

APPROACHES FOR GEOSPATIAL PROCESSING OF FIELD-BASED HIGH-THROUGHPUT PLANT PHENOMICS DATA FROM GROUND VEHICLE PLATFORMS

X. Wang, K. R. Thorp, J. W. White, A. N. French, J. A. Poland

ABSTRACT. *Understanding the genetic basis of complex plant traits requires connecting genotype to phenotype information, known as the “G2P question.” In the last three decades, genotyping methods have become highly developed. Much less innovation has occurred for measuring plant traits (phenotyping), particularly under field conditions. This imbalance has stimulated research to develop methods for field-based high-throughput plant phenotyping (HTPP). Sensors installed on ground vehicles can provide a huge amount of potentially transformative phenotypic measurements, orders of magnitude larger than provided by traditional phenotyping practice, but their utility requires accurate mapping. Using geospatial processing techniques, sensor data must be consistently matched to their corresponding field plots to establish links between breeding lines and measured phenotypes. This article examines problems and solutions for georeferencing sensor measurements from ground vehicle platforms for field-based HTPP. Using three case studies, the importance of vehicle heading for sensor positioning is examined. Three corresponding approaches for georeferencing are introduced based on different methods to estimate vehicle heading. For two of the cases, approaches to develop a field map of plot areas are addressed, where the issue is to ensure that sensor positions are correctly assigned to plots. Two solutions are proposed. One uses a geographic information system to design a field map before planting, while the other adopts an algorithm that calculates plot boundaries from the georeferenced sensor measurements. An advantage of the latter approach is the accommodation of irregular planting patterns. Using the algorithm to calculate the plot boundaries of a winter wheat field in Kansas, 98.4% of the calculated plot centers were within 0.4 m of the surveyed plot centers, and all of the calculated plot centers were within 0.6 m of the surveyed plot centers. While multiple options and software tools are available for geospatial processing of field-based HTPP data, they share common problems with sensor positioning and plot delineation. The options and tools presented in this study are distinguished by their practicality, accessibility, and ability to rapidly map phenotypic data.*

Keywords. *Field-based high-throughput plant phenotyping, Georeferencing, Ground vehicles.*

In the 21st century, decoding the genetic basis of complex plant traits (phenotype) from its genetic composition (genotype) has become a great challenge (Furbank and Tester, 2011). Spectacular advances in genotyping methods have resulted in a wealth of genomic information, whereas methods for collecting plant phenotypes have improved little (Cobb et al., 2013; Araus and Cairns, 2014). This discrepancy has led to the development of field-based, high-throughput plant phenotyping (HTPP) methods, which seek to rapidly and accurately measure plant characteristics

in the field using proximal sensing and computing tools. By using HTPP platforms to measure phenotypes for diverse genotypes in the field, plant growth characteristics can be associated with underlying genetic mechanisms.

To achieve field-based HTPP, mobile platforms integrated with sensing units have been designed and developed (Busemeyer et al., 2013; Andrade-Sanchez et al., 2014). Potential field-based phenotyping platforms include ground vehicles, gantry cranes, cable robots, and aerial vehicles. Each platform has advantages and disadvantages for field-based HTPP; however, ground vehicles have a leading role for current field-based HTPP because of their high payload, operational simplicity, and flexibility to move (White et al., 2012).

A typical ground-based platform for field-based HTPP incorporates a suite of sensing units, Global Navigation Satellite System (GNSS) receivers, and a data acquisition system, all mounted on a high-clearance field vehicle, such as a tractor (Andrade-Sanchez et al., 2014), a sprayer, or a flexible, low-cost cart (White and Conley, 2013). Depending on phenotyping requirements, different types of proximal sensing units, such as infrared thermometers (IRT), active spectral sensors, and displacement sensors (ultrasonic transceivers), can be installed on the ground vehicle. The GNSS receiver provides both Coordinated Universal Time (UTC) information and the antenna’s geographic position, which are

Submitted for review in August 2015 as manuscript number ITSC 11502; approved for publication by the Information, Technology, Sensors, & Control Systems Community of ASABE in June 2016.

Mention of company or trade names is for description only and does not imply endorsement by the USDA. The USDA is an equal opportunity provider and employer.

The authors are **Xu Wang, ASABE Member**, Research Associate, Department of Plant Pathology, Kansas State University, Manhattan, Kansas; **Kelly R. Thorp, ASABE Member**, Agricultural Engineer, **Jeffrey W. White**, Research Plant Physiologist, and **Andrew N. French**, Research Physical Scientist, USDA-ARS U.S. Arid-Land Agricultural Center, Maricopa, Arizona; **Jesse A. Poland**, Assistant Professor, Department of Plant Pathology, Kansas State University, Manhattan, Kansas. **Corresponding author:** Xu Wang, 4024 Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506; phone: 785-532-3253; e-mail: xuwang@ksu.edu.

logged to derive the sampling time and position. The data acquisition system controls the sampling rate, collects measurements from the sensing unit, and saves raw data to storage devices for subsequent processing. When processed data are matched to field plots of the target breeding lines or cultivars, the correspondence between phenotypes and genotypes can be assessed.

Field-based HTPP systems collect large quantities of sensing measurements, sometimes in the megabyte to gigabyte range, over experimental fields planted with diverse crop cultivars. For example, with a sampling rate of 5 Hz, a field-based HTPP platform will collect 18,000 measurements from one sensor in an hour. Depending on the field size, the vehicle traveling speed, and the sensor type, millions of measurements can be generated by a field-based HTPP system over a growing season. Two essential steps in processing the raw phenotypic data are to (1) use GNSS positioning data, sensor offset distances, and vehicle heading to calculate sensor positions via coordinate transformations and (2) locate sensor measurements within the correct field plot to link sensor measurements (phenotypes) with corresponding information on the experimental plot entry (genotypes). This process requires information on sensing positions and field boundaries for each plot.

Accurate geospatial data processing, however, can be problematic for HTPP systems for two reasons: sensors are not co-located with GNSS antennae, and field planting layouts may not be well-defined. The first problem arises because of the sensor displacement from the GNSS antenna, which means that sensor positions cannot be computed without knowledge of platform azimuthal heading. As shown in figure 1, the antenna can be in a fixed position, yet the sensor positions are not identical. Thus, sensor positions must be calculated using the vehicle heading and the offset distances between the sensors and the GNSS antenna. This procedure is termed “georeferencing” of sensor measurements. Vehicle “heading” is the direction the front of the vehicle is pointing, while vehicle “course” is the over-the-ground travel path. This distinction is important because GNSS receivers record vehicle course, while vehicle heading is needed for georeferencing. In some cases, HTPP vehicles may operate with identical heading and course, but this may not be the case for all implementations of HTPP platforms. For example, tractor

platforms typically travel in serpentine patterns, while gantry systems reverse directions but do not turn. The positioning problem is further complicated when sensors are used with off-nadir view angles; however, this study does not consider such cases. Thus, the first objective of this study was to compare different approaches for georeferencing sensor measurements based on different approaches to obtain the vehicle heading. Three case studies of ground vehicle platforms for field-based HTPP are described to show how vehicle configuration impacts the heading calculation and the georeferencing algorithms.

The second problem, field layout uncertainty, arises because of differences between field design and actual field conditions. To reference sensor data within a given field plot, a field map must be generated to delineate plot boundaries with unique identification codes (IDs) for each plot. The plot ID links treatment plots with information for breeding lines, which may include land races, populations, cultivars or other materials sown in the plot, and also management practices for each plot. If the positions of breeding lines do not match the planned design, due for example to emergence and planting inconsistencies, sensor data cannot be meaningfully matched with the corresponding plants. Thus, the second objective of this study was to introduce two approaches for generating maps of plot boundaries. One approach uses a GIS to design the field plot layout and save plot boundaries in the shapefile format, while the other computes plot boundaries from georeferenced phenotypic data. The advantages and disadvantages of each plot boundary generation approach are compared.

METHODS

GENERAL STEPS FOR GEOREFERENCING

In order to achieve HTPP, multiple sets of sensors are installed on the ground vehicle to simultaneously collect phenotypic data from multiple crop rows, which can contain different breeding lines. Each set may include different types of sensors to measure different plant traits. Georeferencing sensor data involves calculating each sensor’s absolute position based on its relative position to the onboard GNSS antenna and the vehicle heading. For this purpose, the relative position was defined as the *X* and *Y* offsets in meters from the GNSS

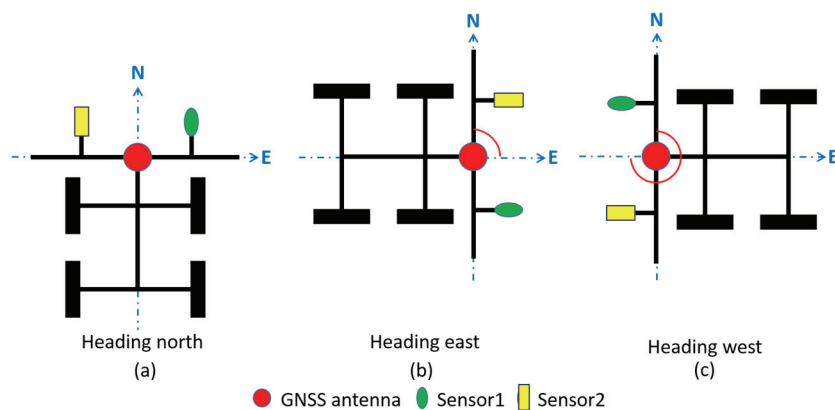


Figure 1. Effect of heading angle on sensor positions. Top views of a ground vehicle HTPP platform are shown for (a) north, (b) east, and (c) west heading angles. While the sensor positions are fixed relative to the GNSS antenna, their absolute positions depend on the heading angle.

antenna to the sensor. The X and Y offsets were derived from a local vehicle coordinate system with the origin at the GNSS antenna and the positive Y axis in the direction of the vehicle heading. The vehicle's heading angle was defined as the clockwise angle between the northward cardinal direction and the positive Y axis of the vehicle's local coordinate system (i.e., north = 0° , east = 90° , south = 180° , and west = 270°). Figure 2 shows an example of the sensor and the GNSS antenna layout on the vehicle with the X and Y offsets indicated.

To incorporate sensor offsets with the GNSS antenna positions, the latitude and longitude provided by the GNSS receiver must first be transformed to a planar coordinate system, such as the Universal Transverse Mercator (UTM) system. UTM was chosen based on its popularity and convenience, but other coordinate systems using conformable projections are equally viable (e.g., Lambert conic or Albers equal area). This transformation simplifies trigonometric calculations by permitting treatment of geospatial data using rules of the Euclidean distance. Although distortions are expected in the conversion from geodetic to planar coordinates, the distortion level is negligible compared to the plot size, so the effect is insignificant at the scale of field-based HTPP experiments.

While the sensor position relative to the GNSS antenna is fixed, its absolute position in the UTM coordinate system changes with the vehicle heading (fig. 1). The relationship between the GNSS antenna position and the absolute sensor position is:

$$\begin{cases} X_{Sabs} = X_G + X_S \cos(\alpha) + Y_S \sin(\alpha) \\ Y_{Sabs} = Y_G - X_S \sin(\alpha) + Y_S \cos(\alpha) \end{cases} \quad (1)$$

where

(X_{Sabs}, Y_{Sabs}) = UTM coordinate of the sensor's absolute position

(X_G, Y_G) = UTM coordinate of the GNSS antenna's position (fig. 2a)

X_S and Y_S = X and Y offset distances from the sensor to the GNSS antenna (fig. 2a)

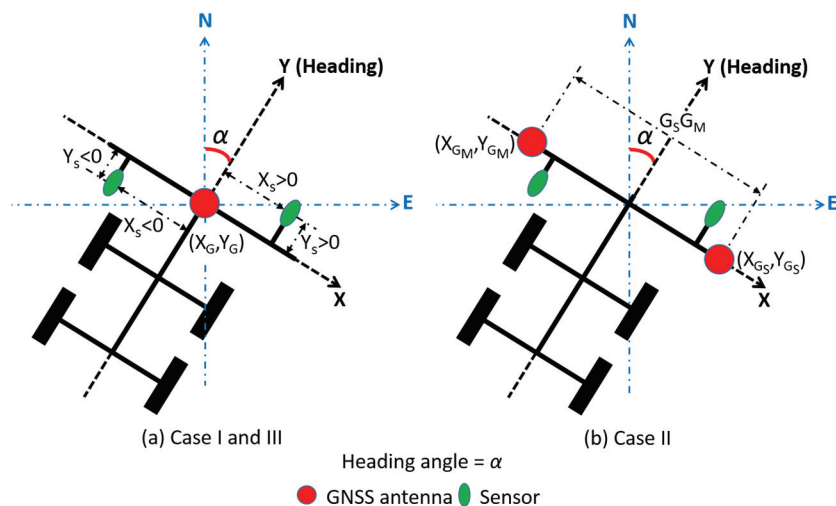


Figure 2. Local coordinate system describing sensor positions relative to the GNSS antenna with ground vehicle HTPP platform viewed from above. Positive offset values indicate that the sensor is forward or right of the GNSS antenna, while negative offset values indicate that the sensor is backward or left of the GNSS antenna. The georeferencing approaches introduced in this article are independent of the GNSS antenna's location on the ground vehicle.

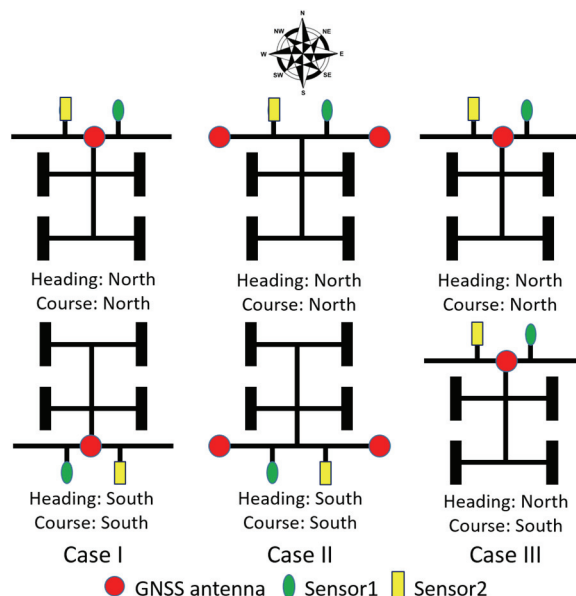


Figure 3. Top views of vehicles used in the three case studies: single receiver on a turning-platform (case study I), dual receivers on a turning platform (case study II), and single receiver on a non-turning platform (case study III).

α = vehicle's heading angle, measured clockwise from true north (fig. 2a).

Different methods are available to obtain the vehicle heading based on the deployment of the GNSS antenna on the vehicle. Field vehicles are commonly equipped with a single GNSS antenna, from which information on vehicle course can be obtained. If vehicle heading is equal to vehicle course (i.e., the vehicle nose always points in the direction of travel), vehicle course can be used in lieu of vehicle heading information. However, this condition is not met for HTPP platforms that are difficult or unable to turn in the field, which can lead to opposing vehicle heading and course directions. Case studies I and III (fig. 3) demonstrate these two conditions, respectively. Some dual-antenna input

GNSS receivers (e.g., Trimble BX982 GNSS) can output the vehicle heading by using the positioning information from two antennae to both position and orient the vehicle. In case study II, as two GNSS antennae are installed at each end of the sensor boom (fig. 3), a vector between the two antenna positions can be calculated. Because the vehicle heading remains perpendicular to the vector, the heading angle can thus be acquired. Detailed georeferencing algorithms will be described in each of the three case studies.

GENERAL STEPS FOR OBTAINING PLOT BOUNDARIES

Once phenotypic measurements are georeferenced, plot boundary information is needed to locate measurements within in each plot (fig. 4). Because plot boundaries delineate the area planted to individual breeding lines, phenotypic measurements can then be linked to genotypic data. Depending on the plot's shape, plot boundaries can be described in different ways. For example, rectangular boundaries can be established using the four corner coordinates, which is likely the most common layout for field-based HTPP studies. However, non-rectangular plot shapes are also possible, a

common example being experiments under center-pivot irrigation systems. In this case, arc-shaped plot maps can be established using angle and radius data. Robust tools for plot map generation can quickly accommodate diverse experimental designs and plot shapes.

To develop field plot boundary maps, geographic coordinates of the plot perimeter must be obtained. One approach is to survey sufficient field points to delineate the perimeter of each plot. A drawback of this approach is the time and labor required to survey experiments with potentially thousands or tens of thousands of plots, although not every plot needs to be surveyed if plot boundary coordinates can be calculated from key survey points and known plot geometry and arrangement. Another data source that can assist field map generation is the precision planting system, if used. When equipped with a precision GNSS rover, planters can record planting positions with centimeter-level accuracy. By importing those coordinates into a GIS, points with planting positions belonging to the same plot can be grouped together and used to calculate plot boundaries. This method may

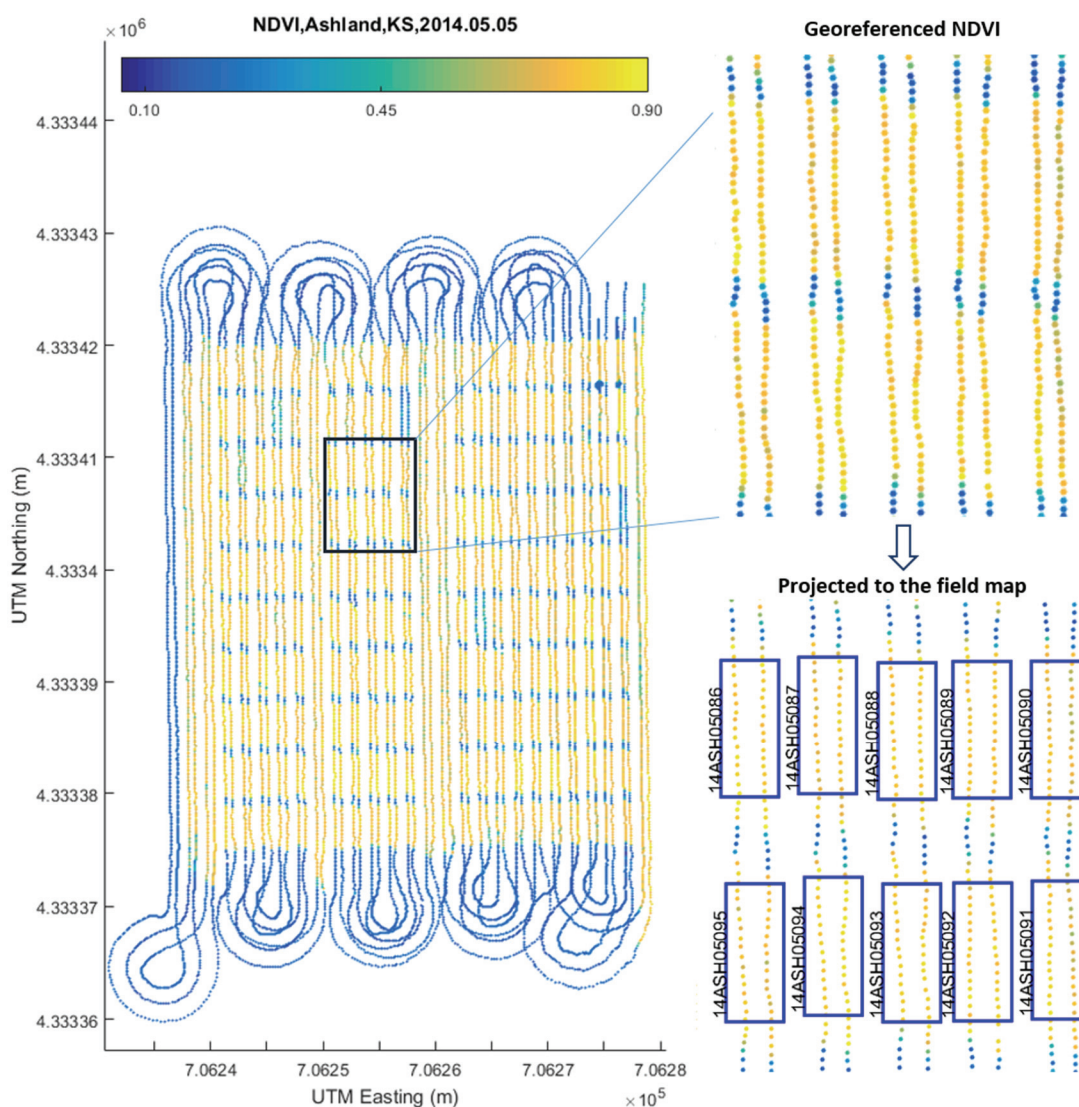


Figure 4. Georeferencing and projecting sensor readings to plot boundaries: (left) raw in-field sensor positions converted to UTM coordinates, (upper right) a closer view of georeferenced NDVI measurements, and (lower right) NDVI measurements matched with plot polygons.

avoid the need for survey. However, some planting systems only record the positions where planters are triggered to drop seeds. Due to the mechanics of the planter, there may be a delay between triggering and dropping the seed in the soil, which results in an offset between the triggering position and the true plot boundary. If the offset distance is known and consistent, it can be accounted for in the calculation of plot boundary positions. Digital orthophotos of the study area from the U.S. Geological Survey (USGS), for example, are another source of information to assist establishment of field plots boundaries prior to planting.

Georeferenced phenotypic measurements not only quantify plant traits but also provide information on spatial distribution of the plants themselves. Therefore, plot boundaries can also be defined by clustering measurements based on their similarity while also considering their spatial proximity. In figure 4, the continuous dots represent locations of normalized difference vegetation index (NDVI) measurements, with large values indicating the presence of plants, which thus can be clustered to delineate individual plots. Using coordinates of the measurements and the plot geometry (e.g., plot width and length), plot boundaries can be calculated. In this article, an approach to generate a field map in a GIS before planting is described in case study I, while case study II demonstrates an algorithm to calculate plot boundaries by post-processing of georeferenced phenotypic measurements.

CASE STUDIES

CASE STUDY I

A mobile, field-based HTPP sensor platform (fig. 5) was developed to measure cotton phenotypes at the Maricopa Agricultural Center in Maricopa, Arizona (33.068° N, 111.971° W, 360 m above mean sea level) and used over multiple field seasons. A sensor boom was mounted on the front of a high-clearance tractor (Lee Avenger, LeeAgra, Inc., Lubbock, Tex.). The tractor moved at 1.5 m s⁻¹ on av-

erage. The sensing unit consisted of four sensor sets. Each sensor set contained two IRTs (SI-131, Apogee Instruments, Inc., Logan, Utah), one ultrasonic sensor (Pulsar dB 3, Pulsar Process Measurement Ltd., Worcestershire, U.K.), and one multispectral reflectance sensor (Crop Circle ACS-430, Holland Scientific, Lincoln, Neb.) to measure canopy temperature, crop height, and canopy spectral reflectance, respectively. In addition to the sensors, an RTK GNSS receiver (R6, Trimble, Westminster, Colo.) was installed at the center of the sensor boom. Software for collecting and recording sensor data was deployed on dataloggers (CR3000, Campbell Scientific, Inc., Logan, Utah). Sensor measurements and the GNSS antenna positions were recorded simultaneously in the datalogger at 5 Hz. Georeferencing was conducted by post-processing the raw data transmitted from the datalogger to the computer. A plug-in for the open-source Quantum GIS environment (QGIS, www.qgis.org) was developed to georeference the sensor measurements and locate sensor measurements within plot boundaries.

Within the QGIS plug-in, a preprocessing script conducted georeferencing by (1) converting latitude and longitude coordinates to the UTM coordinate system and (2) calculating coordinate transformations from the RTK GNSS antenna to the sensor positions depending on the vehicle heading. Figure 6 shows the user interface of the preprocessor plug-in. To use the plug-in, a user first inputs the column numbers of key data values in the raw data file. This allows flexibility to interpret data files of diverse formats. The required *X* and *Y* offsets from each sensor to the GNSS antenna can be measured on the vehicle and provided as input to the plug-in. Because the tractor's heading was always identical to its course (i.e., always moving in a forward direction), the vehicle heading was derived from its course, which was acquired from the "track angle" field in the National Marine Electronics Association (NMEA) data string GPRMC output from the GNSS receiver. With vehicle heading and sensor offset distances available, the absolute sensor positions were calculated using equation 1.



Figure 5. Field-based HTPP platform on a Lee Avenger tractor. Four rows of sensor sets are mounted on a hydraulically positioned bar. The GNSS antenna is located at the bar midline.

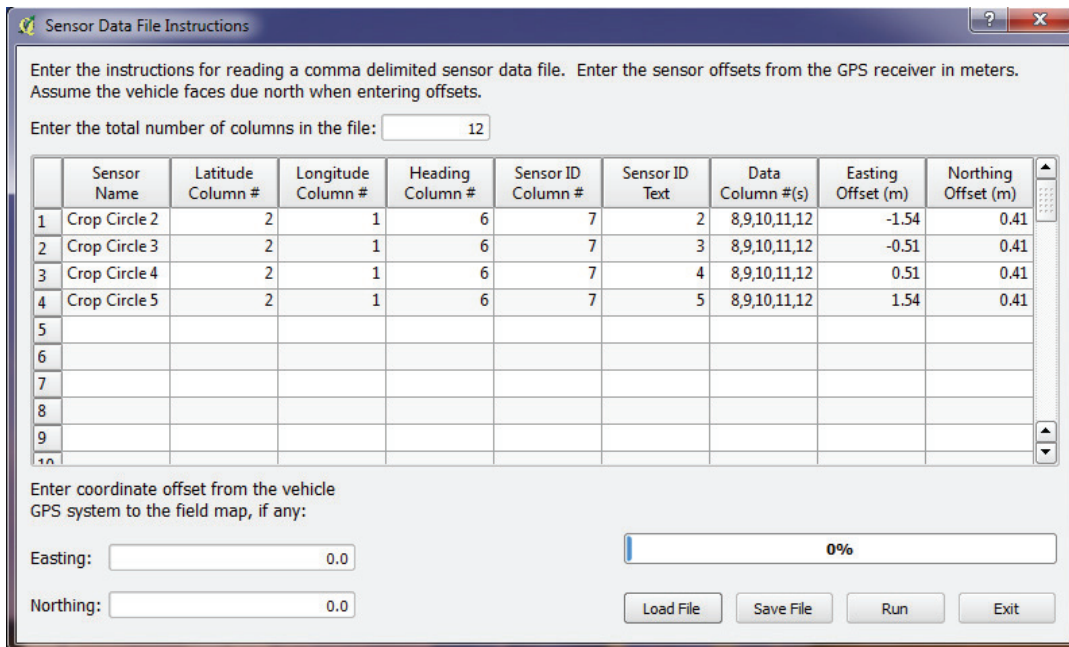


Figure 6. Graphical user interface for the Preprocessor tool within the HTPP Geoprocessor plug-in for Quantum GIS. The Preprocessor converts latitude and longitude coordinates to the UTM coordinate system and calculates sensor positions from the GNSS antenna position depending on vehicle heading.

After phenotypic measurements were georeferenced, each data point was located within the plot boundaries using the Geoprocessor tool (fig. 7) to implement a spatial intersection function within QGIS. In Arizona, field experiments were planned using ArcGIS (ver. 10.2, ESRI, Redlands, Cal.) prior to planting. Digital orthophotos of the Maricopa Agricultural Center from the USGS provided georeferenced information on the field areas available for experiment planning. Coordinates of the corners of the experiment and of other important features (e.g., the locations of buried drip irrigation lines) were surveyed. Using these data, field exper-

iments were planned using ArcGIS tools to generate polygons representing plot boundaries. From this approach, the “A-B lines” (i.e., vehicle courses defined by a smooth and mathematically predefined line between two points, the offset distance between two lines, and the integer of the line number) were generated for use with a tractor auto-guidance system to prepare beds for cotton planting. For field-based HTPP data processing, an added advantage was that the plot map required for locating sensor measurements within plot boundaries was developed during the preplant planning stage. After the georeferenced sensor measurements were loaded into QGIS and represented as a shapefile of points, these field maps, represented as polygon shapefiles, were used with the QGIS Geoprocessor plug-in to identify the sensor measurements that were collected within the boundaries of each field plot. By manually assigning a plot ID as an attribute of each plot map polygon, the Geoprocessor appended the plot IDs to the appropriate measurements in the sensor data layer, hence completing the geospatial processing of the data. The Geoprocessor also calculated summary statistics, such as the mean and standard deviation of phenotypic data collected within each plot, which were appended to the plot boundary layer as polygon attributes.

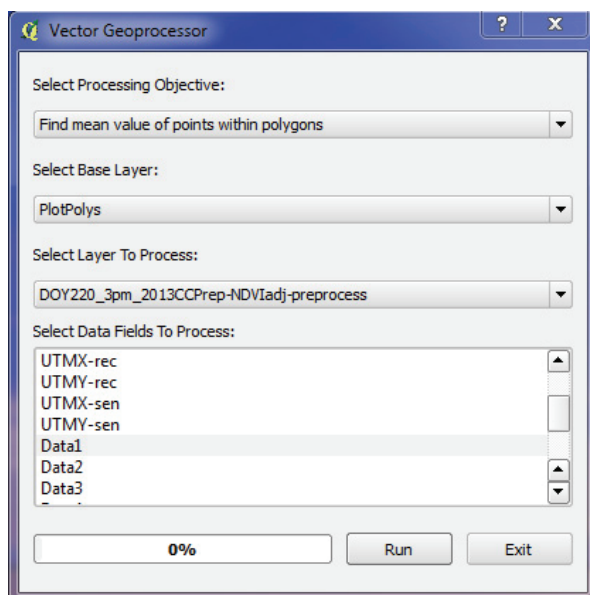


Figure 7. Graphical user interface for the Geoprocessor tool within the HTPP Geoprocessor plug-in for Quantum GIS. The Geoprocessor analyzes sensor data delimited by plot map polygons and assigns plot identifier information to each data point.

Figure 8 shows an example result of georeferenced cotton phenotypic data in QGIS. The point layer described NDVI values obtained from the four Crop Circle sensors on 8 August 2013. The polygon layer described the plot boundaries, which contained unique cotton cultivars planted in two rows with 1.02 m spacing. The attribute table (inset in fig. 8) demonstrated the final processing result, where sensor data positions (UTMX-sen and UTMX-sen columns) and measurement data (NDVI column) were joined with plot ID information (BARCODE column). The entry for this plot was “MD-DC” as provided with the plot ID and also in a separate data column (not shown). The average of NDVI point data

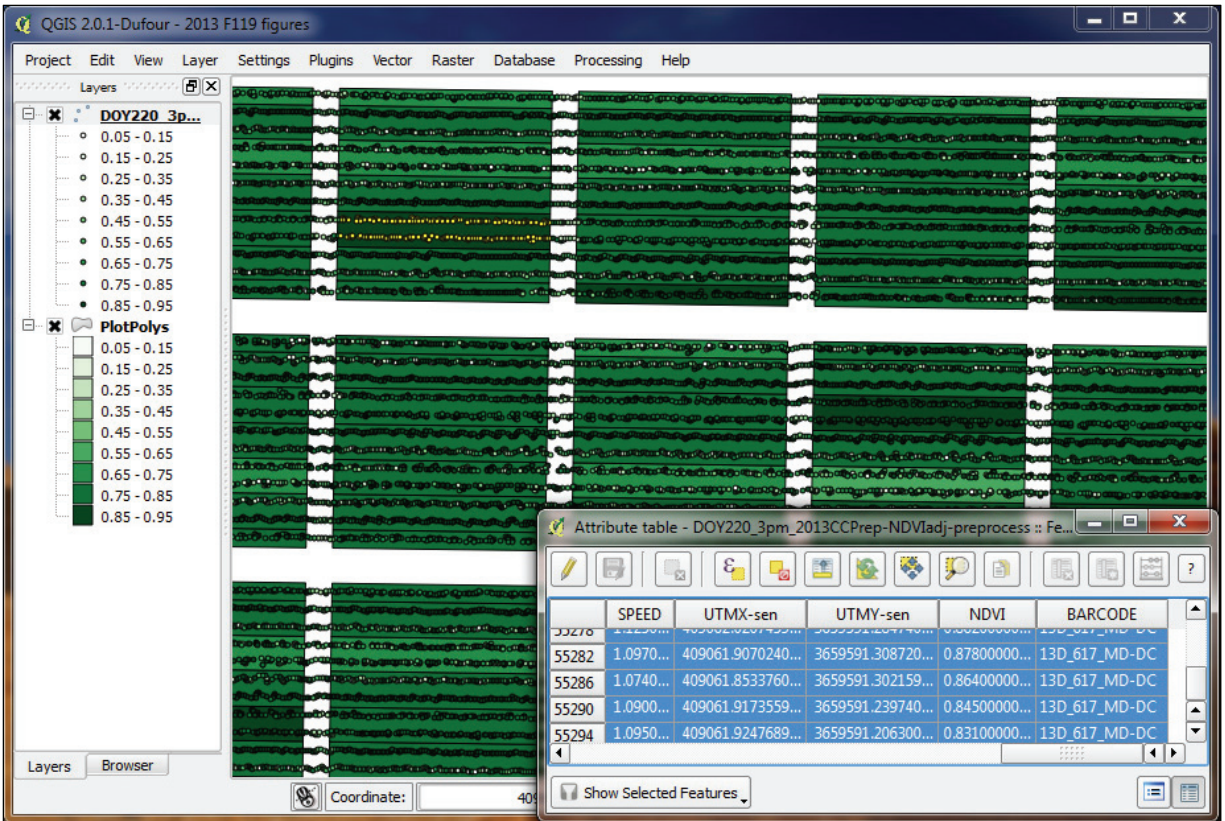


Figure 8. Geoprocessing result for Crop Circle NDVI data collected over cotton with two rows per plot on 8 August 2013 in Maricopa, Arizona. The average of NDVI point data contained by each plot is represented in the plot map color scheme. Plot ID information (BARCODE column) is appended to each record in the attribute table for the Crop Circle point data (bottom right).

from within each plot was represented in the plot map color scheme. Sensor data points highlighted in yellow were collected from a plot with distinctly high NDVI values on this date, as demonstrated by the higher average NDVI value (0.86) of points contained within the plot.

The geospatial data processing approach adopted in case study I was straightforward. However, because a single GNSS receiver could not provide vehicle heading information, sensor positions were subject to greater error when vehicle speed was low. Because the track angle (vehicle course) output from the GNSS receiver is based on a history of recorded positions, noise in the position measurements led to substantial error in track angle calculations when the tractor was stopped or initiating at the start of a row. Subsequently, the calculation of sensor positions (eq. 1) was affected by using these inaccurate data for vehicle heading, and the preprocessor produced circular sweeps in the data layer as though the tractor was spinning around the location of its single GNSS (fig. 9). Because this issue led to positioning of sensor data in the wrong plot, it is a particularly serious problem, but one that can be easily mitigated with more direct vehicle heading measurements. Case study II demonstrates one potential, although more expensive, solution in which vehicle headings were obtained using data from two GNSS receivers on the vehicle.

CASE STUDY II

A phenotyping mobile unit (PheMU) (Barker et al., 2016)

was developed to measure winter wheat phenotypes at Kansas State University (KSU) in Manhattan, Kansas (39.133° N, 96.619° W, 314 m above mean sea level), as shown in figure 10. The PheMU equipment was retrofitted on a high-clearance sprayer (MudMaster, Bowman Manufacturing Co., Inc., Newport, Ark.). The sprayer's maximum moving speed was 2.2 m s⁻¹; however, to control sensor data quality, the moving speed was reduced to around 0.5 m s⁻¹ to reduce vehicle bouncing. The height of the sensor boom was adjustable to collect sensor measurements from crops in different heights. On the boom, three sensor sets were mounted for simultaneous data collection from up to three different rows of crops. Each sensor set contained two IRTs (15:1 CX-SF15-C8 and 2:1 CTH-SF02 IRT, Micro-Epsilon, Raleigh, N.C.), one ultrasonic sensor (U-GAGE Q45U, Banner Engineering Corp., Minneapolis, Minn.), one multispectral crop canopy sensor (Crop Circle ACS-470, Holland Scientific, Lincoln, Neb.), and one GreenSeeker crop sensing system (Trimble, Westminster, Colo.). Each sensor set measured canopy temperature, crop height, and spectral reflectance. In addition to the sensors, two GNSS antennae (AG25, Trimble, Westminster, Colo.) were installed at each end of the sensor boom and connected to two RTK GNSS receivers (FmX integrated display, Trimble, Westminster, Colo.). A top view of the PheMU's sensor and GNSS antenna layout is shown in figure 3 (case study II). Software for collecting and logging sensor data was developed using LabVIEW (National Instruments, Austin, Tex.) and deployed on a lap-

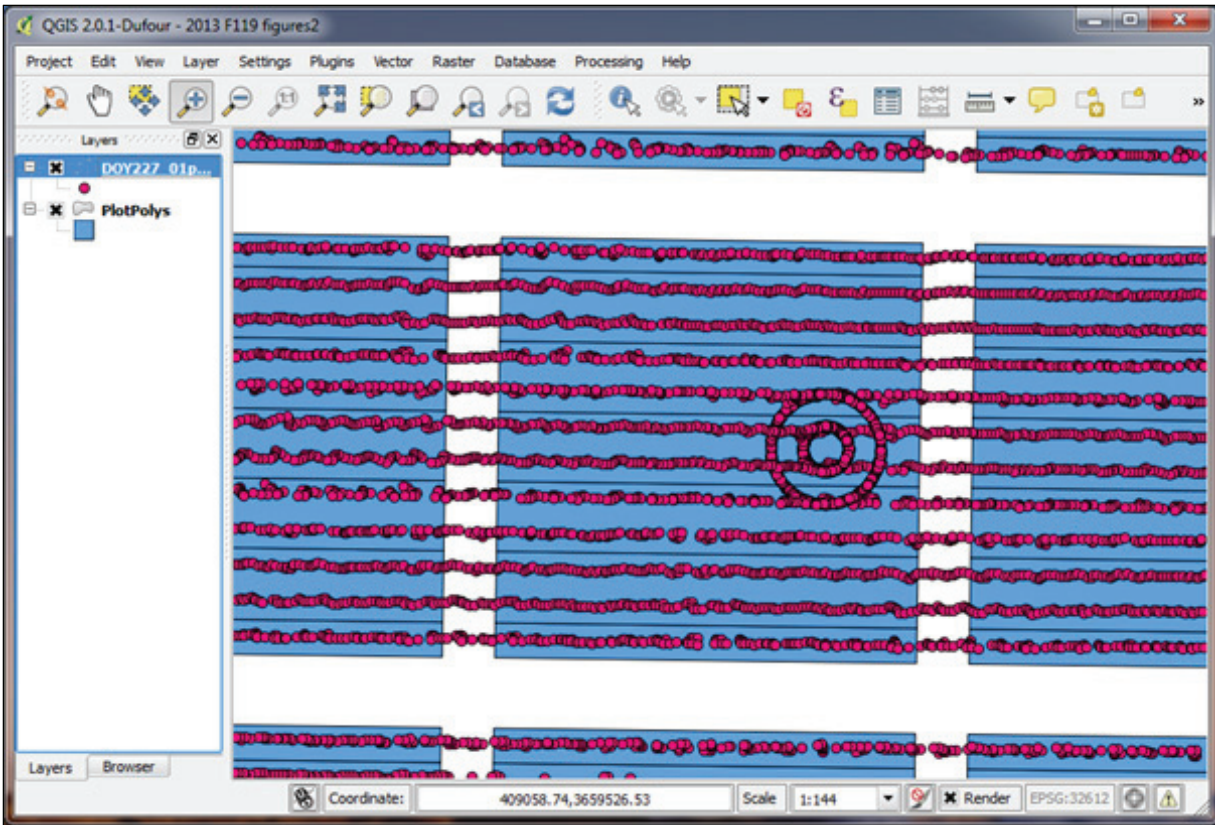


Figure 9. Circular sweeps in the data layer due to error in the heading calculation when the vehicle stops.



Figure 10. Field-based HTPP platform on a MudMaster sprayer. Three rows of sensor sets are mounted on a hydraulically positioned bar. Two GNSS antennas are located at each bar end.

top computer. Measurements from the sensors and GNSS receivers were recorded at 10 Hz. Raw data files were saved on the laptop computer during data collection and uploaded to a server computer for preprocessing. To implement georeferencing and import georeferenced sensor measurements to a field-based HTPP data table in a database, a preprocessing workflow using Java programs was designed to perform similar operations as the preprocessor plug-in for

QGIS (case study I). Field plot IDs with boundary coordinates were imported to a plot map table. Users were able to query the phenotypic data within any field plot by joining the HTPP data table and the plot map table.

In case study II, two RTK GNSS antennae were installed at each end of the sensor boom. Using two coordinates provided by each RTK GNSS receiver at every sampling interval, a vector was derived from one antenna (as the primary)

to the other (as the secondary). The vehicle heading was always perpendicular to this vector. Knowing the orientation of the primary and secondary antennae on the boom, the vehicle heading was calculated. In case study II, the vector direction was from the west side GNSS antenna (primary) to the east side antenna (secondary) with a vehicle heading of 0° (due north), hence generating equation 2:

$$\begin{cases} \cos(\alpha) = (X_{GS} - X_{GM}) / |G_S G_M| \\ \sin(\alpha) = (Y_{GS} - Y_{GM}) / |G_S G_M| \end{cases} \quad (2)$$

where

(X_{GM}, Y_{GM}) = UTM coordinate of the primary GNSS antenna's position (fig. 2b)

(X_{GS}, Y_{GS}) = UTM coordinate of the secondary GNSS antenna's position (fig. 2b)

$|G_S G_M|$ = distance from the primary GNSS antenna to the secondary antenna (fig. 2b)

α = vehicle's heading angle, measured clockwise from true north (fig. 2b).

By using equation 2 together with equation 1, the absolute position of each sensor was calculated. Using two RTK GNSS receivers, the vehicle heading could be calculated directly based on positioning information only and independent of GNSS outputs for vehicle course. Depending on the accuracy of the RTK GNSS receiver, the heading accuracy might be affected by the distance between the two antennae.

Unlike case study I, field maps were not generated using a GIS. The wheat fields rotated among different locations on the research farm, and the field layout changed each field season. Hence, a new field map was created each season. Although the planter was navigated by an RTK GNSS and seed drop was triggered by the GNSS, the true plot locations still had an offset from the triggering position recorded by the planter. Facing these limitations, an approach was developed to calculate plot boundaries by processing the georeferenced phenotypic data. The approach was based on two fundamental conditions: (1) obvious features could be extracted from the sensor measurements to distinguish between data from inside and outside of plots and (2) different plots were spatially separated by bare soil. Figure 11 shows a set of georeferenced NDVIs collected by the PheMU at the KSU Ashland Bottoms Experimental Farm in Kansas. Field plots at this location were planted in a range (row due east-west) by column (row due north-south) layout. The PheMU moved along columns, generating sensor measurements as shown in figure 11. The yellow dots represent high NDVI values that were collected inside individual plots, while the blue dots represent low NDVI values that were located over bare soil outside of plots. According to the range by column design, there was an alley (e.g. bare soil) between two plots in the same column, hence separating two plots, as shown by the two black polygons in figure 11. However, vegetation (weeds or misplanting) might grow in the alley and have similar traits as the crops (e.g., canopy spectral reflectance or plant height), causing an oversized plot, as shown by the green polygon in figure 11. In this case, further processing was required to separate the oversized plot into two plots of the correct size.

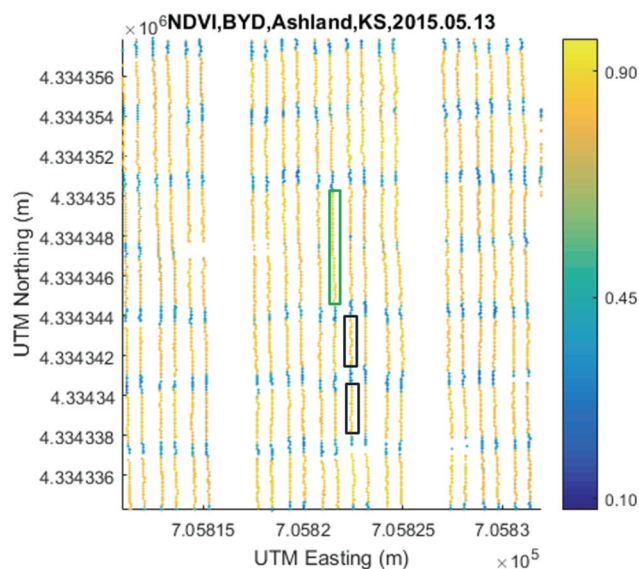


Figure 11. Plot boundaries shown by georeferenced NDVIs.

If the two conditions are satisfied, the algorithm can be applied to calculate the plot boundaries automatically from sensor measurement data. A flow diagram of the algorithm is shown in figure 12. First, a threshold value to differentiate measurements inside and outside the plots was applied to filter out the out-of-plot data (fig. 13b). The threshold value was determined using a histogram of phenotypic measurements. Different threshold values were tested to increase the

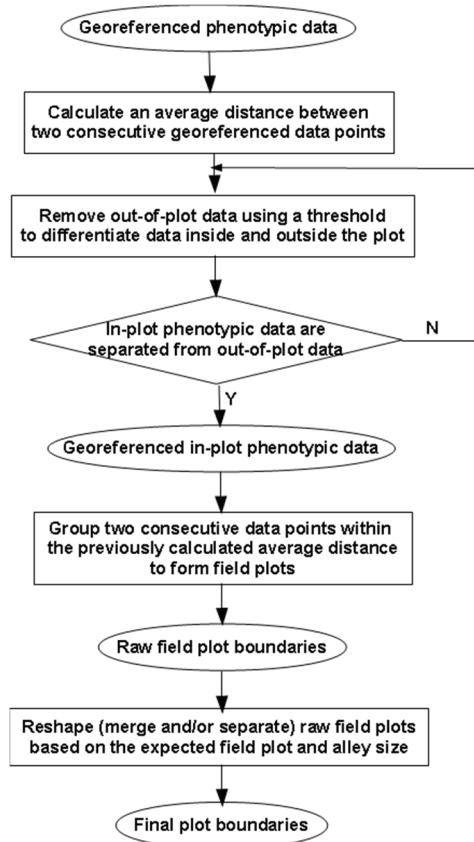


Figure 12. Flow diagram of the algorithm used to calculate plot boundaries by analyzing georeferenced phenotypic data.

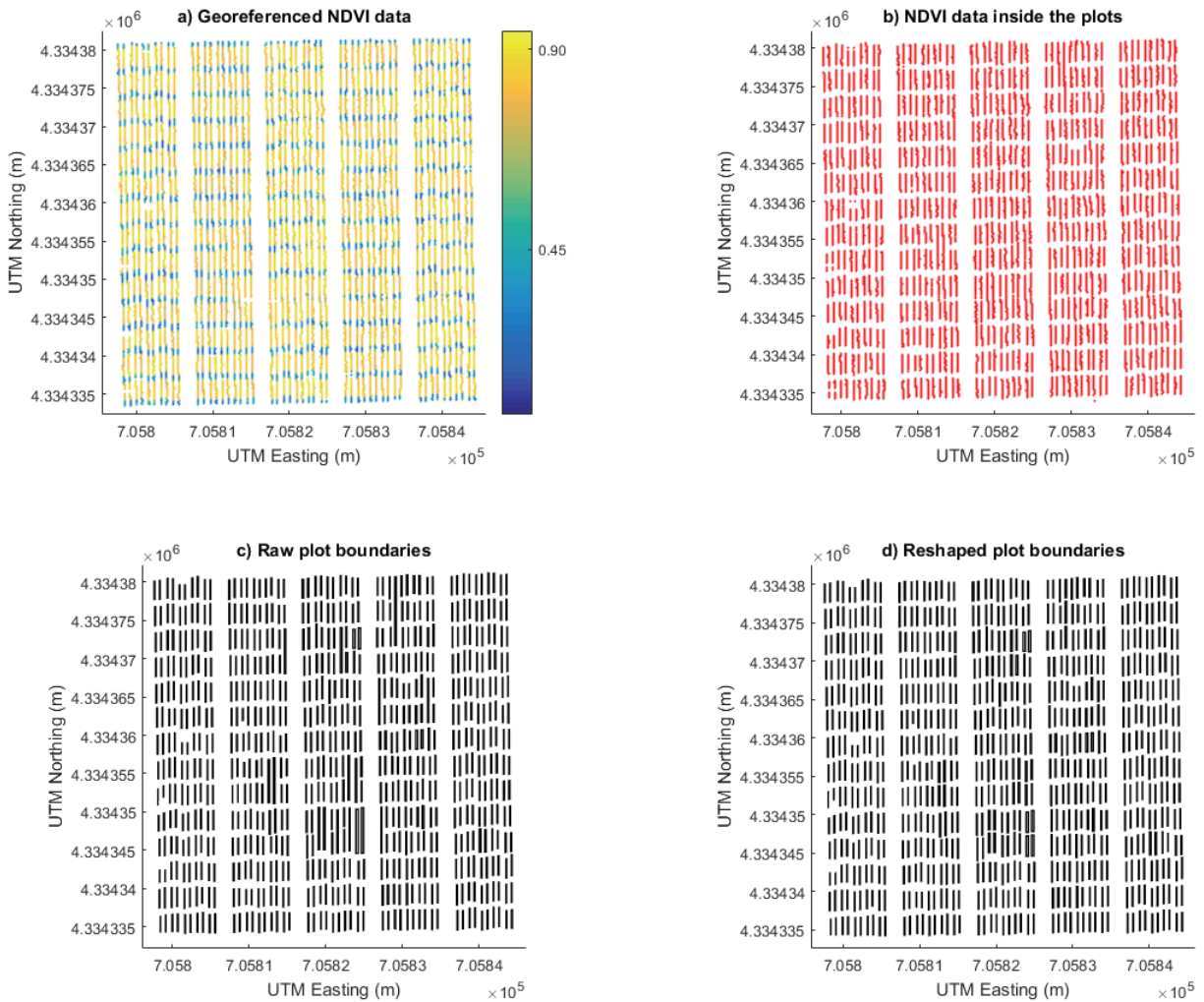


Figure 13. Procedure used to calculate plot boundaries from georeferenced measurements: (a) georeferenced GreenSeeker NDVI data collected over wheat with three rows per plot on 13 May 2015 at Ashland Farm in Manhattan, Kansas, (b) georeferenced NDVI data inside the plots after removing out-of-plot data, (c) raw plot boundaries generated by clustering georeferenced data in each plot area, and (d) final plot boundaries by reshaping the raw plot boundaries.

precision of marked plot boundaries. Because the measurements were collected continuously, two consecutive measurements were close in proximity depending on the vehicle velocity. This average distance between consecutive measurements was calculated before filtering. After removing the out-of-plot data, a distance gap between consecutive plots existed, which was larger than the distance between consecutive points within a plot. In this condition, any two data points with point-to-point distance within this range were grouped; hence, clusters of measurements were generated to form field plots (fig. 13c). Because it was possible that measurements inside the field plot (e.g., low NDVI values from bare soil or from diseased or dead plants) could be removed, several tiny plots might exist in one field plot area. These subplots needed to be merged as one plot. Conversely, joined (oversized) plots needed to be separated to the true plot size. Both merging and separation were performed based on the expected field plot and alley size. Final plot boundaries are shown in figure 13d.

To test the above approach of finding plot boundaries, outputs of the algorithm for two data sets of georeferenced

NDVI measurements were compared to surveyed plot boundaries. In the first test, plot boundaries of a winter wheat field in Kansas were overlaid on plot boundaries surveyed using an RTK GNSS receiver (fig. 14), while the second test compared calculated plot boundaries of a cotton field in Arizona with plot boundaries predefined in ArcGIS (fig. 15). A distribution of the Euclidean distances from the calculated plot centers to the surveyed and predefined plot centers is shown in figure 16. Using the algorithm for the georeferenced NDVI from the winter wheat field in Kansas, 98.4% of the calculated plot centers were within 0.4 m of the surveyed plot centers, and all of the calculated plot centers were within 0.6 m of the surveyed plot centers. For the georeferenced NDVI from the cotton field in Arizona, 84.6% of the calculated plot centers were within 0.4 m of the predefined plot centers, while 1.1% of the calculated plot centers were farther than 1 m from the predefined plot centers. The large offset distances arose from the plot boundaries generated from georeferenced measurements collected in field plots with incomplete sensor measurements or where the crop planting or emergence was incomplete.

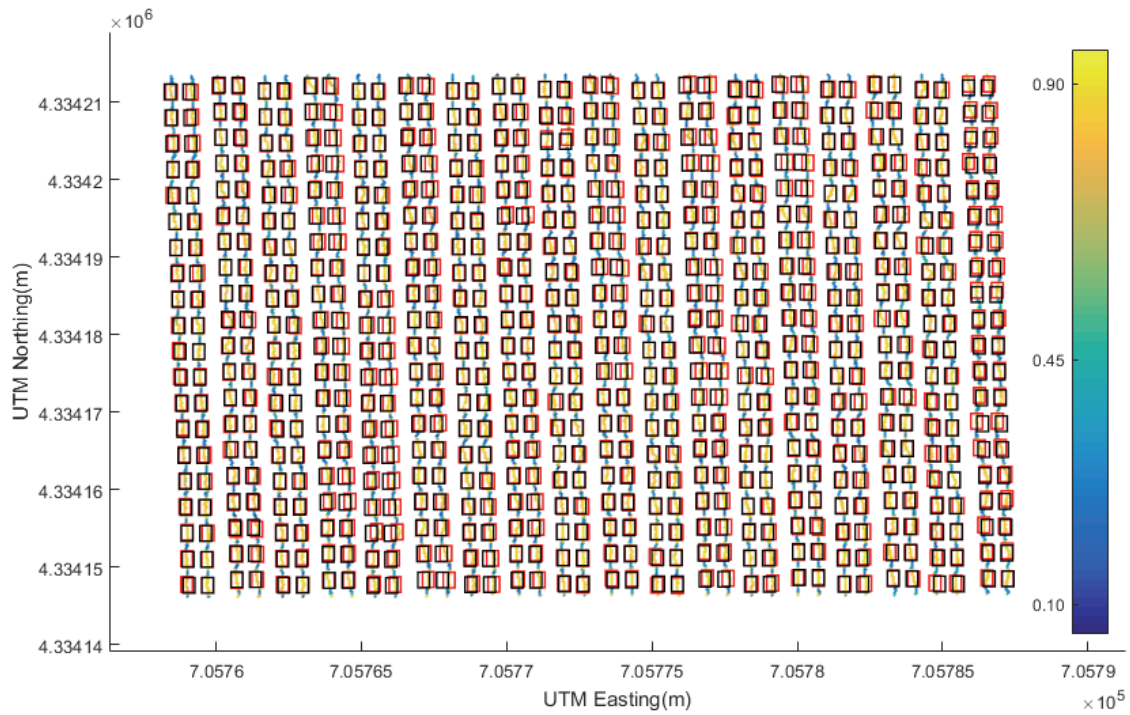


Figure 14. Comparison of calculated versus surveyed plot boundaries for a winter wheat field in Kansas. Calculated boundaries (red rectangles) from analysis of georeferenced NDVI are overlaid on surveyed plot boundaries (black rectangles).

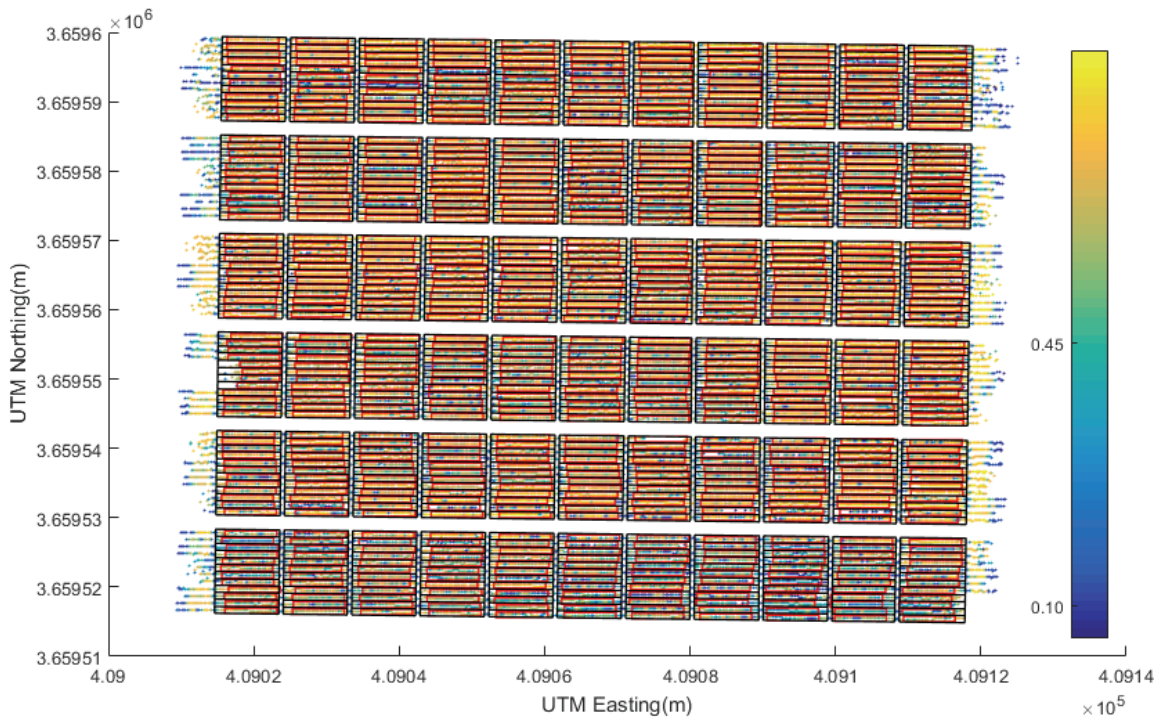


Figure 15. Comparison of calculated versus predefined plot boundaries for a cotton field in Arizona. Calculated boundaries (red rectangles) from analysis of georeferenced NDVI are overlaid on plot boundaries (black rectangles) predefined in ArcGIS.

Different from the preplanned field map developed in ArcGIS, the plot boundaries calculated by the above algorithm only delineated areas where measurements were taken and therefore may not delineate the whole field plot. As shown in figure 17, for example, the black polygons show the calculated plot boundaries, while the blue polygons show the real field plots with plot IDs. It is likely that only black polygons

are found after field-based HTPP data collection, but those polygons do not delineate the whole field plot area. This condition happens when the sensors travel through different paths over a multi-row plot during different experiments. A feasible solution for this case is to generate a field map by overlaying plot boundaries calculated from multiple field-based HTPP data collection outings and calculating the outermost bounda-

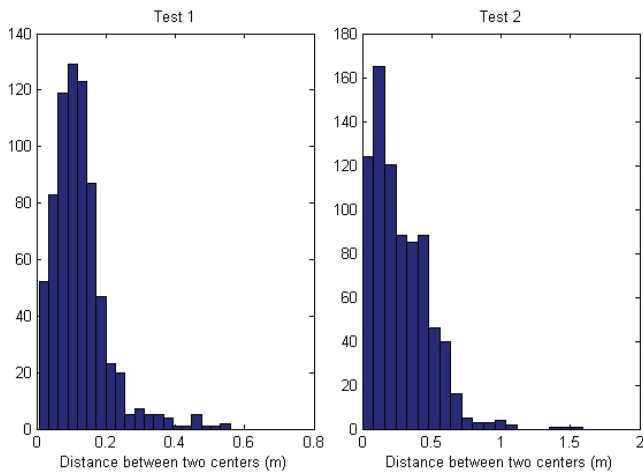


Figure 16. Distribution of distances between calculated plot centers and surveyed (test 1, Kansas) and predefined (test 2, Arizona) plot centers.

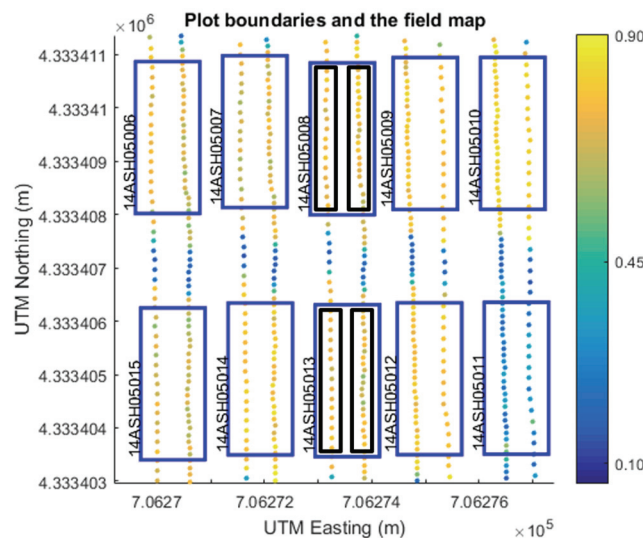


Figure 17. Calculated plot boundaries (black) compared with actual plot boundaries (blue) for the Kansas wheat study.

ries. Once generated, the field map can be applied to all data sets. For example, phenotypic data collected at different crop stages (e.g., seedling stage and lodging stage) may not have sufficient features to clearly reflect the plot boundaries, which limits use of the above algorithm. In this case, the complete field map can be applied to analyze plot-level data. In addition, the field map generated using this method could be more accurate than the predefined field map in GIS in some special conditions, such as imprecise planting due to malfunctions or operator error, causing portions of plots to be bare. In this case, phenotypic data collected over non-plant material would still be accounted for within the predefined field map. On the other hand, the plot boundaries calculated from the measurements themselves would only contain measurements above the defined threshold (fig. 13).

CASE STUDY III

A low-cost, cart-based platform was developed for field-based HTPP at the U.S. Arid Land Agricultural Research Center in Maricopa, Arizona (White and Conley, 2013). The cart was built from two bicycles and additional framing to



Figure 18. Field-based HTPP platform on a low-cost cart.

support sensors (fig. 18). The sensing unit contained three IRTs (SI-111, Apogee Instruments, Inc., Logan, Utah), one ultrasonic sensor (943-F4Y-2D, Honeywell International, Inc., Morristown, N.J.), and two multispectral sensors (Crop Circle SCS-470s, Holland Scientific, Lincoln, Neb.) to measure canopy temperature, crop height, and canopy spectral reflectance, respectively. A GPS receiver (A101, Hemisphere GNSS, Inc., Scottsdale, Ariz.) with Wide Area Augmentation System (WAAS; a correction signal to improve positioning accuracy to better than 3 m and available in North America) was installed for georeferencing sensor measurements. Software for collecting and recording sensor data was deployed on a datalogger (CR3000, Campbell Scientific, Inc., Logan, Utah). Measurements from the sensors and GPS receivers were recorded at 5 Hz. The cart can be moved at a regular walking speed of about 1.0 m s^{-1} .

Unlike the previous two vehicle platforms, the cart was steered manually. To avoid shadows from entering the sensor fields of view, the cart heading was constant (due south) while traversing the field. On one pass, the cart was pushed, meaning the vehicle heading and course were equal. On the next pass, the cart was pulled in the reverse direction, meaning the vehicle heading was opposite to the course. Thus, the cart was operated similarly to a gantry system, traversing the field without changing heading. In this and similar cases, using information on vehicle course would lead to erroneous sensor positions because the vehicle heading angle would be 180° different from its course angle on alternating passes. One solution is to log additional data to mark the passes where course and heading are equal, so that course can be adjusted by 180° when appropriate. After calculating absolute positions of each sensor measurement, plot-level data were analyzed using the predefined field map in GIS. For example, in case study III, one Crop Circle multispectral sensor had a Y offset of 1.32 m to the GPS antenna. During a phenotyping experiment in a wheat field in Arizona, the cart was operated by maintaining a southward heading, while the cart course alternated between southward (pushing the cart) and northward (pulling the cart) directions in consecutive passes. Thus, the vehicle course and heading were 180° different for alternating passes. Using equation 1 with cart heading information, the NDVI measure-

ments were correctly georeferenced, and the data overlaid accurately on the plot map (fig. 19a). However, if the georeferenced NDVI was calculated only using the vehicle course, the data were not correctly georeferenced for passes with opposing vehicle course and heading. Therefore, the measurements were assigned an erroneous plot ID or located out of the field boundary, causing data loss (fig. 19b).

DISCUSSION

GEOREFERENCING

Georeferencing is the first step for processing field-based HTPP data. As long as an offset exists between the sensor and the GNSS antenna on the vehicle platform, georeferencing must be performed to calculate the sensor's absolute position. A key, but easily overlooked, step for georeferencing is to obtain the vehicle heading. The sensor's absolute position errors due to errors from heading measurements are shown in equation 3:

$$\left\{ \begin{array}{l} \Delta X_{Sabs} = \sqrt{X_S^2 + Y_S^2} \\ \quad \times \left[\sin\left(\alpha + \sin^{-1}\left(X_S / \sqrt{X_S^2 + Y_S^2}\right) + \Delta\alpha\right) \right. \\ \quad \left. - \sin\left(\alpha + \sin^{-1}\left(X_S / \sqrt{X_S^2 + Y_S^2}\right)\right) \right] \\ \Delta Y_{Sabs} = \sqrt{X_S^2 + Y_S^2} \\ \quad \times \left[\cos\left(\alpha + \sin^{-1}\left(X_S / \sqrt{X_S^2 + Y_S^2}\right) + \Delta\alpha\right) \right. \\ \quad \left. - \cos\left(\alpha + \sin^{-1}\left(X_S / \sqrt{X_S^2 + Y_S^2}\right)\right) \right] \end{array} \right. \quad (3)$$

where

ΔX_{Sabs} and ΔY_{Sabs} = errors of the sensor's absolute position
 X_S and $Y_S = X$ and Y offset distances from the sensor to the GNSS antenna (fig. 2a)

α = vehicle's heading angle, measured clockwise from true north (fig. 2a)

$\Delta\alpha$ = heading error, measured clockwise from true north.

As vehicles are typically equipped with only one GNSS receiver, the approach for georeferencing described in case study I is more common, but issues with GNSS noise at low travel speeds must be overcome (fig. 9). One solution to this issue is to eliminate measurements when the vehicle velocity is below a speed threshold, but this results in loss of raw data. Another solution is to estimate the vehicle heading at the current position from a set of previous positions (i.e., using a longer course). This solution works when vehicles travel straight. An alternative is to obtain the heading from additional sensors. An inertial guidance sensor can provide roll-pitch-yaw motions of the sensor boom for linking this information with positions provided by GNSS to perform georeferencing. As high-precision GNSS receivers become less expensive, a dual-antenna GNSS receiver becomes another potential option to provide both vehicle position and heading in real-time. The constraints of each approach to obtain vehicle heading are shown in table 1.

FIELD MAP GENERATION

In case studies I and III, a field map was designed in GIS and used to develop coordinates for tractor navigation during field prep operations. The predefined field map could thus be used to locate georeferenced measurements in the correct field plot and could be used over the whole season. If the

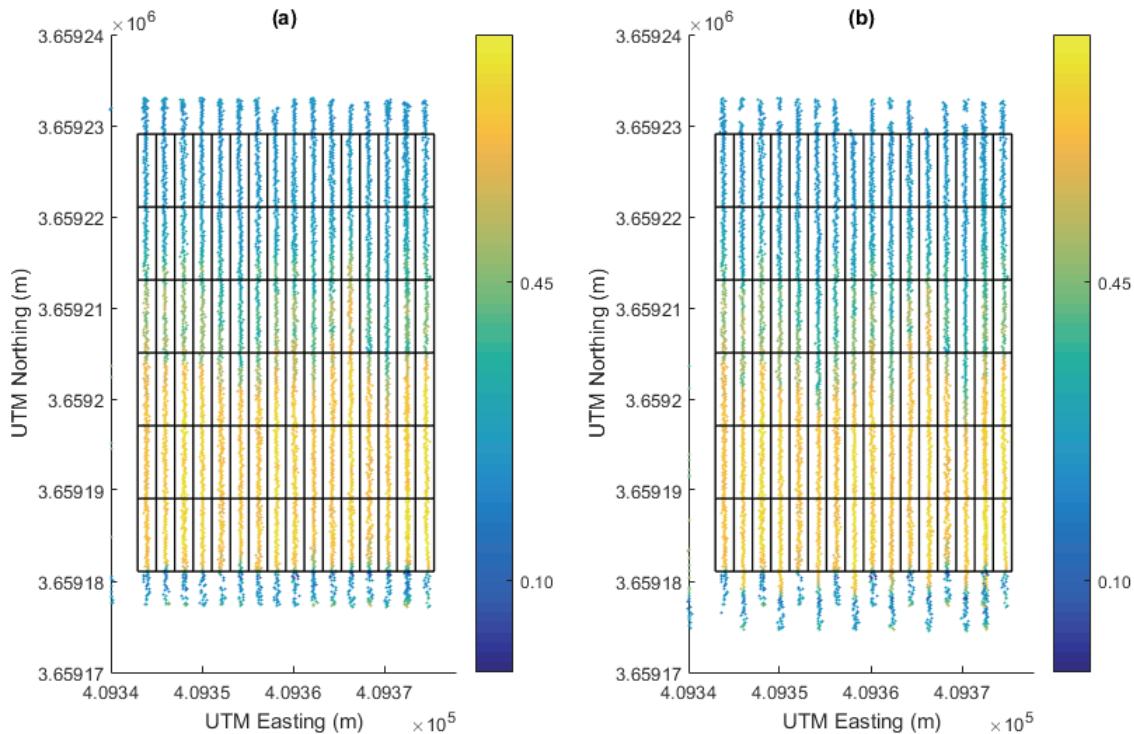


Figure 19. Comparison between georeferenced NDVI of a wheat field in Arizona calculated using (a) vehicle heading data converted from the vehicle course and (b) vehicle course data. Grids in black represented field plot boundaries. Note that in (a) survey extents are aligned, while in (b) they are staggered.

Table 1. Approaches to obtain vehicle heading and constraints.

| Type of GNSS Receiver | Approaches to Obtain the Vehicle Heading | Constraints |
|-----------------------|--|--|
| Single antenna | Use track angles from GPRMC strings | Noise at low travel speeds |
| Single antenna | Calculate time series averaging from more historic positions | Inaccurate when vehicles travel non-straight |
| Single antenna | Use yaw angles from an inertial guidance sensor | Affected by magnetic interferences |
| Dual antenna | Calculate heading angles from two antennae's positions | High expenses |

field map was not generated, the algorithm introduced in case study II could calculate field plot boundaries based on the georeferenced data. However, this algorithm is only useful if the two conditions mentioned in case study II are met. Therefore, field-based HTPP teams are encouraged to develop field-surveying expertise as a component of their data processing workflow.

In all three case studies, the field plots were designed with rectangular shapes. Irregular plot shapes (e.g., arc-shaped plots to accommodate the path of a center-pivot system) are common at some research stations and for on-farm research. For field-based HTPP technology to be applied in those fields, methods to describe the plot shape need to be improved in both field map generation approaches.

All three ground vehicles in the case studies used GNSS for georeferencing. Although RTK GNSS can offer centimeter-level positioning information, the accuracy is relative and relies on the RTK base station measurements. If the base station's position can be surveyed and has centimeter-level accuracy, the calculated absolute sensor positions will have the same level of accuracy. If the base station position cannot be measured, it should be located at an identical position during planting and collecting of field-based HTPP data, so the positioning information maintains the same level of accuracy and relative correction. In addition, the same GNSS, including the base station and the rover receiver, should be used for field surveying, field prep operations, and HTPP data collection. Otherwise, the sensor measurements may not align with the field map.

In the last two decades, remote sensing technology has been used for precision agriculture research. High-resolution cameras held by unmanned aerial vehicles are able to acquire crop images that can be stitched to generate a georeferenced orthomosaic image of a complete field using commercial software (e.g., Agisoft). Potentially, a field map can be generated from this georeferenced orthomosaic image through image processing methods. The resulting field map should be more accurate than the field map generated from sensor measurements because images provide more spatial information. In addition, compared to the predefined field map in GIS, the field map generated from this georeferenced orthomosaic image should reflect a more realistic plot distribution when there are issues of stand density or unforeseen planting errors relative to the intended plot locations. Therefore, this approach will be investigated in further research.

CONCLUSIONS

Precision agriculture and proximal sensing technologies are central to developing the research area of field-based HTPP measurements. How to efficiently and precisely process large phenotypic data sets remains a major challenge.

Geospatial data processing is essential for linking sensor outputs to treatments, including varieties and management practices that were imposed at a given location in the field. The approaches introduced in this article provide solutions to two common problems: accurately georeferencing sensor data and locating the data within plot boundaries. Sensor positions can be mapped in different but equally viable ways depending on the platform and GNSS equipment. For single-antenna systems, we suggest either time series averaging or an additional electronic compass to maintain accurate heading information at all travel speeds and platform courses. For dual-antenna systems, a formula has been provided to compute accurate headings. Delineating plot boundaries can be done programmatically using GIS to define all plot corners prior to planting, or it can be done in-season by using the sensor data to identify plot edges. The latter approach could be especially valuable when actual plant emergence patterns differ greatly from original designs or when researchers lack the resources needed for GIS layout. The three case studies described in this article demonstrate the HTPP geospatial data processing solutions used for two years in Arizona and Kansas that have accelerated the link between plant traits and plant variety. These approaches can have broad application to a range of ground vehicles for field-based HTPP.

ACKNOWLEDGEMENTS

This work was supported through the NSF Plant Genome Research Program (PGRP) (Award No. IOS-1238187), the USDA Agricultural Research Service, the Kansas Wheat Commission and Kansas Wheat Alliance, and the U.S. Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051).

REFERENCES

- Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., ... White, W. (2013). Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.*, *41*(1), 68-79. <http://dx.doi.org/10.1071/FP13126>
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.*, *19*(1), 52-61. <http://dx.doi.org/10.1016/j.tplants.2013.09.008>
- Barker III, J., Zhang, N., Sharon, J., Steeves, R., Wang, X., Wei, Y., & Poland, J. (2016). Development of a field-based high-throughput mobile phenotyping platform. *Comput. Electron. Agric.*, *122*, 74-85. <http://dx.doi.org/10.1016/j.compag.2016.01.017>
- Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., ... Ruckelshausen, A. (2013). Breedvision: A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors*, *13*(3), 2830-2847. <http://dx.doi.org/10.3390/s130302830>
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch,

- S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.*, *126*(4), 867-887. <http://dx.doi.org/10.1007/s00122-013-2066-0>
- Furbank, R. T., & Tester, M. (2011). Phenomics: Technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.*, *16*(12), 635-644. <http://dx.doi.org/10.1016/j.tplants.2011.09.005>
- White, J. W., & Conley, M. M. (2013). A flexible, low-cost cart for proximal sensing. *Crop Sci.*, *53*(4), 1646-1649. <http://dx.doi.org/10.2135/cropsci2013.01.0054>
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., ... Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Res.*, *133*, 101-112. <http://dx.doi.org/10.1016/j.fcr.2012.04.003>