# COTTON YIELD MONITOR VALUES DRIFT OVER TIME

**R. G. Trevisan**
**A. O. G. Barbosa**
**University of Illinois at Urbana Champaign**
**Urbana, IL**
**L. S. Shiratsuchi**
**LSU Ag Center**
**Baton Rouge, LA**
**N. F. Martin**
**University of Illinois at Urbana Champaign**
**Urbana, IL**

## Abstract

On-farm precision experimentation is an important resource to understand the spatial variation of crop response to management practices and thus improve agronomic decisions. High-quality yield data is fundamental for getting real insights from this type of experiments. Previous experimental projects have demonstrated that yield data quality affects the outcome of on-farm trials. Many reports have also shown that cotton yield monitors are sufficiently affected by varietal properties and lint turnout to alter the inferences made from data with multiple varieties. They also indicate that errors were correlated with the time of harvest, which can introduce artificial trends in the yield maps. Thus, there is great interest in finding new methods to improve yield monitor information quality. This work aims to investigate the existence of temporal trends affecting cotton yield monitor errors and develop methodologies to improve yield data quality. A data-fusion technique was developed using high-frequency data from the cotton mass flow sensors and the low frequency information available for each module in the new cotton harvester models. Errors up to 20% were observed comparing the extreme conditions of the day. The yield errors were correlated with cotton moisture, air relative humidity and time of harvest. For the new machine models, moisture and weight measured for each module can be used to correct the data and improve resulting yield maps. This data fusion could be implemented on the real-time yield mapping software or in post-processing. For older machines, in which this data is not available, post-processing using quadratic regression with the time of day can be used with a similar performance. This procedure represents an important step to improve the use of sensor-based yield data to evaluate on-farm trials.

## Introduction

Many cotton producers have focused on improving techniques to increase crop productivity by adopting higher levels of technology, being placed amongst the earliest adopters of precision agriculture methods. The potential of using precision agriculture tools to conduct on-farm precision experimentation (OFPE) is of special interest in this sector. OFPE is becoming an important resource to understand the spatial variation of crop response to management practices and thus improve agronomic decisions. High-quality yield data is fundamental for getting real insights from this type of experiments. Previous research studies have demonstrated that yield data quality affects the results of OFPE. Many reports have also shown that cotton yield monitors are sufficiently affected by varietal properties to alter the inferences made from data with multiple varieties. Thus there is great interest in finding new methods to improve yield monitor information quality.

Cotton yield monitors are designed to measure crop yield without any contact with the product, while seed cotton is transported through an air stream from the header to the basket. To achieve that, cotton yield monitors rely on remotely sensed physical properties to estimate cotton mass flow, reducing possible errors associated with sensor maintenance. However, external environmental conditions during harvest can affect the interaction between the measured properties and yield, introducing other sources of temporal trends.

Main commercially available systems are based either on optical or microwave sensors. In both cases, yield is inferred based on the volume of material traveling through an air stream. Estimating cotton`s mass from volume can be challenging since its density depends on fiber characteristics and the level of compaction. To prevent errors caused by compaction, measurements are taken during seed cotton transportation through the air stream. The yield values estimated by the mass flow sensors have the advantage of the high frequency and small error in the area estimate. Nevertheless they are indirect sensors and the accuracy of the results depends on the calibration process and the model used to account for bias inherent to their working principle. Moreover, variations on moisture and fiber characteristics can affect the speed and the density of the cotton flowing in the ducts, and therefore the accuracy of the yield mapping process. In a study comparing observed weights to yield monitor predictions over six site-years, the errors were found to be significantly different with regard to variety and lint turnout, but they didn't find a consistent method to account for such errors (Taylor et al. 2014).

In a studies conducted to compare yield monitor-estimated weights from a module-building picker to observed weights in replicated variety trials, the errors were found to be correlated with the time of harvest (Stewart et al. 2008; Vories et al. 2017). Other authors found significant differences in seed cotton yields were detected between the observed (weighed modules) and estimated (yield monitor) values. They highlighted the need for more research to achieve a better understanding of the factors affecting cotton yield monitor calibration and allow more on-farm studies to improve site-specific recommendations for cotton (Vories et al. 2018).

Another possible way to estimate cotton yield is measuring every basket load or module of cotton (Ge et al. 2009). This produces a low resolution map and was done mainly for research purposes. However, there are newer cotton picker machines available that include a scale to weight the module after it is built. The load cells in the scale are less sensitive to short-term variations and provide a more generic reading, based on mass rather than volume. The main limitation and source of errors for this method lies on the estimation of the harvested area associated to each module. In the basket type harvesters, the machine has to stop to unload and all cotton harvested is weighted at once. The module building harvesters have only a small basket that works as a temporary accumulator. When this accumulator is full, the cotton is transferred to the module building mechanism and the module is built by adding layers and compacting the module in an intermittent process. Because of that, not all cotton accumulated in the temporary accumulator of the machine may be used to build the current module. In this situation, the area of contribution will be overestimated for the current module and underestimated for the subsequent module. This has a large impact on the accuracy of individual modules, however these errors are non-cumulative and will be diluted if averaged over a few modules.

In summary, the cotton mass flow sensors are precise but not necessarily accurate, as their data has a low dispersion but their accuracy depends on the stability of the relationship between the properties being measured and the property being estimated. The module weight data provides accurate estimates for cotton yield, as long as the scale is well calibrated, but the resolution and the precision are inferior due to the low frequency of measurements and the errors in the area estimate. Due to these differences, the direct comparison of the yield values is not sufficient to evaluate their accuracy. Combining both types of measurements can be an opportunity to better understand the sources of errors and output generate detailed and accurate yield maps. Therefore, this work aims to investigate the existence of temporal trends affecting cotton yield monitor errors and develop methodologies to improve yield data quality.

## Material and Methods

Yield data was collected in a 527 ha field during the 2018 cotton season in Mato Grosso state, Brazil. Cotton was planted in January and harvested in August. Yield data was recorded by a John Deere CP690 cotton picker with harvest doc yield monitor. The yield monitoring system is composed of one microwave sensor for each of the six rows, a DGPS system for georeferencing the data, and a cab monitor that records the data every second.

This model of harvester also has a scale to measure the weight of each round module. This additional data was used to develop and validate the methodology proposed. The time, area, weight, moisture, and coordinates were recorded when the module was built. The time information was used to merge both data sources. All points collected after one module was finished were assigned to the next one, determining the area contributing to that module. The average yield registered by the mass flow sensors of all the points contributing to that module was used as the yield value estimated by the sensors. The weight measured by the scale was converted to dry weight, discounting the moisture content. Seed cotton yield is usually reported at an unknown moisture concentration, nevertheless standardizing the weights was done to provide unbiased estimates of the differences between the two yield estimates.

The difference between the yields estimated by the two methods was considered as the yield error. The yield errors were evaluated regarding their temporal distribution and their correlation with crop and environmental variables. Weather information was collect from the official meteorological system (National Institute of Meteorology - INMET) from the Juara-A914 station, located about 200 km from the farm. Although the weather station was located far from the field, the latitude and elevation of both places are similar and the data was still considered valid for the comparisons. Linear regression models were fitted between the yield errors and the auxiliary variables. The models were then used to predict the mass flow sensor error at each point and this value was subtracted from the raw values to remove the predicted trends. All procedures described in this paper were developed using the R programing language (R Core Team, 2018). All figures were prepared using QGIS geographic information system (QGIS Development Team 2018).

## Results and Discussion

The field was harvest in a period of 13 days, from August 7, 2018 to August 19, 2018. Over 430,000 data points were recorded, with an average density of 816 points per hectare. In the same period, 1,033 modules were produced, with an average of 2,400 kg and 0.5 ha each. The data from the modules was joined with the yield monitor points using the timestamp information. This process is represented in Figure 1, where random colors were assigned to each module ID and attributed to all the yield points from that module.

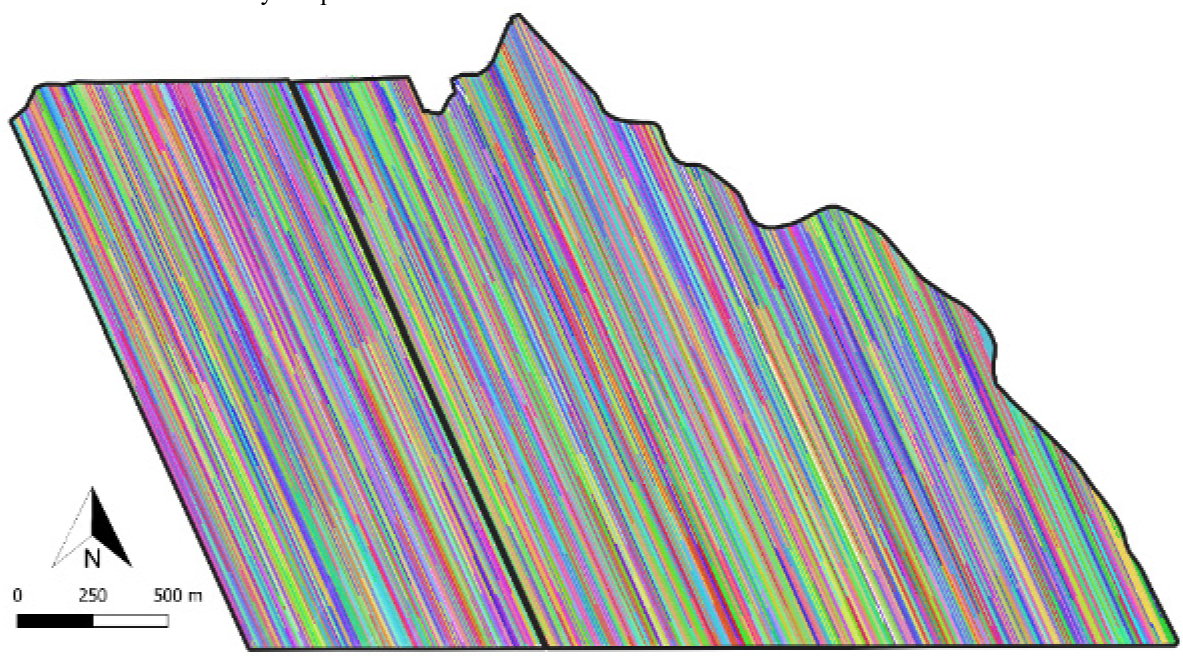


Figure 1. Area coverage for each cotton module, totaling 1033 modules in a field with 527.0 ha.

The exploratory data analysis showed a consistent trend in the variation of cottonseed moisture during the day (Figure 2). A similar trend was observed between the time of harvest and the total weight of the module, with heavier modules when the cotton moisture is higher. The difference in weight is not caused only by the extra water, as the dry weight presents the same trend. The same variation was observed in the area of each module, and as a consequence the yield estimated by the weights and the area of each module shows no correlation with the time of the day (Figure 3).

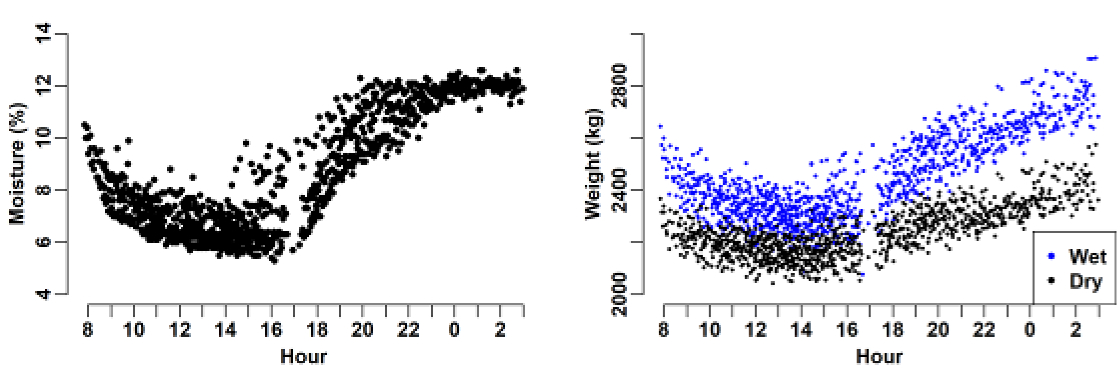


Figure 2. Daily variation of harvested cotton moisture and module weight.

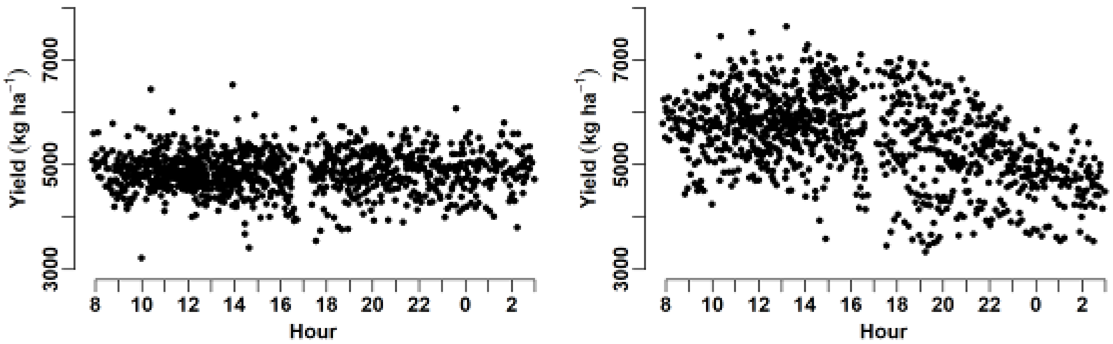


Figure 3. Daily variation in yield estimated by module weight and by the cotton flow sensors.

Conversely, variation of the yield estimated by the cotton mass flow sensors presented a clear trend, with higher values in the afternoon and lower values during mornings and evenings. This trend shows evidence that yield values drift over time. As this measurements were taken during 13 days at different parts of field, it's unlikely that the trend is induced by any correlation between the spatial variability of the true yield with the time of harvest. The fact that the yield estimates by the module weight shows no trends reinforces this assumption.

The higher density of the module when the cotton has a higher moisture concentration may be due to better compaction of the module, probably because of the higher resistance to be compacted when the cotton fiber is dryer. If the density of the cotton in the module is changing, the density of the cotton plume traveling through the ducts where the microwave sensor could also change, and contribute to the observed drift.

The map with the spatial distribution of the raw yield values recorded by the cotton mass flow sensors (Figure 4) reveals the presence of alternating stripes of high and low yields following the direction that the field was harvested. This direction was the same used for all other operations, such as planting and spraying, therefore it is possible that some other operation influenced the yield in this pattern. One example would be the soil moisture at planting, which usually is excessively high in the morning or soon after a rain, and becomes more adequate as the time passes. Another example is the application of herbicides for which the crop is not resistant, but partially tolerant. If the leaves are wet, the product is more likely to cause injuries to the plant, and there is a high correlation between time of the day and leaf wetness. The possibility of factors affecting yield in this pattern makes it harder to distinguish what is true variation and what can be related to sensor errors.

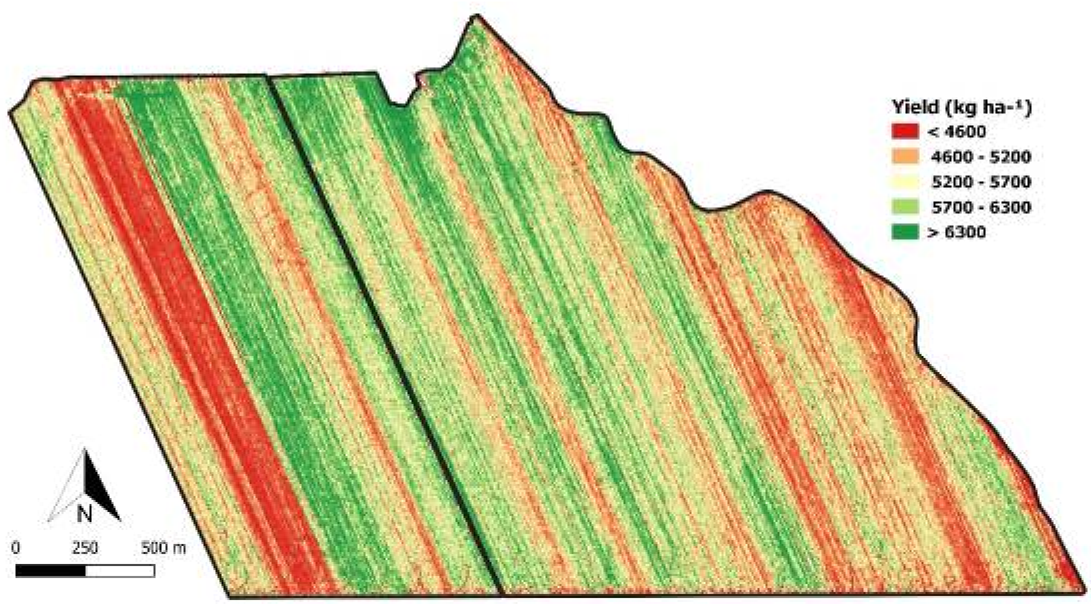


Figure 4. Spatial distribution of raw yield values recorded by the cotton mass flow sensors.

Table 1. Performance of harvest and environmental variables to correct yield temporal drift.

| Variable (V) | Equation | RMSE | R-squared |
|---|---|---|---|
| Moisture | $Yield = Yield\_Obs - 369.34 - 31.77*V + 8.36*V^2$ | 385.7 | 0.32 |
| Time | $Yield = Yield\_Obs + 571.51 - 120.90*V + 4.79*V^2$ | 394.8 | 0.29 |
| Temperature | $Yield = Yield\_Obs - 1860.67 + 105.01*V - 1.19*V^2$ | 412.8 | 0.22 |
| Humidity | $Yield = Yield\_Obs + 696.96 - 11.21*V + 0.01*V^2$ | 417.1 | 0.20 |

The regression analysis used to compare the weather information with the harvest data revealed that cottonseed yield errors could be partially explained by the variation in cotton moisture, time of harvest, air temperature and relative humidity (Table 1). The relationship between yield error and moisture concentration was almost linear, while the other three parameters had quadratic relationships, which is expected since they are cyclic and a short period can be approximated by a parable (Figure 6). The correlations make it possible to construct models with any of the data available to correct the temporal drift. The cotton moisture would be the most accurate source of correction, but is only available in the newer machine models. The relative humidity could be the second choice if the information is available in a location close to the field. The simpler and more readily available information is the time of day, which can be used with similar accuracy. It is important to mention that the weather conditions in this region during the time of the year in which the harvest was performed and the data collected are similar across different days due to the drought season. Under more variable weather conditions using the time of harvest is likely to have a lower performance when compared to the other variables.

The map with the spatial distribution of the moisture concentration recorded for each module (Figure 5) reveals the presence of similar trend to the one observed in Figure 4, with low moisture concentration matching the stripes with high yields and high moisture in the parts with lower yields. The harvest is usually started at about 8:00 am when the dew from the night is evaporating fast and the moisture is around 9%. Between 9:00 am and 5:00 pm the moisture stays around 7% and then starts to increase until it reaches 11% at about 8:00 pm. The harvesters only stop when the moisture goes over 12%, usually around 3:00 am. In this field, the moisture pattern reflects solely the harvest pattern, and not spatial variability due to variations in crop maturity.

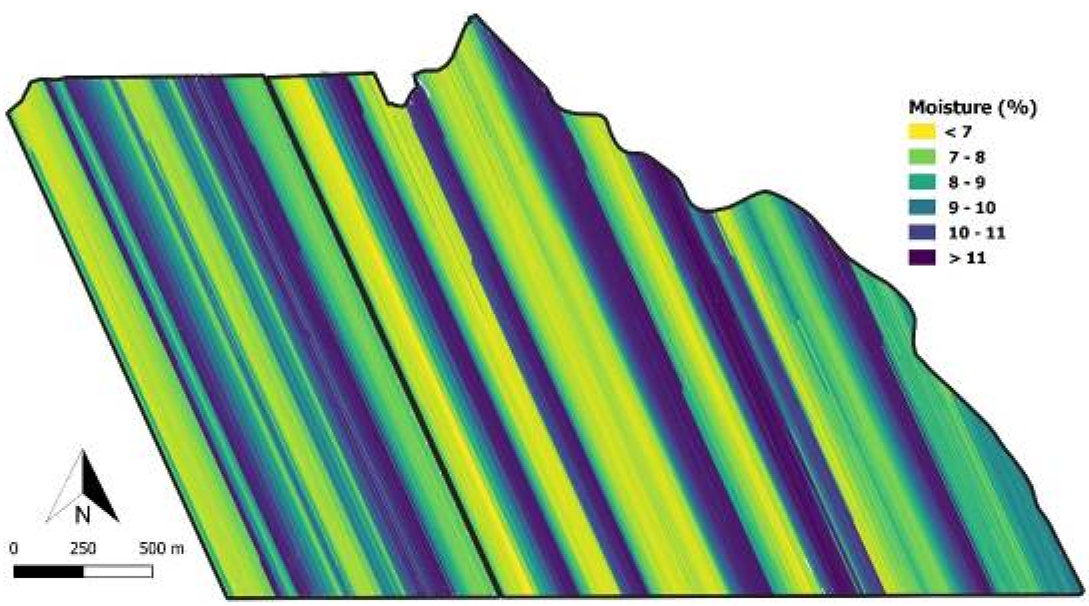


Figure 5. Spatial distribution of cottonseed moisture at harvest.

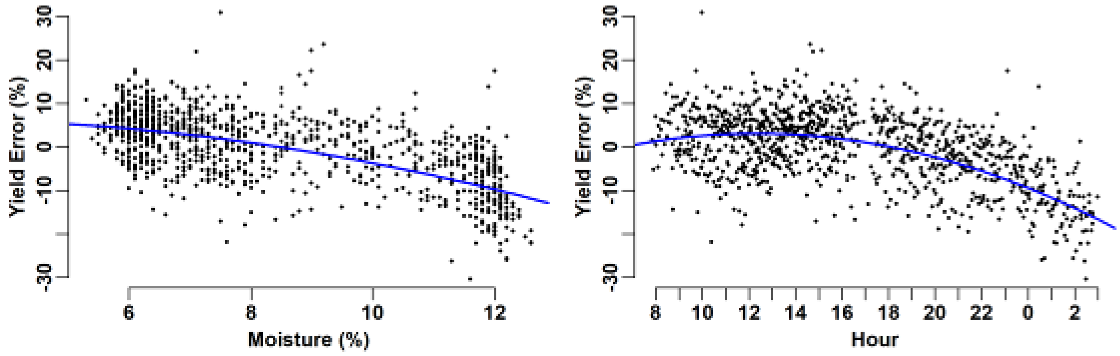


Figure 6. Correlation of yield error with the cottonseed moisture concentration and the time of harvest.

For the newer machines, moisture and weight can be used to correct the data and increase the quality of the yield maps. This data fusion could be implemented in the yield mapping software. For older machines were this data is not available, simple regression with time of day can be used with similar performance. The trends observed could explain the difficulty to calibrate different machines harvesting the same field. If there is a significant difference in the time between the calibrations of each one, a systematic error will be introduced. This is also important for using this data to evaluate on-farm trials, especially in side by side comparisons, or with few number of repetitions. When comparing side-by-side or nearby passes (Figure 7), the parts of the field that are harvested in the beginning and the end of the day or between 5:00 pm and 8:00 pm presents the higher errors, which can be up to 20%.

The resulting yield map after applying the correction based on the moisture concentration equation is shown in Figure 8. The spatial distribution of yield values is more representative of the variability expected for the field with less evident parallel stripes caused by the temporal drift of sensors values and the harvest pattern. The low r-squared values shown in Table 1 are mainly a consequence of the errors introduced by the cotton retained in the temporary accumulator, adding a sensor to monitor its volume over time could further improve this methodology and also improve the value of this data to the traceability of cotton production. These finds also suggest that additional information and sensor fusion could improve the quality of yield mapping in other systems such as the grain yield mapping. Although daily trends are less likely to be significant in the impact plate measuring systems, errors in calibration are also common and could be minimized with information of the total weight when the grain tank is filled.

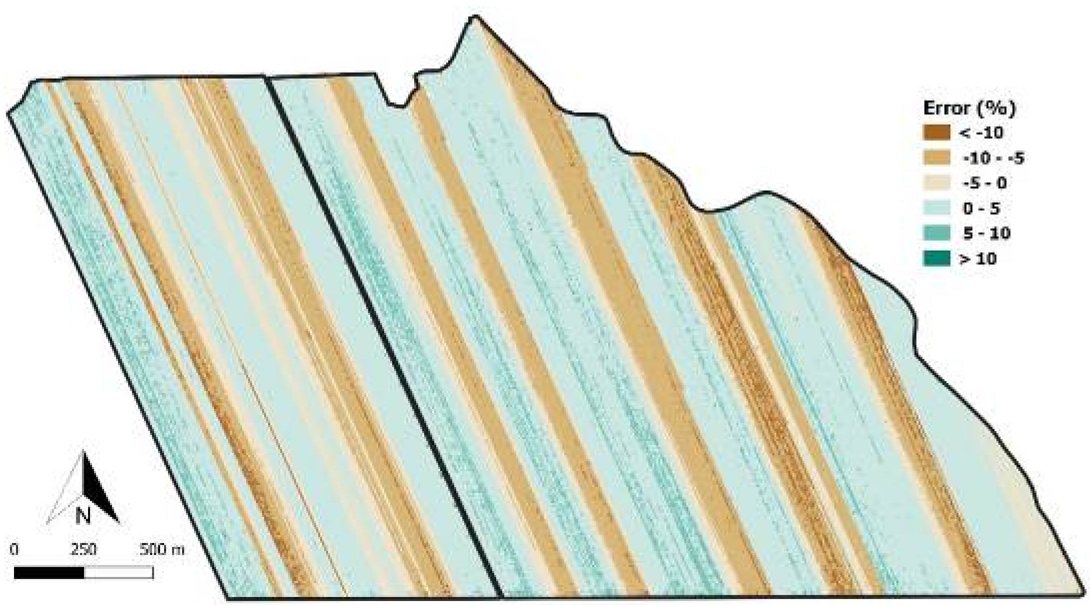


Figure 7. Spatial distribution of the cotton mass flow sensors yield errors.

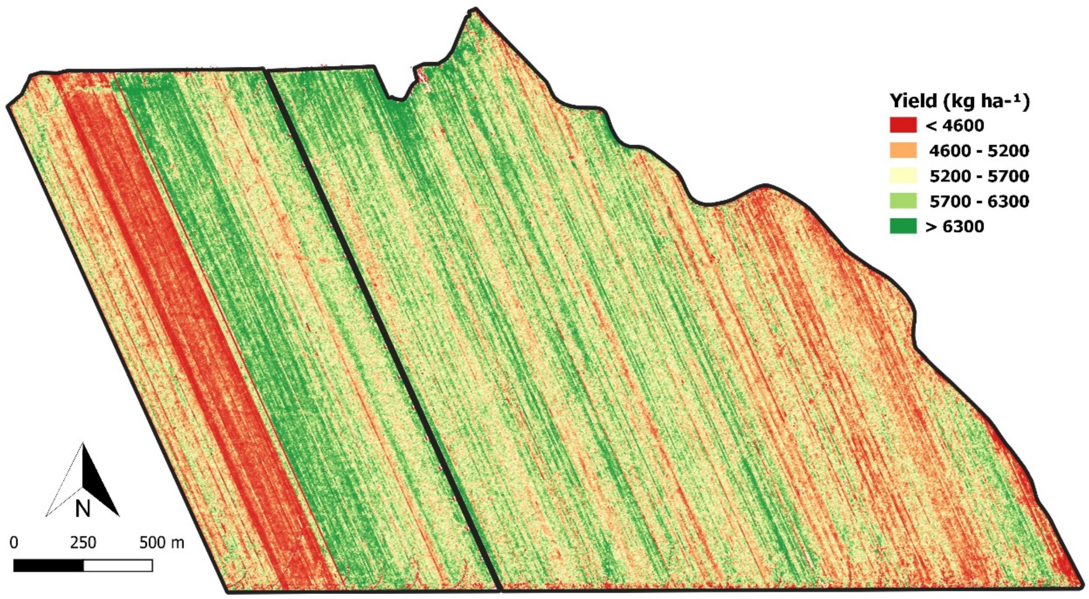


Figure 8. Spatial distribution of yield values after accounting for temporal drifts.

## Conclusions

Errors up to 20% were observed in side-by-side passes harvested in the extreme weather conditions of the day. The errors were correlated with cotton moisture, time of harvest, air temperature and relative humidity. For the new machine models, moisture and weight of each module can be used to correct the data and increase the quality of the yield maps. This data fusion could be implemented on the real-time yield mapping software or in post-processing. For older machines, in which this data is not available, post-processing using quadratic regression with the time of day can be used with a similar performance. This procedure represents an important step to improve the use of sensor-based yield data to evaluate on-farm trials.


## Acknowledgements

The authors would like to acknowledge Terra Santa Agro for providing the datasets used in this work.

## References

Ge, Y., Thomasson, J. A., Sui, R., & Morgan, C. L. S. (2009). A module-specific post-processing calibration method to improve cotton yield mapping. *Computers and Electronics in Agriculture*, *68*(2), 161–167. doi:10.1016/j.compag.2009.05.007

QGIS Development Team. (2018). QGIS Geographic Information System, Version 3.2. http://qgis.org

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. https://www.r-project.org/

Stewart, A. M., Deville, S. F., & Woolam, B. W. (2008). Comparison of on-farm cotton variety trial results when using yield monitors vs. weigh wagons. In *Proc. Beltwide Cotton Conference* (pp. 69–71)
.

Taylor, R., Porter, W., Boman, R., Osborne, S., Henderson, W., Buschermohle, M., & Barnes, E. (2014). Using yield monitors to evaluate cotton variety tests. In *Proc. Beltwide Cotton Conf.* (pp. 494–498). New Orleans, LA.

Vories, E. D., Jones, A., Stevens, G., & Meeks, C. (2018). Variety Effects on Cotton Yield Monitor Calibration. In *Proceedings of the 14th International Conference on Precision Agriculture* (pp. 1–7).

Vories, E. D., Jones, A., & Sudduth, K. A. (2017). Impact of Variety on Cotton Yield Monitor Calibration. *2017 ASABE Annual International Meeting*, 1–6. doi:10.13031/aim.201700685