ARABLE

Arable Mark 2 Core Measurements

An Accuracy Comparison

December 2020

arable.com

TABLE OF CONTENTS

INTRODUCTION

Arable is a data and analytics company that powers better decisions in agriculture through an integrated approach to hardware, software, and data science. At the center of Arable's solution is the Arable Mark 2, an all-in-one weather station and crop monitor that collects climate and plant data for actionable insights in all growing conditions. The device features a rugged design for field durability, global cellular connectivity, and a robust sensor suite that measures over 40 plant and climate data streams. It includes an acoustic disdrometer for measuring rainfall, as well as upwelling and downwelling shortwave radiometers, longwave radiometers, 6-band spectrometers, and GPS.

The design and hardware components of the Mark 2 were chosen with an eye towards accuracy and durability in the field, but also with an understanding that the core measurements—as well as any derived agronomic features—would be enhanced by applying machine learning (ML) solutions. Machine learning falls within the field of data science and broadly refers to the study of computer algorithms that build mathematical models based on sample data, known as "training data," in order to make predictions or decisions [1]. In particular, Arable's use of ML focuses on the calibration of core measurements to improve accuracy as well as leveraging predictive analytics for relevant agronomic features.

This paper details how our ML solutions improve core measurements and how we validate their performance in the field. This validation process entails co-locating Mark 2 devices with gold-standard, research-grade reference instruments located at AmeriFlux sites [2] or similar in North America, Europe, and Australia. These co-located deployments allow us to compare Mark 2 measurements with localized ground truth to establish true accuracy metrics within fields across a range of climate zones. We also co-locate with two commercial-grade weather stations, the Davis Vantage Pro2 [3] and METER ATMOS 41 [4], to determine how the Mark 2 compares to similar devices on the market.

In the Method section, we outline our process for training ML models and testing their performance using both the reference instruments and commercial-grade devices. In the Results section, we show sets of comparative plots and performance metrics for each of the core measurements and provide an interpretation of those results. In the Conclusion section, we summarize the results and discuss how the Mark 2 and overall Arable system will improve over time using our ML solutions.

METHOD

Arable's ML models require training data to build them, as well as testing data to validate their performance. As such, Arable has invested heavily in collecting vast amounts of this reference data: to date, Arable has collected over 70 million data points across 45 measurements at 35 different sites. These controlled research sites collectively make up our calibration/validation network, referred to as Arable's "Cal/Val network."

Through the Cal/Val network, Arable leverages in situ, high resolution, research-grade datasets for Mark 2 sensor calibrations and analytics. This effort started in late 2018 with the collection of air temperature and rainfall data and, in 2019, expanded to include measurements across the full sensor suite. The network currently covers 11 Köppen-Geiger climate zones [5] as shown in Figure 1 below. The locations were selected based on data quality, maximum geospatial spread, and ability to work with Arable. This network of high-quality data streams not only allows for regular calibration updates to the Mark 2 system, but also enables us to engineer sensor hardware with shorter development timelines, test new features, and jump-start novel analytics with third-party sensor integrations.

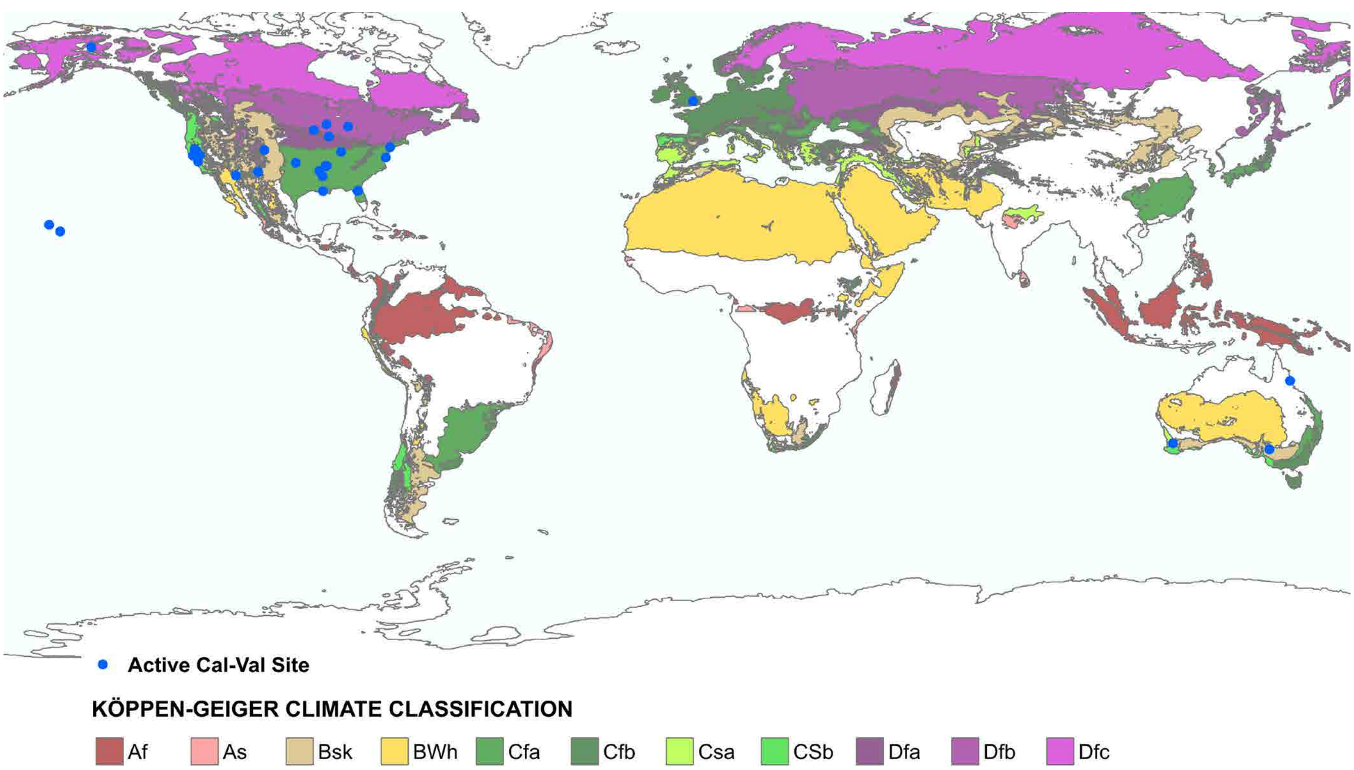


Figure 1. Map of Arable's Cal/Val network. Blue dots indicate the location of specific sites, with Köppen-Geiger climate zones differentiated by color.

The Cal/Val network is associated with a wide range of institutions that provide us with data sampled using high-accuracy, high-precision instruments. The NASA Goddard Space Flight Center [6] is a world-class research and development center in precipitation science. This is one of four sites where the Mark 2 is co-located with a laser disdrometer, providing access to radar and disdrometer precipitation data for rainfall calibration. The National Renewable Energy Laboratory (NREL) [7] is another world-class research site known for spectroradiometry. We have worked with NREL since 2017, validating our shortwave, longwave, and direction of incident light measurements. A large percentage of our partners are members of the AmeriFlux Network [2], which is instrumented with gold-standard sensors for climate and ecological research. Arable also partners with researchers managing meteorological towers from universities and other meteorological networks, including the University of Hawaii's Lyon Arboretum, University of Florida's Field and Fork, Rutgers University, and the Atmospheric and Oceanic Sciences department at the University of Wisconsin. The network spans four levels of specialized data:

(1) **Rainfall**. Drop size and velocity distributions and total rainfall measured with laser disdrometers.

(2) **Meteorology**. Temperature, humidity, pressure, rainfall, and wind at scientific-grade quality.

(3) **Spectroradiometry**. Broadband shortwave and longwave radiation, as well as highly resolved environmental spectrometry, from gold-standard sensors.

(4) **Eddy covariance measurements**. Whole-ecosystem exchange of water vapor (evapotranspiration), carbon dioxide (photosynthesis and respiration) and radiation (shortwave and longwave energy) along with weather drivers.

We use the data collected from these Cal/Val sites to align Mark 2 measurements with gold-standard reference measurements over time. This provides us with continuous data streams across multiple measurements, allowing us to identify any discrepancies between raw Mark 2 recordings and localized ground truth. This data is used to train ML models that "learn" the corrected or calibrated measurements coming from the gold-standard reference instruments. The ML models are then applied to incoming data in real time to obtain a more accurate and reliable core measurement base.

This process is iterative in the sense that these models are updated on a regular basis as we collect more data. As the models improve with more training data, they are able to discern more complex patterns and, in turn, produce predictions that are better and more accurate. This is the true power of ML: we can keep improving results by adding data to the models and releasing new software without having to update the hardware.

To ensure the highest level of accuracy, we not only use our Cal/Val network to build the models, but also to extensively test and confirm their continued performance. This paper provides a glimpse into the world of this validation process, with a focus on how the Mark 2 compares against other commercial weather station devices using the reference devices as a baseline for performance.

For the results discussed in this paper, we use data from two Cal/Val sites, which are named by location: Hilo (in Hawaii, USA) and Santa Rosa (in California, USA). At both of these sites, we use an OTT Parsivel² [8] as the gold-standard reference for precipitation. This is a high-quality laser disdrometer that counts water droplets and classifies them according to their diameter and velocity, thereby deriving a precise precipitation rate and amount. At the Santa Rosa site, we additionally use a Vaisala HUMICAP® HMP155 [9] with a MeteoShield (naturally aspirated helical radiation shield) [10] as the reference for air temperature and relative humidity, along with a Kipp & Zonen CNR4 net radiometer [11] as the reference for solar radiation. These are research-grade instruments that provide meteorological inputs for studies at the finest research facilities around the world.

In addition, we chose to co-locate with two commercial-grade weather stations, Davis Vantage Pro2 [3] and METER ATMOS 41 [4], at each of these sites. The Vantage Pro2 is a conventional weather station that provides some agronomically-relevant features such as evapotranspiration. The ATMOS 41 is a similarly-priced weather station that is one of several offerings from METER in their line of environmental solutions. At both sites, the devices are positioned at a distance of 1-2 meters apart from each other, at a two-meter height above ground level. They are cleaned regularly and maintained according to manufacturer specifications. In the following section, we provide an in-depth analysis of the performance of these devices across a variety of measurements.

RESULTS

This section compares the Arable Mark 2 to both the Davis Vantage Pro2 and METER ATMOS 41 across the following measurements:

Precipitation

Air temperature

Relative humidity/vapor pressure deficit, as a substitute for relative humidity since the ATMOS 41 does not provide relative humidity directly.

Three metrics are used to evaluate performance:

MAE: Mean Absolute Error

RMSE: Root Mean Square Error

MBE: Mean Bias Error

All results were computed on an hourly timescale, noting that for precipitation an accumulated percentage error is also computed to demonstrate how errors compound over time.

PRECIPITATION

In Figures 2 and 3 below, we show precipitation amounts for the Mark 2, Davis Vantage Pro2, and METER ATMOS 41 (y-axis) versus the gold-standard reference OTT Parsivel[2] (x-axis). Each point corresponds to an hourly value and the black identity line shows where the precipitation amounts would be equal (no error or discrepancy from the gold-standard reference). Points falling above or below that line correspond to instances of overestimation and underestimation of precipitation, respectively, noting that the further the point falls from the line, the larger the error. In contrast, a tight scatter centered about the line represents minimal errors and a close match to the gold-standard reference.

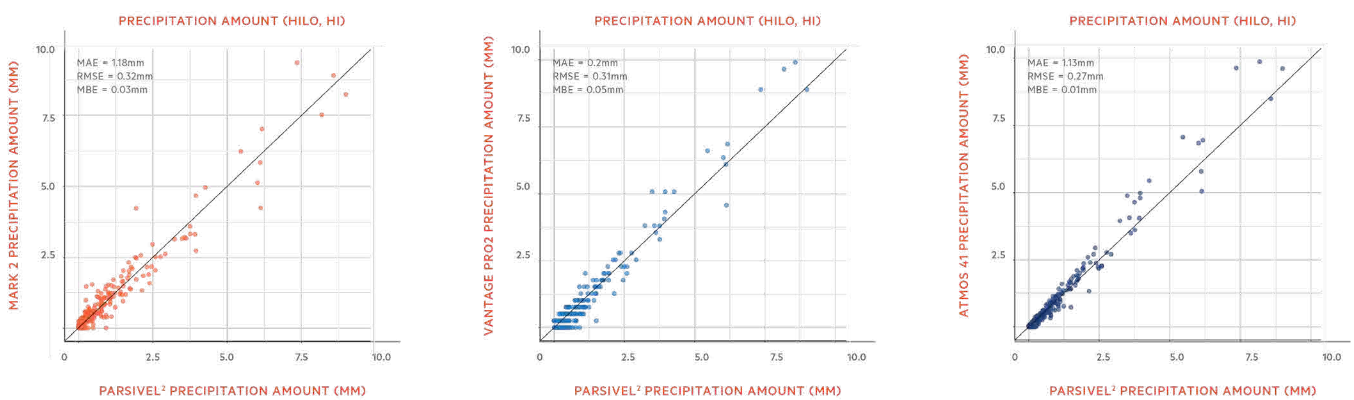


Figure 2. Precipitation comparison at the Hilo site using data from August 2020 through October 2020.

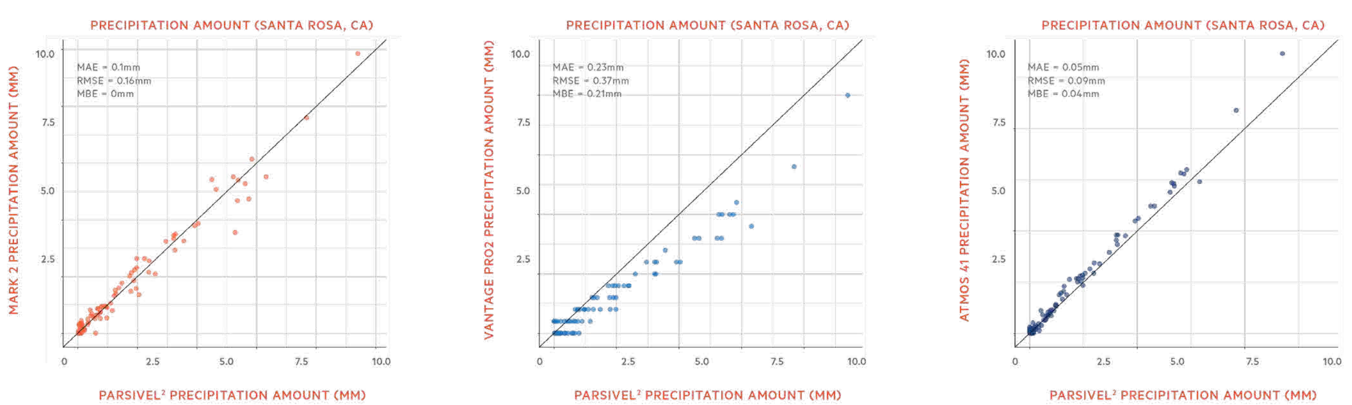


Figure 3. Precipitation comparison at the Santa Rosa site using data from April 2020 through June 2020.

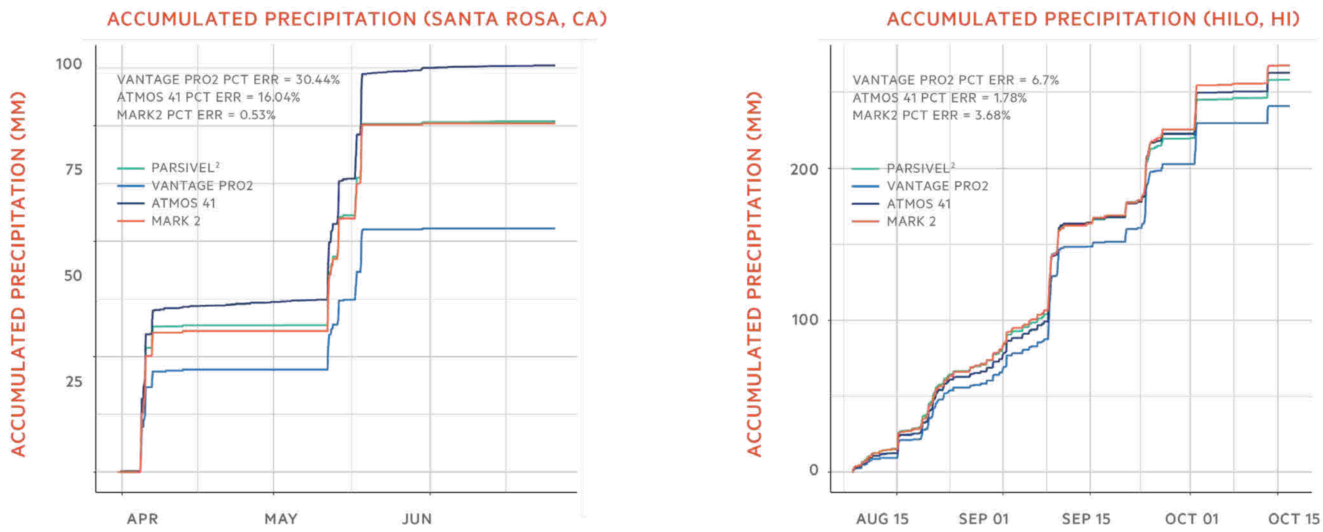


Figure 4. Accumulated precipitation comparison using the same data as in Figures 2 and 3.

As seen in Figure 3, the ATMOS 41 has the lowest hourly errors across all metrics with the exception of MBE at the Santa Rosa site, where the Mark 2 has zero bias on average. The Vantage Pro2 consistently underestimates, most evident at the Santa Rosa site where the data points are clearly off-center producing an MBE of -0.21 mm. In this context, MBE can be considered the most important metric since it is often of interest to understand how the errors compound over time. For this reason, we also plot the accumulated precipitation for the time frames studied, as shown in Figure 4 above.

The results of Figure 4 are consistent with those of Figure 3 in that the overall percentage errors match what we would expect given their corresponding MBEs. At the Santa Rosa site, the MBEs are ranked from best to worst as follows: Mark 2 (0 mm), ATMOS 41 (0.04 mm), and Vantage Pro2 (-0.21 mm); meanwhile, the percentage errors follow the same ranking: Mark 2 (0.53%), ATMOS 41 (16.04%), and Vantage Pro2 (30.44%). The results from Hilo can be broken down in a similar manner with the Mark 2 and ATMOS 41 swapping first and second place. At both sites, the Vantage Pro2 does not seem to be as accurate, due to its trend of underreporting.

This exemplifies some of the common issues that arise when measuring precipitation with traditional instrumentation. The Vantage Pro2 records precipitation using a conventional tipping bucket, which is one of the most common methods used today. However, even though widely used, this method—and rain gauges in general—are subject to many different sources of error, yielding amounts less than the actual precipitation reaching the ground by up to 30% or more, according to the World

Meteorological Organization (WMO) [12]. The magnitude of these errors are highly dependent on weather conditions, especially wind speed and precipitation intensity, and can be caused by [ibid.]:

- Systematic wind field deformation above the gauge orifice (2-10%)
- Wetting loss on the internal walls of the collector as well as wetting loss in the container when it is emptied (2-15% in summer and 1-8% in winter)
- Evaporation from the container (0-4%)
- In- and out-splashing of water (1-2%)
- Systematic mechanical and sampling errors
- Dynamic effects errors (5-15%)
- Random observational and instrumental errors

Arable's solution to measuring precipitation is an entirely novel approach that avoids some of the traditional sources of error described above. The Mark 2 uses a patented acoustic disdrometer that captures the sound of rainfall [13]. The process of deriving precipitation amounts can be broken into two main steps:

1. Audio data is collected and analyzed to determine whether the source of sound is rainfall, or something else.
2. If determined to be rainfall, the audio data is transformed such that individual droplets are identified, binned according to an energy range scheme, and then mapped to a corresponding diameter bin. From the distribution of droplets across diameter bins, we can derive a total precipitation amount.

Both steps in this process are enhanced by using classification and regression ML algorithms that enable us to more accurately and consistently arrive at the correct precipitation amount. Although these models have already achieved significant success, often outperforming other commercial-grade weather stations, they will continue to evolve and further improve over time by leveraging Arable's dynamic data and ML platform.

AIR TEMPERATURE

Figure 5 shows air temperature for the Arable Mark 2, Davis Vantage Pro2, and METER ATMOS 41 (y-axis) versus the reference Vaisala HUMICAP® HMP155 with helical radiation shield (x-axis).

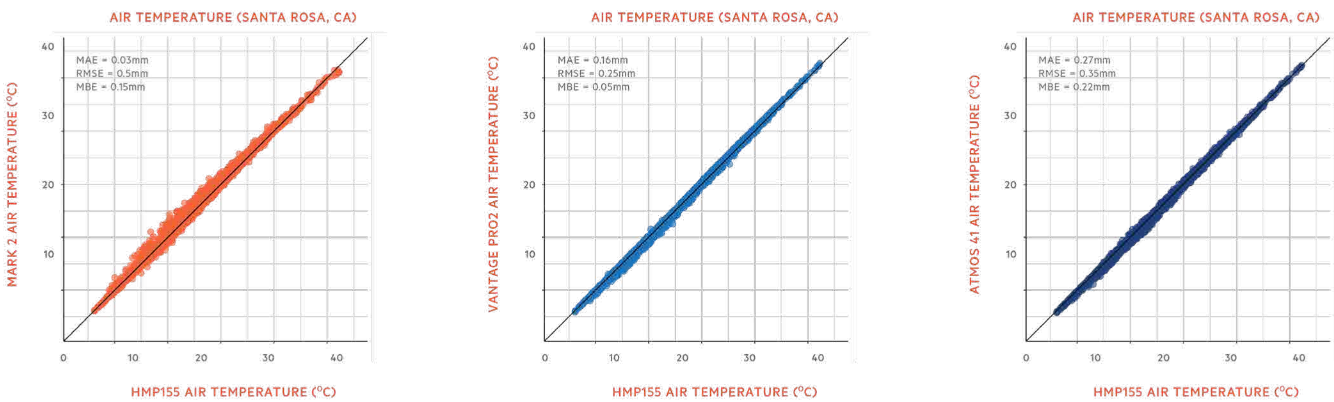
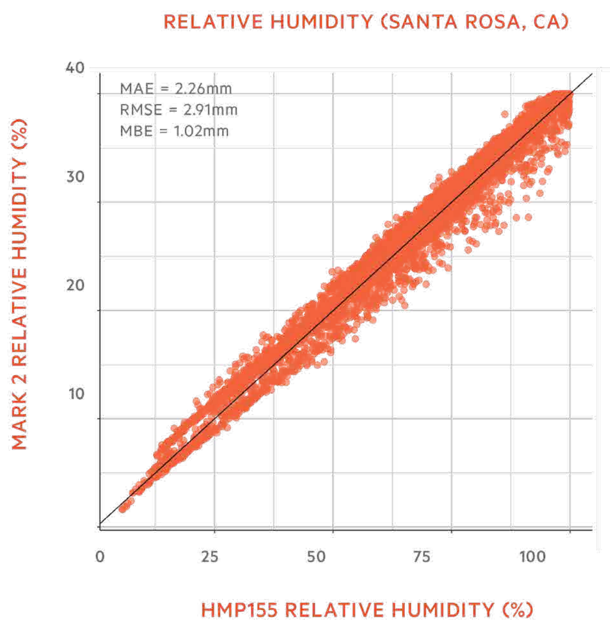


Figure 5. Air temperature comparison at the Santa Rosa site using data from October 2019 through October 2020.

As seen from Figure 5, all three devices perform very similarly across the three metrics, with MAE ranging from 0.16° C (Vantage Pro2) to 0.3° C (Mark 2). Notably, the 0.03° C difference between the Mark 2 and ATMOS 41 is negligible in terms of any physical significance. Furthermore, with all MAEs at or below 0.3° C, all devices exhibit sufficiently accurate temperatures, with errors that are unlikely to have a negative impact on agronomic applications. Also important to note is that at specific temperature ranges, these results can shift slightly. For example, at and near freezing temperatures, the Mark 2 MAE is only 0.14° C, while the Vantage Pro2 and ATMOS 41 MAEs are 0.17° C and 0.32° C, respectively.

RELATIVE HUMIDITY/ VAPOR PRESSURE DEFICIT

Figure 6 shows relative humidity for the Arable Mark 2 and Davis Vantage Pro2 (y-axis) versus the reference Vaisala HUMICAP® HMP155 with helical radiation shield (x-axis). As seen from the figure, the Mark 2 clearly outperforms the Vantage Pro2 with reduced errors that are up to 9x better. It should be noted that the maximum relative humidity value observed by the Vantage Pro2 is only 91%, even though this data covers an entire year and certainly contains some hours of rain.



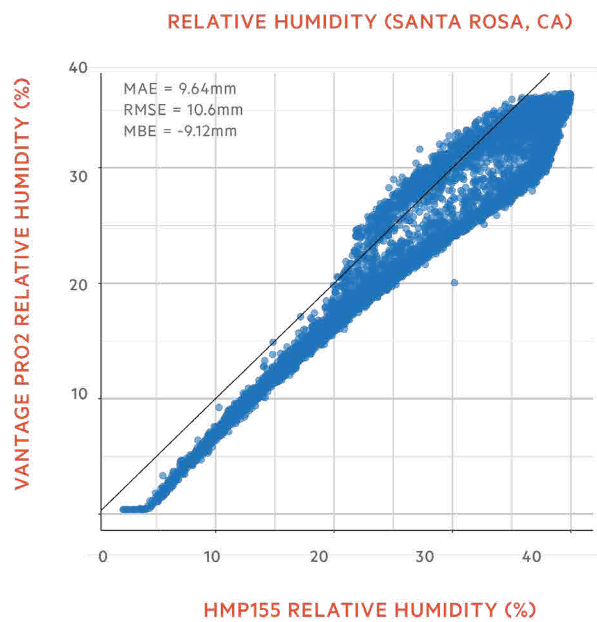


Figure 6. Relative humidity comparison at the Santa Rosa site using data from October 2019 through October 2020.

Figure 7 shows vapor pressure deficit for the Mark 2 and ATMOS 41 (y-axis) versus the reference Vaisala HUMICAP® HMP155 with helical radiation shield (x-axis). The ATMOS 41 does not provide relative humidity directly, but does provide vapor pressure deficit—a function of air temperature and relative humidity. Hence, this comparison is indirectly testing the accuracy of relative humidity, although notably the errors of air temperature will also be present. As evident from the figure, the Mark 2 slightly outperforms the ATMOS 41 with MBE deviating the most from -0.01 kPa (Mark 2) to -0.06 kPa (ATMOS 41).

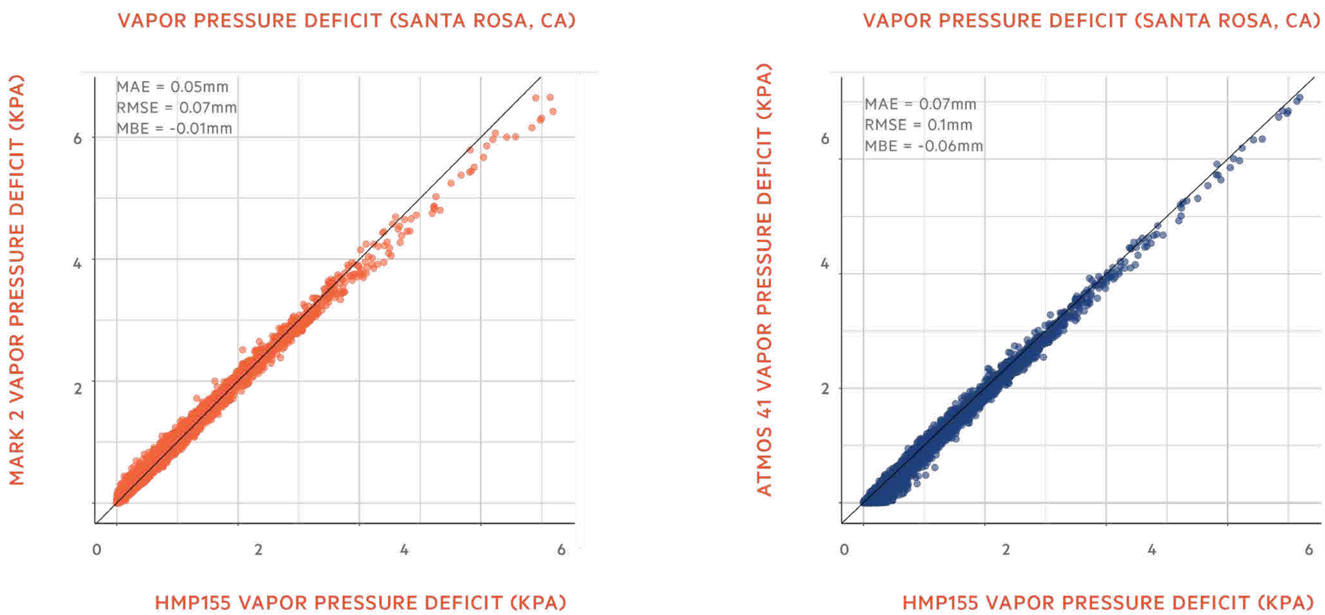


Figure 7. Vapor pressure deficit comparison at the Santa Rosa site using data from October 2019 through October 2020.

CONCLUSION

Based on the results outlined in this paper, it is clear that the Arable Mark 2 is a competitive alternative to traditional weather stations on the market in terms of data accuracy. The Mark 2 was shown to compare well against and sometimes outperform the Davis Vantage Pro2 and METER ATMOS 41 across air temperature, precipitation, and relative humidity/ vapor pressure deficit using gold-standard reference devices as a benchmark. In particular, the Mark 2 was the most accurate of the three within the freezing range of air temperature, across all relative humidity/ vapor pressure deficit measurements, and at one of the two rainfall sites studied. At this site, the ATMOS 41 and Vantage Pro2 lagged behind the Mark 2 with percentage errors that were roughly 32x and 60x worse, respectively. We attribute this success to our novel approach to measuring precipitation that avoids some of the typical sources of error that traditional methods are subject to.

As a leader in weather and crop monitoring technology and data analytics, Arable is committed to providing powerful, yet affordable tools to stakeholders across the agricultural spectrum. Our development cycles are not constrained to simply newer and more expensive hardware, but are also focused on regular software releases that contain automated ML model updates that maximize performance across all of our feature offerings. This is not only a more flexible and sustainable way of developing and building technologies, but it is less expensive and allows us to produce more affordable products.

The key takeaway is that our innovative strategies, particularly our use of ML, propel the continual enhancement of existing and future Arable products. As we collect more data through the Cal/Val network, our accuracy improves, allowing us to build smarter models that facilitate better data-driven agricultural decisions. Put simply, we get better and better over time. In this paper, we have seen the success that these models have already achieved to date, noting that they will only continue to improve and set new standards for weather and crop monitoring. While this paper specifically covers the application of ML to our core measurement base, these core measurements are the foundation of our derived agronomic models which are key to practical, informed agronomic decision-making.

ACKNOWLEDGEMENTS

Arable would like to thank Rob Hamnet and Susan Cordell of the US Forest Service for helping us deploy and maintain instrumentation in Hilo, HI. This site has been invaluable for our collection of rainfall data since 2018. Arable would also like to extend thanks to the many other partners, researchers, and collaborators that contribute to the larger Cal/Val network. We may not have directly used data from all sites in this analysis, but your support has nonetheless been paramount to Arable's research and development efforts and continued success.

REFERENCES

[1] Wikipedia contributors. (2020, November 19). Machine learning. Retrieved November 19, 2020, from Wikipedia, The Free Encyclopedia website: https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=989487893

[2] AmeriFlux: Measuring carbon, water and energy flux across the Americas. (n.d.) https://ameriflux.lbl.gov/

[3] Vantage Pro2. (2019, October 23). https://www.davisinstruments.com/vantage-pro2/

[4] Weather station for research. (2016, October 25). https://www.metergroup.com/environment/products/atmos-41-weather-station/

[5] Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15(3), 259–263.

[6] Garner, R. (2015). NASA's Goddard Space Flight Center. https://www.nasa.gov/goddard

[7] National Renewable Energy Laboratory (NREL) home page. (n.d.). https://www.nrel.gov/

[8] OTT Parsivel² Laser Present Weather Sensor. (n.d.). https://www.otthydromet.com/en/p-ott-parsivel-laser-present-weather-sensor/70.210.002.3.0

[9] HUMICAP® Humidity and Temperature Probe HMP155. (n.d.). https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/weather-stations-and-sensors/hmp155

[10] RADIATION SHIELDS — BARANI DESIGN Technologies. (n.d.). https://www.baranidesign.com/radiation-shields

[11] CNR4 Net Radiometer. (n.d.). https://www.kippzonen.com/Product/85/CNR4-Net-Radiometer

[12] World Meteorological Organization (WMO). (2018). Guide to Instruments and Methods of Observation: Volume I – Measurement of Meteorological Variables, pp. 221–223. Retrieved from https://library.wmo.int/doc_num.php?explnum_id=10179

[13] Wolf, L. A., Siegfried, B. J., Smith, A. L. (2018). Disdrometer having acoustic transducer and methods thereof. U.S. Patent No. 10578772B2. Washington, DC: U.S. Patent and Trademark Office.