Algorithms to predict moisture content of grain using relative humidity time-series

Charles B. Delahunt[1,2], Wenbo Wang[1], Simon Ghionea[1], Andrew Miller[1], Austin Chan[2], Anjali Sehrawat[1], Courosh Mehanian[1], Michael Friend[1]
[1]*Global Health Labs, Bellevue, WA, USA (formerly at Intellectual Ventures Lab / Global Good Research)*
[2]*University of Washington*, Seattle, WA, USA
Corresponding authors: charles.delahunt@ghlabs.org, mfriend262@gmail.com

***Abstract*—Post-harvest losses to grain crops are conservatively estimated at 10-20% (ranging up to 40%) in many countries. In particular, grains must be properly dried to avoid spoilage, harmful mycotoxins from mold, and financial loss. Smallholder farmers can thus greatly benefit from a means to assess Moisture Content (MC) in their grain. We describe a two-step algorithm, with very low computational cost, that calculates MC with high accuracy, using Relative Humidity (RH) and Temperature (T) time-series. The time-series do not need to reach equilibrium state, enabling fast (12-minute) time-to-result. The algorithm first curve-fits the RH time-series to estimate asymptotic RH, in order to leverage the physics of the RH-T-MC equilibrium relationship. It then uses regression to estimate MC to within $\pm$1% on $\geq$95% of samples over a wide range of ambient RH-T conditions, on both Lab and Field samples of 10 different grains.**

***Index Terms*—grain drying, moisture content, relative humidity, machine learning, Chung-Pfost**

I. Introduction

A. Use-case

Post-harvest losses to grain crops are conservatively estimated at 10-20% (ranging up to 40%) in Sub-Saharan Africa [1]. Inadequate drying and storage practices are a significant problem, since high temperature and humidity are often factors at time of grain harvest. Smallholder farmers have qualitative methods for determining moisture content (such as biting), but these methods are subjective and harm the farmer when negotiating the sale of their grains: If the grain is too wet, the buyer will either reject the grain or charge a penalty; if it is too dry, the farmer loses income on sales by weight. In stored grain, insufficient drying can lead to fungal growth that reduces nutritional value, produces harmful mycotoxins, and reduces economic value of the grain [1]–[3].

B. Abbreviations

Moisture Content (%): MC
Temperature (°C): T
Relative Humidity (% or fraction): RH
Estimate of equilibrium RH (asymptotic RH): $\widehat{RH}$
Neural Network: NN

C. Predicting MC using RH and T

Devices that measure MC of grain using capacitance or infra-red spectra are too expensive for this use-case [4]. A lower-cost approach is to estimate grain MC indirectly by measuring the RH of a closed volume of air which has come into equilibrium with a sample of grain. Of RH-measuring devices, some are low cost but non-quantitative, instead giving a "go-no go" reading (*e.g.* [5], [6]); another device is promising, but still does not meet accuracy and cost specs for our use case [7].

Our lab researched market needs and developed a device to deliver: *(i)* quantitative estimates of grain MC; *(ii)* high accuracy; *(iii)* fast time-to-result; *(iv)* long device life; and *(v)* low cost (bill of materials ≈9 USD, all-in cost to farmer ≈20 USD), which tightly constrains computational power.

This paper focuses solely on the algorithms to predict MC using device output; description of the device itself is beyond our scope. The performance specs relevant to the algorithms are detailed in section II-B. The device outputs RH and T time-series, which however do not reach equilibrium state within the time-to-result allowed. The algorithms seek to accurately estimate grain MC using these time-series.

Approaches to estimation of MC using RH and T, in equilibrium state, include *(1)* physics-based equations, and *(2)* machine learning (ML) methods.

1) Physics-based equations: There are several equations of the form $MC = f(RH, T)$ linking equilibrium-state variables MC, RH, and T, with parameters depending on grain type and RH-T range. These equations seek to model some aspect of the physics of the system, and are often further modified to better fit empirical isotherm data. Examples of such equations include those of Chung-Pfost, Henderson, Oswin, and GAB [8]. Given proper conditions and knowledge of the key grain- and condition-dependent parameters, these equations can be in principle used to predict MC, via calculation or look-up tables [9], [10].

However, several factors impede direct application of these equations: They require equilibrium RH and T readings; equation fitting parameters for actual grains or grain subtypes can differ significantly from those listed in the standards [11]; and the optimal equation to use varies by grain [11]. Also, the equations and look-up tables give insufficient accuracy for our use-case performance specs [7], [11], [12].

2) ML regression methods: We can bypass modeling the physics (in a sense) by applying ML regression methods.

Given a dataset of samples with features X and target variable y, $\{X_i, y_i\}$, we parametrize (train) a model on a training set, then apply the trained model to other samples. Model quality is judged by its predictive accuracy on test sets relative to a ground truth MC as determined by a high accuracy bench device (the GAC 2500, Dickey-John Corp, precision $\pm 0.1\%$).

Examples of models include: *(i)* The simple but powerful Linear Model, which assumes a relationship $y_i = \sum_j \beta_j x_j + \epsilon$, where ϵ is a noise term; *(ii)* Neural nets (NNs), which are universal function approximators and effective all-purpose prediction models. They have been used to predict MC of grains [13], [14] and (on synthetic data) [15].

D. Scope of this paper

This paper describes ML algorithms to predict grain moisture, using time-series of RH and T as features. The time-series data reflect behavior of RH and T in a small airspace above the grain sample, after the device is closed and as the grain-air system moves towards equilibrium. Crucially, the system does not need to reach equilibrium before making an accurate MC estimate.

Contributions of this paper include: *(i)* Performance requirements for a grain MC predictor to be useful in the field; *(ii)* an accurate, low-cost algorithm for MC prediction that uses RH and T time-series far from equilibrium state; and *(iii)* some bright ideas that did not work.

II. Methods

A. Datasets

Experiments addressed 10 crops (maize, coffee, beans, ground nuts (*e.g.* peanuts), sorghum, rice, cowpeas, soybean, sesame, and millet) across the full range of environmental conditions in the target market.

Data consisted of ground-truth MC (as determined by the GAC 2500) plus time-series of RH and T for grain samples. Time-series were 60 minutes long, but since allowable time-to-decision for test samples was 17 minutes after device closure, the models (with one exception, see section II-F) used only the first 17 minutes (later reduced to 12 minutes). Two sets of data were collected: Lab and Field.

1) Lab-generated data: Lab data was collected under a range of RH-T conditions, for 6 crops: Cowpeas (n = 335), millet (n = 274), rice (n = 334), sesame (n = 234), sorghum (n = 335), and soy (n = 335). Dry grains were rewetted to a range of MCs, then processed in the device inside a controlled climate chamber set to the desired RH-T. Ambient RH-T combinations covered all expected (year-round) field conditions, and in fact covered more conditions than we encountered in the field.

2) Field-generated data: Field data was collected at several locations in Uganda, for grains with various moisture contents, at various ambient RH-T conditions, for five crops (driven by availability of recent harvests): Bean (n = 941), coffee (n = 105), maize (n = 352), nut (n = 428), and sorghum (n = 339). Beans were of 4 types of different sizes, including Mung beans which were differed considerably in size from the others. Data was also collected for small numbers ($n < 50$) of cowpeas, millet, rice, sesame, and soybean. RH-T combinations were limited by weather conditions, so they did not cover as wide a range of RH-T conditions as did the lab-generated data.

Fig. 1 shows a typical distribution of RH-T conditions for Lab and Field data, showing overlap in a central region of RH-T conditions and the lack of Field data in two corner regions.

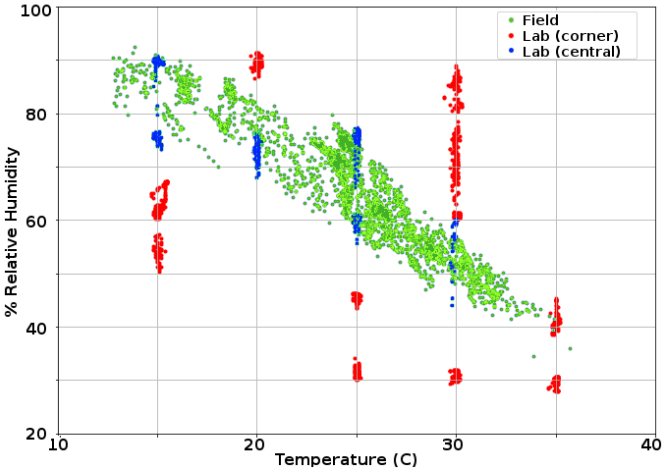

Fig. 1. Distribution of ambient RH-T conditions of collected samples (Maize). x-axis: Temperature, y-axis = RH (relative humidity). Green: Field data. Blue: Lab data, overlapping with field (central region). Red: Lab data for conditions absent in the field data (corner regions).

B. Accuracy and other use-case requirements

Accuracy specs were driven by the need for the farmer to have highly accurate estimates of MC near the optimal drying target, for selling into the formal market: Drying must not stop too early, to avoid rot and mycotoxins, as well as price penalties. Drying should stop as soon as possible once below the optimal target, to maintain maximum selling weight.

Each grain had a target MC [2], shown in Table I. For MCs close to this target (the "critical" region), acceptable error was under $\pm 1\%$ raw MC (compare specs in [16]). For MCs far above or below the target MC (the "non-critical" region), acceptable error was larger. Fig. 2 illustrates the acceptable error bands. The key Figure of Merit was % accuracy, *i.e.* the percent of samples that had error within acceptable bounds. The algorithm was expected to have within-bounds error on 95% of samples in a test set. Note that this goal can be easier or harder depending on the distribution of MCs in the test set, because samples with MC in the critical region are more challenging due to the tighter error specs.

Time-to-result requirement was $\leq$17 minutes from lid closure (12 minutes sufficed as it turned out).

TABLE I
Target moisture content (%) by grain.

Coffee	11	Bean	14
Nuts	7.5	Maize	13.5
Cowpeas	12	Millet	14
Sesame	6	Rice	14
Sorghum	13.5	Soy	14

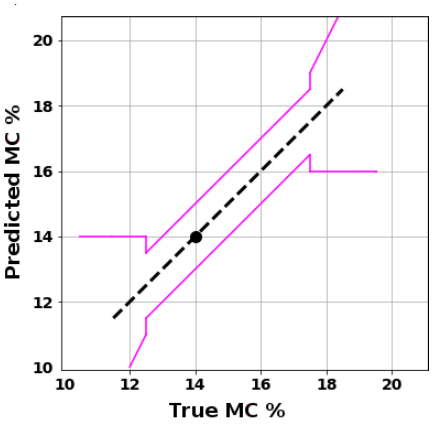

Fig. 2. Definitions of acceptable error bands. Black dot shows the target MC, $\widehat{mc}$. Magenta lines show acceptable error at each true MC. x-axis is true MC, y-axis is predicted MC. For true MCs between $\widehat{mc}$ - 2% and $\widehat{mc}$ + 3.5%, raw predicted error should be under 1% (critical region). For true MCs outside this range, the acceptable error increases (non-critical region).

Processing hardware was tightly constrained by cost (8 bit, 2 kB RAM, 32 kB flash program memory).

C. Algorithm overview

We implicitly assume that $MC \approx f(RH, T)$, with RH the most important input variable. RH also had the most variable time-series (the most dramatic changes after device closure) in our device. Because time-to-result constraints prevent access to equilibrium state RH, we need to estimate an asymptotic value $\widehat{RH}$.

The approach is as follows:

1) A quadratic ax^2+bx+c is fitted to the RH timecourse, starting at 5 to 10 minutes post-lid-closure and ending 17 (or 12) minutes post-lid-closure.

2) The peak (trough) of the fitted quadratic serves as an estimate of $\widehat{RH}$.

3) A regression model predicts MC using 5 features: $a, b, c, \widehat{RH}$, and ΔT. The model can be linear, or a simple NN (*e.g.* 2 hidden layers with 3 units each).

4) For training, only samples with MC in the critical region close to the target MC (see Fig. 2) are used to train the algorithm.

A key driver of our algorithm design was the low computing power available. Thus we did not try feeding the time-series into a large NN (see section IV). Code was developed in Python, including the scikit-learn library [17], [18].

In the following sections, we describe in detail the curve-fitting step, the regression step, and some ideas that did not pan out.

D. (Step 1) Curve fitting

In general, RH time-series exhibited a short transient followed by a roughly exponential increase or decrease (depending on the initial ambient RH-T conditions and the grain MC) towards an equilibrium state (for example RH curves see Fig. 3). Equilibrium usually took much longer to reach than 17 minutes, so a challenge was to estimate the equilibrium RH, $\widehat{RH}$, given a partial trajectory. Equilibrium was usually reached before 60 minutes, so we had a reasonable direct proxy for $\widehat{RH}$ in our training data. But this was unavailable in test samples.

Although the RH curves appeared roughly exponential, fitting an exponential gave poor results (the fitting often did not converge, and its asymptotic value underestimated $\widehat{RH}$). We found that fitting a quadratic curve $q(t)$ was very robust and enabled more accurate estimates of $\widehat{RH}$, which was taken to be the value at the zero derivative: $\widehat{RH} = q(t) \mid \frac{dq}{dt} = 0$.

We used a 5 second time-step, and a Levenberg-Marquardt algorithm for fitting. Tests with a version in C showed this stayed within our chip's resources.

Some samples displayed longer transients after lid closure, so a 5 (or even 10) minute delay before curve-fitting greatly improved $\widehat{RH}$ estimates. Thus we ignored the first third of our data, and fitted RH between t = 5 minutes and t = 17 (or 12) minutes. Fig. 3 shows the improved $\widehat{RH}$ estimates due to this delay.

Because $\widehat{RH}$ was a more important regression variable than T, and because T time-series were generally less active, fitting a curve to T time-series did not yield benefit.

E. (Step 2) Regression

The fitted quadratic parameters $\{a, b, c\}$ and the estimate $\widehat{RH}$ gave four features for regression. For a fifth T-based feature, we found that the variable $\Delta T = T_{end} - T_{start}$ was a superior proxy for T.

We trained a different model for each grain, consistent with physics equation parameters being different for each grain. Linear models and simple 2 layer NNs (*e.g.* 3 units per layer) were roughly equivalent, one or the other performing slightly better on different grains. The small size of the NNs allowed the constrained hardware to run them on test samples.

Based on linear model β coefficients (standardized features), the most important features were $\widehat{RH}$ and quadratic parameters a, b, while ΔT had almost no effect (compare the GAB equation, which has no T dependence).

We note that an automated grid search did not discover an optimal NN architecture. We suspect this is because we chose

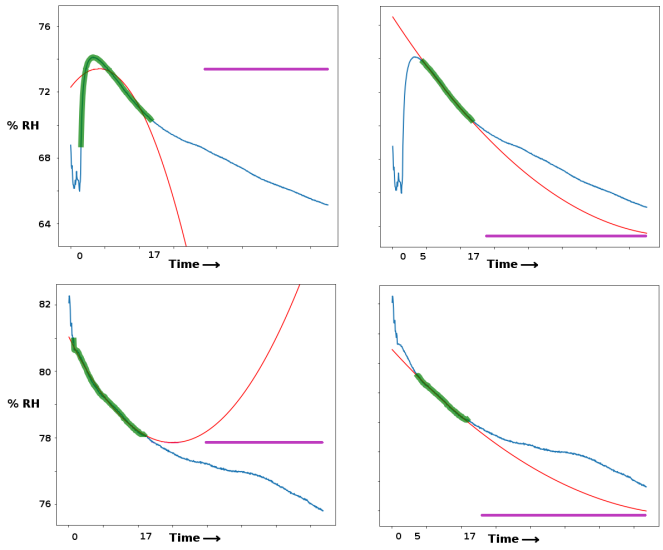

Fig. 3. Effect delaying the start of the RH time window. Two coffee samples (top and bottom). Left: No delay. Right: 5 minute delay. Blue = raw time-series, green = fitted window, red = fitted quadratic, thick magenta bar = estimate of $\widehat{RH}$ at the zero derivative of the quadratic. The delay gives a much more accurate estimate of $\widehat{RH}$ in many cases.

an unsuitable figure-of-merit to rank architectures during the search.

We found it advantageous to impose a constraint on the training data, which reflected our use-case's particular accuracy requirements (cf section II-B). Higher prediction accuracy was required for MCs close to the target MC, and these samples were thus most important for the model to fit closely. We therefore removed from training all samples with MC outside the critical region, so that the model saw only critical region samples during training. This improved test set results, because the critical samples were more accurately predicted, while non-critical samples, though ignored by training, were still predicted within spec, sometimes because of their looser error allowance.

An aside: Grain MC exhibits hysteresis, *i.e.* different rates of absorption and desorption [11], [19], [20]. Thus, different linear models might be indicated for cases where RH is increasing vs. decreasing, assuming sufficient cases of each type. We did not explore this avenue.

F. Ideas that did not work

Three promising methods failed to yield benefit. We describe them briefly here, and follow up in the Discussion.

1) Physics-based equations: As noted earlier, various equations potentially offer straight-forward prediction of MC using $\widehat{RH}$ and T as inputs.

The equations require fitted parameters (generally denoted A, B, C) specific to each grain type and Temperature regime, which based on the literature would not be sufficiently accurate. However, the equations can be manipulated to give functions of the form $MC = f(g_1(RH,T), g_2(RH,T), ...)$ where $f(...)$ has somewhat simple form, by extracting intermediate functions $g(RH,T)$. Examples include:
(i) Oswin's equation (using [9], where parameter $C = 2$)

$$RH = \left[\left(\frac{a+bT}{MC}\right)^2 + 1\right]^{-1} \tag{1}$$

can be rearranged to give a linear function

$$MC = az + b(Tz), \text{ where } z = \frac{1}{\sqrt{\frac{1}{RH}} - 1} \tag{2}$$

(ii) The GAB equation does not include T, and can be written as a linear function $\frac{1}{MC} = \beta_0 + \beta_1 RH + \beta_2 \frac{1}{RH}$.

(iii) If T is assumed fixed (motivated by the minor importance of T in the regression models), Chung-Pfost can be rewritten as a linear function of $z = ln(ln(RH))$.

However, using these intermediate variables as features yielded poor results, for both linear and NN models.

2) Two-stage regression for improved $\widehat{RH}$: Although we were restricted to a 17 minute time-series in test samples, we had 60 minute time-series for training data, and $RH(60)$ was much closer to (or at) equilibrium state. We thus tried a two-stage regression, where we used one model to predict $RH(60)$ based on features derived from the 17 minute time-series, and a second model (as described above) to predict MC using the predicted $RH(60)$ value as the new $\widehat{RH}$ feature. This method yielded much more accurate estimates of $\widehat{RH}$, as shown in Fig. 4. However, the improved $\widehat{RH}$ feature did not improve test set accuracy.

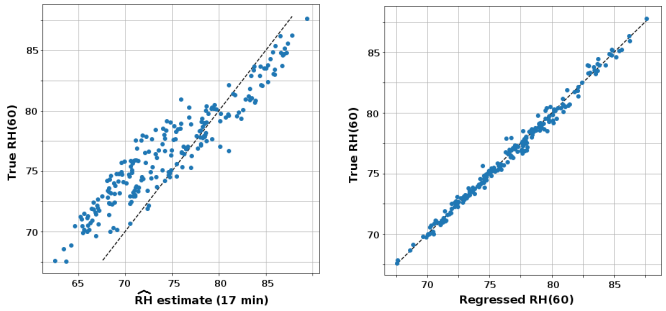

Fig. 4. Effect of regressing from 17-minute features to $RH(60)$. **A:** $RH(60)$ vs 17-minute $\widehat{RH}$. **B**: $RH(60)$ vs prediction from the 17-minute features. The 17-minute features very reliably estimate $RH(60)$.

3) Dewpoint conversions: Psychrometric charts [20], [21] enable an $\{RH, T\}$ pair to be mapped via dewpoint conversions to a different $\{RH', T_0\}$ pair, where RH' is uniquely determined by T_0. This raised the possibility of transforming RH time-series such that all T values were normalized to some T_0. This would be useful, for example, in linearizing the Chung-Pfost equation above. However, the wide range ambient $\{RH, T\}$ conditions led to similarly wide ranges

of $\{RH, T\}$ values in time-series. Due to the constraints of dewpoint conversion by psychrometric chart, we could not normalize all pairs to a single T_0. Further, even among subsets of data that could be normalized to a single T_0, final MC prediction accuracy was impaired.

III. Results

We report on four experiments that tested generalization of models: (1) Within dataset (Lab→Lab or Field→Field); (2) to new datasets (Lab→Field, Field→Lab); (3) to new RH-T conditions (Central→Corner); and (4) to the special case of unequal starting temperatures for device and grain.

To calculate accuracy for within-dataset experiments, we divided the data into 10 random folds. Each fold was excluded from training and treated as a test set in turn, and the 10 test set results were afterwards parsed into critical and non-critical MCs and combined to give accuracy statistics (10-fold cross-validation). For all other experiments, the training and test sets were naturally separate, and accuracy was reported for the test set. We reported separate accuracies for critical and non-critical region samples (critical region accuracy is most important functionally).

Experiments were run on each grain separately. For each grain, samples were curve-fitted, then either a linear or NN ($\{3,3\}$ architecture) model was trained with 5 features. Because the two model types gave largely equivalent results, we did not report which was used in each case.

A. Within dataset accuracy (Experiment 1)

This experiment assessed the cross-validated accuracy over datasets of samples with the same collection method (*i.e.* Lab-only or Field-only).

For all grains except Soy, prediction accuracy matched or exceeded performance specs. Soy, Beans, and Coffee were borderline, depending on how accuracy is calculated (*e.g.* on just critical region, on all samples, or on a fixed blend of critical and non-critical samples). For all grains, accuracy on samples in the non-critical region was close to perfect. Table II shows percent accuracy by grain for samples with MC inside the critical region, and outside (non-critical). Fig. 5 show scatterplots of estimated vs true MC by grain for Rice (top) and Beans (bottom).

B. Generalization Lab→Field (Experiment 2)

This experiment assessed whether a model trained on one type of data (*e.g.* Lab-generated) could generalize well to a different type (*e.g.* field generated). Only three grains (Maize, Nuts, and Sorghum) had large field sets and thus allow solid conclusions about generalization, and also allow Field→Lab generalization experiments (coffee had no Lab data). Results for other grains have limited strength due to small numbers of Field samples. Generalization from Lab to Field was varied but overall poor: Accuracy dropped between 0% and 20%.

Generalization in the Field→Lab direction was poor also. Note that the Field→Lab direction includes two kinds of generalization: Field→Lab, and Central→Corner region (cf Experiment 3 below). Table IV shows generalization accuracy in both directions, as well as numbers of Field samples. Fig. 6 shows scatterplots for model generalization from Lab→Field and from Field→Lab.

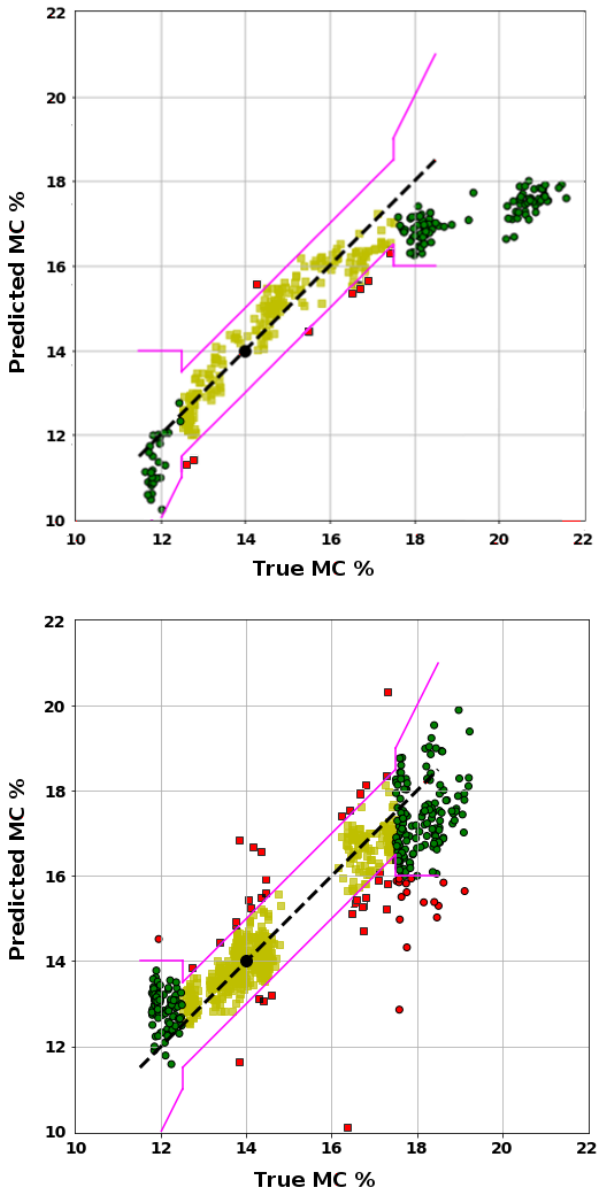

Fig. 5. Scatterplots of predicted MC vs true MC for Rice (top) and Beans (bottom). Yellow squares: critical region samples within spec (red squares: outside spec). Green circles: non-critical samples within spec (red circles: outside spec). Rice (top) shows an unusually strong bleed of non-critical samples into the looser error bounds. Beans (bottom) had one of the widest spreads, due to diversity of bean types.

C. Generalization Central to Corner regions (Experiment 3)

Field data was not available for the full grid of expected ambient RH-T combinations. A central region of the $\{RH, T\}$ space was well-covered, while corner regions were not (see Fig. 1). Note that "central-corner" is a split based on ambient

TABLE II
Prediction accuracy by grain for sample MCs in the critical region and in the non-critical region. "Bean" refers to Beans without Mung, which had considerably different size (accuracy for Beans including Mung in parentheses).

Accuracy (%)	Bean	Coffee	Cowpeas	Maize	Millet	Nuts	Rice	Sesame	Sorghum	Soy
Critical	93 (90)	94	96	97	99	96	95	97	99	91
Non-critical	99	100	100	100	100	100	100	100	100	100

TABLE III
Accuracy using a window ending at 12 minutes. "Bean" refers to Beans without Mung (accuracy for Beans including Mung in parentheses). Accuracy is almost identical to the 17 minute case.

Accuracy (%)	Bean	Coffee	Cowpeas	Maize	Millet	Nuts	Rice	Sesame	Sorghum	Soy
Critical	92 (89)	91	96	97	99	95	95	96	99	90
Non-critical	99	100	100	100	100	100	100	100	100	100

TABLE IV
Generalization accuracy, when trained on one type of data (Lab or Field) and tested on the other type. "Lab Critical" values are the same as in Table II. Only three grains (Maize, Nuts, and Sorghum) had sufficient data in both Lab and Field to draw solid conclusions, shown in bold font.

Accuracy (%)	Cowpeas	**Maize**	Millet	**Nuts**	Rice	Sesame	**Sorghum**	Soy
Lab Critical	96	97	99	96	95	97	99	91
Lab as Training	97	99	99	99	97	97	99	94
Field as Holdout	76	**82**	100	**84**	98	100	**92**	89
Field as Training	-	98	-	96	-	-	98	-
Lab as Holdout	-	**89**	-	**73**	-	-	**80**	-
# Field samples	33	352	27	428	43	21	100	18

RH-T conditions, distinct from the "critical-noncritical" split based on MC. We wished to know whether a model trained on the existing Field data (containing central region samples only) might effectively handle future corner samples in the field. This experiment assessed whether a model trained on central region samples generalized to corner region samples.

The Lab data was divided into central and corner regions as shown in Fig. 1. An algorithm was trained on samples from the central region only and tested on samples from the corner region. Accuracy on samples from the central region was assessed (by 10-fold cross-validation) as a control. Generalization was good. A model trained only on central RH-T region samples had little degradation in accuracy on corner region samples (0% to 6%). Table V shows percentage accuracy by grain on central and on corner regions, for an algorithm trained on central region samples.

D. Unequal device and grain temperatures (Experiment 4)

During field data collection, we observed that device and grain might sometimes start at different temperatures (in particular, if the grain was in the sun and the device was in the shade prior to loading). This was therefore a potentially important operating condition to add to the performance specs.

Small datasets simulating this unequal temperature case were generated in the Lab for Maize (n = 25) and Nuts (n = 21). This experiment assessed whether models trained on samples with equal device-grain temperatures could accurately predict MC of samples with unequal device-grain temperatures. We trained models to predict true MC on Lab samples with equal device-grain temperatures, then applied these models to the Lab samples with unequal temperatures (Δ = 5 or 10 °C).

The models did not generalize well on these small test sets. Test set accuracy for Nuts was 76%, for Maize 64% (with another 13% very close to the acceptable boundary), vs training set accuracies of 98% (10-fold cross-validation). Generalization was slightly better for samples with a temperature Δ = 5 °C than for those with Δ = 10 °C. Table VI gives the accuracy results.

IV. Discussion

We presented details of a two-step algorithm for predicting MC of grains. The algorithm acted on time-series for RH and T that had not reached the equilibrium state normally required for MC prediction, and had low computational cost. The algorithm matched or exceeded demanding performance specs (MC error $\leq$1% raw, on 95% of samples), given test samples from the same general distribution (*i.e.* Lab-generated or Field-generated). However, performance did not generalize well to samples from different distributions (*e.g.* Lab→Field).

A. Two-step algorithm

Our algorithm had two main steps: First, it fitted a quadratic to the RH time-series, a low-cost and effective way to estimate asymptotic RH, $\widehat{RH}$, based on the system response from time t = 5 minutes to t = 17 (or 12) minutes, as it moved towards but was still far from an equilibrium state.

TABLE V
Percentage accuracy on corner region samples, of algorithms trained on the central regions. Generalization to corner regions was good for most grains.

	Cowpeas	Millet	Rice	Sesame	Sorghum	Soy
Central sample size	115	106	113	78	112	115
Corner sample size	224	168	226	159	227	221
Central accuracy (%)	97	98	97	100	96	90
Corner accuracy (%)	92	95	92	97	97	84

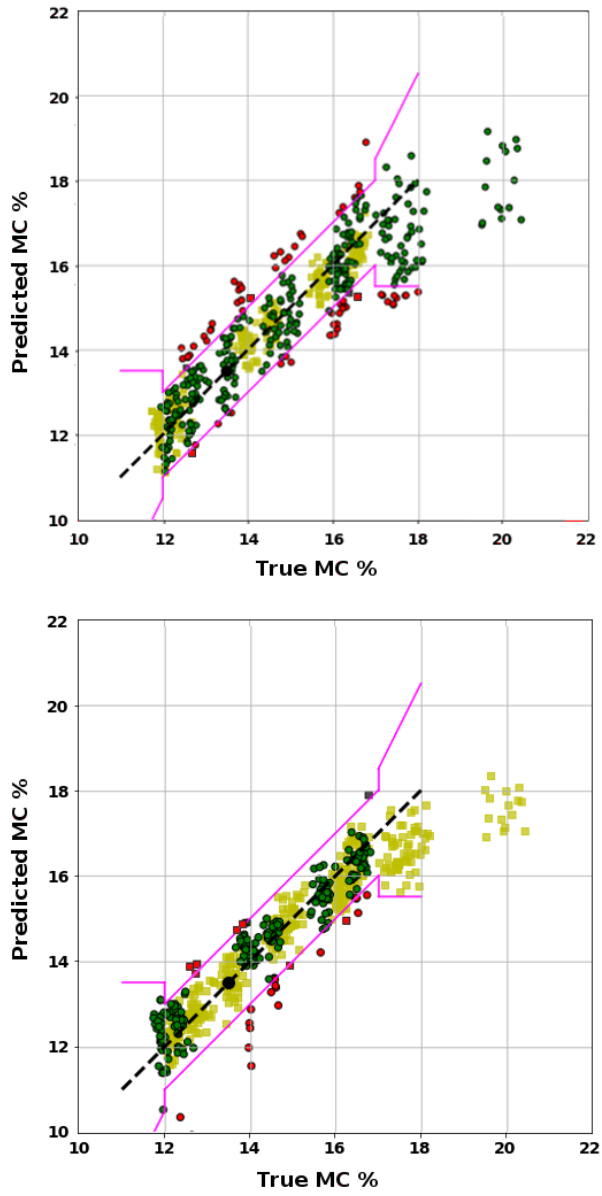

Fig. 6. Scatterplots of estimated vs true MC for Maize. x-axis = true MC. y-axis = predicted MC. Top: Field data predicted using models trained on Lab data. Bottom: Lab data predicted using model trained on Field data. Yellow squares are training set predictions (10-fold CV). Green circles are test samples within spec; red circles are test samples outside spec.

TABLE VI
Accuracies (as percentages) on training set (10-fold CV) and holdout set (*i.e.* samples with unequal grain-device temperatures) for Maize and Nuts. For Maize, an additional 13% of holdout samples were very close to the acceptable boundary. Numbers in parentheses are results parsed by samples with Δ = 5 vs Δ = 10. Generalization was better on samples with Δ = 5.

Accuracy (%)	Maize	Nuts
On training (CV)	99	98
On holdout	**64** (67, 62)	**76** (80, 73)
# holdout samples	25 (12, 13)	21 (10, 11)

Second, regression used this $\widehat{RH}$, as well as T and the quadratic fit parameters, as input features to simple models (linear or small NN). Because $\widehat{RH}$ is of central importance to the target variable MC, isolating it for use as a feature enabled a simple regression model to make accurate predictions. That is, it successfully leveraged physics-based priors.

B. Generalization from Lab→Field

From a practical point of view, the relevant direction for model generalization in Experiment 2 (section III-B) is Lab→Field, since it may be easier to generate Lab data but the device is deployed only in the Field. Degraded test set accuracy is typical when train and test samples are not drawn from the same distribution. The results suggest that the lab set-up did not reproduce field conditions closely enough to generate a stand-alone training set (i.e needing no field data), given the models used and the field performance spec of 95% accuracy. For example, exact species differed (lab-grown in U.S. vs field-grown in Uganda), and Lab samples were rewetted, which may have affected their behavior [11], [22].

We lacked Field data from corner regions of the $\{RH, T\}$ space of ambient conditions (the conditions were never encountered). Experiment 3 (section III-C) showed good generalization from central to corner region samples on Lab samples. But a small drop in accuracy might have outsized impact in terms of hitting the 95% accuracy spec on Field sample sets drawn purely from corner RH-T regions. Note that results of Experiment 2 (section III-B suggest that including Lab-generated samples with corner region conditions might not fully substitute for Field data from corner regions.

C. Ideas that did not work

Somewhat surprising was the failure of direct physics-based methods. We had expected these to improve regression

performance by constraining, in realistic ways, the solution space to be searched by the optimizer.

Sole use of physics-based equations, of form $MC = f(RH, T)$, gave much worse results than a regression model using RH, T, and curve parameter features. We hypothesize that these equations depend on conditions crucially distinct than those we had. That is, our data (especially from the Field) was better modeled by $MC = f(RH, T) + \epsilon(RH, T)$, where $\epsilon(RH, T)$ represents noise not captured by the physics-based function $f(RH, T)$, but still sufficiently dependent on RH and T to allow a regressor (such as a function-approximating NN) to accurately encode it. The loss in generalization from Lab→Field data (and vice versa) suggests that the $\epsilon(RH, T)$ terms were different for Lab and Field datasets.

Dewpoint conversions, as a means of normalizing temperatures T to some T_0, offered the promise of focusing predictive algorithms on RH as a feature. The wide range of RH-T conditions precluded use of a single T_0. However, even within subsets of the data that allowed normalization to a single T_0 the conversion was not helpful, for reasons still unclear to us.

Leveraging our readings of RH at 60 minutes (during training) to better predict $\widehat{RH}$ via an intermediate regressor gave no additional benefit. This was perhaps because the regression model implicitly did this step anyway, or because of the already high intra-data type accuracies (cf section III-A).

Given the (justifiable) popularity of deep NN methods, we might have tried to input the raw time-series into a deep NN, with convolutions or recurrent connections to accommodate the time dependencies. However, this strategy had three drawbacks: *(i)* The memory to store such a model, and the computation required to apply it to a sample in the field, would far exceed our hardware and battery life constraints; *(ii)* deep NNs typically require large amounts of training data, which in this case is expensive to collect, especially in the field; and *(iii)* such a model would not fully leverage the physical fact that grain MC is closely related to equilibrium RH and T.

D. Unequal temperature use-case

The possibility that grain and device might start at different temperatures, *e.g.* grain laying in the sun being loaded into a device that had been stored in the shade, was observed while collecting field data. We do not know how vital this use-case is, and whether it might be negotiated out of the product specifications. One might perhaps argue that this situation could be excluded in the instruction manual; but effective deployments must accept certain realities of field use. One of the benefits of field work is exposure to such unforeseen but vital usage details, and we are currently evaluating how to proceed in light of this new use-case.

The number of samples in this experiment was small ($\sim$25 per grain) making the result noisy. However, poor generalization from equal-temperature training samples to unequal-temperature test samples may have been due to *(i)* the method by which in-lab data was generated, or *(ii)* intrinsic limitations of our algorithmic approach:

(i) This batch of data was generated at a later date than the rest (since it was a response to field work observations), and some changes in processing method might have occurred.

(ii) MC is a function of grain moisture diffusion, which is strongly dependent on T and RH. In the equal-temperature case, T has minimal change over the 17 minute timecourse, while unequal-temperature cases can show high fluctuations in T. Given this difference, poor generalization is not surprising.

ACKNOWLEDGMENTS

We gratefully acknowledge funding support from the Bill and Melinda Gates Foundation Trust, via the Global Good Fund.

REFERENCES

[1] "Missing food: The case of postharvest grain losses in Sub-Saharan Africa," World Bank, 2011.
[2] J. Hayma, "The storage of tropical agricultural products," Agrodok 31, r1995.
[3] K. Bradford, P. Dahal, J. V. Asbrouck, K. Kunusoth, P. Bello, J. Thompson, and F. Wu, "The dry chain: Reducing postharvest losses and improving food safety in humid climates," *Trends in Food Science and Technology*, 2018.
[4] P. R. Armstrong and M. Weiting, "Design and testing of an instrument to measure equilibrium moisture content of grain," *Applied Engineering in Agriculture*, 2008.
[5] UC Davis, USAID. Drycard. [Online]. Available: https://horticulture.ucdavis.edu/drycard
[6] T. Tubbs, C. Woloshuk, and K. Ileleji, "A simple low-cost method of determining whether it is safe to store maize," *AIMS Agriculture and Food*, 2017.
[7] P. Armstrong, S. McNeil, N. Manu, A. Bosomtwe, J. Danso, E. Osekre, and G. Opit, "Development and evaluation of a low-cost probe-type instrument to measure the equilibrium moisture content of grain," *Applied Engineering in Agriculture*, 2017.
[8] S. Basu, U. Shivhare, and A. Mujumdar, "Models for sorption isotherms for foods: A review," *Drying Technology*, 2006.
[9] *ASAE D245.5: Moisture relationships of plantbased agricultural products*, ASAE Std., r2017.
[10] M. Uddin, P. Armstrong, and N. Zhang, "Accuracy of grain moisture content prediction using temperature and relative humidity sensors," *Applied Engineering in Agriculture*, 2006.
[11] D. Garg, R. Larson, and D. Maier, "Equilibrium moisture content (EMC) relationships of three popcorn grain varieties and their incorporation into in-bin grain conditioning strategies," *9th Int'l Working Conf on Stored Product Protection*, 2006.
[12] P. Armstrong, M. Casada, and J. Lawrence, "Development of equilibrium moisture relationships for storage moisture monitoring of corn," *Applied Engineering in Agriculture*, 2012.
[13] R. Chayjan, "Modeling of sesame seed dehydration energy requirements by a soft-computing approach," *Australian J of Crop Science*, 2010.
[14] R. Amiri-Chayjan and M. Esna-Ashari, "Comparison between mathematical models and artificial neural networks for prediction of sorption isotherm in rough rice," *Int Agrophysics*, 2010.
[15] J. Zhang and G. Mittal, "Prediction of equilibrium moisture content of various foods using a neural network," *Int'l J. Food Properties*, 2013.
[16] *Moisture meters for cereal grain and oilseeds Specification*, Uganda National Bureau of Standards Std., 2013.
[17] Python. [Online]. Available: https://www.python.org/
[18] scikit-learn. [Online]. Available: https://scikit-learn.org/stable/
[19] A. Denlove and A. Ade-John, "Moisture sorption isotherms of some Nigerian food grains," *J. Stored Product Research*, 1985.
[20] B. Bala, *Drying and Storage of Grains, 2nd Ed.* Wiley, 2017.
[21] S. Sadaka and R. Bautista, "Grain drying tools: Equilibrium moisture content tables and psychrometric charts FSA1074," *U of AR Div of Agr*.
[22] R. Bartosik and D. Maier, "Study of adsorption and desorption equilibrium relationships for yellow dent, white, and waxy corn types using the modified Chung-Pfost equation," *Trans ASABE*, 2007.