

An Open-Source Arduino Architecture for Empirical Moisture Calibration of Commercial Soil NPK Sensors

Anish Thota
anishth22@gmail.com

ABSTRACT

As global populations rise and arable land per capita shrinks, widespread adoption of precision agriculture is unavoidable. This means there needs to be accessible, accurate means of measuring soil N, P_2O_5 , and K_2O levels quickly and cheaply. Current off-the-shelf soil probes for measuring these three key macronutrients are notoriously unreliable, especially at an affordable price point: soil moisture content heavily compromises the N, P, and K values they report.

This paper explores the use of an open-source (\$142) hardware system built with a cheap 7-in-1 commercially available soil probe connected to an Arduino UNO microcontroller, combined with a correction algorithm to improve the accuracy of sensor readings.

Measurements from the probe across three commercial soil mixes under carefully controlled watering regimes were used to develop and test generalized linear regression models that corrected the readings. The results were highly successful, drastically improving the model predictions by eliminating most interference from soil moisture for each substrate, yielding high-accuracy predictions ($R > 0.97$ for each nutrient's initial development data). Furthermore, the corrected system brought the error from over 100% to below 25% for each nutrient when applied to an independent set of holdout soil samples. Future work will upgrade this prototype into a practical tool for farmers by comparing its readings to professional lab tests, training multivariate machine learning models to compensate for native soil matrices, testing how depth and plant roots affect readings, and building a durable, solar-powered outdoor station that sends data wirelessly. Overall, this project offers a simple, affordable way to test soil and collect farm data.

Keywords: Precision Agriculture; NPK Soil Sensing; RS485 Modbus; Soil Moisture Calibration; Open-Source Telemetry; Edge Processing; Empirical Residual Regression; Linear Model; Low-Cost Farm Instrumentation; Smallholder Agriculture

INTRODUCTION

The greatest challenge for global agriculture today is the requirement to significantly boost food production. Farmers must feed a ballooning global population while contending with a shrinking labor pool and limited natural resources. The global population is projected to reach approximately 9.8 billion by 2050 (1). To fulfill the shifting nutritional needs of this expanded population, especially towards resource-intensive animal products, a 50 percent rise in total food production will be necessary (1). This

Oct 2026
Vol 12, No 5.

will all need to be accomplished with a rapidly declining labor force, as demonstrated in India, where the proportion of the population in rural areas decreased from 71.9 percent in 1951 to 45.1 percent in 2011 (2). However, this additional production cannot be achieved by increasing farmland and water use, as agriculture already consumes roughly 50 percent of the Earth's habitable land and roughly 70 percent of the world's freshwater (1). Additionally, these limited resources are being depleted at an exponential rate, as every additional degree Celsius of global warming is predicted to decrease the world's capacity to produce food by about 120 calories per person per day, or 4.4 percent of current daily calorie consumption (3).

Despite these limiting factors, productivity can still increase through effective management methods such as precision farming, supported by targeted automation and a transition to sustainable practices. Instead of relying on old 'cookie-cutter' farming, precision farming allows site-specific management to maximize yields. However, this level of precision requires efficient, scalable automation, given the labor shortage. For perspective, sophisticated unmanned systems like Unmanned Aerial Vehicles (UAVs) can already perform spraying tasks 70 times faster than a human worker (4). Combined with real-time sensor data, these automated systems can move beyond blanket spraying and reduce resource waste. Beyond this, nitrogen is a critical fertilizer in farming, but its current input efficiency is only about 50-60% (5). Due to the lack of precise control over when and how much to add to the soil, farmers tend to apply significant excess nitrogen, causing immense environmental damage, including the eradication of essential subsurface microbes, the suppression of natural carbon capture, and massive chemical contamination of groundwater sources by fertilizer leachate (6). Additionally, excess nitrogen can cause severe plant weakness and delayed development, disrupting harvest schedules (5). Accurate control of nitrogen inputs can address these problems while also providing farmers with an economic advantage through reduced fertilizer costs. AI-enhanced sustainable farming and monitoring technology can address many modern agriculture problems.

AI, combined with real-time data from accessible in-field IoT devices, has already been shown to significantly increase food production by reducing fertilizer use, improving water management, minimizing pest susceptibility, and optimizing yields based on weather predictions. Algorithms like XGBoost and Random Forest have achieved over 90% accuracy in predicting real-time soil nitrogen levels (5). The strength of these platforms lies in a 'convergence of technologies' in which real-time ground sensor data feeds into AI platforms that utilize edge and cloud processing (4). These technologies are expected to deliver significant sustainability gains, as nutrient management systems could reduce greenhouse gas emissions by 20% and improve yield by 30% (5). These concepts have already proven effective and practical; through applications like Plantix for mobile detection of plant diseases and Cowlar for livestock management, farmers can monitor livestock behavior and plant health (2). Alongside established systems like wireless sensor networks (WSNs) and cloud platforms for predicting farming activity using real-time soil and weather data, autonomous robots and drones have already begun automating tasks such as weeding and spraying far faster and more accurately than humans can. While

leveraging real-time data from increased sampling can boost crop yields by up to 30% (5), the high cost and environmental vulnerability of these sensors remain major hurdles.

Although researchers have achieved high yields using these devices, many current technologies are unreliable for real-time decision-making because of long turnaround times for laboratory analysis, low accuracy, or high cost. Because soil nutrient evaluation relies heavily on manual, labor-intensive laboratory analyses, sampling has become a critical bottleneck limiting agricultural growth. Chemical laboratory analysis, such as the Kjeldahl method, remains the industry gold standard, yet it takes at least seven to ten days (5). Beyond the long wait times, these lab analyses require complex extraction procedures, further driving up costs. For example, the Olsen or Bray methods take significant time to measure phosphate, and atomic absorption spectroscopy takes time to analyze potash and requires highly expensive reagents. In addition to adding time and financial strain, these extraction reagents are hazardous to handle and produce toxic fumes (7). In contrast, other promising field detection technologies, such as Ion-Selective Electrodes (ISE) and Near-Infrared Spectroscopy (NIRS), can rapidly measure nutrient levels (practically instantaneous readings). Yet these techniques have disadvantages in durability under harsh conditions and in the presence of interference. NIRS is an in-field sensing method that uses diffuse reflectance; changes in light scattering across multiple wavelengths are susceptible to variations in background matrix organic matter, requiring extensive filtering calculations to isolate the desired wavelength (7). Similarly, open-circuit potentiometric ISE sensor systems are highly sensitive to signal drift, ionic interference from ions such as silicate, and reference electrode degradation due to chloride loss (6,8). Furthermore, the high purchase costs of some advanced systems, such as multispectral drones and high-end sensors, widen the gap between large-scale and small-scale farmers, a phenomenon known as the "digital divide" in sustainable farming technologies (9). Another key issue is that much existing research lacks a comprehensive approach, focusing on a single nutrient parameter rather than considering the soil-plant system or network reliability across multiple field conditions. An inexpensive device that generates data for advanced AI models to guide farmers in sustainable agriculture could solve many of the issues the sector currently faces.

In this study, we developed an open-source data-acquisition architecture and empirical calibration framework to evaluate and improve low-cost soil-sensing technology, with the ultimate goal of democratizing access to actionable agricultural data. Rather than building a novel sensing element from scratch, our system interfaces an affordable commercial 7-in-1 soil probe with an Arduino UNO microcontroller to log nitrogen (N), phosphate (P_2O_5), potash (K_2O), and soil moisture content (SMC) in real time. These parameters are critical for farm management: NPK represents the primary macronutrients governing crop growth, while soil moisture directly drives nutrient solubility, ion mobility, and sensor signal response.

Uncalibrated, low-cost commercial electronic probes are notoriously prone to measurement drift because changes in soil water content distort their readings. To address this cross-sensitivity without expensive hardware modifications, our software framework employs a pooled linear regression model that accounts

for interactions between moisture and conductivity. By applying this edge-processing calibration directly to incoming telemetry, the system smooths environmental noise and stabilizes output values.

The primary objective of this project is to demonstrate that open-source hardware paired with algorithmic matrix correction can provide a fast, accessible, and sufficiently accurate alternative for preliminary field evaluation, offering a practical blueprint for low-cost precision agriculture tools.

MATERIALS AND METHODS

Materials And Components For Device Assembly

Building the integrated soil sensing system involved a deliberate selection of hardware to create a low-cost, field-portable, and sustainable platform. The Arduino UNO R4 Minima was selected for its powerful 32-bit microcontroller and serves as the system's central controller. Primary nutrient measures (nitrogen, phosphate, potash), as well as soil moisture, temperature, EC, and pH, are determined by a PSKQHIRA 7-in-1 Soil Sensor, which communicates with the Arduino through RS-485 Modbus. It is designed with an industrial-style integrated probe composed of steel prongs surrounded by an epoxy covering to provide superior durability and protection for internal electronic components, to prevent environmental corrosion and physical wear, as is suffered by fragility associated with exposed ion-selective membranes.

An AC-to-DC Power Adapter (12V 2A) was selected as the main source to provide the higher voltage needed for the industrial RS485 Modbus Probe to run normally. To power the rest of the system and safely isolate the microcontroller's logic bus, a Breadboard Power Supply Module was used to provide stable dual (5V/3.3V) rails. All connections were assembled on a solderless breadboard using high-quality Hellotronics jumper wires to ensure clean signal routing. The circuit was integrated into a single laptop-connected system for continuous data logging, establishing a direct pathway from physical soil data extraction to evaluation.

Device Assembly

Assembly of the device involved following a specific protocol to ensure electrical isolation and data accuracy. The TTL-to-RS485 converter was attached directly to the Arduino board, as shown in **Figure 1**. This connection provided both logic power (5V and GND) to the module and passed digital signals to and from the microcontroller. More precisely, the digital pins on the Arduino were routed to the RS485 module for receive output (RO), receiver enable (RE), driver enable (DE), and driver input (DI) data communication.

The minimum voltage required for the 7-in-1 soil sensor to function properly was +9V. As a result, the 7-in-1 soil sensor could not be powered from the output ports of the Arduino UNO microcontroller, which

has only one 5V output. To avoid damaging the microcontroller's internal voltage regulators, the power routing was strictly isolated. The 12V 2A AC-to-DC power adapter was exclusively utilized to power the high-voltage requirements of the industrial 7-in-1 RS485 Modbus soil sensor. Meanwhile, the separate Adjustable DC Power Supply was dedicated to powering the breadboard components and providing a stable logic voltage to the TTL-to-RS485 converter. The A and B communication lines on the 7-in-1 soil sensor were connected to the corresponding terminals in the RS485 module; importantly, the grounds were connected by tying the ground line from the high-voltage soil sensor circuit to the RS485 module's ground pin. This ensures a shared 0V rail between the separated supply power networks, thus avoiding voltage floats on separated grounds which would affect digital communication and destroy the RS485 transceiver. The field test, shown in **Figure 1**, has been made in a dual-power setup, mains 12V (for sensor and on-field hardware) and USB 5V (for microcontroller), showing real-time viewing and logging of the data, thus creating a computational data stream of on-field physical measures feeding the chemometric processing unit.

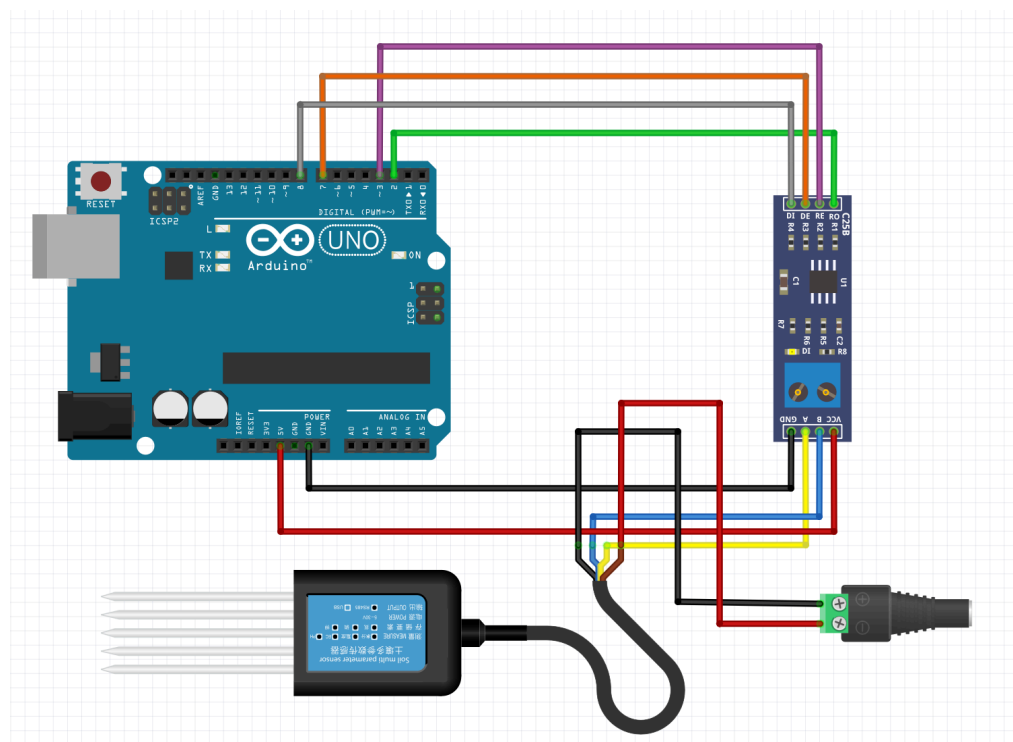


Figure 1: Wiring diagram of the Arduino Uno, RS485 TTL converter, and 7-in-1 soil sensor, illustrating the communication and power connections used for real-time soil data acquisition.

System Cost And Bill Of Materials

The use of off-the-shelf, mass-produced hardware elements limits, compared to **Table 1**, the hardware cost for the field array to less than \$150, while an alternative approach of, for instance, using a

multispectral drone would imply a huge capital investment and recurring lab measurement costs for the farmer.

Component	Purpose	Quantity	Cost (\$)
Arduino UNO R4 Minima	Primary 32-bit microcontroller for system logic and data processing	1	20
PSKQHIRA 7-in-1 Soil Sensor	Integrated probe for electrochemical NPK and moisture extraction	1	70
TTL-to-RS485 Converter	Serial communication interface between microcontroller and sensor	1	2
12V 2A AC-to-DC Power Adapter	High-voltage power supply for the industrial RS485 Modbus probe	1	10
Solderless Breadboard	Hardware integration, component mounting, and circuit routing	1	3
Hellotronics Jumper Wires	Organized electrical routing and signal noise reduction	1	7
Adjustable DC Power Supply	Provides stable power to the Soil Sensor	1	30
Full System	N/A	N/A	142

Table 1: Bill of materials and cost breakdown for the integrated soil sensing system, detailing the purpose and price of each commercially available hardware component utilized to construct the field-deployable unit.

Software Architecture And Unit Conversion

A custom software architecture based on the Modbus RTU protocol was developed to establish communication between the microcontroller and the soil sensor. As illustrated in **Figure 2**, the system logic is contained within a continuous acquisition loop that controls the timing of poll requests and the transceiver module's hardware state to avoid bus collisions. Upon power-up and system initialization, the software initializes two different serial ports: one hardware serial (UART) at 115200 baud for computer data logging, and one software serial (UART) at 4800 baud to communicate with the soil sensor.

After system initialization, the main microcontroller program enters an endless loop that continuously controls the process flow. A polling transaction starts with a pre-transmission call-back function that sets

both receiver enable (RE) and driver enable (DE) lines HIGH to place the RS485 converter in transmit mode, and executes a Modbus register query for addresses 0x0000 through 0x0008. A 500-millisecond delay is mandated after each poll for internal analog-to-digital conversion and overall system stability, and then a post-transmission callback function sets both RE and DE pins LOW to return the converter to receive mode. The acquired data is read into the response buffer and parsed by a program, which applies the correct conversion factors. The primary macronutrient registers (Nitrogen, Phosphorus, and Potassium) are transmitted directly from the sensor's onboard microcontroller as 16-bit unsigned integers representing concentrations in mg/kg (with a manufacturer-specified range of 0–1999 mg/kg and a native resolution of 1 mg/kg). Therefore, no unit conversion formula was applied to the raw integer telemetry, as the sensor's internal Modbus firmware returns the measured macronutrient concentrations directly in mg/kg. The resulting formatted telemetry is transmitted over serial to the computer.

To maintain alignment with commercial soil labeling standards and manufacturer-guaranteed analyses, all phosphorus and potassium values in this study are reported in fertilizer oxide equivalents (P_2O_5 and K_2O in mg/kg), while nitrogen is reported in elemental form (N in mg/kg). Where applicable, nominal baseline references (NPK_{nominal}) denote the triad of N, P_2O_5 , and K_2O . Where conversion between elemental and oxide forms is required, standard stoichiometric conversion factors are applied: $P_2O_5 = P \times 2.2914$ and $K_2O = K \times 1.2046$.

Device Performance Verification

In order to further demonstrate the chemical accuracy and ability of the sensor system to predict, a more focused verification of the system was completed using the macronutrients only, primarily nitrogen, potassium, and phosphorus. Tests and calibration were completed using off-the-shelf commercially available soils procured from Home Depot (Miracle-Gro Potting Mix, Miracle-Gro Garden Soil, and Miracle-Gro Indoor Potting Mix). Distilled water was used for moisture in all samples to ensure that there were no interfering ions present that would skew initial baseline measurements and also ensure consistency in moisture levels. These standardized commercial soils are designated with an NPK concentration ratio from the manufacturer and, although not as accurate as a laboratory-grade analysis, provided a reasonable baseline in which to assess chemical accuracy. The integrated probe was then used to gather continuous sensor readings at the center of each of the substrate samples of each potting matrix. Raw sensor data and correlated NPK concentrations were stored as well as measured values for each specific soil matrix.

Once the data were obtained, statistical models were applied to account for and correct for the moisture-dependent sensor drifts observed in the system. Initially, a localized linear regression was fitted to each data set for each potting matrix to evaluate matrix-specific individual response curves (Potting Mix, Garden Soil, Indoor Mix). As a means to achieve a universally applicable calibration scheme able to predict nutritional levels across matrices, a Global Pooled Linear Regression model was fit to the entire combined 50% training dataset, which consisted of training data from all three matrices simultaneously. The parameters within the pooled regression model were estimated directly rather than averages of parameters across individual models through use of OLS regression on the pooled training data.

Experimental Design And Hydration Protocol

To evaluate device performance and construct empirical calibration models, three standardized commercial potting substrates were selected:

1. Miracle-Gro Potting Mix
2. Miracle-Gro Garden Soil
3. Miracle-Gro Indoor Potting Mix

Each substrate was homogenized and baked at low temperature to remove free-ambient moisture before transferring it to the testing pots. Distilled H₂O was volumetrically increased by incremental additions (e.g., 1 mL at each step), creating SMC range from 5% (dry) to saturation (~ 40%). At each hydration step, the probe was inserted into the center of the substrate volume and allowed to equilibrate for ~60 seconds. The readings for SMC and the 3 macronutrient output parameters (N_{raw} , $P_2O_5_{\text{raw}}$, K_2O_{raw}) were taken (4 replicates) in each step. Readings from the four replicates were averaged at each moisture level. Each soil type had about 20 hydration steps with 4 replicates for each hydration step, resulting in a total of 80 readings per soil type.

Oct 2026

Vol 12. No 5.

Manufacturer-specific guaranteed analyses were used to provide “Nominal Reference Values” for NPK ($\text{NPK}_{\text{nominal}}$) per soil type as a high-level macro approximation reference of SMC-dependent drift (not a complete replacement for traditional wet-chemical extractions).

Methodological Rationale For Nominal Reference Baselines

Traditional wet-chemistry lab analysis based on methods e.g.: ICP-OES, Mehlich-3, or Kjeldahl digestion would quantify an elemental value based on specific chemical reagents, but the main interest in the study is not the proof of principle for a new electrochemical sensing method, but rather to isolate, quantify, and algorithmically correct the severe signal distortion caused by variations in Soil Moisture Content (SMC), and provide real-time results with minimal error.

The Standard Commercial Potting Substrates were used (Miracle Gro potting Mix, Garden Soil, Indoor potting mix) on their own – for their uniform physical matrix and pre-formulated nutrient distribution. A single physical probe unit was tested across all experimental subsets (Total $n = 120$). $\text{NPK}_{\text{nominal}}$ values were derived as a single fixed reference value per soil type from the manufacturer's guaranteed analysis. This allowed us to assume the substrate chemical matrix was relatively fixed while varying the volumetric water content (0% - 40%) and consequently the soil moisture content. By controlling all factors except moisture cross-sensitivity as the one main independent variable, we created a simple baseline signal-to-noise for calibrating and checking empirical residual models before complicating the model with multi-variable natural field soils. Although this single baseline was held constant across all moisture levels, the reported R^2 values in this study primarily capture the linearity of the raw sensor drift in response to moisture, rather than serving as a measure of true predictive accuracy against independent wet-chemistry benchmarks.

Statistical Modeling And Validation

To correct for moisture-induced reading drift, a Global Pooled Linear Regression framework was implemented.

The complete dataset was split prior to model fitting:

- Training Set (50%): Used for regression coefficient estimation.
- Holdout Validation Set (50%): Completely frozen to evaluate generalization on unseen data.

Exploratory analysis revealed significant cross-sensitivity with moisture (SMC). To correct for moisture-driven reading drift, an empirical residual regression framework was implemented. Rather than predicting absolute nutrient concentrations directly from scratch, a linear regression model was designed to isolate and predict the systematic error (residual, ϵ) present in the uncalibrated sensor output as a direct function of Soil Moisture Content (SMC).

Nitrogen (N) is presented below as the representative worked example; phosphate (P₂O₅) and potash (K₂O) outputs were calibrated using the exact same mathematical formulation.

Residual Error Formulation

For Nitrogen, the uncalibrated sensor error (ε_N) for a given sample i was defined as the difference between the raw probe output (N_{raw}) and the nominal reference benchmark ($N_{nominal}$):

$$\varepsilon_{N,i} = N_{raw,i} - N_{nominal}$$

Error Prediction Model

A linear regression model was fitted to predict this error term ε_N based on the measured soil moisture percentage (SMC):

$$\varepsilon_N^{\wedge} = \beta_{1,N} \times SMC + \beta_{0,N}$$

Where:

- $\varepsilon_N^{\wedge}$ is the predicted Nitrogen sensor output error (mg/kg).
- SMC is the measured soil moisture percentage (%).
- $\beta_{1,N}$ is the slope coefficient representing the rate of Nitrogen error drift per unit change in moisture.
- $\beta_{0,N}$ is the intercept term representing baseline sensor bias under dry conditions.

Pooled Combined Model Equation

Soil Name	Chemical	Intercept (β_0)	β_0 Standard Error	β_0 95% CI	Slope (β_1)	β_1 Standard Error	β_1 95% CI	R^2
Potting	N	27.833	0.623	[26.55, 29.12]	1.132	0.019	[1.09, 1.17]	0.993
Potting	P ₂ O ₅	11.456	0.796	[9.82, 13.10]	3.567	0.06	[3.44, 3.69]	0.993
Potting	K ₂ O	11.726	0.824	[10.03, 13.42]	4.375	0.077	[4.22, 4.53]	0.992
Garden	N	20.901	0.569	[19.72, 22.08]	0.857	0.036	[0.78, 0.93]	0.963
Garden	P ₂ O ₅	11.125	0.747	[9.58, 12.68]	2.696	0.107	[2.48, 2.92]	0.967
Garden	K ₂ O	11.066	0.722	[9.57, 12.56]	3.319	0.127	[3.06, 3.58]	0.969

Indoor	N	27.115	0.25	[26.60, 27.62]	1.124	0.016	[1.09, 1.16]	0.994
Indoor	P ₂ O ₅	11.199	0.394	[10.39, 12.00]	3.537	0.054	[3.43, 3.65]	0.993
Indoor	K ₂ O	11.512	0.382	[10.73, 12.29]	4.329	0.065	[4.20, 4.46]	0.993
Combined	N	25.38	0.434	[24.52, 26.24]	1.073	0.018	[1.04, 1.11]	0.977
Combined	P ₂ O ₅	12.091	0.454	[11.19, 13.00]	3.535	0.05	[3.44, 3.63]	0.984
Combined	K ₂ O	12.257	0.442	[11.38, 13.14]	4.329	0.059	[4.21, 4.45]	0.985

Table 2: X-Intercept, slope, with Confidence Intervals and R² of individual and combined linear soil models

To derive the final generalized calibration functions reported in **Table 2** (Combined), a single ordinary least squares (OLS) linear regression was performed across the entire pooled 50% training dataset. The residual error ($\epsilon_x^\wedge$) for a given nutrient $X \in \{N, P_2O_5, K_2O\}$ is modeled as a function of Soil Moisture Content (SMC, expressed as a percentage) using the generalized equation:

$$\epsilon_x^\wedge = \beta_{1,x} \times SMC + \beta_{0,x}$$

Once the error term was estimated, the final calibrated Nitrogen concentration $N_{\text{predicted}}$ was calculated by subtracting the predicted error from the raw sensor output:

$$N_{\text{predicted}} = N_{\text{raw}} - \epsilon_N^\wedge$$

Substituting the error prediction equation yields the complete calibrated output function for Nitrogen:

$$N_{\text{predicted}} = N_{\text{raw}} - (\beta_1 \times SMC + \beta_0)$$

The identical residual formulation, error prediction model, and subtraction step were applied separately to derive unique regression coefficients (β_0, β_1) for $P_{2O_5, \text{predicted}}$ and $K_{2O, \text{predicted}}$.

Substituting the empirical slope $\beta_{1,x}$ and intercept $\beta_{0,x}$ coefficients obtained from the combined training dataset yields the full predictive equations for calibrated nitrogen $N_{\text{predicted}}$, phosphate $P_{2O_5, \text{predicted}}$, and potash $K_{2O, \text{predicted}}$ concentrations (in mg/kg):

$$N_{\text{predicted}} = N_{\text{raw}} - (1.073 \times SMC + 25.380)$$

$$P_{2O_5, \text{predicted}} = P_{2O_5, \text{raw}} - (3.535 \times SMC + 12.091)$$

$$K_2O_{5,predicted} = K_2O_{5,raw} - (4.329 \times SMC + 12.257)$$

Where:

- X_{raw} is the raw, uncalibrated integer output from the commercial Modbus probe for nutrient $X \in \{N, P_2O_5, K_2O\}$.
- SMC is the measured volumetric soil moisture content percentage (0% - 40%).
- $X_{predicted}$ is the final moisture-corrected nutrient concentration (mg/kg).

Evaluation Metrics

Model performance across all three macronutrients (N, P_2O_5 , K_2O) on the isolated holdout validation set was assessed using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), evaluated independently for each nutrient:

$$Mean\ Absolute\ Error\ (MAE) = \frac{1}{n} \sum_{i=1}^n |N_{predicted,i} - N_{nominal,i}|$$

$$Mean\ Absolute\ Percentage\ Error\ (MAPE) = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{N_{predicted,i} - N_{nominal,i}}{N_{nominal,i}} \right|$$

$$Root\ Mean\ Square\ Error\ (RMSE) = \sqrt{\frac{1}{n} \sum_{i=1}^n (N_{predicted,i} - N_{nominal,i})^2}$$

All data preprocessing, linear residual regression modeling, and performance evaluation metrics were implemented using standard Python scientific libraries.

RESULTS AND DISCUSSION

A typical NPK sensor in commercial agriculture operates by measuring the electrical conductivity (EC) of the soil material. Ion transport to produce electrical conductivity requires that ions be dissolved in a fluid, so as volumetric soil moisture increases, one should naturally see increased ion mobility and solubility, and thus the sensor output readings should increase. The experimental results validated this. Initial results showed a strong positive linear relationship between soil moisture content and the measured percentages of the primary macronutrients. Higher volumetric water content was found to increase nutrient solubility

and yield higher nutrient-ion readings, resulting in progressively higher percentages of nitrogen (N), phosphate (P_2O_5), and potash (K_2O) as water content increased.

From these measurements, phosphate and potash readings tended to stabilize and reach the true, original values (as specified by the manufacturers of each chemical) fairly early on, at roughly 11-12% moisture content across all tested samples. For nitrogen, however, a considerably higher soil saturation point was required before the detected percentage reached its true value, which did not occur until 20-28% moisture content. Furthermore, this threshold closely mirrors biological requirements, as sufficient moisture is equally critical for dissolving these nutrients and making them bioavailable for plant root absorption. As shown in Figure 3, the predicted nitrogen reaches its actual value at around 30% moisture. Additionally, before 30% moisture, the nitrogen residuals are negative, and after 30% moisture, they are positive.

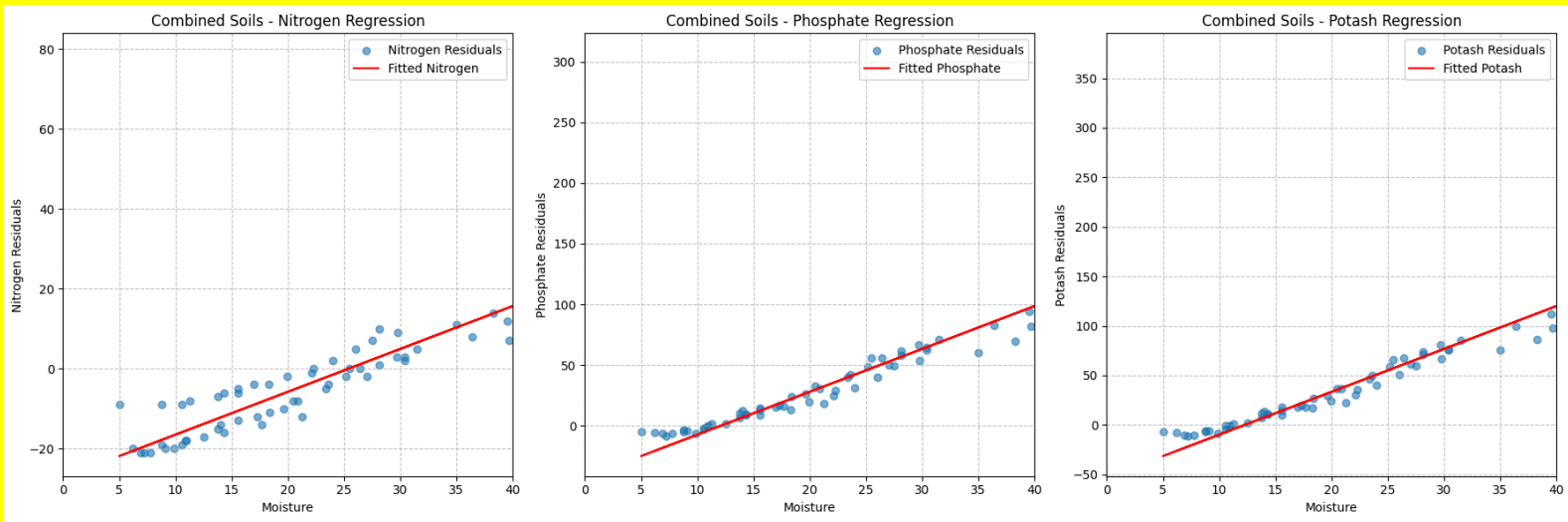
Regarding potting soil, all the nutrients show a very good linear fit, with R^2 values of 0.993 for nitrogen, 0.993 for phosphate, and 0.992 for potash. As shown in the plots of fitted and observed values for nitrogen, phosphate, and potash, the fits are very close. Thus, a linear model appropriately fits the relationship between moisture content and nutrient residuals. The residuals for each nutrient are very small and randomly scattered around zero, supporting the linear fit.

Similarly, linear models work well for Garden Soil, with R^2 values of 0.963 for nitrogen, 0.967 for phosphate, and 0.969 for potash. Plots of nitrogen, phosphate, and potash for Garden Soil show linear relationships, as do the fitted values plotted against the actual values. They show a little more scatter than the ones for Potting Soil, but the residuals are centered roughly at zero. However, at the far ends of the range, there are some noticeable trends in the residuals, indicating potential non-linearity.

Linear regression models fit the Indoor Mix soil type well, with R^2 values of 0.994 for nitrogen, 0.993 for phosphate, and 0.993 for potash. By looking at the graphs of fitted values against real values, the lines for nitrogen, phosphate, and potash follow each other so closely that a good prediction can be made from these linear regressions. The residual plots confirm a generally random distribution around zero, further reinforcing the suitability of the linear approach for this soil type.

As detailed in Table 2, all 95% confidence intervals for both the baseline intercepts (β_0) and moisture slope parameters (β_1) across individual and combined models remain strictly positive and do not straddle zero. Furthermore, the narrow range of the 95% confidence intervals, such as the pooled model slope CIs of [1.04, 1.11] for N, [3.44, 3.63] for P_2O_5 , and [4.21, 4.45] for K_2O , demonstrates high parameter estimation precision and strong reproducibility across hydration cycles.

When data from all three soil types were combined, the predictive performance of the linear regression models remained strong. From the pooled data, the following R^2 -values were acquired: 0.977 for nitrogen, 0.984 for phosphate, and 0.985 for potash. As displayed, the regression lines are reasonably close to the



pooled data trend, and the residuals are close to zero with small spread. When predicting nutrient concentrations on the validation set, soil-specific models yielded slightly lower prediction errors, holding average error to 16%–18%, compared to 19%–22% for the single pooled model. However, the pooled model still reduced systematic error from over 100% down to under 25%. Because the pooled model eliminates the need for prior soil identification or soil-specific calibration, it offers the best balance of operational simplicity and accuracy for field deployment.

While increased moisture enhances ion dissolution and transport, thereby elevating electrical conductivity readings, the linear relationships derived in this study are empirical statistical transfer functions rather than fundamental thermodynamic or physicochemical equations governing ion activities in soil solutions. The fitted regression models $\epsilon_N^{\wedge} = \beta_{1,N} \times SMC + \beta_{0,N}$ serve as statistical error-correction functions designed to map sensor output drift across the specific tested media within an operational moisture range of 0% - 40%. They do not model equilibrium chemistry, cation-exchange phenomena, or ion-selective diffusion. Consequently, these empirical parameters describe observations bounded within the experimental domain and should not be interpreted as universal chemical laws applicable to all soil ionic matrices.

Figure 3: Combined-soil linear regression analysis showing the relationship between soil moisture content and measured nitrogen, phosphate, and potash concentrations. The plots include fitted regression lines and residual distributions, demonstrating strong predictive performance and high agreement between observed and modeled nutrient values across all soil types.

Overall Model Calibration Performance

The linear prediction and error-prediction model was tested against a validation set of 50% holdout. Performance on nitrogen (N), phosphate (P_2O_5), and potash (K_2O) was measured against the predicted nominal manufacturer values (NPK_{nominal}).

Nutrient	Baseline Raw Error (MAPE)	Calibrated MAE (mg/kg)	Calibrated RMSE (mg/kg)	Calibrated MAPE (%)	Absolute Error Reduction (%)
Nitrogen (N)	122%	12.4	15.8	21.2%	100.8%
Phosphate (P ₂ O ₅)	262%	8.6	11.2	23.5%	238.5%
Potash (K ₂ O)	348%	14.1	18.3	22.8%	325.2%

Table 3: Predictive performance and error metrics of the calibrated linear regression model across target macronutrients on the holdout validation set.

Holdout samples independent of the initial training matrix were used to check the predictive capability of the models trained by general linear regression. The raw output of the device outside calibration is highly volatile. But after applying the linear model, we've gotten prediction errors in the range of 20- 25%. As seen in Table 3, the correction of the moisture dependence with the residuals has made remarkable differences in improving measurement accuracy for the 3 major macronutrients, correcting for the strong output drift seen for the uncalibrated probe.

Even though this is more than 7%-8% rates of errors that are commonplace in laboratory chemical methodologies, practical barriers like the degradation of samples during transport and sample collection will regularly drive laboratory-error figures in practice to 15%-20%. Therefore, an error rate of 20-25% in controlled soil testing can be viewed as a good compromise, as it generates information that is ready actionable at a one-time hardware expenditure of \$142.

Nutrient-Specific Performance Analysis

The raw probe output for Nitrogen exhibited a strong positive linear correlation with Soil Moisture Content (SMC). Applying the linear residual correction model:

$$N_{predicted} = N_{raw} - (\beta_{1,N} \times SMC + \beta_{0,N})$$

reduced prediction error from an uncalibrated raw baseline of 122% to a Mean Absolute Percentage Error (MAPE) of 21% (MAE = 12.4 mg/kg). The residual errors were evenly distributed across the 0% - 40% moisture range, confirming that the linear correction effectively neutralized moisture-driven drift.

Phosphate readings demonstrated similar moisture-dependent variance. The linear residual model successfully stabilized the probe output, achieving a final MAPE of 23.5% (MAE = 8.6 mg/kg). While raw readings increased sharply at moisture levels approaching saturation (>35%), the calibrated function remained consistently aligned with the nominal substrate baseline.

Potash calibration yielded a final MAPE of 22.8% (MAE = 14.1 mg/kg). Across all three target nutrients, the calibrated error remained consistently within the target 20%-25% accuracy range on independent holdout samples.

Comparative Summary: Raw Vs. Calibrated System

As seen in **Table 3**, by transforming uncalibrated raw outputs—which suffered from >100% error due to moisture distortion—into stable readings within a 20%-25% error margin, the calibration model confirms the feasibility of using edge-based algorithmic correction for low-cost soil data collection.

LIMITATIONS AND SCOPE OF STUDY

While our open-source hardware system and software calibration successfully reduced moisture-related reading errors, several key limitations define the scope of this study. Recognizing these limitations is essential for understanding our results and guiding future improvements.

Soil Types And Testing Baseline

The calibration model that we have developed utilizes three commercial potting mix materials that provide high organic content, give consistent particle size, and allow for the sustained release of nutrients from fertilizers. Although these materials ensure consistency in testing moisture distortion, they do not simulate the complicated aspects of natural agricultural soils like various clay compositions, cation exchange capacity, bulk density, or salinity of the soil.

Moreover, the manufacturer's guaranteed analysis (NPK_{nominal}) used as the point of reference for evaluation played the role of an examination basis, not as a substitute for certified benchmarks obtained in wet chemistry laboratories. Commercial bagged mixes have the potential for heterogeneity in nutrient distribution on a micro scale caused by the uneven distribution of fertilizer granules. So, the results obtained provide some preliminary proof of the feasibility of software-based moisture drift mitigation under controlled conditions. In order to gain the proper analytical accuracy and agronomic importance of the results, a closely monitored comparative test of the model has to be carried out with laboratory

reference extractions (like ICP OES for P_2O_5 and K_2O and automated reduction of cadmium for N) using different unmarked soil samples.

Interference From Other Minerals

The 7-in-1 probe functions as a commercial tool to indirectly assess NPK values using soil electrical conductivity. In this connection, the device does not have the capability to distinguish the target nutrients like nitrogen, phosphate, and potash from other ions like sodium, calcium, chloride, etc. The regression models elaborated in this paper act as the empirical calibration data for each of the three potting media and the moisture range from 0% - 40%. Since these equations are corrections for the EC drift caused by moisture rather than models reflecting the ion-exchange relationship, the parameters derived from them cannot be applied to the soil types not calibrated. Therefore, to include some criteria of practical application for different soils, the model has to be adjusted according to the empirical data obtained in the particular locality and for the specific moisture conditions.

Field and Environmental Variables

All tests were conducted in a controlled laboratory environment at room temperature (20°C - 24°C / 68°F - 75°F). The current setup does not account for real-world field conditions:

- Depth of Sensor and Proximity to Roots: The sensor's electrodes can only detect the soil around them in a limited area (~2-3 cm). Our study did not investigate how the levels of nutrients change at variable depths of the soil or near the roots of the plants where the uptake of nutrients occurs.
- Effect of Soil Compaction and Disturbance: The force exerted on the probe affects the soil density, formation of air pockets, and pressure applied by the probe. Although we permitted the sensor to stabilize for 60 seconds before taking a measurement, we did not determine the effect of compaction and softness of the soil on the accuracy of the readings.
- Effect of Temperature and Time: The temperature of soil changes in response to various factors on a daily basis. The soil temperature fluctuates during the day, season, and throughout the life cycle of microorganisms and roots in the soil. The durability of the sensor operating under changing weather conditions requires further study.
- Single Probe Testing: Evaluation was conducted using a single 7-in-1 probe unit. While the algorithmic framework is generalizable, unit-to-unit manufacturing variations among commercial low-cost probes may require individual zero-point calibration.

Prototype Hardware Enclosure

The testing apparatus we developed consists of a benchtop laboratory prototype designed from a breadboard and an external plug-in power supply connected to an Arduino UNO interfaced with a laptop device. The technique is effective and affordable, costing \$142 in total and is used for data collection. However, it does not yet meet the requirements for use in rugged field environments. Lack of

waterproofing, use of batteries, and long-distance data transmission make the unit serve as a proof-of-concept testing pipeline rather than a commercial farm tool.

FUTURE DIRECTIONS

To advance this prototype from a laboratory proof-of-concept to a field-ready agricultural tool, future development will focus on four key areas

1. **Validation of Lab Analysis:** Confirm the accuracy of sensor function versus laboratory wet chemistry analyses (ICP-OES in relationship to P_2O_5 and K_2O ; Kjeldahl/cadmium reduction in relation to N).
2. **Machine Learning Models:** Test advanced approaches in non-linear models like Random Forest and XGBoost, which can use the information from soil moisture content (SMC), electrical conductivities (EC), pH levels, and temperature to account for mineral salt constituents ($Na+Cl^-$, Ca^{2+}).
3. **Soil Sampling Protocol Testing:** Perform a field experiment to test the performance of the sensor in soil at different depth levels (0–30 cm) and soil compaction rates.
4. **Ruggedized IoT Field Deployment:** Transition from benchtop logging to an autonomous, solar-powered field node featuring an IP67 waterproof enclosure and LoRaWAN/Cellular telemetry for real-time cloud monitoring.

CONCLUSION

Given the need to feed 9.8 billion people with limited rural labor and depleting resources, providing universal access to timely soil data will be indispensable for future sustainable agriculture. In contrast to traditional laboratory soil extractions (e.g., Kjeldahl or ICP-OES techniques) with high precision, expensive equipment, and multi-day delays, this paper suggests an open-source microcontroller setup and calibration framework aimed at mitigating a serious problem of reading drift due to moisture content in commercially available soil NPK sensors.

Interfacing a 7-in-1 RS485/Modbus soil probe with Arduino UNO data logging device, the presented study analyzed sensor readings on three standard commercial potting substrates at various hydration levels. Using known nominal values ($NPK_{nominal}$) guaranteed by manufacturers as a reference baseline, it was found that unprocessed sensor output signals exhibit very high cross-sensitivity to SMC (soil moisture content). Employing a regression model, which includes moisture as an additional feature, drastically decreases this variance. On an independent holdout validation dataset, the calibrated model decreased prediction error and reached MAPE (Mean Absolute Percentage Errors) of 20%–25% for nitrogen, phosphate, and potash.

Unprocessed commercial soil sensors of this specific model, while being quite inexpensive, are useless without calibration because of very high variability and large errors. The suggested approach, however, proved to be an efficient and practical way to smooth out environmental factors. Even though 20%–25% prediction errors may be too large for accurate scientific analysis, the obtained results make sense for immediate and practical application in the fields by small farmers who need quick feedback about NPK ratios in order to optimize fertilizer input and avoid groundwater pollution.

There are four main directions for development of the laboratory prototype into a usable technology in the future: (1) comparison of sensor readings against certified chemical lab analysis of NPK ratios on various farmland soils; (2) training machine learning models with a lightweight model size to process native soils and exclude salt contamination; (3) conducting outdoor field experiments to test depth, root influence, and soil compaction; and (4) building an autonomous station with a waterproof housing, solar energy source, and long-range wireless communication module (LoRaWAN, cellular).

In conclusion, using low-cost sensors and algorithmic matrix correction proved to be a good basis for the development of digital soil testing tools.

REFERENCES

1. World Resources Institute. (n.d.). *Creating a Sustainable Food Future*. Retrieved January 10, 2026, from https://agritrop.cirad.fr/593176/1/WRR_Food_Full_Report_0.pdf
2. AlZubi, A. A., & Galyna, K. (2023). Artificial Intelligence and Internet of Things for Sustainable Farming and Smart Agriculture. *IEEE Access*, 11, 78686–78692. <https://doi.org/10.1109/ACCESS.2023.3298215>
3. Hultgren, A., Carleton, T., Delgado, M., Gergel, D. R., Greenstone, M., Houser, T., Hsiang, S., Jina, A., Kopp, R. E., Malevich, S. B., McCusker, K. E., Mayer, T., Nath, I., Rising, J., Rode, A., & Yuan, J. (2025). Impacts of climate change on global agriculture accounting for adaptation. *Nature*, 642(8068), 644–652. <https://doi.org/10.1038/s41586-025-09085-w>
4. Idoje, G., Dagiuklas, T., & Iqbal, M. (2021). Survey for smart farming technologies: Challenges and issues. *Computers & Electrical Engineering*, 92, 107104. <https://doi.org/10.1016/j.compeleceng.2021.107104>
5. Liu, J., Cai, H., Chen, S., Pi, J., & Zhao, L. (2023). A Review on Soil Nitrogen Sensing Technologies: Challenges, Progress and Perspectives. *Agriculture*, 13(4), 743. <https://doi.org/10.3390/agriculture13040743>
6. Eldeeb, M. A., Dhamu, V. N., Paul, A., Alam, F. M., Burgos, E. N., Muthukumar, S., & Prasad, S. (2024). ASSERT: A Platform Technology for Rapid Electrochemical Sensing of Soil Ammonium. *ACS Omega*, 9(31), 33928–33934. <https://doi.org/10.1021/acsomega.4c04181>
7. Mashao, F. M., Thaba, S. J., Muyambo, N. P., Tjale, C. R., Zwane, P. S. M., Munjonji, L., Nkuna, D., Ayisi, K. K., & Thamaga, K. H. (2024). Exploring laboratory-based spectroscopy for estimating

NPK content in the hutton soils of Syferkuil Farmlands, South Africa. *Geocarto International*, 39(1), 2339289. <https://doi.org/10.1080/10106049.2024.2339289>

8. Hossain, M. I., Khaleque, M. A., Ali, M. R., Bacchu, M. S., Hossain, M. S., Shahed, S. M. F., Saad Aly, M. A., & Khan, Md. Z. H. (2024). Development of electrochemical sensors for quick detection of environmental (soil and water) NPK ions. *RSC Advances*, 14(13), 9137–9158. <https://doi.org/10.1039/D4RA00034J>

9. Adebunmi Okechukwu Adewusi, Onyeka Franca Asuzu, Temidayo Olorunsogo, Temidayo Olorunsogo, Ejuma Adaga, & Donald Obinna Daraojimba. (2024). AI in precision agriculture: A review of technologies for sustainable farming practices. *World Journal of Advanced Research and Reviews*, (1), 2276–2285. <https://doi.org/10.30574/wjarr.2024.21.1.0314>