nature food

ARTICLES

https://doi.org/10.1038/s43016-021-00416-4

Point-of-use sensors and machine learning enable low-cost determination of soil nitrogen

Max Grell, Giandrin Barandun, Tarek Asfour, Michael Kasimatis, Alex Silva Pinto Collins, Jieni Wang and Firat Güder ✉

Overfertilization with nitrogen fertilizers has damaged the environment and health of soil, but standard laboratory testing of soil to determine the levels of nitrogen (mainly NH_4^+ and NO_3^-) is not performed regularly. Here we demonstrate that point-of-use measurements of NH_4^+, combined with soil conductivity, pH, easily accessible weather and timing data, allow instantaneous prediction of levels of NO_3^- in soil ($R^2 = 0.70$) using a machine learning model. A long short-term memory recurrent neural network model can also be used to predict levels of NH_4^+ and NO_3^- up to 12 days into the future from a single measurement at day one, with $R^2_{NH_4^+} = 0.60$ and $R^2_{NO_3^-} = 0.70$, for unseen weather conditions. Our machine-learning-based approach eliminates the need for dedicated instruments to determine the levels of NO_3^- in soil. Nitrogenous soil nutrients can be determined and predicted with enough accuracy to forecast the impact of climate on fertilization planning and to tune timing for crop requirements, reducing overfertilization while improving crop yields.

There is a global effort to find practices for food production that can sustainably feed the population, which is expected to surpass ten billion people by 2050[1]. The Haber–Bosch process enabled inexpensive nitrogen-based fertilizers to feed the booming population, with a >600% increase in their use in the past 50 years[2,3]. Increased fertilization has, however, come with a great environmental cost. Approximately 12% of the available arable land is now degraded, of which >240 Mha (~926,000 mi^2 or four times the area of France or Texas) is chemically degraded—that is, contaminated with heavy metals and/or acidified, especially from nitrogen fertilizers, which interfere with nutrient mobility and uptake by plants[4–6]. Overfertilization has visibly destroyed ecosystems by the leaching of excess NO_3^- into surface waters, which causes eutrophication, giving rise to dead zones such as in the Gulf of Mexico[7]. Overfertilization also impacts the soil microbiome[8,9]. Furthermore, N fertilization seems to shift the relative abundance of certain microbial communities in soil, with important implications for C cycling and ecosystems[10].

The application of fertilizers is poorly understood and largely varies between regions and countries; for example, eight times more is applied per hectare in China than in Australia[11]. Farmers across the globe typically rely on guidelines from their governments, fertilizer suppliers or family know-how when deciding the economic optimum rate of fertilization to ensure maximum profit. Professional agronomists generally advise along guidelines and look at yields from previous years to estimate fertilizer requirements; they may also take soil samples for laboratory testing prior to sowing. Laboratory testing, however, is an expensive and slow process and hence is not performed regularly. Soil nitrogen (soil-N) is crucial for high yields, and nitrogen fertilizer is the most frequently applied fertilizer. The optimal application rate is highly variable, however, since soil-N fluctuates widely with the properties of soil and weather over short timescales. Benchmark guidelines are unable to account for these variations. With the lack of data concerning the current and future nitrogen levels in soil, farmers tend towards overfertilization to protect yields, an environmentally and economically inefficient practice[12–14].

The measurement of soil-N is important for optimizing the use of nitrogen fertilizers and enabling spatio-temporal variable-rate fertilization. Indirect spectroscopic precision farming technologies such as crop canopy sensors (for example, near-infrared spectroscopic cameras) can be used to approximate the N requirements of plants[15–17]. Commercially available options include the CropSpec from Topcon Totalcare[18], SunScan from Delta-T Devices[19] and N-Sensor from Yara[20], which can be mounted on a tractor for large-area monitoring. Indirect spectroscopic techniques, however, do not measure the levels of nitrogen in soil; instead, they measure green light from the leaves of plants (related to nitrogenous compounds) to indirectly estimate the levels of N fertilizer required. Machine learning (ML) algorithms are suitable for calibrating spectra (for example, near-infrared) to soil-N[21]. Spectroscopic methods require plant mass (for example, leaves), so the measurements cannot be performed until after germination and growth. Fertilizer, however, is usually applied just before seeds are sown; hence, spectroscopic techniques rarely help in-season and only complement national guidelines. Using ion-selective membranes, the levels of nitrogen in soil (mainly in the form of NO_3^- and NH_4^+) can be directly detected electrochemically[22]. Such sensors can be integrated into Internet-of-Things-type remote sensors that can provide continuous data streams concerning levels of nitrogen in soil (for example, Teralytic[23]). To provide spatio-temporal resolution, however, many units would need to be deployed to fields[24]. Statistical models using ML are therefore well suited to filling in missing soil data[25] and forecasting them into the future[26]. Given that each costly sensor node is not disposable, they would require collection before harvest (that is, they are labour intensive) and would be susceptible to theft. They also require infrastructure investments to create a wireless network with access points. With challenges such as large investment requirements, sector heterogeneity, data ownership and privacy, user acceptance, and lack of interoperability, the adoption of Internet-of-Things systems for soil sensing has been slow[27–29]. Ion-selective electrochemical sensors can also be produced in a small point-of-use (PoU) formfactor (for example, Horiba LAQUAtwin and ELIT 8021). These commercially available sensors demonstrate

Department of Bioengineering, Imperial College London, London, UK. ✉e-mail: guder@ic.ac.uk

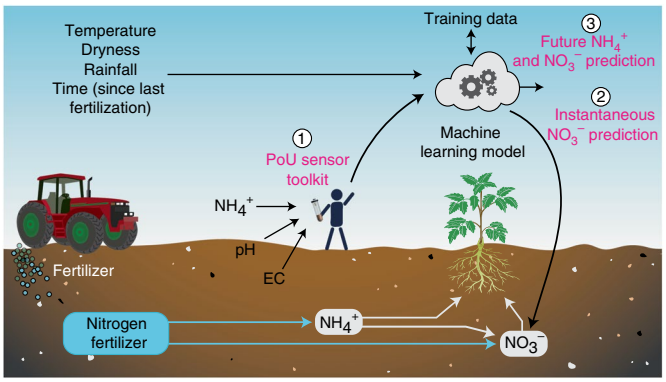

Fig. 1 | Summary of the method. Nitrogen fertilizer is the backbone of modern agriculture and is typically applied as urea or ammonium nitrate (we used ammonium nitrate in this work), fuelling the soil-N cycle. These fertilizers produce NH_4^+ and NO_3^- in soil, available to be taken up by plants. We have combined PoU electrical measurements (including a new soil NH_4^+ sensor) with an ML model to quantify difficult-to-measure soil nutrients (such as NO_3^-) and forecast them into the future. This provides information concerning the dynamics of soil-N instantaneously and into the future to guide fertilization (reducing overfertilization, understanding the effect of weather and ensuring enough plant-available nitrogen to maximize crop yield) without laboratory measurements. At scale, this model could use a bare minimum of readily available input data to quantify and predict crucial outputs (for example, soil macronutrients) in highly complex systems (such as soil).

high accuracy for NO_3^- ($R^2=0.96$)[30] and NH_4^+ ($R^2=0.98$)[31]; however, they are delicate and relatively expensive (Horiba LAQUAtwin NO_3^- sells for ~US$350; each electrode is ~US$150), and they require sample preparation and on-site calibration[32]. Optical methods such as near-infrared spectrometry are nearing commercialization for soil-N at the PoU ($R^2=0.72$ to 0.92)[33]; however, they also require on-site calibration and struggle with low concentrations[34]. Enzymatic biosensors are being researched, with reported $R^2=0.99$, although they are delicate and have short lifetimes[35].

In this work, we demonstrate a new approach for determining crucial but difficult-to-measure N levels in soil. We combine a new type of gas-phase NH_4^+ sensor (to eliminate matrix effects due to the complex sample)[36,37], simulated climate data (that is, rainfall and temperature), time passed since fertilization (that is, number of days) and off-the-shelf soil pH and conductivity sensors with a statistical ML model to instantaneously and accurately determine levels of NO_3^- in soil. We demonstrate that the N levels in soil can also be predicted into the future using a long short-term memory (LSTM) recurrent neural network over a 12-day period. With this new approach (Fig. 1), fertilization can be provided more precisely to improve yields while preventing overfertilization and environmental degradation.

Results and discussion

Disposable PoU NH_4^+ sensor. To measure levels of NH_4^+ in soil, we developed an electrical PoU sensor to accurately determine soil NH_4^+ ($R^2=0.85$; limit of detection, 3 ± 1 ppm; tested up to 144 ppm) at low cost with a large dynamic range (Fig. 2a). Each sensing module (that is, cartridge) consisted of only a container, a disposable, chemically functionalized paper-based electrical gas sensor (chemPEGS, Fig. 2b) and 1 ml of 15 M NaOH, costing <US$0.10 (Supplementary Video 1)[38–41]. The chemical functionalization (10 µl of 0.025 M H_2SO_4) of the chemPEGS was the best compromise between precision and measurement time (Supplementary Fig. 1) and builds on the paper-based electrical gas sensors described by Barandun et al.[37]; the addition of H_2SO_4 makes the sensor more specific for the sensing of ammonia in comparison with our previous work. Sensing of NH_3 with the chemPEGS is susceptible to interferences from other water-soluble alkaline gases; however, because NH_3 has the highest water solubility and is the dominant water-soluble alkaline gas species in our samples due to the fertilizer (ammonium nitrate), the signal generated by the chemPEGS largely originates from NH_3. There is a decrease in ionic impedance during neutralization on the chemPEGS, which is measured electrically using our homemade electronics[37]. An alternating voltage (10 Hz, 4 V amplitude) was supplied across the chemPEGS, and the current passing through was measured as a voltage with a transimpedance amplifier, amplified with a gain resistor (Supplementary Information and Supplementary Fig. 2). As the H_2SO_4 neutralization continued, the impedance of paper increased slowly, and the time it took to complete (gradient reaching zero) or slow dramatically (step change in gradient, >80% in <10 minutes) was used as the analytical signal shown in Fig. 2c (see Supplementary Fig. 3 for the raw data). We evaluated the stability of the chemPEGSs and determined that they can be stored in dry conditions for 16 days without negatively impacting performance (Supplementary Fig. 4). Before measuring unknown concentrations, we calibrated the sensor in a range of concentrations of NH_4^+ from 4.5 to 144 ppm, in soil fertilized with NH_4NO_3; the calibration curve (log-log) is shown in Fig. 2d. Our measurements were compared to external laboratory measurements with a score of $R^2=0.85$ (Supplementary Fig. 5; see Supplementary Fig. 6 for the Bland–Altman plot). This is below the reported scores of ion-selective electrochemical sensors mentioned above, but our sensing mechanism needs only simple and robust components for matrix-free sensing and is therefore likely to offer more dependable results in-field. We also measured a calibration curve for NH_4NO_3 in water (no soil) and verified that our soil measurements are indeed from NH_4^+ alone (Supplementary Fig. 1).

Time-dependent nitrogen dynamics in soil. Time-series data concerning the dynamics of soil-N were collected over short timescales (<20 days) in experiments simulating soil in a field (Fig. 3). To reduce complexity, we did not grow any plants, and we investigated the nitrogen dynamics due only to microbial activity, run-off and volatilization (escape of NH_3(g)). We placed 5.1 kg of soil (Westland Top Soil, unfertilized) in 15 l plastic pots and stored them in the laboratory without covering their tops. In each experiment, we controlled the environmental conditions in two ways: (1) adding a controlled amount of water to simulate rainfall and (2) passing an electrical current through a resistive heating wire (nichrome wire), wrapped around the containers, to control temperature uniformly. We kept the soil type (sandy loam, sieved) and amount of fertilizer added fixed for all experiments (the fertilizer NH_4NO_3 was added in the beginning of each experiment to produce a concentration of 120 ppm, approximately equal to 241 kg ha^{-1} NH_4NO_3 or 85 kg ha^{-1} nitrogen; the calculation is shown in the Supplementary Information). Experiments were performed for eight sets of environmental conditions spanning arid (1 mm d^{-1} rainfall) to tropical (10 mm d^{-1} rainfall) with temperatures ranging from 19–21 °C (temperate) to 26–34 °C (warm). Measurements of soil temperature, rainfall, pH, electrical conductivity (EC) and NH_4^+ were made in our laboratory (Güder Research Group). Levels of pH, EC and NH_4^+ were also measured in an external laboratory (NRM, Cawood Scientific), as were dryness and NO_3^-, for comparison and training of the ML model. When building the ML model, we relied on the dryness values provided by the commercial laboratory, but dryness is highly correlated with rainfall and temperature (in our dataset, temperature and rainfall predict dryness using linear regression with $R^2=0.86$; see Supplementary Fig. 7 and the equation in the Supplementary Information) and hence can be estimated using these two metrics without needing further analytical measurements.

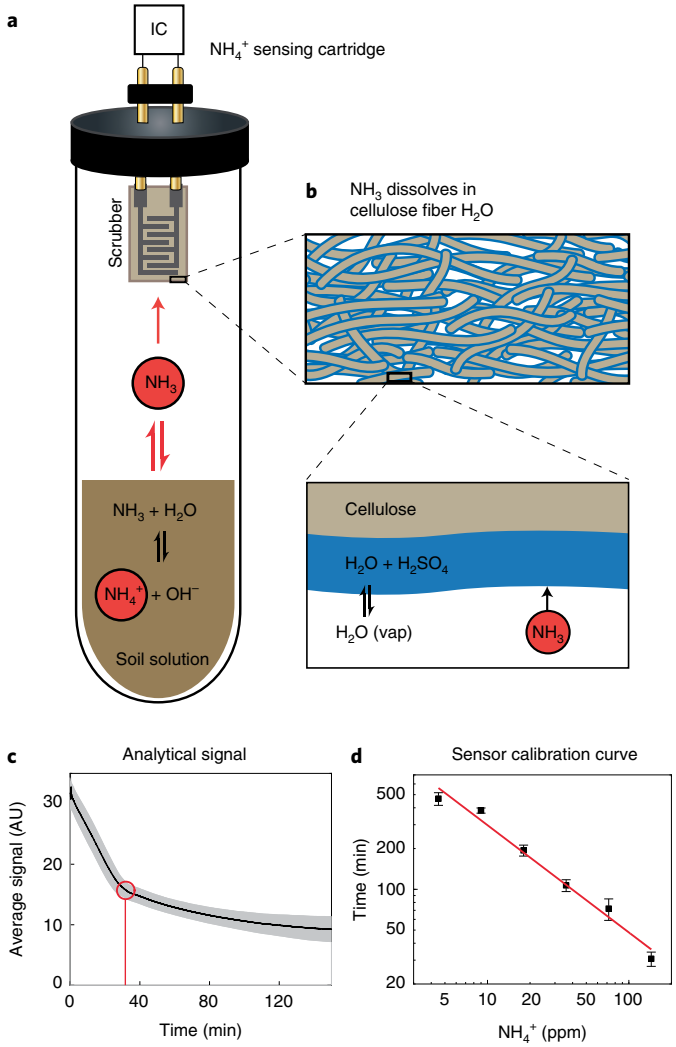

Fig. 2 | PoU soil ammonium sensor. a, A gas-phase NH_4^+ sensor cartridge, consisting of a container, 1 ml of 15 M NaOH and a disposable chemPEGS that acts as a scrubber of soil NH_4^+, is connected to an integrated circuit (IC) to perform impedance analysis. **b**, The volatilized NH_3(g) dissolves in the layer of water adsorbed on the chemPEGS, neutralizing the H_2SO_4 and increasing ionic impedance, which was measured electrically. The neutralization reaction draws out the remaining NH_4^+ from the soil solution to maintain the equilibrium of NH_3 in the headspace. The time it took for neutralization to slow dramatically or to complete was used as the analytical signal (see Supplementary Fig. 3 for the raw data and mathematical criteria). **c**, An example signal from 144-ppm soil NH_4^+, with the analytical signal circled in red and the error shown in grey (calculated by the standard deviation of $n = 5$ measurements). **d**, We calibrated the sensor in a range of concentrations of NH_4^+ from 4.5 to 144 ppm in soil fertilized with NH_4NO_3. Each point had $n = 5$ repeats, measured simultaneously with chemPEGSs from the same batch of fabrication, with error bars corresponding to the standard deviations.

Dynamics of soil NH_4^+. In all time-dependent soil experiments, the level of NH_4^+ dropped rapidly over time, levelling out after about a week, independent of the environmental conditions. Temperature played a considerable role only in the case of 1 mm d^{-1} rainfall, in which the NH_4^+ levels settled at ~50 ppm in warm conditions, compared with ~0 ppm in temperate conditions. In all other scenarios, temperature and rainfall only slightly affected the NH_4^+ dynamics, without large differences in the trends. Decreasing levels of NH_4^+ result from multiple processes, such as nitrification (that is, the conversion of NH_4^+ to NO_2^- to NO_3^-) or environmental losses (leaching or volatilization), that run in parallel; however, the extent of each process might vary with environmental and soil conditions. Soil dehydration tends to limit nitrification by restricting substrate supply to microorganisms and lowering the activity of enzymes[42], which may explain the retention of NH_4^+ at higher temperatures (and low rainfall). This observation is further supported by the fact that the levels of NO_3^- were lower in warm conditions than in temperate conditions.

Dynamics of soil NO_3^-. Nitrification is a complex, aerobic microbial process affected by temperature, moisture, levels of O_2, pH and the availability of NH_4^+, among other things (for example, nitrifier populations)[6]. We observed that at 1 mm d^{-1} rainfall, the level of NO_3^- increased compared with the initial (day zero) concentration, but at 3 mm d^{-1}, it remained relatively unchanged in both warm and temperate conditions. At 5 mm d^{-1} rainfall in warm conditions, the levels of NO_3^- only slightly increased towards the end of the experiment. In temperate conditions, the concentration of NO_3^- nearly halved, with a rapid drop after day 10. Under heavy rainfall (10 mm d^{-1}), the concentrations of NO_3^- dropped towards zero in a linear manner over the course of the experiments. From these experiments, it could be concluded that the optimum point for maximum nitrification and retention of NO_3^- in soil occurs in temperate and drier conditions, which are consistently more favourable than warm and wetter conditions. The reasons behind these trends may differ, however, depending on the conditions. While the run-off caused by heavy rainfall (that is, 10 mm d^{-1}) may physically leach NO_3^- away (the excess water was pouring out from the bottoms of the pots), less rainfall (3 or 5 mm d^{-1}) may hinder the penetration of O_2 into the soil (that is, waterlogged soil) and thereby reduce nitrification, especially if the climate is temperate so that not enough water is removed from the soil to allow oxygenation[43]. The optimal temperatures for nitrification are typically reported between 24 and 27 °C (ref. [44]), in line with our observations. In the experiments where the dryness of soil did not increase, temperature did not have a large effect, evidenced by the first four days of the experiments with 3 and 5 mm d^{-1} rainfall. Dryness (that is, rainfall plus temperature) therefore seems to be a more important factor in determining the levels of NO_3^- than temperature alone.

Dynamics of soil EC and pH. EC and pH were measured to investigate their correlation with soil-N under different environmental conditions. Due to technical difficulties, we were unable to complete the EC and pH measurements for all samples in a single day and therefore missed measurements that were to be performed in our laboratories. Nevertheless, we did not observe any major trends in pH or EC regardless of rainfall or temperature except for the experiments with 1 and 10 mm d^{-1} rainfall. At 1 mm d^{-1} rainfall, the EC only slightly increased, and pH slightly decreased over time. Ammonium-based fertilizers are known to acidify soil and therefore decrease pH[45,46]. With an increase in the concentration of mobile NO_3^- ions in soil, EC is also known to increase[46]. When the rainfall was increased to 10 mm d^{-1}, however, the run-off leached out ionic species from the soil, in turn reducing the EC of soil without affecting pH. The EC and pH measurements performed in our laboratory and externally did not correlate to the degree we expected (see the discussion below Supplementary Fig. 7 in the Supplementary Information).

Predicting levels of nitrogen in soil using ML. Retention, conversion or loss of nutrients added to soil is a complex function of rainfall, temperature, pH, microorganism populations, soil type and other factors. This complexity makes it difficult (if not impossible) to create deterministic models to understand the relationship between

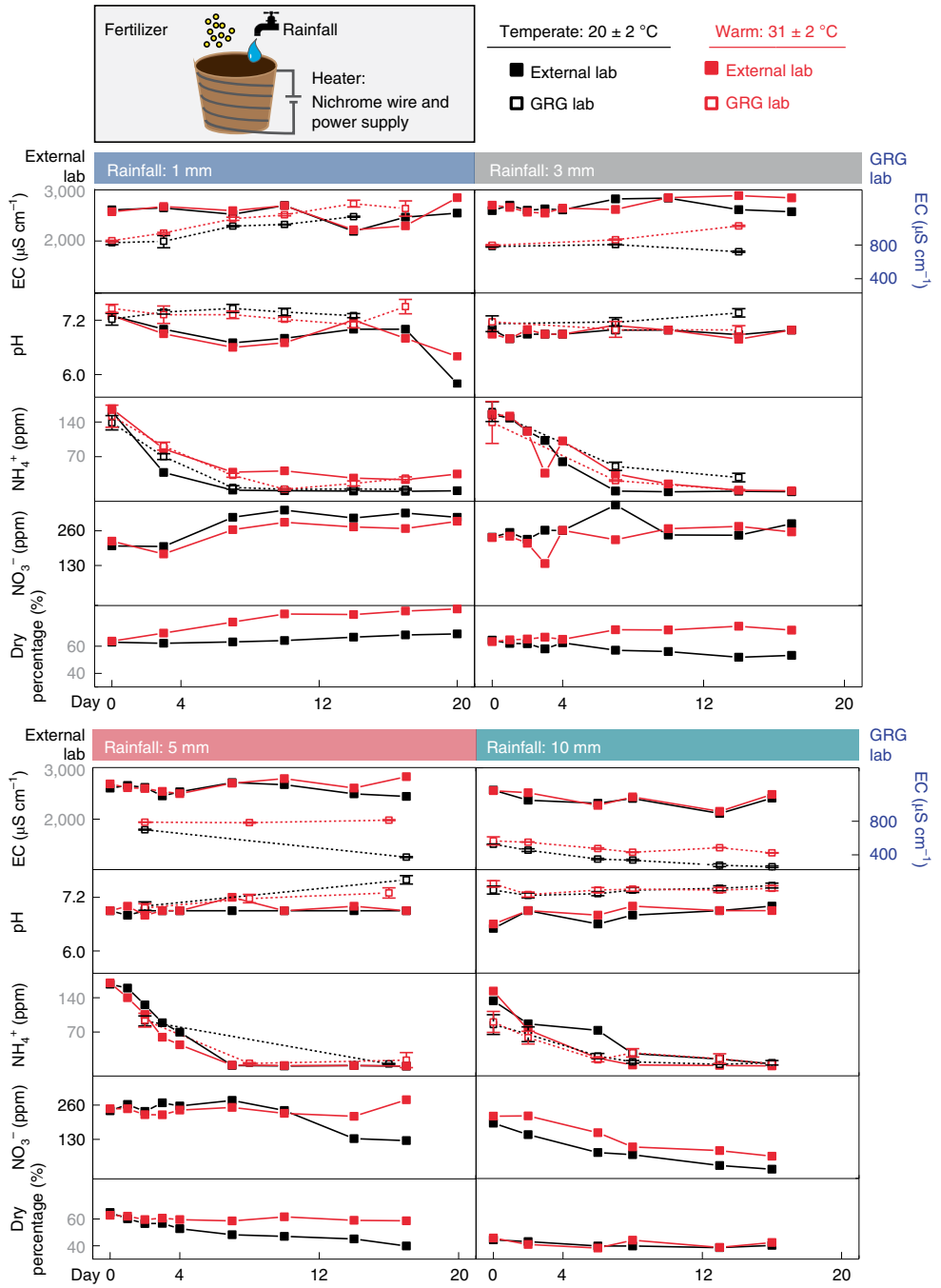

Fig. 3 | Soil chemistry and nitrogen dynamics. Time-series data of soil-N dynamics were measured over short timescales (<20 days), where each time series corresponds to soil under different environmental conditions. We controlled rainfall and temperature by adding a controlled amount of water and passing current through a resistive heating wire (top). The environmental conditions span arid (1 mm rain per day) to tropical (10 mm rain per day) and temperate (20 ± 2 °C) to warm (31 ± 2 °C). Initial fertilization with NH_4NO_3 was fixed at 120 ppm. Measurements of soil temperature ($n = 3$), rainfall, pH ($n = 5$), EC ($n = 5$) and NH_4^+ ($n = 5$) were made in our laboratory (GRG, Güder Research Group), with errors corresponding to the standard deviation. Measurements of pH, EC, dry percentage, NH_4^+ and NO_3^- were measured in an external laboratory ($n = 1$) for comparison and training of the ML model.

nitrogenous species and their levels in soil after some time, even if the initial concentrations are known. We have therefore attempted to create a statistical model using (existing) ML approaches to predict levels of hard-to-measure NO_3^- in soil using information on weather (that is, rainfall and temperature), time since fertilization, pH, EC and NH_4^+ (Supplementary Video 2).

Using supervised ML, we attempted to predict the levels of NO_3^- in soil instantaneously and those of both NH_4^+ and NO_3^- into the future (see Supplementary Fig. 8 for the ML prediction process flow). We used the measurements performed by the external laboratory (Fig. 3) as a training set (the data processing for ML is described in the Supplementary Information). The performance of the model was then tested either with data from the external lab or with data generated by the PoU sensors in our lab as inputs. Training data matching the same environmental conditions (temperature and rainfall) as the test inputs were removed, so the model was always tested on unseen environmental conditions. Features were ranked in order of importance (by XGBoost; Fig. 4a, top left), where soil

dryness, time since fertilization and NH_4^+ were the most important. We compared combinations of features, regressors and tuning parameters exhaustively (by grid search) to find the best general regressor to estimate the levels of NO_3^- instantaneously (see Supplementary Fig. 9 for R^2 scores for each set of environmental conditions). We determined that the *K*-nearest-neighbours (Knn) algorithm, trained on all seven features and tuned with number of neighbours (k) = 14, leaf size = 1 and power parameter (p) = 1, can predict instantaneous levels of NO_3^- with $R^2 = 0.63$ using the external lab results for training and our lab PoU sensors for test input (Fig. 4a, top right; Bland–Altman plots for Fig. 4 are shown in Supplementary Fig. 12). Using the same model, but with the external lab results as test inputs, removes the impact of inaccuracy of our lab PoU sensors, resulting in $R^2 = 0.68$ (Supplementary Fig. 10). We also determined that the XGBoost regressor produced the best predictions for dryer soils and Knn for wetter. Taking our lab PoU sensors as test inputs, and using the optimal regressor and tuning for each set of environmental conditions, offers even better performance, giving an optimized score of $R^2 = 0.70$ (score averaged across each set of environmental conditions; Fig. 4a, bottom left). The same process, but with the external lab results as test inputs, produces an optimized best-case scenario score of $R^2 = 0.86$ (Fig. 4a, bottom right). This score for predicting levels of NO_3^- in soil is comparable to direct measurements using optical ($R^2 = 0.83$; Fourier-transform infrared spectroscopy) or electrochemical methods ($R^2 = 0.96$; ion-selective electrodes) as reported in the literature[31,47]. This result was pleasantly surprising given that no additional hardware was required for determining levels of NO_3^- with high accuracy. We also tuned a Knn model ($k = 11$, leaf size = 3, $p = 1$) to predict levels of NO_3^- in soil using only the most basic inputs—days since fertilization, rainfall and temperature (that is, requiring no soil sensors at all)—which yielded $R^2 = 0.54$ (Supplementary Fig. 11). We attempted to validate the same model for NO_3^- prediction on a different soil type (55% clay, 1% sand, 44% silt) with a new time series of external laboratory data. The result was a positive correlation ($R^2 = 0.31$) between measured and predicted NO_3^-; however, more data are needed to train the model to understand nitrogen dynamics in different soil types (see Supplementary Fig. 13 for the data and further discussion).

Determining the concentration of NH_4^+ and NO_3^- in soil at any given moment is important (as described above), but from an operational point of view, it would also be useful to know what the levels of soil-N (that is, NH_4^+ and NO_3^-) will be in the future from a single measurement to create a precise schedule for future fertilization. Soil, however, introduces a memory effect: nutrient levels today depend on the nutrient levels and other factors from yesterday (property X at time t will be a function of X at $t-1$). Forecasting of soil-N into the future must therefore consider time and sequence of data, and possess a degree of memory, for multiple correlated features. Using the time-series dataset generated by the external lab, we trained an LSTM recurrent neural network model (another supervised ML algorithm) to forecast NH_4^+ and NO_3^- into the future for unseen environmental conditions. We tuned the model using grid search, minimizing root mean squared error using time lag and model hyperparameters (training epochs, batch size and number of neurons, defined after Supplementary Fig. 15). The optimal tuning was time lag, 1; epochs, 50; batch size, 3; and number of neurons, 3. The dataset was first concatenated into one multivariate time series. Each time series was then removed sequentially, and the model was trained to predict the removed time series from the remaining data. The models were retrained for each desired forecast time (1–12 days into the future; longer time periods were inaccurate), but the same tuning parameters were used for all environmental conditions. Comparing predicted to real values over the 12-day period gives scores of $R^2_{NH_4+} = 0.60$ and $R^2_{NO_3-} = 0.70$ using only the initial concentrations of NH_4^+ and NO_3^- on day zero, which demonstrates efficacy even with our limited training dataset (Fig. 4b, with R^2 plots in Supplementary Fig. 15 and Bland–Altman plots in Supplementary Fig. 16). In essence, by measuring NH_4^+, EC and pH in the field and gathering other environmental data from public sources, one can estimate the levels of NO_3^- for today and levels of both NH_4^+ and NO_3^- into the future.

Conclusions

In this study, we demonstrated that it is possible to estimate the levels of hard-to-measure nitrogen levels in soil using easily accessible soil and climate data and ML models. This strategy allows one to determine and predict levels of nitrogen (NH_4^+ and NO_3^-) in soil, both instantaneously and into the future. We have produced a soil nitrification dataset that provides enough temporal resolution (about a three-day measurement frequency), for a range of conditions, to train an ML model. The strength of our approach is that we primarily use inexpensive and easily accessible tools for the soil measurements (pH and EC meter, with the exception of the chemPEGS developed in this work) and publicly available weather data (rainfall and temperature; in this study, we simulated weather in a controlled manner) to estimate the levels of soil-N through ML. The method presented is high performance, such that the concentration of instantaneous soil NO_3^- can be estimated using PoU inputs with $R^2 = 0.70$ and using external laboratory inputs with $R^2 = 0.86$ (comparable to existing high-performance NO_3^- sensors) without the need for additional hardware. Using an LSTM model, the levels of NH_4^+ and NO_3^- can also be forecasted 12 days into the future, for unseen environmental conditions, with $R^2_{NH_4+} = 0.60$ and $R^2_{NO_3-} = 0.70$. Furthermore, the paper-based, disposable, gas-phase NH_4^+ sensors (that is, chemPEGS) developed in this work could also be used alone at the PoU without the ML model or other sensors if instantaneous detection of NH_4^+ is needed alone.

The approach presented in this work has the following three weaknesses. First, the supervised ML algorithms used for the prediction of soil-N require a training dataset, meaning that prior measurements and climate data are needed to make the estimation algorithms work. This problem could be partially resolved by using data for soil-N already published in the literature to create a training dataset. A training dataset could also be created using the PoU sensor toolkit described in this work in addition to occasional measurements of soil NO_3^- in an external laboratory. We expect that the performance of the algorithms will increase over time as more data are generated using the sensors and laboratory measurements. Although leave-one-out cross validation was used for training, the lack of a validation dataset may have resulted in overfitting of the hyperparameters to the weather conditions in Fig. 4a (bottom left and bottom right).

Second, the chemPEGS (for measuring NH_4^+ at the PoU) is expected to be cross-sensitive to other alkaline gases and currently takes a long time to perform a measurement (30–450 min for 144–4.5 ppm NH_4^+). The chemPEGS, however, demonstrated sufficient performance for measuring soil NH_4^+, as it is most sensitive to NH_3(g) due to its high water solubility. The time it takes to produce a result could also be reduced by measuring the rate of change of the impedance during neutralization or training a predictive ML model on short measurement times. The sensitivity of the chemPEGS could also be improved by using a lower concentration of H_2SO_4 (refs. [37,38,48]).

Third, the dataset generated in this work is limited (sparse) and does not include various scenarios such as sudden changes in weather and different types of fertilizers (for example, urea). The current work also does not include crops, which would draw nitrogen from the soil and affect nitrogen dynamics. The limited dataset may explain the unusual LSTM tuning for batch and number of neurons (both 3) and the poor performance in clay soil. Further work is needed to create a model to predict nitrogen uptake by plants.

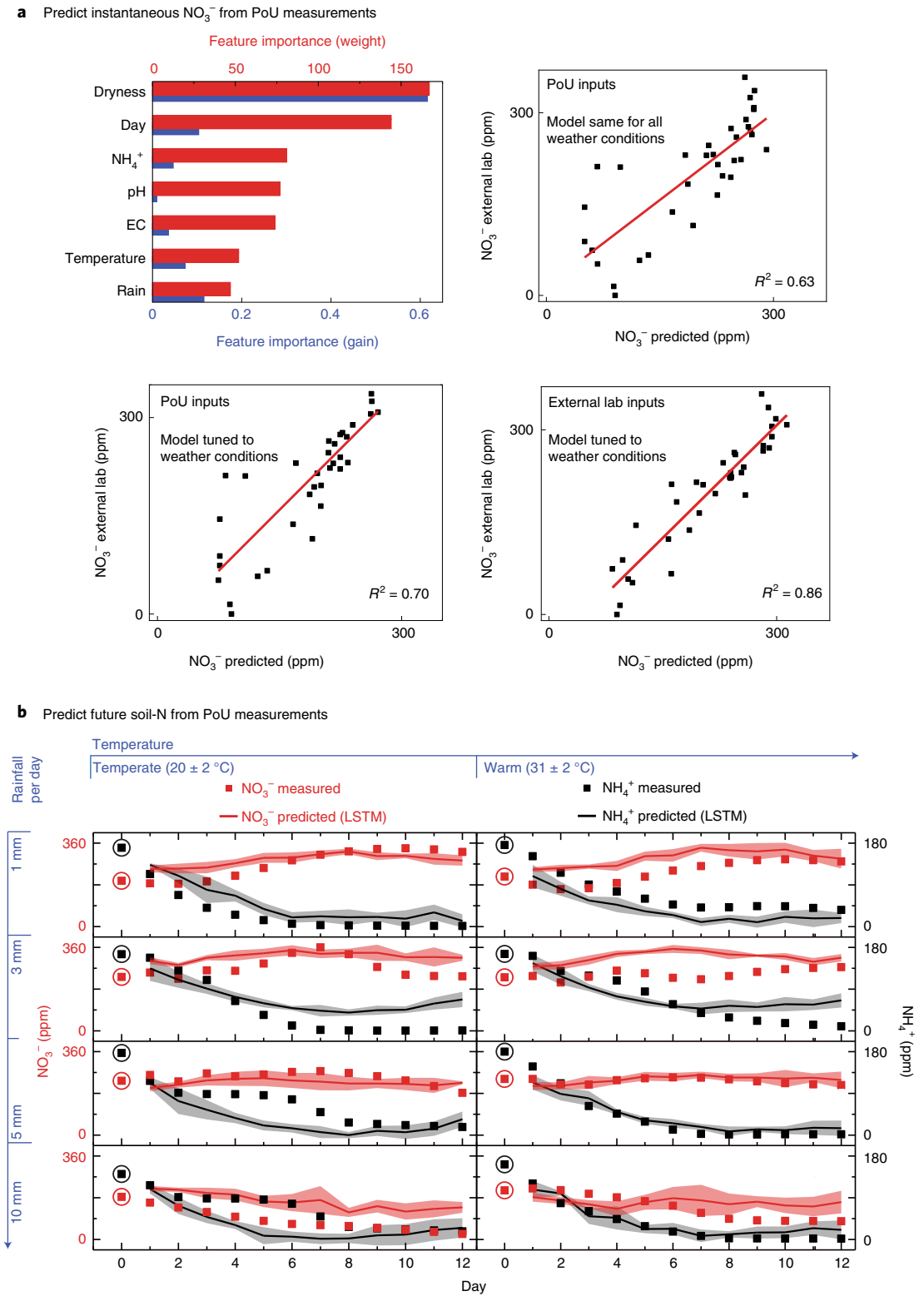

Fig. 4 | ML model to predict soil nitrate. a, Instantaneous soil NO_3^- was predicted with instance-based and ensemble learning regressors. Features were first ranked by importance to the XGBoost model, calculated by weight (the number of times the feature occurs in the trees) and gain (each feature's contribution to each tree) (top left). The best-performing regressor for all environmental conditions is Knn, which predicts NO_3^- using only PoU sensors from our lab as test inputs with $R^2 = 0.63$ (top right). Optimizing the model for each set of environmental conditions (the regressor and tuning for each are shown in Supplementary Fig. 9) improves the score to $R^2 = 0.70$ (bottom left). Using the external lab data as test inputs (removing any inaccuracy from our PoU sensors) gave optimized predictions of NO_3^- with $R^2 = 0.86$ (bottom right). **b**, The time-series dataset was used to train an LSTM recurrent neural network model to forecast NH_4^+ and NO_3^- into the future, also for unseen environmental conditions. The models were retrained for each desired forecast time (1–12 days into the future, with shading corresponding to the standard deviation of seven predictions from the model, with the same initial conditions), and comparing predicted to real values over the 12-day period gives scores of $R^2_{NH_4+} = 0.60$ and $R^2_{NO_3-} = 0.70$.

The impact of this work is that growers can instantly determine crucial soil nutrients using only PoU measurements and weather data and can forecast the levels of nutrients into the future to build better fertilization plans. This would ensure that appropriate nutrients are present when needed by the crops. This approach could enable precision farming of a new caliber (with substantially lowered capital investment), reducing fertilizer requirements, soil degradation and eutrophication while improving crop yields. Furthermore, we hope this approach will extend to complex media other than soil, where simple chemical measurements and easily accessible data, combined with ML, can be used to predict crucial outputs in health care, food and environmental monitoring.

Methods

Soil experiments. Top-soil with sandy loam texture (69% sand, 2.00–0.063 mm diameter; 25% silt, 0.063–0.002 mm diameter; 6% clay, <0.002 mm diameter; density, 774 $g l^{-1}$; measured in NRM Laboratories, part of Cawood Scientific) was purchased from Westland and used in the experiments without further modifications. The soil was heat sterilized, but microbiota are expected to survive. For the soil experiments performed in our laboratory, the water-soluble compounds and small particles were extracted from the soil samples by mixing 100 ml of diH_2O with 100 g of soil for approximately one minute (simulating a PoU scenario; longer mixing may enhance extraction) and then pressing out the solution with a potato press (VonShef) for a few seconds. We used the soil solution for the pH, EC and NH_4^+ measurements in our laboratory. Soil samples (200 g) for the measurements at the external laboratory (NRM) were extracted from the soil pots and stored in Ziploc bags (placed inside a cool box along with a cooling element), which were collected and analysed within two days. Different from our method of handling, the external lab used a soil-to-water ratio of 1:2.5, as they dried the samples before processing to improve consistency (we did not do this, which caused issues surrounding unmatching results between the external measurements and the measurements performed by our group). The levels of soil-N were measured colorimetrically by the external laboratory. NH_4^+ was reacted with alkaline hypochlorite and phenol to form indophenol blue. Sodium nitroprusside acted as a catalyst in the formation of indophenol blue, which was measured at 640 nm. NO_3^- was reduced to nitrite using cadmium in an open tubular cadmium reactor. A diazo compound formed between nitrite and sulphanilamide, which was coupled with *N*-(1-naphthyl)ethylenediamine dihydrochloride to give a red azo dye, measured at 540 nm. For all soil experiments, soil was weighed into pots of 5.1 kg and fertilized with 51 ml of 0.665 M (12,000 ppm) NH_4NO_3 while mixing thoroughly, resulting in soil at approximately 120 ppm NH_4NO_3.

Fabrication of the chemPEGS. ChemPEGSs with carbon electrodes (no. C2130925D1 conductive carbon ink, 55/45 wt% with no. S60118D3 diluent from GWENT Group) were screen-printed on chromatography paper (Whatman, grade 1 chromatography paper, 20 cm × 20 cm, 0.18 mm thickness) and dried overnight at room temperature to remove excess organic solvents from the electrodes. The design of the electrodes consists of three interdigitated electrodes with 1 mm spacing between each finger, which was optimized to increase sensitivity—that is, we wanted to keep the sensors as resistive as possible up to the level where our electronics could handle the low currents. To contain the H_2SO_4 added in a certain region within the paper, wax was deposited around the electrode area to provide a hydrophobic barrier. We printed the wax designs using a Xerox ColorQube 8580 printer on Office Depot transparent acetate sheets and then heat-transferred them to the chemPEGS substrate with a Vevor HP230B heat press (180 °C). We then drop-casted 10 µl of 0.025 M H_2SO_4 on the chemPEGS before use, to neutralize the ammonia gas. An array of batch-processed chemPEGS sensors is shown in Supplementary Fig. 17.

Gas-phase measurement of NH_4^+ with the chemPEGS. To operate the sensor module (that is, the cartridge), a soil solution was created by pressing 100 ml of deionized water through 100 g of soil. A 5 ml soil solution was injected into each container. In the container, the solubilized soil $NH_4^+(aq)$ is in equilibrium with solubilized $NH_3(aq)$ (equation (1)), which is in equilibrium with volatilized $NH_3(g)$ in the headspace of the container under Henry's law (equation (2)):

$$NH_4^+(aq) + OH^-(aq) \leftrightarrow NH_3(aq) + H_2O \quad (1)$$

$$NH_3(aq) \leftrightarrow NH_3(g) \quad (2)$$

When the pH is increased to 14 by the concentrated NaOH solution, the equilibrium shifts towards $NH_3(aq)$ and ultimately $NH_3(g)$. The $NH_3(g)$ in the headspace of the container once again dissolves in the layer of water adsorbed on the paper sensors (described by Barandun et al.[37]) and then neutralizes the H_2SO_4 in the paper, causing an increase in the ionic impedance (mainly due to the neutralization of highly mobile H^+ ions) of the sensor in a concentration-dependent manner (Fig. 2b)[36,49–51]. The neutralization of NH_3 in the chemPEGS (equation (3)) draws out more NH_4^+ from the soil solution to maintain equilibrium; hence, the paper sensor acts as a scrubber of soil NH_4^+:

$$2NH_4^+ + 2OH^- + 2H^+ + SO_4^{2-} \rightarrow 2NH_4^+ + SO_4^{2-} + 2H_2O \quad (3)$$

The cartridges comprised a gold-plated card-reader (Midland Ross CD6734 BRN 34 Way IDC Ribbon Cable Card edge Connector) inserted into the screw cap of a 50 ml centrifuge tube (VWR) and sealed with a glue gun. Using the electrical contacts on the back (left outside the tube), the card-reader was connected to a custom-built printed circuit board containing electronics that can apply a 10 Hz, 4 V_{p-p} signal across the chemPEGS. The chemPEGS was inserted into the card-reader and placed in a sealed centrifuge tube containing 1 ml of 15 M NaOH. The current produced as a result of the voltage applied was converted to a voltage again by an operational amplifier-based transimpedance amplifier (with a ratio defined by a gain resistor); the voltage signal was subsequently recorded by an Arduino Due using its onboard Analog-to-Digital Converter and communicated to a nearby PC over a serial link. Once the electrical signal (the current passing through the sensor) stabilized (that is, the paper substrate reached an equilibrium with the humidity inside the tube), 5 ml of soil solution was injected into the tube using a syringe, and changes in the electrical current were recorded as the analytical signal.

Calibration of the chemPEGS for measuring soil NH_4^+. The increase in impedance (calculated using Ohm's law) of the chemPEGS was measured over time during the neutralization of H_2SO_4 that was drop-casted previously. Calibration was first performed without soil for a range of NH_4NO_3 concentrations (4.5, 9, 18, 36, 72, 144 and 287 ppm) and a range of H_2SO_4 concentrations (0.1, 0.05 and 0.025 M), as shown in Supplementary Fig. 1. Calibration was attempted by comparing the concentration of NH_4NO_3 to the total change in the impedance of the chemPEGS and the time taken for the impedance to stop increasing or slow dramatically (Supplementary Fig. 3). A concentration of 0.025 M for H_2SO_4 gave the fastest and most precise measurements (Supplementary Fig. 1). Calibration in soil solution was performed for a range of concentrations (4.5, 9, 18, 36, 72 and 144 ppm) of NH_4NO_3 spiked in the soil sample and then extracted. A calibration curve was fitted as NH_4^+ (ppm) = (time [minutes] × 5.43 × 10^{-4})$^{-1.26}$ with $R^2 = 0.96$ (Fig. 2d).

Control of rainfall and temperature. Rainfall was fixed at 1, 3, 5 or 10 mm d^{-1}, implemented by adding a daily equivalent of 57 ml, 172 ml, 286 ml or 573 ml to a pot area of 573 cm^2 (the pots were watered every two days). Temperature was controlled by wrapping pots containing soil with nichrome wire (purchased from Amazon) and applying a 36 V potential, resulting in an electrical current of 1.5 Å supplied from two Tenma 72-8350A power supplies in series. Soil temperature was measured at three points (centre, edge and in between) and averaged to estimate the temperature of the soil periodically, using a Silverline 469539 Pocket Digital Probe Thermometer.

Measurement of EC and pH of soil. Using a Hanna Instruments HI5222-type benchtop EC/pH meter, we measured the pH and EC of the solution extracted from the samples of soil. Each sample was measured five times, and the readings were averaged to reduce error.

ML model. All computational work was performed using Python (v.3.6) in the PyCharm integrated development environment. For modelling and optimization, we used the following core packages: Keras API for Tensorflow (LSTM model), Scikit-learn (ensemble and Knn regressors), XGBoost, pandas and NumPy.

Statistics and reproducibility. The sample sizes for the measurements from sensors made in our lab were $n = 5$ unless otherwise stated, with error calculated by standard deviation. No statistical method was used to predetermine sample sizes, but they are similar to those reported in similar work[37]. All data shown in Fig. 3 were used for training the ML models with no exclusions. Since only physical measurements were made, the experiments were not randomized, and the investigators were not blinded to allocation during the experiments and outcome assessment.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data are available from the corresponding author upon reasonable request.

Received: 14 December 2020; Accepted: 28 October 2021;
Published online: 13 December 2021

References

1. *World Population Prospects 2019* https://population.un.org/wpp/Download/Files/1_Indicators(Standard)/EXCEL_FILES/1_Population/WPP2019_POP_F01_1_TOTAL_POPULATION_BOTH_SEXES.xlsx (United Nations Department of Economic and Social Affairs, 2019).

2. Mosheim, R. *Fertilizer Use and Price* https://www.ers.usda.gov/webdocs/DataFiles/50341/fertilizeruse.xls?v=5260.4 (United States Department of Agriculture—Economic Research Service, 2019).
3. Alexandratos, N. & Bruinsma, J. *World Agriculture Towards 2030/2050* http://www.fao.org/3/ap106e/ap106e.pdf (Food and Agriculture Organization of the United Nations, 2012).
4. Tetteh, R. N. Chemical soil degradation as a result of contamination: a review. *J. Soil Sci. Environ. Manage.* **6**, 301–308 (2015).
5. Osman, K. T. *Soil Degradation, Conservation and Remediation* (Springer Netherlands, 2014); https://doi.org/10.1007/978-94-007-7590-9
6. Beeckman, F., Motte, H. & Beeckman, T. Nitrification in agricultural soils: impact, actors and mitigation. *Curr. Opin. Biotechnol.* **50**, 166–173 (2018).
7. Smith, V. H., Tilman, G. D. & Nekola, J. C. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ. Pollut.* **100**, 179–196 (1999).
8. Zheng, F. et al. Mineral and organic fertilization alters the microbiome of a soil nematode *Dorylaimus stagnalis* and its resistome. *Sci. Total Environ.* **680**, 70–78 (2019).
9. Wang, Q. et al. Impact of 36 years of nitrogen fertilization on microbial community composition and soil carbon cycling-related enzyme activities in rhizospheres and bulk soils in northeast China. *Appl. Soil Ecol.* **136**, 148–157 (2019).
10. Zhao, Z. B. et al. Fertilization changes soil microbiome functioning, especially phagotrophic protists. *Soil Biol. Biochem.* **148**, 107863 (2020).
11. Rosas, F. *Fertilizer Use by Crop at the Country Level (1990–2010)* Working Paper No. 12-WP 535 (Center for Agricultural and Rural Development, Iowa State University, 2012); http://publications.iowa.gov/13990/1/12-WP_535.pdf
12. Zhu, Q., Schmidt, J. P., Lin, H. S. & Sripada, R. P. Hydropedological processes and their implications for nitrogen availability to corn. *Geoderma* **154**, 111–122 (2009).
13. Shanahan, J. F., Kitchen, N. R., Raun, W. R. & Schepers, J. S. Responsive in-season nitrogen management for cereals. *Comput. Electron. Agric.* **61**, 51–62 (2008).
14. Tremblay, N. et al. Corn response to nitrogen is influenced by soil texture and weather. *Agron. J.* **104**, 1658–1671 (2012).
15. Cilia, C. et al. Nitrogen status assessment for variable rate fertilization in maize through hyperspectral imagery. *Remote Sens.* **6**, 6549–6565 (2014).
16. Kitchen, N. R. et al. Ground-based canopy reflectance sensing for variable-rate nitrogen corn fertilization. *Agron. J.* **102**, 71–84 (2010).
17. Stone, M. L. et al. Use of spectral radiance for correcting in-season fertilizer nitrogen deficiencies in winter wheat. *Trans. ASABE* **39**, 1623–1631 (1996).
18. *CropSpec Crop Canopy Sensors* http://topconcare.com/en/agriculture/crop-monitoring-technology/crop-canopy-sensors/ (Topcon Totalcare, 2021).
19. *SS1 SunScan Canopy Analysis System* https://delta-t.co.uk/product/sunscan/ (Delta-T Devices, 2021).
20. *N-Sensor* https://www.yara.co.uk/crop-nutrition/farmers-toolbox/n-sensor/ (Yara, 2021).
21. Morellos, A. et al. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* **152**, 104–116 (2016).
22. Kim, H. J., Hummel, J. W. & Birrell, S. J. Evaluation of nitrate and potassium ion-selective membranes for soil macronutrient sensing. *Trans. ASABE* **49**, 597–606 (2006).
23. *World's First Wireless NPK Soil Sensor* (Teralytic, 2021).
24. Shaw, R., Lark, R. M., Williams, A. P., Chadwick, D. R. & Jones, D. L. Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture. *Agric. Ecosyst. Environ.* **230**, 294–306 (2016).
25. Hengl, T. et al. Soil nutrient maps of sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* **109**, 77–102 (2017).
26. Zaman, B. & McKee, M. Spatio-temporal prediction of root zone soil moisture using multivariate relevance vector machines. *Open J. Mod. Hydrol.* **4**, 80–90 (2014).
27. Brewster, C., Roussaki, I., Kalatzis, N., Doolin, K. & Ellis, K. IoT in agriculture: designing a Europe-wide large-scale pilot. *IEEE Commun. Mag.* **55**, 26–33 (2017).
28. Dincer, C. et al. Disposable sensors in diagnostics, food and environmental monitoring. *Adv. Mater.* **31**, e1806739 (2019).
29. Burton, R. *Nitrate Sensing in the Soil* https://www.cambridgeconsultants.com/insights/nitrate-sensing-in-the-soil (Cambridge Consultants, 2016).
30. Tully, K. L. & Weil, R. Ion-selective electrode offers accurate, inexpensive method for analyzing soil solution nitrate in remote regions. *Commun. Soil Sci. Plant Anal.* **45**, 1974–1980 (2014).
31. Choosang, J. et al. Simultaneous detection of ammonium and nitrate in environmental samples using on ion-selective electrode and comparison with portable colorimetric assays. *Sensors (Basel)* **18**, 3555 (2018).
32. Shaw, R., Williams, A. P., Miller, A. & Jones, D. L. Assessing the potential for ion selective electrodes and dual wavelength UV spectroscopy as a rapid on-farm measurement of soil nitrate concentration. *Agriculture* **3**, 327–341 (2013).
33. Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J. & Skjemstad, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59–75 (2006).
34. Sinfield, J. V., Fagerman, D. & Colic, O. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Comput. Electron. Agric.* **70**, 1–18 (2010).
35. Xuejiang, W. et al. Conductometric nitrate biosensor based on methyl viologen/Nafion®/ nitrate reductase interdigitated electrodes. *Talanta* **69**, 450–455 (2006).
36. Wongchoosuk, C. et al. Electronic nose for toxic gas detection based on photostimulated core-shell nanowires. *RSC Adv.* **4**, 35084–35088 (2014).
37. Barandun, G. et al. Cellulose fibers enable near-zero-cost electrical sensing of water-soluble gases. *ACS Sens.* **4**, 1662–1669 (2019).
38. Grell, M. et al. Autocatalytic metallization of fabrics using Si ink, for biosensors, batteries and energy harvesting. *Adv. Funct. Mater.* **29**, 1804798 (2018).
39. Ainla, A., Hamedi, M. M., Güder, F. & Whitesides, G. M. Electrical textile valves for paper microfluidics. *Adv. Mater.* **29**, 1702894 (2017).
40. Hamedi, M. M. et al. Integrating electronics and microfluidics on paper. *Adv. Mater.* **28**, 5054–5063 (2016).
41. Glavan, A. C. et al. Analytical devices based on direct synthesis of DNA on paper. *Anal. Chem.* **88**, 725–731 (2016).
42. Stark, J. M. & Firestone, M. K. Mechanisms for soil moisture effects on activity of nitrifying bacteria. *Appl. Environ. Microbiol.* **61**, 218–221 (1995).
43. Nguyen, L. T. T. et al. Impacts of waterlogging on soil nitrification and ammonia-oxidizing communities in farming system. *Plant Soil* **426**, 299–311 (2018).
44. Taylor, A. E., Giguere, A. T., Zoebelein, C. M., Myrold, D. D. & Bottomley, P. J. Modeling of soil nitrification responses to temperature reveals thermodynamic differences between ammonia-oxidizing activity of archaea and bacteria. *ISME J.* **11**, 896–908 (2017).
45. Hao, T. et al. Impacts of nitrogen fertilizer type and application rate on soil acidification rate under a wheat–maize double cropping system. *J. Environ. Manage.* **270**, 110888 (2020).
46. Smith, J. L. & Doran, J. W. in *Methods for Assessing Soil Quality* Vol. 49 (eds Doran, J. W. & Jones, A. J.) 169–185 (Soil Science Society of America, 1997).
47. Rogovska, N., Laird, D. A., Chiou, C. P. & Bond, L. J. Development of field mobile soil nitrate sensor technology to facilitate precision fertilizer management. *Precis. Agric.* **20**, 40–55 (2019).
48. Ortega, L., Llorella, A., Esquivel, J. P. & Sabaté, N. Paper-based batteries as conductivity sensors for single-use applications. *ACS Sens.* **5**, 1743–1749 (2020).
49. Güder, F. et al. Superior functionality by design: selective ozone sensing realized by rationally constructed high-index ZnO surfaces. *Small* **8**, 3307–3314 (2012).
50. Güder, F. et al. Paper-based electrical respiration sensor. *Angew. Chem. Int. Ed.* **55**, 5727–5732 (2016).
51. Maier, D. et al. Toward continuous monitoring of breath biochemistry: a paper-based wearable sensor for real-time hydrogen peroxide measurement in simulated breath. *ACS Sens.* **4**, 2945–2951 (2019).

Acknowledgements

We thank EPSRC (grant no. EP/R010242/1), Innovate UK (grant no. 33486), Cytiva (formerly General Electric Healthcare Life Sciences) and Imperial College, Department of Bioengineering, for their generous support. F.G. and M.G. thank the Imperial College Centre for Processable Electronics and the Centre for Doctoral Training in Plastic Electronics, S. Yaliraki from Imperial College Department of Chemistry, T. Bell from Imperial College Department of Life Sciences, C. Mace from Tufts University Department of Chemistry and T. Schaul from (Google) DeepMind Technologies Ltd. F.G. also acknowledges Agri Futures Lab. M.K. acknowledges EPSRC DTP (grant no. 1846144). A.S.P.C. acknowledges BBSRC DTP (grant no. 2177734).

Author contributions

M.G. designed and performed most of the experiments and wrote the manuscript. G.B. designed the homemade electronics. T.A. produced the supplementary videos and iterated on Fig. 1. M.K. performed the soil measurements and maintenance. A.S.P.C. and J.W. performed measurements with the homemade electronics. F.G. designed the experiments and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43016-021-00416-4.

Correspondence and requests for materials should be addressed to Firat Güder.

Peer review information *Nature Food* thanks Arben Merkoçi, Anna Herland and Ramses Martinez for their contribution to the peer review of this work.

nature research

Corresponding author(s): Dr Firat Guder

Last updated by author(s): Sep 30, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Arduino IDE (1.8), Google Sheets
Data analysis	Machine learning computation performed using Python (3.6) in PyCharm IDE (2019). Packages: Keras API for Tensorflow (LSTM model), Scikit-learn (ensemble and Knn regressors), XGBoost, pandas and NumPy. Plotting and R2 fitting performed with Origin (2018).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We have gathered a new dataset using point-of-use soil sensors, some of which we have developed. We use machine learning regressors to predict difficult-to-measure soil nitrogen from measured data, then train a LSTM machine learning model to predict soil nitrogen into the future.
Research sample	We made measurements in top soil (sandy loam) to simulate a field. We varied temperature and rainfall to simulate a variety of weather conditions (arid to tropical, temperate to warm).
Sampling strategy	Samples were taken from pots of soil in our laboratory, with constant mass for each sample. 5 repeat measurements of NH4, EC and pH were made on each soil sample every time. Each soil temperature measurement was made 3 times, once on the edge, once in the middle and once half way between.
Data collection	All measurements made in our laboratory were performed by Max Grell, with data recorded in Google Sheets and as text files by Arduino IDE. Samples were also analyzed by an external laboratory.
Timing and spatial scale	Measurements were made every few days, but more frequently at the start of each experiment. The variable frequency was chosen to better capture changes in soil nutrients. There was no spatial dimension.
Data exclusions	Data were not excluded
Reproducibility	Leave-one-out cross-validation was used for all machine learning predictions. Repeats of n>3 were used for each measurement (described above)
Randomization	N/A, our experiments were built and controlled in-house
Blinding	N/A, there were no subjects to hide placebos from

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging