Published online November 22, 2019

The Plant Phenome Journal

Original Research

In-Field Whole-Plant Maize Architecture Characterized by Subcanopy Rovers and Latent Space Phenotyping

Joseph L. Gage,* Elliot Richards, Nicholas Lepak, Nicholas Kaczmar, Chinmay Soman, Girish Chowdhary, Michael A. Gore, and Edward S. Buckler

Core Ideas

- Subcanopy rovers enabled 3D characterization of thousands of hybrid maize plots.
- Machine learning produces heritable latent traits that describe plant architecture.
- Rover-based phenotyping is far more efficient than manual phenotyping.
- Latent phenotypes from rovers are ready for application to plant biology and breeding.

Collecting useful, interpretable, and biologically relevant phenotypes in a resource-efficient manner is a bottleneck to plant breeding, genetic mapping, and genomic prediction. Autonomous and affordable subcanopy rovers are an efficient and scalable way to generate sensor-based datasets of in-field crop plants. Rovers equipped with lidar can produce three-dimensional reconstructions of entire hybrid maize (*Zea mays* L.) fields. In this study, we collected 2103 lidar scans of hybrid maize field plots and extracted phenotypic data from them by latent space phenotyping. We performed latent space phenotyping by two methods, principal component analysis and a convolutional autoencoder, to extract meaningful, quantitative latent space phenotypes (LSPs) describing whole-plant architecture and biomass distribution. The LSPs had heritabilities of up to 0.44, similar to some manually measured traits, indicating that they can be selected on or genetically mapped. Manually measured traits can be successfully predicted by using LSPs as explanatory variables in partial least squares regression, indicating that the LSPs contain biologically relevant information about plant architecture. These techniques can be used to assess crop architecture at a reduced cost and in an automated fashion for breeding, research, or extension purposes, as well as to create or inform crop growth models.

The cost to genotype a population of plants has become increasingly affordable since the advent of next-generation sequencing, leaving the acquisition of high-quality phenotypic data as a limiting step in conducting genetic mapping studies and training genomic prediction models. Sensors and imaging devices with the ability to capture terabytes of data, rather than providing a solution to the phenotyping bottleneck, can compound the problem by producing hundreds or thousands of non-independent phenotypes for downstream analysis or, at worst, producing massive datasets that are never used to their full potential. Novel data acquisition methods require custom computational methods for extracting interpretable, useful, and biologically relevant traits from this deluge of digital data.

The massive quantity of data produced by image- and sensor-based phenotyping methods has created new challenges regarding data management and analysis (Omasa et al., 2007; Houle et al., 2010; Minervini et al., 2015). Many analysis methods for novel phenotyping methods are laboratory oriented, site specific, or cannot be directly applied to crops in situ (Araus et al., 2018). To be useful, phenotypes derived from new imaging and sensor methods need to be applicable in breeding programs or useful for gaining biological insights (Cobb et al., 2013). Many recently developed high-throughput phenotyping platforms are immobile (Virlet et al., 2017; Bai et al., 2019) or mounted on tractors (Busemeyer et al., 2013; Andrade-Sanchez et al., 2014; Salas Fernandez et al., 2017; Sun et al., 2018). Such solutions are viewed with some skepticism by plant breeders, who prioritize flexible, affordable approaches to survey large populations of germplasm in numerous growing environments (Araus et al., 2018). An ideal phenotyping system

J.L. Gage, N. Lepak, and E.S. Buckler, USDA–ARS, Ithaca, NY 14853; J.L. Gage and E.S. Buckler, Institute for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; E. Richards, Dep. of Mechanical Engineering, Drexel Univ., Philadelphia, PA 19104; E.S. Buckler, N. Kaczmar and M.A. Gore, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell Univ., Ithaca, NY 14853; C. Soman, EarthSense, Inc., Champaign, IL 61820; G. Chowdhary, Dep. of Agricultural and Biological Engineering, Univ. of Illinois, Champaign, IL 61801.

Supplemental material online.

© 2019 The Author(s). This is an open access article distributed under the terms of the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Plant Phenome J. 2:190011 (2019)
doi:10.2135/tppj2019.07.0011

Received 29 July 2019.
Accepted 5 Nov. 2019.
*Corresponding author (jlg374@cornell.edu).

Abbreviations: LSP, latent space phenotype; PC, principal component; PCA, principal component analysis; PLSR, partial least squares regression.

should have high mobility and flexibility, high sensor capacity, be easily scaled up, and be cheap (Fig. 1). Unmanned aerial vehicles (UAVs) and ground rovers meet most of those needs. However, UAVs can be difficult to fly in poor weather (Yang et al., 2017) and are limited in their ability to characterize plant architecture below and within the canopy, especially late in the growing season (Sun et al., 2018). Equipped with RGB cameras and lidar, rovers can be inexpensive relative to existing automated phenotyping systems, making them scalable and flexible enough to characterize large populations in many locations. Because they are small enough to fit between conventionally planted rows of maize, they can be used to characterize plant architecture from a perspective that is not accessible by most existing phenotyping systems (Mueller-Sim et al., 2017; Higuti et al., 2018; Kayacan et al., 2018; Stager et al., 2019). Lidar sensors enable three-dimensional reconstruction of field plots, which can be used to characterize plant architecture and biomass distribution.

Three-dimensional reconstructions of in-field crop plots, though information rich, also present challenges for analysis. Methods for analyzing image or sensor data are often created for a specific purpose or study and can be difficult for external groups to apply to newly acquired datasets. These issues can occur due to threshold or parameter specification, image settings or modality, or ad hoc solutions that do not translate well to new datasets (Ubbens et al., 2019). Latent space phenotyping, described by Ubbens et al. (2019), leverages the flexibility of machine learning to address some of the inherent difficulties in analyzing and extracting phenotypes from large, complex, sensor- or image-based datasets. Rather than develop methods to measure human-defined traits such as plant height or leaf number, Ubbens et al. (2019) described latent space phenotyping as a way to generate phenotypes that are an "abstract learned concept of the response to treatment, inferred automatically from the image data using deep learning" (Ubbens et al., 2019). Described another way, latent space phenotyping allows a computer to process images or sensor-based data in such a way as to produce novel traits that distinguish individuals based on treatments. As an example, Ubbens et al. (2019) used latent space phenotyping on RGB images of well-watered and water-limited *Setaria viridis* (L.) P. Beauv. plants from a biparental population to create and evaluate computer-generated traits that distinguish the two treatments based on characteristics learned from the images. They used those traits for quantitative trait locus (QTL) mapping and were able to reproduce discovery of a QTL related to water use (Ubbens et al., 2019).

In this study, we expanded their definition to also include principal component analysis (PCA), which provides a way to create latent phenotypes without building machine learning models. We evaluated both PCA and a machine learning model called an autoencoder, both of which generate latent space phenotypes (LSPs) that, rather than distinguishing among treatments, distinguish among different cultivars of hybrid maize. This approach to latent space phenotyping is different from that originally outlined by Ubbens et al. (2019). Their approach to latent space phenotyping uses machine learning models to embed plant images into a latent space and subsequently predict treatment or control status from the embedding. This end goal of predicting a treatment vs. control status enables the model to learn latent representations of the images that contain features that are specifically informative of the difference between the treatment individuals and the control individuals. The methods we developed in this study, however, are a form of dimensionality reduction without label predictions at the end. Because we are only reducing the dimensionality of the images, and not predicting specific traits, our dimensionality reduction is likely to capture major differences between the images related to numerous different traits.

Principal component analysis is a method of dimensionality reduction that creates latent variables from linear combinations of the original data. These latent variables, or principal components (PCs), are ordered, orthogonal to each other, and greedy such that early PCs capture more variability than later PCs. By using PCA, an 8192-dimensional image (64 pixels by 128 pixels), where many of the dimensions are correlated with each other (e.g., adjacent pixels), can be compressed to a reduced-dimensional representation described by a chosen number of independent PCs. As such, PCA serves as a way to both reduce dimensionality and impose independence between variables.

Autoencoders are a type of unsupervised neural network that can be trained to learn a reduced representation of their input (Rumelhart et al., 1986; Baldi, 2012). Conceptually similar to various image compression methods, this reduced representation

	Mobility	Cost	Sensor Capacity	Scalability
Manual Measurement	High	Medium	Low	Medium
Gantry	None	High	High	Low
Tractor-Mounted	Medium	High	High	Low
UAV	High	Low	Medium	High
Rover	High	Low	Medium	High

Fig. 1. Comparison of phenotyping methods. To be effectively implemented in breeding programs or large-scale biological studies, phenotyping technologies need to be mobile, cheap, and scalable while enabling collection of relevant data by onboard (or in-hand) sensors. Rovers and unmanned aerial vehicles (UAVs) meet most of those requirements. For within- and below-canopy phenotyping, rovers have an advantage over UAVs, which can only characterize trials from above.

can be used to reconstruct, with some loss of information, the original input. The difference between the original input and the reconstructed version can be used as a heuristic to train the network. Depending on their architecture, autoencoders can produce results similar to PCA. By adding convolutional layers and nonlinear activations to the autoencoder, however, it can capture different relationships between image elements than PCA, which is based entirely on linear combinations of pixel values.

One of the main challenges in producing LSPs from lidar data is to extract signal while ignoring noise. Data from each plot consist of hundreds of thousands of data points, and the data are inherently noisy due to rover position and velocity during data acquisition, plant movement, and sensor error. Useful LSPs will contain information about plant architecture and plot-level biomass distribution regardless of rover position, wind speed and direction, or other factors that contribute noise. Because of this, the objective of the autoencoder in this study was no longer to accurately recreate the input but to recreate the elements of the input that are relevant to plant biomass while excluding or "ignoring" signal that comes from unwanted sources of variability.

In this study, we evaluated both PCA and an autoencoder as methods to extract LSPs in an unsupervised manner from in-field lidar scans of hybrid maize plots. We calculated heritability for the LSPs and compared them with phenotypes that were measured manually, showing that LSPs have similar heritability to traditionally evaluated phenotypes. Latent space phenotypes were then used as explanatory variables to predict manually measured phenotypes, demonstrating that LSPs, though not directly interpretable, contain information about plant architecture and biomass distribution. By demonstrating heritability and relevance to plant architecture and biomass distribution, we show that LSPs are a useful tool with applications to plant breeding and biology.

Materials and Methods

Germplasm and Manual Phenotyping

The field experiments used in this study were the 2018 Genomes to Fields NYH2 and NYH3 locations. Genomes to Fields is a multiyear, multi-location coordinated experiment that was planted at 31 unique locations across the United States and Canada in 2018 (Gage et al., 2017; AlKhalifah et al., 2018; Lawrence-Dill et al., 2019). The NYH2 and NYH3 locations used for this study consisted of 1600 two-row plots containing 890 unique maize hybrids planted in a modified randomized complete block design in Aurora, NY. The plots were 5.33 m long with 0.76-m space between rows and 1.07-m alleys between ranges. The field was laid out as two experiments, NYH2 and NYH3, which were each 25 plots (50 rows) wide and 32 plots (ranges) long. Twenty hybrid checks and 84 photoperiod introgression lines crossed to LH123Ht were planted in both NYH2 and NYH3. The remainder of each field consisted of three inbred biparental populations sharing a common parent (Mo44 × PHW65, PHN11 × PHW65, PHW65 × MoG; Gage et al., 2018), crossed to either PHT69 (in NYH2) or LH195 (in NYH3). Manual measurements were recorded for days to silk, days to anthesis, plant height, ear height, stand count, root lodging, stalk lodging, grain moisture, plot weight, test weight, total leaf count, and leaf number up to the primary ear. All traits except the last two were evaluated according to the Genomes to Fields standard operating procedure (https://www.genomes2fields.org/resources/), with plant height and ear height measured on a single representative plant per plot. Leaf count traits were performed at maturity by counting leaves on one representative plant per plot.

Rover Design and Sensors

The TerraSentia rover is a compact, autonomous, and teachable robot designed by Earthsense, Inc. (www.earthsense.co) for high-throughput phenotyping beneath the canopy of agricultural crops. It is battery operated and has an onboard Intel i7 computer with 500 GB SSD and 16 GB RAM as well as a separate autonomy computer. The rover is equipped with three RGB cameras and two planar Hokuyo UST-10LX lidar that record continuously as it drives through a field (Kayacan et al., 2018). Furthermore, it fits between rows in standard maize fields and can autonomously follow rows using an embedded lidar (Higuti et al., 2018).

Data collection by the rovers took place after all lines had reached physiological maturity (flowering) but before harvest. All data were collected on 4, 7, 12, 19, or 20 Sept. 2018. Rovers were driven along a single column of the field at a time, passing between the rows of two-row plots such that the data recorded on both sides of the rover at any given time corresponded to the same hybrid and the same plot. Data belonging to discrete plots within each column of the field were separated based on manually recorded time stamps of the video feed from the side-facing cameras. Incomplete passes of the field, and plots that were not driven through continuously, were manually removed from the dataset. The cleaned dataset (including border plots; numbers in parentheses exclude border plots) consisted of 2103 (1972) discrete plot-level records from 1153 (1083) unique plots (many plots were driven through and recorded more than once) representing 698 (697) unique hybrids. Border plots were included in the process of calculating LSPs because they contain useful data that can be used during PCA and autoencoder training. They were not included in subsequent heritability calculations or predictions of manually measured phenotypes.

Cleaning and Preprocessing Lidar Data

Lidar data for each plot were recorded every 25 ms. Each time point was recorded as a vector of distances at 1080 discrete angles around the lidar unit. These can be processed as a two-dimensional matrix with 1080 rows and as many columns as time points within each plot, where each entry corresponds to a measured distance. The distances and angles at which the measurements were taken can also be converted from polar coordinates to Cartesian coordinates for processing as a point cloud. Because the Hokuyo UST-10LX is a planar lidar, the data recorded at any given time

point correspond to a two-dimensional "slice" within a plane perpendicular to the motion of the rover. A three-dimensional representation of each plot can be created by arranging the slices along a third dimension corresponding to their time stamps as the rover drives through the plot (Fig. 2).

Lidar data cleaning and processing were performed in R (R Core Team, 2018) using the ash (Scott et al., 2009), EBImage (Pau et al., 2010), mmand (Clayden, 2014), slidaRtools (github.com/niknap/slidaRtools), and tidyverse (Wickham, 2017) packages. Three-dimensional point clouds were cleaned by first removing points outside of the plot of interest. Because lidar measures distance to the nearest object in a straight line, many data points correspond to plants in adjacent plots. We removed any points outside the plot of interest by thresholding the absolute horizontal distance from the rover according to Otsu's method (Otsu, 1979). Following this step, points belonging to the ground were identified by performing a morphological opening on the image with a 65 by 65 square kernel, which leaves only points belonging to the ground. The ground points were removed from the original image and the matrix converted into three-dimensional point cloud format. The point cloud was aggregated into voxels at a resolution of 5 by 5 mm within each time point. Voxels containing only a single lidar data point and with no neighbors in three-dimensional space were assumed to be noise and were subsequently removed.

Cleaned, three-dimensional lidar data for each distinct point cloud were then compressed into a two-dimensional density map. First, the two rows of the plot were merged by reflecting one row of the plot around the vertical axis. This effectively lined all plants from both rows of the plot into a single row. Second, data points were binned along the length of the plot (represented by the time dimension; Fig. 2) into 128 by 64 bins (vertical by lateral), and two-dimensional density was calculated by average shifted histogram (Scott, 2015). This results in an image that resembles a single plant, representing the average distribution of all plant surfaces in the plot. Applying this compression to all plot-level point clouds resulted in 2103 density maps that were subjected to dimensionality reduction to extract LSPs (Ubbens et al., 2019).

Creating Latent Space Phenotypes

Two dimensionality reduction methods were applied to the plant surface density maps: PCA and an autoencoder. To format the data for PCA, each plant surface density map was vectorized, then row vectors from all observations were bound together into a matrix. The density values were logarithmically transformed, then centered and scaled column-wise (pixel-wise) before performing PCA. The PC scores for each plant surface density map were used as PCA-based LSPs.

One of the primary goals of the autoencoder was to extract an encoded representation of each plot that represents plant architecture and biomass distribution while ignoring variability that is attributable to differences in rover position, wind direction, or other factors that introduce noise. In some senses this resembles the goals of denoising autoencoders, which can be trained to remove noise from images, but in this case we did not know ahead of time which elements of the lidar data constituted noise and which constituted signal nor did we know the distribution of noise to simulate it. Because many plots were driven more than once by the rover, we were able to use those technical replications to identify features of a density map that are persistent between technical replications while ignoring those that are variable due to noise. The autoencoder was constructed to take two plant surface density maps as input images, representing repeated data collections on the same plot. Both technical replication images are encoded separately, but their encodings are averaged before being decoded, to train the autoencoder to identify a reduced representation that can be used to reconstruct both images. Figure 3 has a schematic of the autoencoder model described below.

For the encoding step, 128- by 64-pixel plant surface density maps are put through a series of convolutional layers combined with average pooling to reduce the size of the image. Average pooling was chosen over max pooling because lidar data are not always perfectly

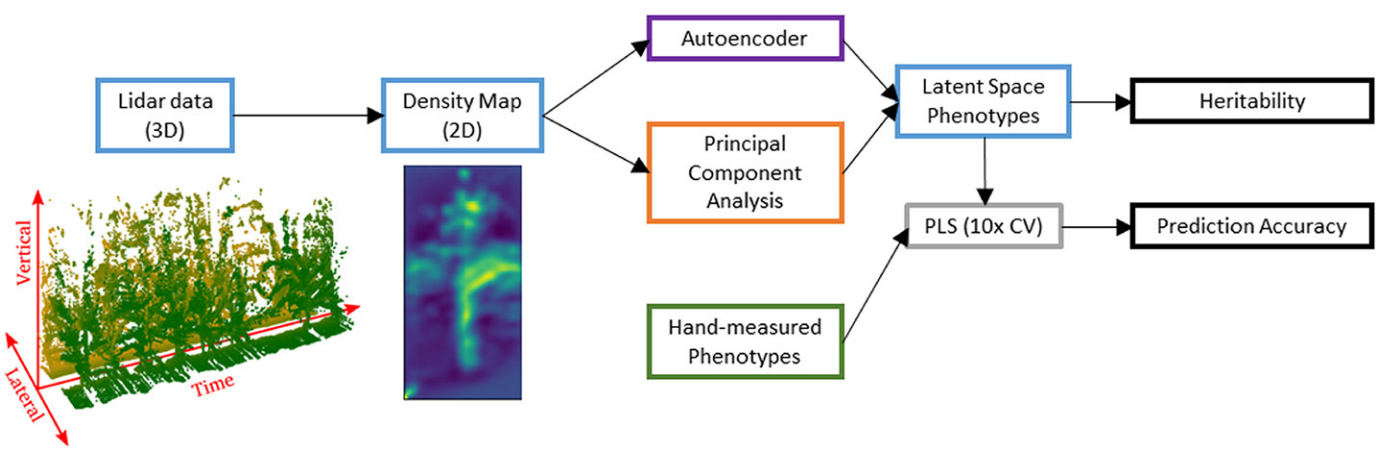

Fig. 2. Schematic of this study's workflow. Three-dimensional lidar data are compressed into two-dimensional density maps. Examples of a lidar point cloud and density map can be seen below their respective labeled boxes. The density map images are subjected to dimensionality reduction by principal component analysis and an autoencoder, resulting in latent space phenotypes (LSPs) that contain information about plant architecture and biomass distribution. Then heritabilities were calculated for the LSPs, and the LSPs were used to predict manually measured phenotypes.

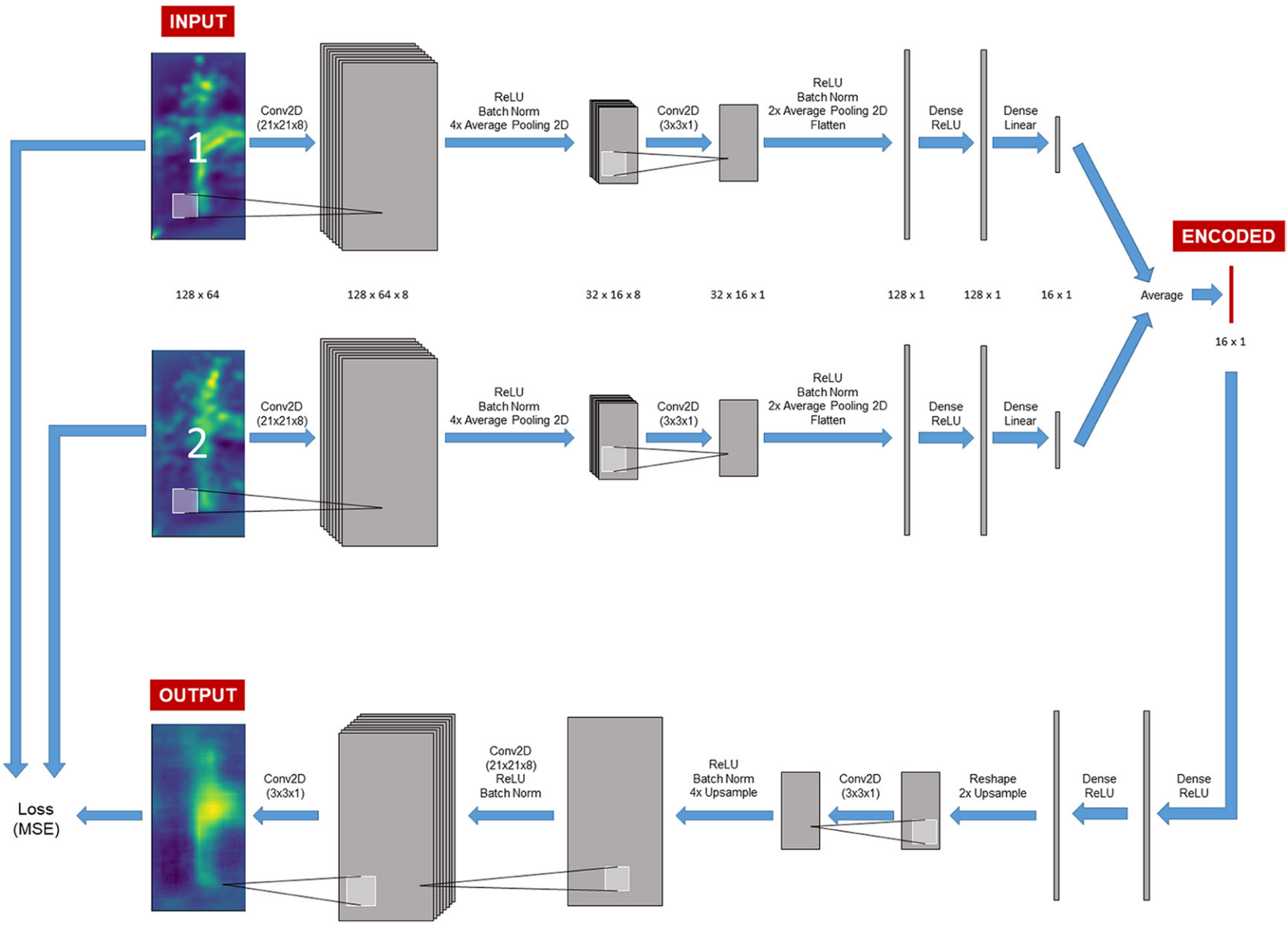

Fig. 3. Diagram of the autoencoder used for extracting latent space phenotypes (LSPs). Many field plots were driven by the rover more than once, producing technical replications of data collection in those plots. Technical replications were used to train the autoencoder to identify a signal that is persistent in both images while ignoring noise caused by rover position, wind, or other stochastic influences. Both technical replications are put through identical encoding pipelines, where they are convolved twice and put through two dense layers to produce a 16-dimensional encoding. Before decoding, the encoded representations of the technical replicates are averaged. Decoding proceeds through two dense layers and three convolutional layers before producing the output image. Loss was calculated as the mean squared error (MSE) between the two input images and the output. After model training, the input was changed from two different technical replication images to duplicates of a single density map to produce unique encodings for each uniquely acquired density map.

aligned from plot to plot, and averaging or smoothing improved the consistency of results between technical replications. The resulting 32 by 16 matrix is flattened and passed through a series of dense layers to produce the vector of 16 encoded values. To decode, the encoded values are passed through a series of dense layers before being reshaped into a 32 by 16 image, which is subsequently convolved and upsampled to produce the output. L2 regularization ($\lambda = 0.01$) was applied to the 16-value encoded vector.

Of the 2103 plot-level observations recorded, 1386 were technical replications of 693 unique plots. These technical replications were augmented by rotating and adding noise, then used to train the autoencoder (Fig. 4). Twenty percent of the training data, split by plots, was used as a validation set to assess overfitting. By forcing technical replications to have the same encoded values during training, the network learns to ignore noise or perturbations that occur during repeated sampling. The loss function during model training was the mean squared error between the decoded image and the original plant surface density image for each of the technical replicates. The autoencoder was trained with a batch size of 256 and trained for 200 epochs, at which point validation set loss was stabilized.

After training, all original data were encoded. Unlike in training, replicated plots were encoded separately, allowing different encodings for each technical replication. Code for creating and training the model, as well as an HDF5 of the final model, can be found at https://bitbucket.org/bucklerlab/p_lidar_lsp. Lidar data can be found at https://doi.org/10.25739/zxp6-g188.

Evaluating Latent Space Phenotypes

Latent space phenotypes were evaluated based on two metrics: their heritability and how well they predict manually measured phenotypes. If LSPs are heritable, it means there is genetic variation for these novel traits and they can be effectively selected on in an applied breeding program. If they can be used to predict manually measured phenotypes, it means the LSPs are not characterizing meaningless differences between individuals but instead contain information about plant architecture and other biologically important phenotypes.

Heritability was calculated by modeling manually measured phenotypes and LSPs as $y_{ijk} = \mu + G_i + E_j + R(E)_{kj} + \varepsilon_{ijk}$, where y_{ijk} is the manual phenotype or LSP of the ith genotype in the jth experiment in the kth block; μ is the overall mean; $G_i \sim N(0, \sigma_g^2 I)$ is the random effect of the ith hybrid individual; E_j is the effect of the jth experiment (NYH2 or NYH3); $R(E)_{kj}$ is the effect of the kth block nested in the jth experiment; and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2 I)$ is a normally and independently distributed error term. Broad-sense heritability was calculated as

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2 / p_h}$$

where p_h is the harmonic mean of the number of plots in which each hybrid was evaluated (Holland et al., 2003).

Latent space phenotypes were used to predict manually measured phenotypes by partial least squares regression (PLSR). Partial least squares regression models were built using the plsr() function in the R (R Core Team, 2018) package pls (Wehrens and Mevik, 2007) with 10-fold cross-validation by hybrid and five latent variables. For each manually measured trait, separate predictions were performed using either the first 16 PCA LSPs or all 16 autoencoder LSPs. Accuracy of the PLSR predictions was measured as the Pearson's correlation between predicted and manually measured phenotypes.

We also used PLSR to see how well non-architectural traits can be predicted from manually measured architectural and

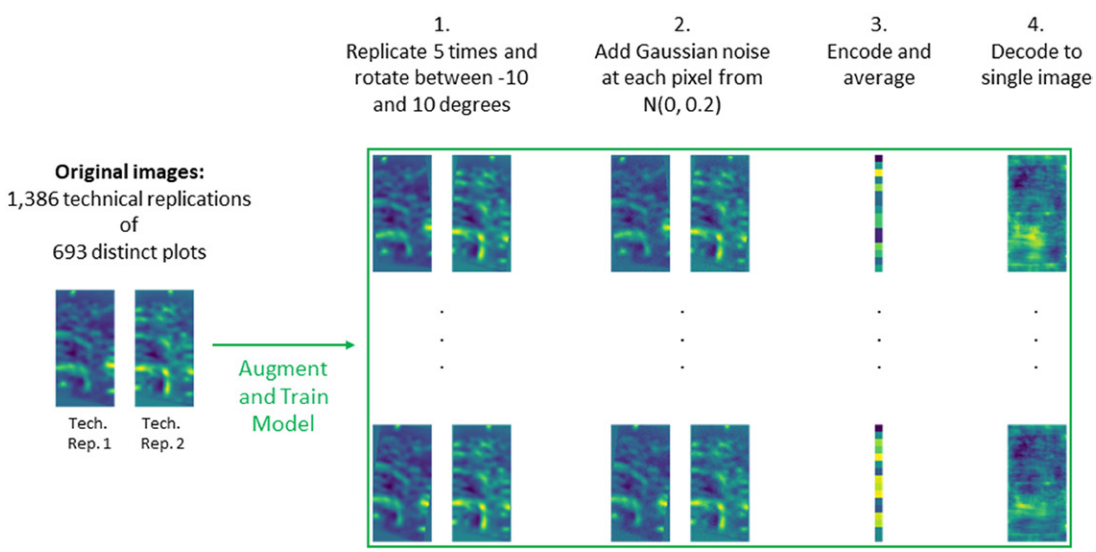

Fig. 4. Image augmentation for model training. Distinct field plots that were recorded more than once represent technical replications. Because any two technical replications represent the same plot, they should have the same encoded values. Hence, for any given plot in the training set, the two technical replications were processed independently through the encoding portion of the network, but their encoded values were averaged before being decoded into a single image representing the plot. Density maps for training plots were replicated and augmented by random rotation and addition of noise.

biomass distribution traits alone. We used plant and ear height, total leaf number, ear leaf number, stand count, stalk lodging, and root lodging as explanatory variables to predict days to anthesis, days to silking, grain moisture, test weight, and plot weight using five latent variables and the same 10-fold cross-validation scheme as above.

Because PCA and autoencoders can be prone to encoding only the major differences between images, we wanted to check that they were not simply encoding differences in overall density between images. To test this, we evaluated the relationship between heritability of each LSP and its correlation with the sum of pixel intensity for each image. In addition, we checked for heritability in the LSPs independent of image intensity: we used simple linear regression to remove the effects of total pixel intensity from the LSPs and then calculated heritability for the residuals using the same method described above.

Results and Discussion

Driving the rover through hybrid trials yielded 2103 three-dimensional point cloud representations of 1153 unique two-row plots containing 698 hybrid genotypes. Visual inspection of the point clouds reveals that they captured individual plant architecture well (see Fig. 2 for an example point cloud). Previous to this study, there was some concern that because lidar can only measure objects within "line of sight" of the sensor, the middle to upper canopy would be poorly represented due to occlusion. Rather, examination of individual plots revealed reconstruction of upper leaves and in some cases even tassels. By overlaying the two rows of a plot and calculating the density along the rows of the plot, we produced a two-dimensional "image" showing the distribution of plant surfaces for an average maize plant in that plot (see Fig. 2 for an example density map). The LSPs extracted from those plant surface distribution images by PCA or an autoencoder were evaluated for their heritability and whether they can be used to predict manually measured phenotypes.

Latent Space Phenotype Heritability

We calculated the heritability of the LSPs to assess whether they are characterized by enough genetic variation to be useful for biological studies or as selection metrics in breeding programs (Fig. 5). Heritability for manually measured phenotypes ranged from 0.26 (stand count) to 0.78 (plant height). The PCA-based LSPs had higher heritability in the earlier PCs, which decreased rapidly to near zero after a dozen PCs. This pattern was expected because the earlier PCs capture greater amounts of variability than later PCs. The first 16 PCs ranged in heritability from 0.00 (PC9 and PC12) to 0.44 (PC2). The 16 LSPs produced by the autoencoder method do not have a logical ordering the way that PCs do. To reference them, they will be arbitrarily named ENC1 through ENC16. Autoencoder-based LSPs ranged in heritability from 0.01 (ENC16) to 0.40 (ENC1). The high-end heritabilities of the LSPs are comparable (± 0.1) to many of the manually measured traits and high enough to be effectively selected on in a breeding program. These heritability measurements are from a relatively small sample in a single environment and are anticipated to rise as the rovers are deployed in replicated field trials.

Because the LSPs generated in this study are dimensionality reductions of the plant surface density images, they are likely to characterize the most predominant differences between images. A major source of variation between images was the simple sum of pixel intensities. We found a nonzero relationship between the LSP correlation with the sum of pixel intensities and LSP heritability (Supplemental Fig. S1), indicating that more heritable LSPs do tend to capture some aspect of total image intensity. After accounting for image intensity, LSPs still had nonzero heritabilities similar to those shown in Fig. 5 (Supplemental Fig. S2). Together, these results indicate that the LSPs are capturing information related to the overall intensity of the plant surface density maps but that there is still heritable variation for the distribution of plant surfaces in the images. For the remainder of this discussion, we will consider the original (not corrected for image intensity) plant surface density images. Although overall image intensity may have technical causes related to sensor performance and rover velocity, it also may contain information about the total biomass accumulation in different field plots.

Predicting Manually Measured Traits with Latent Space Phenotypes

Although the LSPs are heritable, that alone does not mean that they contain useful information about plant architecture. To determine whether the LSPs are capturing elements of biomass distribution and plant architecture, we used PLSR to predict each manually measured trait with either the first 16 PCA-based LSPs or all 16 autoencoder-based LSPs (Fig. 6). Prediction accuracy with PCA-based LSPs ranged from 0.25 (stalk lodging) to 0.89 (plant height), whereas prediction accuracy with autoencoder-based LSPs ranged from 0.24 (stalk lodging) to 0.84 (plant height). Similar prediction accuracies from autoencoder-based and PCA-based LSPs may be due to a limit in useful information contained in the plant surface density images. However, PCA-based LSPs are more predictive of manually measured traits if all PCA-based LSPs (rather than the first 16 PCs) are used for prediction. It is unsurprising that plant height showed the highest prediction accuracy, as it is straightforward to identify from the plant surface density images. Other architectural and agronomic traits, such as stand count, lodging, and leaf counts, demonstrated lower yet nonzero prediction accuracies. This could be for a few reasons: first, these traits do not show themselves as clearly as plant height in the image of plant surface distribution that was used as input for creating the LSPs; second, they are more difficult to measure manually and, as such, the measured values may be inaccurate due to human error during counting. Some traits, such as lodging, stand count, and all grain-related traits, were not yet fully determined when the rovers were driven through the field. For example, lodging and stand counts were

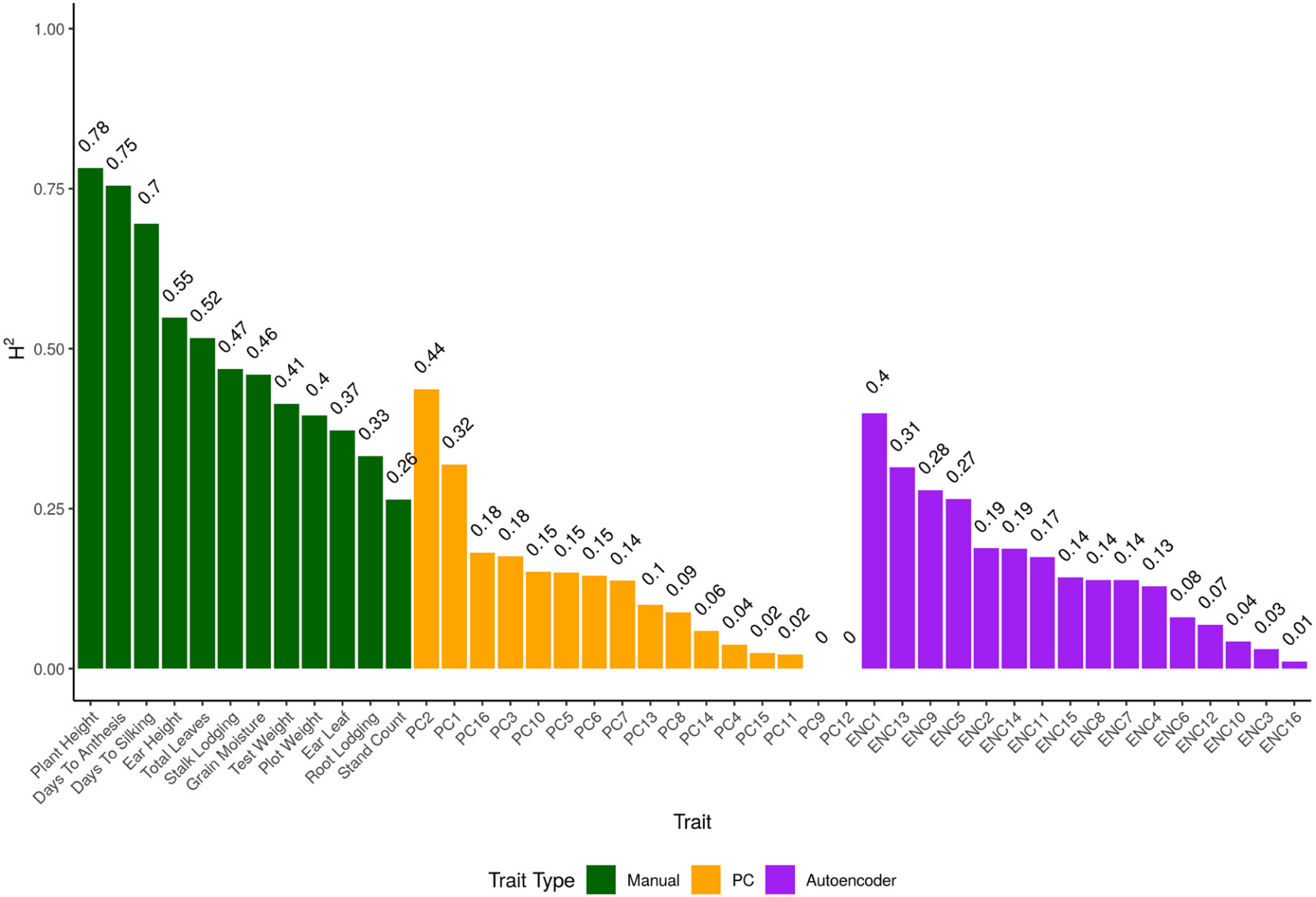

Fig. 5. Comparison of heritability for manually measured traits and latent space phenotypes (LSPs) based on autoencoder and principal component (PC) analysis. Manually measured traits (green) show a range of heritabilities. The higher heritability LSPs (orange and purple) have heritabilities similar to mid- and low-heritability manually measured traits. Heritability values are shown above the respective phenotype's bar. All heritability measurements are from hybrid maize in a single environment.

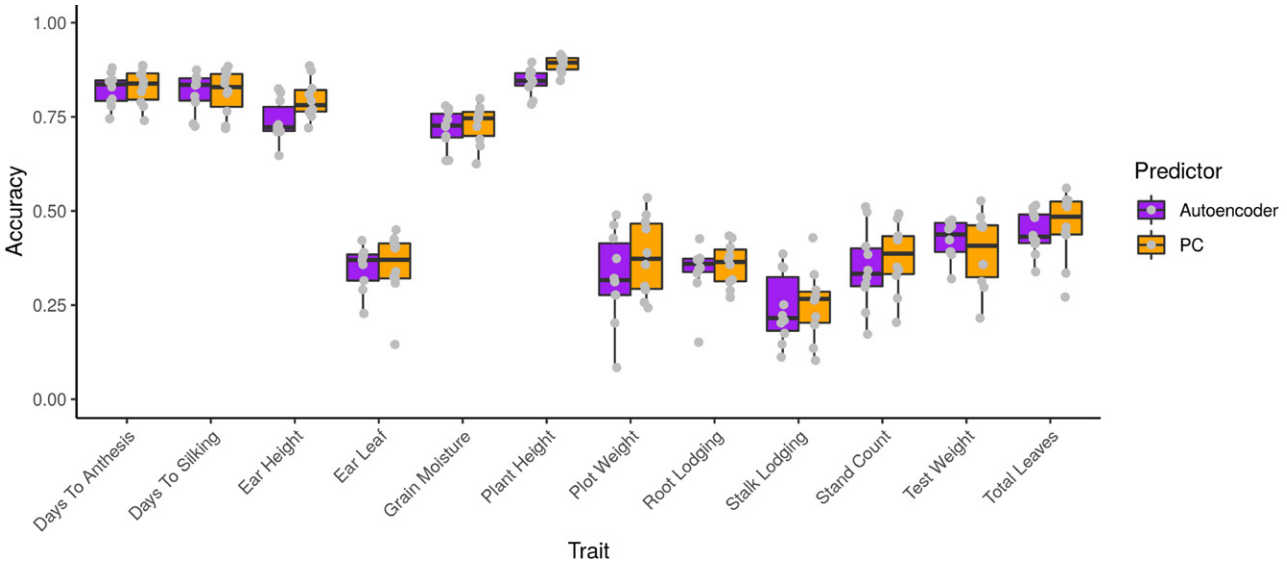

Fig. 6. Ability of latent space phenotypes (LSPs) to predict manually measured traits: LSPs based on autoencoder and principal component (PC) analysis (purple and orange, respectively) were used to predict all manually measured traits by partial least squares regression. The LSPs contain enough information about plant architecture to accurately predict architectural traits (plant and ear height, ear leaf, total leaves, stalk and root lodging, and stand count) as well as non-architectural traits. Accuracy reported is the Pearson correlation between predicted values and manually measured values.

taken several weeks after the rover data were recorded and may have changed in the intervening time. Generally, traits that were heritable and already determined when rover data was collected (e.g., flowering traits, plant and ear height) had higher prediction accuracies from the LSPs than traits that had low heritability (e.g., leaf counts) or may have continued to change in the weeks after rover data were collected (e.g., stand count) (Supplemental Fig. S3). As such, poor prediction accuracies may be due more to noisy manual measurements than to a lack of information captured by the LSPs.

Latent space phenotypes were also predictive of non-architectural traits (grain moisture, test weight, plot weight, and flowering time) due to inherent correlations between them and agronomic or plant architectural traits. To demonstrate this, we used height, stand count, lodging, and leaf count traits as the explanatory variables in PLSR models to predict the non-architectural traits. The prediction accuracies of PLSR models trained on architectural traits are similar to the prediction accuracies of PLSR models trained on LSPs, showing that information about plant architecture is also indicative of non-architectural traits of agronomic importance (Supplemental Fig. S4). Similarity in accuracy when predicting non-architectural traits with architectural traits vs. LSPs shows that LSPs are capturing a similar quantity of information about plant architecture as the manually measured traits.

It should be noted that using LSPs to predict manually measured phenotypes is only meant to demonstrate that the LSPs contain signal related to plant architecture and performance. Higher prediction accuracy and better precision for such predictions can probably be achieved by creating algorithms or models that predict such traits directly from the lidar data.

Advantages, Disadvantages, and Future Prospects for Rover-Based Phenotyping

By showing that LSPs are both heritable and contain information about plant architecture, we have demonstrated that rover-based phenotyping and LSPs are promising for application to plant biology and crop breeding efforts. Subcanopy rovers can be made to operate semi-autonomously, reducing the number of person-hours necessary to obtain quality phenotype data. The amount of time needed to manually gather phenotype data on a large, field-grown population is a function of the number of traits being recorded, the amount of time needed to measure and record each data point, the number of individuals helping with data collection, and the number of plots to be measured. The amount of time needed to measure a field by rover, on the other hand, is a function of the amount of time needed to drive through each plot and the number of plots to be measured. Because the lidar enables three-dimensional reconstruction of each plot, all plant architectural information is recorded by default. The rover traverses the field at a rate of approximately 10 s per 5.33-m plot. At that rate, it can evaluate 2000 two-row plots in <6 h, easily achievable in a standard work day, and requires supervision by one person. If, instead, the same single person were to measure and record phenotypes manually on the same 2000 plots, more time would be needed. Assuming an experienced researcher can collect phenotypic data at a rate of 20 s per plot, it would take between 11 and 12 h to phenotype 2000 plots. This is twice as long as it takes to evaluate the field by rover, and results in only a single data point per plot. If the researcher in this example is being paid US$15 per hour, all of the rover traits presented here could be phenotyped for a cost of US$0.04 per plot (for the researcher supervising the rover), whereas phenotyping a single trait manually would cost twice as much (US$0.08) per plot. Considering the fact that the rover records

enough data in a single pass to produce dozens of traits, whereas the cost of manually phenotyping scales linearly with the number of traits to be recorded, rover-based phenotyping quickly becomes orders of magnitude cheaper than manual phenotyping.

To rephrase this comparison, rover-based phenotyping allows calculation of several LSPs with similar heritability to manually measured traits in the same amount of time needed to phenotype a single manually measured trait. If the purpose of phenotyping is to identify candidate genes via genetic mapping, rover-based phenotyping of LSPs could yield roughly a dozen times as many genetic mapping targets as manual phenotyping for the same investment of time. With the increasing throughput of gene editing technologies, screening such a large number of candidate genes may soon be a less daunting task than it is today (Ramstein et al., 2019). The relative ease of capturing data and processing it into LSPs can facilitate and standardize coordinated phenotyping efforts by research groups across a large number of locations. The same qualities can enable longitudinal evaluation of LSPs, opening the door to including developmental time as a standard axis of variation in experiments, similar to genotypic and environmental variation today.

Though rover-based phenotyping and LSPs show enormous promise for the fields of plant biology and breeding, they do still face some drawbacks. First and foremost, LSPs are of no benefit if they are not interpretable, useful, and applicable to biology or breeding. The results shown in this study provide evidence that LSPs meet these requirements, but more work is needed to prove their utility. The LSPs from autoencoders, in particular, can be difficult to interpret; however, eigenvectors from PCA-based LSPs can be reshaped to visualize the pixels that are up- or down-weighted when creating each latent phenotype (Supplemental Fig. S5), providing some degree of interpretability. The ultimate evidence will come with more data, when LSPs are used to identify candidate genes or as selection criteria in a breeding program. Although rover-based phenotyping has an immense advantage over manual phenotyping with regard to operating cost, it comes with the high up-front cost of purchasing the rovers themselves. In addition to startup costs, rover-based phenotyping is subject to the same pitfall as any other phenotyping system that records large quantities of information: methods need to be developed to analyze and interpret the data. Latent space phenotypes are a convenient way to extract meaningful phenotypes from large quantities of complex and messy data. Scientists interested in faster or more accurate methods for measuring standard traits, such as the manually measured traits in this study, still need to develop reliable methods for doing so, which is a non-trivial task.

Though automated subcanopy phenotyping comes with some challenges, the potential for benefit to plant biology and plant breeding communities is immense and real. The capabilities of subcanopy rovers have been recognized by multiple groups (Mueller-Sim et al., 2017; Kayacan et al., 2018; Stager et al., 2019), and the promise of lidar for in-field characterization of plant architecture has also been recognized (Qiu et al., 2019; Su et al., 2019). By combining novel phenotyping and data analysis techniques, we have demonstrated that rover-based phenotyping by lidar can produce LSPs that are heritable and contain information about plant architecture and plot-level biomass distribution. These techniques will enable in-field high-throughput phenotyping of crops in numerous locations and across developmental time by reducing the cost of collecting high-quality phenotypic data points.

Data Availability

The raw lidar point clouds can be found at https://doi.org/10.25739/zxp6-g188. Code and phenotypic data are available at https://bitbucket.org/bucklerlab/p_lidar_lsp.

Author Contributions

CS and GC, rover design and construction; JLG, ESB, and MAG, study conceptualization; JLG, ER, NL, and NK, data collection; JLG, data analysis; all authors contributed to manuscript preparation or review.

Conflicts of Interest

Authors CS and GC are co-founders and the CEO and CTO, respectively, of EarthSense, Inc.

Acknowledgments

The information, data, or work presented herein was funded in part by the USDA–ARS, Genomes to Fields, and the Advanced Research Projects Agency-Energy (ARPA-E), US Department of Energy, under Award no. DE-AR0000598. The views and opinions of authors expressed herein do not necessarily state or reflect those of the US Government or any agency thereof. The use of trade, firm, or corporation names in this publication (or page) is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the USDA or the ARS of any product or service to the exclusion of others that may be suitable. This manuscript was previously posted to a pre-print server (Gage et al., 2019).

References

AlKhalifah, N., D.A. Campbell, C.M. Falcon, J.M. Gardiner, N.D. Miller, M.C. Romay, et al. 2018. Maize Genomes to Fields: 2014 and 2015 field season genotype, phenotype, environment, and inbred ear image datasets. BMC Res. Notes 11:452. doi:10.1186/s13104-018-3508-1

Andrade-Sanchez, P., M.A. Gore, J.T. Heun, K.R. Thorp, A. Elizabete Carmo-Silva, A.N. French, et al. 2014. Development and evaluation of a field-based high-throughput phenotyping platform. Funct. Plant Biol. 41:68–79. doi:10.1071/FP13126

Araus, J.L., S.C. Kefauver, M. Zaman-Allah, M.S. Olsen, and J.E. Cairns. 2018. Translating high-throughput phenotyping into genetic gain. Trends Plant Sci. 23:451–466. doi:10.1016/j.tplants.2018.02.001

Bai, G., Y. Ge, D. Scoby, B. Leavitt, V. Stoerger, N. Kirchgessner, et al. 2019. NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. Comput. Electron. Agric. 160:71–81. doi:10.1016/j.compag.2019.03.009

Baldi, P. 2012. Autoencoders, unsupervised learning, and deep architectures: Proceeding of the ICML Workshop on Unsupervised and Transfer Learning. Proc. Mach. Learn. Res. 27:37–49.

Busemeyer, L., D. Mentrup, K. Möller, E. Wunder, K. Alheit, V. Hahn, et al. 2013. BreedVision: A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. Sensors 13:2830–2847. doi:10.3390/s130302830

Clayden, J. 2014. mmand: Mathematical morphology in any number of dimensions. R package version 1. R Project for Stat. Comput., Vienna. https://www.r-project.org/nosvn/pandoc/mmand.html

Cobb, J.N., G. Declerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhanc-

ing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126:867–887. doi:10.1007/s00122-013-2066-0

Gage, J.L., D. Jarquin, C. Romay, A. Lorenz, E.S. Buckler, S. Kaeppler, et al. 2017. The effect of artificial selection on phenotypic plasticity in maize. Nat. Commun. 8:1348. doi:10.1038/s41467-017-01450-2

Gage, J.L., E. Richards, N. Lepak, N. Kaczmar, C. Soman, G. Chowdhary, M.A. Gore, and E.S. Buckler. 2019. In-field whole-plant maize architecture characterized by latent space phenotyping. bioRxiv 763342. doi:10.1101/763342

Gage, J.L., M.R. White, J.W. Edwards, S. Kaeppler, and N. de Leon. 2018. Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. Genetics 210:1125–1138. doi:10.1534/genetics.118.301487

Higuti, V.A.H., A.E.B. Velasquez, D.V. Magalhaes, M. Becker, and G. Chowdhary. 2018. Under canopy light detection and ranging-based autonomous navigation. J. Field Robot. 36:547–567. doi:10.1002/rob.21852

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martínez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9–112. doi:10.1002/9780470650202.ch2

Houle, D., D.R. Govindaraju, and S. Omholt. 2010. Phenomics: The next challenge. Nat. Rev. Genet. 11:855–866. doi:10.1038/nrg2897

Kayacan, E., Z. Zhang, and G. Chowdhary. 2018. Embedded high precision control and corn stand counting algorithms for an ultra-compact 3D printed field robot. In: H. Kress-Gazit et al., editor, Robotics: Science and Systems XIV, Pittsburgh, PA. 26–30 June 2018. doi:10.15607/rss.2018.xiv.036

Lawrence-Dill, C.J., P.S. Schnable, and N.M. Springer. 2019. Idea factory: The Maize Genomes to Fields initiative. Crop Sci. 59:1406–1410. doi:10.2135/cropsci2019.02.0071

Minervini, M., H. Scharr, and S.A. Tsaftaris. 2015. Image analysis: The new bottleneck in plant phenotyping. IEEE Signal Process. Mag. 32:126–131. doi:10.1109/MSP.2015.2405111

Mueller-Sim, T., M. Jenkins, J. Abel, and G. Kantor. 2017. The Robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore. 29 May–3 June 2017. IEEE Robotics and Automation Soc., New York. p. 3634–3639.

Omasa, K., F. Hosoi, and A. Konishi. 2007. 3D lidar imaging for detecting and understanding plant responses and canopy structure. J. Exp. Bot. 58:881–898. doi:10.1093/jxb/erl142

Otsu, N. 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9:62–66. doi:10.1109/TSMC.1979.4310076

Pau, G., F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. 2010. EBImage: An R package for image processing with applications to cellular phenotypes. Bioinformatics 26:979–981. doi:10.1093/bioinformatics/btq046

Qiu, Q., N. Sun, Y. Wang, Z. Fan, Z. Meng, B. Li, and Y. Cong. 2019. Field-based high-throughput phenotyping for maize plant using 3D LiDAR point cloud generated with a "phenomobile". Front. Plant Sci. 10:554. doi:10.3389/fpls.2019.00554

Ramstein, G.P., S.E. Jensen, and E.S. Buckler. 2019. Breaking the curse of dimensionality to identify causal variants in breeding 4. Theor. Appl. Genet. 132:559–567. doi:10.1007/s00122-018-3267-3

R Core Team. 2018. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams. 1986. Learning internal representations by error propagation. La Jolla Inst. for Cognitive Science, Univ. of California, San Diego.

Salas Fernandez, M.G., Y. Bao, L. Tang, and P.S. Schnable. 2017. A high-throughput, field-based phenotyping technology for tall biomass crops. Plant Physiol. 174:2008–2022. doi:10.1104/pp.17.00707

Scott, D., A. Gebhardt, and S. Kaluzny. 2009. ash: David Scott's ASH routines. R package version 1-0. CRAN, Vienna. https://rdrr.io/cran/ash/

Scott, D.W. 2015. Multivariate density estimation. 2nd ed. Wiley Ser. Prob. Stat. John Wiley & Sons, Hoboken, NJ. doi:10.1002/9781118575574

Stager, A., H.G. Tanner, and E.E. Sparks. 2019. Design and construction of unmanned ground vehicles for sub-canopy plant phenotyping. arXiv:1903.10608 [cs.RO].

Su, Y., F. Wu, Z. Ao, S. Jin, F. Qin, B. Liu, et al. 2019. Evaluating maize phenotype dynamics under drought stress using terrestrial lidar. Plant Methods 15:11. doi:10.1186/s13007-019-0396-x

Sun, S., C. Li, A.H. Paterson, Y. Jiang, R. Xu, J.S. Robertson, et al. 2018. In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. Front. Plant Sci. 9:16. doi:10.3389/fpls.2018.00016

Ubbens, J., M. Cieslak, P. Prusinkiewicz, and I. Stavness. 2019. Latent space phenotyping: Automatic image-based phenotyping for treatment studies. bioRxiv 557678. doi:10.1101/557678

Virlet, N., K. Sabermanesh, P. Sadeghi-Tehran, and M.J. Hawkesford. 2017. Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. Funct. Plant Biol. 44:143–153. doi:10.1071/FP16163

Wehrens, R., and B.-H. Mevik. 2007. The pls package: Principal component and partial least squares regression in R. J. Stat. Softw. 18(2):1–23. doi:10.18637/jss.v018.i02

Wickham, H. 2017. Tidyverse: Easily install and load 'tidyverse' packages. R package version 1. R Found. Stat. Comput., Vienna.

Yang, G., J. Liu, C. Zhao, Z. Li, Y. Huang, H. Yu, et al. 2017. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. Front. Plant Sci. 8:1111. doi:10.3389/fpls.2017.01111