photometric and spectroscopic surveys will produce accurate and large data sets of precisely measured parameters to be used as templates. At the moment, however, photometric redshifts have become a sort of benchmark for comparing different methods and procedures. Two papers address the problem of obtaining reliable redshift estimates for the radio-selected samples that will be the main target of future surveys such as the Evolutionary Map of the Universe (EMU) to be performed with SKA. These papers use regression—a supervised technique where the target variables are continuous—such as in the case of photometric redshifts (Tagliaferri et al. 2002; Salvato et al. 2019 and references therein), and of stellar formation rates (Stensbo-Smidt et al. 2017; Delli Veneri et al. 2019) in galaxies. This is in contrast to discrete target variables, often called labels, in supervised classification. Both works address a problem commonly affecting all supervised methods, but so far little explored in this context: how performances are affected by biases in the construction of the training sets.

1. In *A Comparison of Photometric Redshift Techniques for Large Radio Surveys*, Norris et al. (2019) study ML techniques—k-nearest neighbors (kNN), random forests (RF) and a multi-layer perceptron with quasi Newton algorithm for optimization (MLPQNA)—for photometric redshift estimation for radio sources, in comparison with a classic template-fitting approach (Le Phare, Arnouts & Ilbert 2011). The study targets the scenario near-future surveys present with tens of millions of radio sources, for which redshift determination is essential but the general lack of spectroscopic redshift measurements (only available for a fraction of the sources) necessitates estimation of the redshift from photometric measurements (available for all sources). The comparative analyses are performed in a multidimensional parameter space that includes low and high ranges of redshift values; quality of photometric data; photometry bands; depth of photometric survey; completeness of training data; stratification by type of source; and more, providing advice about the relative benefits of the techniques under various circumstances.

2. In *Preliminary Results of Using k-nearest-neighbor Regression to Estimate the Redshift of Radio-selected Data Sets*, Luken et al. (2019) use kNN regression to explore a different aspect of the problem. They start from a data set obtained from the fusion of different surveys covering a large wavelength range, from the far infrared to the optical, and degrade it to match the expected shallowness of EMU. Their results confirm the need for a training set providing a good coverage of the parameter space, and provide a strong evidence that there is no need for the training and test set to be extracted from the same region of the sky, thus opening the way to the possibility of using deep, pencil-beam surveys to complement spectroscopic data obtained from shallow wide field surveys.

Machine learning can be a useful tool in the model-based exploration of astronomical data.

1. A nice example of this scenario is *Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes* by Hendriks & Aerts (2019). The authors use a deep-learning network to interpolate between nodes of a huge grid of

3

stellar evolution models. The grid of models is realized as a grid in the parameter space that can be high dimensional, prohibiting any fine grid-based exploration of the model space. The deep network generalizes to models between the grid points and can thus help to adequately parameterize forward asteroseismic modeling. Otherwise, we would be restricted to models corresponding to the (possibly quite sparse) set of grid points. The deep network is trained on a grid of astero-seismological models (parameters) of intermediate- and high-mass stars to predict the frequencies of their coherent oscillation modes. The authors employ a genetic algorithm to identify appropriate regions for the stellar parameters based on the oscillation mode frequencies.

An entirely different topic, adaptive optics (AO), is the subject of one paper.

1. *Experience with Artificial Neural Networks applied in Multi-object Adaptive Optics* by Suárez Gómez et al. (2019) addresses AO with ANNs, specifically MLPs with BP. The paper reviews, in the context of multi-object AOs, the state-of-the-art in artificial intelligence dominated by MLP applications, compared with more traditional methods for fast reconstruction of the wavefront of light after having been deformed by Earth's atmosphere. Simulated atmosphere and on-sky (real data) scenarios are discussed using the CANARY flexible AO demonstration bench and the Complex Atmospheric Reconstructor based on Machine LEarNing (CARMEN) framework involving MLP/BP ANNs. Relative merits of different ANN models for different atmospheric conditions are presented. Extensions of CARMEN to scales such as those of the European Extremely Large Telescope and the Thirty Meter Telescope are then elaborated including computation (recall) speed, aiming at near-real-time reconstruction; concluding with a next generation of CARMEN that further improves the reconstruction by using a CNN.

More general ML issues are targeted by Buchner (2019), and by Vilalta et al. (2019), namely, sampling and transfer learning. In the age of big data, model-based ML performed in the Bayesian framework is often pushed aside for pragmatic reasons. First, while the Bayesian framework is principled and allows for proper treatment of uncertainties (including, e.g., different types of noise models), it unfortunately is computationally hungry. Dealing with huge data sets and many models can quickly become infeasible. Second, it is often the case that point estimates of semi-parametric ML models, such as deep networks, perform empirically very well when trained on adequately large data sets, and are increasingly available in ready-to-use packages. This is fine as long as one's goal is classification of astrophysical objects of interest or predictions on such objects. This is, for example, the case in the paper by Pérez-Carrasco et al. (2019) in this issue, where the authors use a convolutional deep network to classify galaxies into a set of predefined morphological classes. However, we would often like to include, in learning from the data, prior knowledge which is naturally expressed as interpretable physical models. In that case, the use of deep networks can be problematic.

1. *Collaborative Nested Sampling: Big Data versus Complex Physical Models* by Buchner (2019) proposes an innovative and efficient method for sampling in the model space (thus enabling Bayesian framework) in the case of very large data sets. The main principle of this approach is that (especially in the initial stages) the sampling regions can be quite similar across similar data sets, so rejection sampling from the union of such contours can be performed.
2. Transfer learning is another topic of growing importance in ML, useful when expensive learning of a complex model is performed using a data set from one domain and then the model is exported to another. In *A General Approach to Domain Adaptation with Applications in Astronomy*, Vilalta et al. (2019) use the identification of Type Ia supernovae as astronomical template. This represents a problem of paramount importance for the emerging field of time-domain astronomy, and especially for the scientific exploitation of LSST data which is expected to produce 10^{6-7} transients per night, with only a small fraction of Type Ia supernovae among them. The authors show how to transfer learning from spectroscopic data to photometric data. The paper is innovative from the algorithmic point of view (a new approach to domain adaptation that does not depend on the proximity to the source nor on the target distribution) as well as for the astronomical implications: the possibility to increase performance and reduce computational costs in tackling complex problems in very large data sets.

3. Perspectives

In less than a decade, the application of ML methods to astronomical problems has become a widespread practice. This growth has been triggered mainly by the huge, comprehensive data sets provided by modern multi-band, multi-epoch digital sky surveys, as well as by the standardization of heterogeneous data induced by the creation of the Virtual Observatory. It is also clear that the most exciting innovations come from large interdisciplinary groups where the expertise of data scientists and domain experts are combined. In the near future, this kind of virtuous collaboration between experts from different disciplines will probably allow to overcome some of the main problems the astronomical community is currently facing.

One of these problems is that most supervised learning algorithms are not fine-tuned to deal with astronomical data sets, as most assume that all measured features are of the same quality, and that the labels are correctly reflecting the truth Baron (2019). However,

astronomical data are sparse and noisy and, in most cases, the targets in the training sets are derived by human experts, and are thus prone to inconsistencies. This implies that in most applications, supervised learning algorithms perform well when applied to high signal-to-noise ratio data sets, or to (the few) data sets characterised by uniform noise properties. It is therefore imperative to modify existing tools and/or to devise new algorithms, capable of taking into account uncertainties during the model construction and to predict uncertainties based on the intrinsic properties of the objects in the sample and on their measurement uncertainties (Cavuoti et al. 2016). While external domain knowledge and scientific judgment can be infused to weight the reliability of labels (as in the case of citizen-labels of the RGZ), classifiers whetted in other, closely related big data domains such as terrestrial and planetary hyperspectral imaging can be helpful for coping with data-dependent noise, labeling errors, and severely biased training sets (Merényi et al. 2014). Principled construction of classifiers that can handle label noise is increasingly studied in ongoing ML work, e.g., Bootkrajang & Kaban (2014), including incorporation of knowledge about the nature of the process that corrupts the labels, for enhancing the robustness of the classifier estimation.

Another issue that will prove crucial in future applications is knowledge transfer. Models trained on a specific data set usually fail to generalize when applied to another data set with even slightly different properties. For the exploitation of future surveys it will be unthinkable to train algorithms on the final data releases; therefore, algorithms will need to be trained while the data are still being acquired. Transfer learning, which is widely discussed in the ML literature but still seldom applied in astronomy, will facilitate this.

A third aspect that needs to be considered is that the performance of all supervised learning algorithms strongly depends on the number of features that are selected, as most are either not designed to work with hundreds or thousands of features (they will break down mathematically), or an excessive number of features with redundant information decreases the discrimination capability of the algorithm. Traditionally, dimensionality reduction was achieved relying on the choices by experts, but this approach is now being superceded by automatic techniques (D'Isanto et al. 2018; Delli Veneri et al. 2019), and by exhaustive, albeit computationally intensive, approaches (Polsterer et al. 2014). Alternatively, ML can also be applied to derive, for a given classification goal in a many-class scenario in high-dimensional feature spaces, the relative importance of the measured features (Mendenhall & Merényi 2008). In contrast to methods that return reduced (latent) features in transformation spaces (e.g., PCA or deep convolutional networks), this provides (nonlinear) dimensionality reduction in the original feature space, offering a readily interpretable subset of the original physical features. Moreover, computationally efficient algorithms for feature selection are badly needed to cope with the huge data volumes and data streams which are foreseen for the near future.

An extremely important resource for scientific inquiry in astronomy and astrophysics, not represented in this issue, is spectral image data with hundreds or thousands of image planes. These "deep data" are increasingly available from advanced telescope systems, ALMA currently representing the frontier. While "wide" sky surveys such as LSST cover large segments of the sky with high cadence in a few broad optical bands, ALMA and VLA data provide immensely rich compositional, kinematic, and other information for targeted objects such as protoplanetary disks or molecular clouds, offering high potential for discovery. ALMA and VLA also image objects in multiple molecular lines which can be combined (stacked) for further discovery power. However, current standard analysis tools fall short of fully exploiting these data, even from a single molecular line. The large number of image bands (large feature dimension with substantial frequency gaps), complex, irregular cluster structure inherent in such data exceed the capabilities of many unsupervised (clustering, anomaly detection) ML methods. Recent collaborative work that successfully extracts physical phenomena using all channels from multiple stacked ALMA lines of a protoplanetary disk (Merényi et al. 2017) is spearheading developments in this direction.

Finally, we want to emphasize that interpretability is a crucial requirement in many ML applications. One possibility of achieving it is through a creative blending of model-based approaches with data-driven automated learning. For example, one can impose a latent flux from the common source (quasar) as an integral part of the overall machine learning model for resolving the delay in gravitationally lensed images (Cuevas-Tello et al. 2010; Otaibi et al. 2016). Another example is the imposition of a smooth latent low-dimensional structure in the space of interpretable models (as opposed to formulating it in the data space) to discover the pivotal degrees of freedom with physical meaning buried in the observed data (Gianniotis & Tino 2008; Gianniotis et al. 2009). It is important to strive to construct learning methods that, while learning in a data driven manner, are able to provide scientifically interpretable and physically meaningful predictive outputs. A very active line of this research aims at making deep neural networks interpretable (see, e.g., Montavon et al. 2018).

This focus issue—as far as we know—is a first of its kind to appear in an astronomical journal, which posed special challenges. The authors are to be commended for a great effort to make their work accessible to both the astronomy and the machine learning communities while striving to follow the standards of both. We hope that this relatively small collection will encourage similar interdisciplinary issues soon.

Giuseppe Longo, Erzsébet Merényi, Peter Tiňo
Guest Editors

References

Arnouts, S., & Ilbert, O. 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, record, ascl:1108.009

Arviset, C., & Gaudet, S. 2011, arXiv:1106.0291

Banfield, J. K., Wong, O. I., Willett, K. W., et al. 2015, MNRAS, 453, 2326

Baron, D. 2019, arXiv:1904.07248

Bellm, E. 2014, in The Third Hot-wiring the Transient Universe Workshop, ed. P. R. Wozniak et al. (Santa Fe, NM), 27

Bootkrajang, J., & Kaban, A. 2014, Pattern Recognition, 47, 3641

Brescia, M., Djorgovski, S. G., Feigelson, E., Longo, G., & Cavuoti, S. 2017, in Proc. IAU Symp. 325 (Cambridge: Cambridge Univ. Press)

Buchner, J. 2019, PASP, 131, 108005

Carrasco-Davis, R., Cabrera-Vives, G., Förster, F., et al. 2019, PASP, 131, 108006

Cavuoti, S., Amaro, V., Brescia, M., et al. 2016, MNRAS, 465, 1959

Chattophadhyay, T., Fraix-Burnet, D., & Mondal, S. 2019, PASP, 131, 108010

Cuevas-Tello, J. C., Tino, P., Raychaudhury, S., Yao, X., & Harva, M. 2010, Pattern Recognition, 43, 1165

D'Isanto, A., Cavuoti, S., Gieseke, F., & Polsterer, K. L. 2018, A&A, 616, A97

Delli Veneri, M., Cavuoti, S., Brescia, M., Longo, G., & Riccio, G. 2019, MNRAS, 486, 1377

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. 2009, IEEEP, 97, 1482

Galvin, T. J., Huynh, M., Norris, R. P., et al. 2019, PASP, 131, 108009

Gianniotis, N., & Tino, P. 2008, ITNN, 19, 1468

Gianniotis, N., Tino, P., Spreckley, S., & Raychaudhury, S. 2009, Artificial Neural Networks ICANN 2009 (Berlin: Springer), 567

Hendriks, L., & Aerts, C. 2019, PASP, 131, 108001

Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E

Luken, K. J., Norris, R. P., & Park, L. A. F. 2019, PASP, 131, 108003

Mendenhall, M., & Merényi, E. 2008, ITNN, 19, 658

Merényi, E., Farrand, W. H., Taranik, J. V., & Minor, T. B. 2014, Journal on Advances in Signal Processing, 2014:71 doi:10.1186/1687-6180-2014-71

Merényi, E., Taylor, J., & Isella, A. 2017, in IAU Symp. 325, Astroinformatics (Cambridge: Cambridge Univ. Press), 281

Montavon, G., Samek, W., & Miller, K.-R. 2018, DSP, 73, 1

Norris, R. P., Salvato, M., Longo, G., et al. 2019, PASP, 131, 108004

Otaibi, S. A. L., Tino, P., Cuevas-Tello, J. C., Mandel, I., & Raychaudhury, S. 2016, MNRAS, 459, 573

Pérez-Carrasco, M., Cabrera-Vives, G., Martinez-Marin, M., et al. 2019, PASP, 131, 108002

Pesenson, M. Z., Pesenson, I. Z., & McCollum, B. 2010, AdAst, 2010, 350891

Polsterer, K. L., Gieseke, F., Igel, C., & Goto, T. 2014, ASP Conf. Ser. 485, Improving the Performance of Photometric Regression Models via Massive Parallel Feature Selection, in Astronomical Data Analysis Software and Systems XXIII, ed. N. Manset & P. Forshay (San Francisco, CA: ASP), 425

Ralph, N. O., Norris, R. P., Fang, G., et al. 2019, PASP, 131, 108011

Salvato, M., Ilbert, O., & Hoyle, B. 2019, NatAs, 3, 212

Segal, G., Parkinson, D., Norris, R. P., & Swan, J. 2019, PASP, 131, 108007

Stensbo-Smidt, K., Gieseke, F., Igel, C., Zirm, A., & Pedersen, S. 2017, MNRAS, 464, 2577

Suárez Gómez, S. L., González-Gutiérrez, C., Díez, A. E., & Santos, J. D. 2019, PASP, 131, 108012

Tagliaferri, R., Longo, G., Andreon, S., et al. 2002, arXiv:astro-ph/0203445

Vilalta, R., Dhar Gupta, K., Boumber, D., & Meskhi, M. M. 2019, PASP, 131, 108008

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579