ARTICLE OPEN

Check for updates

Real-time detection of ammonium in soil pore water

Rotem Yupiter [1✉], Shlomi Arnon [2], Elad Yeshno[1], Iris Visoly-Fisher[3] and Ofer Dahan[1]

The development of technologies for continuous measurement of nitrogen forms in the soil is essential for optimizing the application of fertilizers in agriculture and preventing water-resource pollution. However, there is no effective commercial technology available for continuous monitoring of ammonium species in soil pore water. This work investigates an approach for real-time measurement of ammonium in soil water using near-infrared transmission spectroscopy and partial least squares regression (PLSR) for spectral analysis. The PLSR model was trained using soil pore water collected from various soils spiked with ammonium to achieve a wide concentration range. The monitoring approach was then validated through transport experiments in a soil column. The results demonstrated capabilities for real-time tracking of the temporal variation in soil ammonium concentration and potential utilization in agronomical or environmental sensing.

INTRODUCTION

The deterioration of natural water resources, such as ground and surface waters, is often attributed to point and non-point sources associated with industrial, urban, and agricultural activities. According to United Nations World Water Development Report 2018[1], one of the leading causes of water-resource deterioration is the excess application of fertilizers in agriculture. Excess fertilizer application is commonly practiced worldwide to prevent nutrient deficiency and consequent yield reduction. Unfortunately, excess nitrogen (N) fertilizer in the soil cannot be utilized by the plants and ultimately leaches down from the root zone through the unsaturated zone (also termed vadose zone) to the ground water[2–5]. Although on a global scale, agriculture is the largest non-point source of N pollution of water resources, high levels of N are also related to other pollution sources, such as landfills, animal husbandry farms and wastewater-treatment facilities[6–9]. The economic and ecological consequences of N pollution are significant. Elevated concentrations of N species, i.e., above drinking-water standards (EU; NO_3^- 50 mg L^{-1}; NO_2^- 0.5 mg L^{-1}; NH_4^+ 0.5 mg L^{-1})[10], are considered hazardous and a health risk, ultimately leading to disqualification of drinking-water wells. In addition, natural discharge of polluted ground water to surface water, such as rivers[11,12], lakes[13] and even oceans[14], can ignite large-scale algal blooms and eutrophication, which threaten aquatic environments and marine life[15–17].

Ammonium (NH_4^+) and nitrate NO_3- are the most readily accessible inorganic N forms for plant uptake[18]. They are limiting factors for crop yield and therefore considered the main components of N fertilizers. While nitrate is a stable mobile anion that is easily transported in the soil, ammonium is a positively charged ion which is subject to sorption. Its mobility and presence in soil pore water is highly affected by its affinity for negatively charged surfaces and lattice vacancies, which are present in natural clay minerals such as vermiculite, montmorillonite, and illite[19,20]. Due to its absorptive attraction to clay minerals, ammonium transport in aerated soils is considered low and strongly dependent on soil-oxidation conditions[21]. Over time, these fixed ions can diffuse and are released from the clays back into the pore water, allowing the roots and microorganisms to use them[22,23]. Under cultivated field conditions, ammonium in the top soil pore water can range from micrograms to hundreds of milligrams per liter, depending on the crop species, fertilizer application, and soil properties[24,25]. Accordingly, achieving proper ammonium balance inside the root zone for maximum yield and minimal loss is a challenging task. While agricultural soils are often aerated, and the main pollution-causing N form is nitrate, under anoxic conditions, ammonium can be a dominant species. Such anoxic conditions may prevail in the presence of high organic loads, as are often found under animal husbandry farms, landfills or mega-industrial sites, or from septic effluents or even flooded conditions, such as those prevailing under flooded rice pads. Anoxic soil conditions may lead to very intensive transport of ammonium across the unsaturated zone into the ground water[26–28]. For example, the anaerobic reducing environment underneath a waste landfill can release ammonium at very high concentrations from the biodegradation of organic waste matter[29]. These ions can then be transported from the waste body to the unsaturated zone following precipitation events, and can reach from dozens to thousands of milligrams per liter in the soil pore water[30]. Similar conditions may prevail under animal husbandry sites, where large quantities of ammonium can leach down from dairy waste lagoons through the soil and deep unsaturated zone[31,32]. In addition, when organic N and ammonium levels exceed 10 mg L^{-1} in water infiltrating through managed aquifer recharge systems, N may leach into the groundwater[33]. Accordingly, real-time detection of ammonium in the soil is of high importance for both optimization of fertilizer application and prevention of water-resource pollution.

Monitoring N species in the soil is key to preventing water-resource pollution. Soil analysis is a broad discipline that includes a variety of tools for managing fertilizers in agricultural areas and for identifying pollutants during their transport through the unsaturated zone[34]. Unfortunately, lack of in situ, rapid, low-cost, robust, and high-resolution technologies for the characterization of soil N makes it difficult to detect these pollutants accurately.

[1]Department of Hydrology & Microbiology, Zuckerberg Institute for Water Research, Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel. [2]Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel. [3]Department of Solar Energy and Environmental Physics, Swiss Institute for Dryland Environmental and Energy Research, Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel. ✉email: rotemyupiter@gmail.com
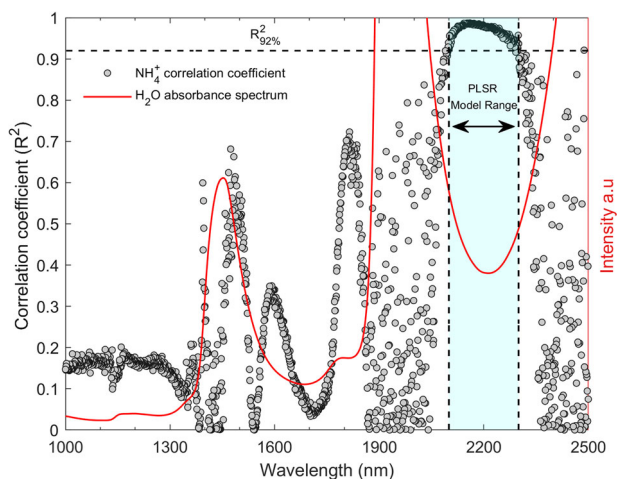
Fig. 1 Defining the prediction model boundaries. Left y axis: correlation values between ammonium concentration and absorbance values for each specific wavelength. Right y axis: water absorption in the NIR spectrum.

Optical spectroscopy based on reflectance analysis offers a good solution as a base technology for real-time soil analyses[35]. The method is rapid, non-destructive, non-polluting, can be done onsite, and is capable of determining several properties simultaneously if broad-wavelength bands are used[36]. However, optical reflectance is not designed to track soil pore water, which is subject to sudden changes in chemical composition and nutrient exchange during different stages of the plant growing phases, irrigation and fertilizer application. Measuring pore water offers a glimpse into these rapid soil processes and is thus better adapted to being a real-time sensor. Tuli et al.[37] used this alternative strategy by measuring nitrate in pore water with UV absorbance spectroscopy in porous cups (suction lysimeters). A similar approach was recently suggested by Yeshno et al.[38], who measured nitrate concentration in a flux of flowing soil pore water obtained continuously from a porous interface in the soil root zone. Their technology was based on UV light absorption measurement in an optical flow cell connected to a soil porewater-sampling system. Special analysis of the absorption spectrum enabled reducing dissolved organic carbon interference of nitrate absorption and determining a site-specific wavelength. The integrated system enabled real-time continuous measurement of nitrate concentration in soil pore water and detection of rapid changes in ion concentration in the macro soil environment. However, the fate of ammonium, which is also a major nutrient for agriculture and a potential pollutant of water resources, remained unknown.

Ammonium has a few spectral signatures in the near infrared (NIR) range (800–2500 nm; 12,500–4000 cm^{-1}) due to overtones and combination bands of fundamental vibrations. However, NIR transmission spectroscopy can be limited when applied to solutions because water molecules (H_2O) have strong and broad absorption bands in the NIR range. Accordingly, other molecular bonds that absorb light at the same wavelengths, such as N–H, C–H and O–H, can be masked by the strong absorption of water. This effect can cause significant analytical interference because transmission spectroscopy is based on quantifying the changes in absorbance intensity. Hence, accurate analysis of water samples with NIR techniques must consider other important parameters that can increase the signal intensity, such as light pathlength and specific wavebands[39].

This manuscript presents a technique for ammonium detection in soil pore water using NIR absorption spectroscopy. We used a specific NIR band (2100–2300 nm) with strong absorption of dissolved ammonium ions. The spectral data were analyzed with the partial least squares regression (PLSR) algorithm to quantify ammonium concentration in soil pore water. The trained model was based on spiked pore water from different agricultural soils. Once the analytical ammonium model was established, the concept was tested and validated in a soil-column experiment. This approach lays the foundation for the development of a monitoring system for continuous real-time measurement of soil ammonium concentration.

RESULTS AND DISCUSSION

Ammonium spectral profile

Determining the correlation coefficient (R^2) between absorption values and the concentration of a specific analyte in a standard solution is a common technique for identifying specific absorption bands. We calculated the R^2 values for a batch of ammonium standard solutions and absorption intensity in the NIR range (Fig. 1). Some R^2 values—around 1450 nm, 1950 nm, and above 2350 nm—showed large variability because of strong water-absorption intensity. These wavelengths are associated with, respectively, the first overtone of the OH-stretching band ($2v_{1,3}$), the OH combination band ($2v_{1,3}+v_2$)[40] and the fundamental vibrations ($v_{1,3}$) of water molecules (H_2O). These water vibrations absorb a large portion of the NIR light going through the sample, and strongly mask other molecules that could potentially be detected. The wavelength band between 2100 and 2300 nm showed the highest correlation coefficients, with $R^2 > 0.92$ ($p < 0.01$) and consequently, was optimal for building a calibration model (noted as PLSR model range in Fig. 1). This band is associated with a strong N–H combination band and has also been observed and used in biological studies of cell culture and fermentation[41–45]. Although the absorption bands at 1490, 1590, and 1800 nm presented $R^2 = 0.61$, 0.35 and 0.72, respectively, they were not included, to make a more compact model. These were associated with overtones and another combination band of N–H[46].

A graphical analysis of the spectrum (2100–2300 nm) divided into groups based on soils from different sites revealed high variability across the groups (Fig. 2). The variability was caused by the diverse chemical composition of the samples collected from various agricultural soils. These variations were both expected and important, as one of our primary goals was to train a model with a large spectral pool. Univariant (single-wavelength) calibration approaches would not be suitable here, because they rely on a single value that might be biased, resulting in mismatched results.

PLSR model and limit of detection

PLSR is a statistical technique used for predictive modeling. The technique is similar to multiple linear regression, but it is particularly effective when there are a large number of predictor variables (also known as "features") and a small number of observations. PLSR aims to find the linear combination of predictor variables that is most strongly correlated with the response variable. This is achieved by constructing a set of "latent variables" (LVs), which are linear combinations of the predictor variables, and using these LVs as predictors in a multiple linear regression model. One advantage of PLSR is that it can handle multicollinearity, where two or more predictor variables are highly correlated. PLSR can identify the most important latent variables and use them to make predictions, rather than being influenced by the noise introduced by multicollinearity.

Here, we used PLSR to predict the concentration of ammonium from spectral analyses of pore water samples. The PLSR model was trained on spectra (2100–2300 nm) of chemically heterogeneous pore water samples ($n = 106$). Given the relatively small number of samples, we assumed that this would be sufficient to build an
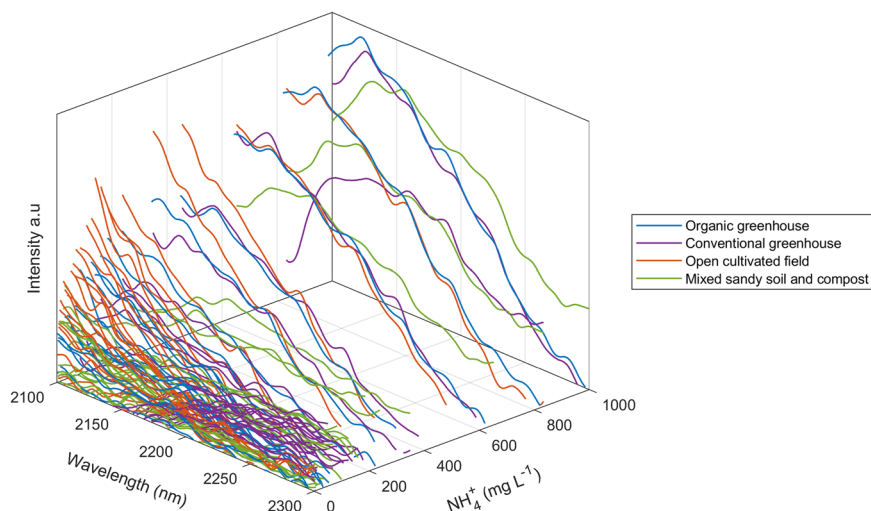
Fig. 2 Spectral data of soil porewater samples used for the training model. Heterogeneous composition of the samples is represented by different spectral signatures at the same concentrations.
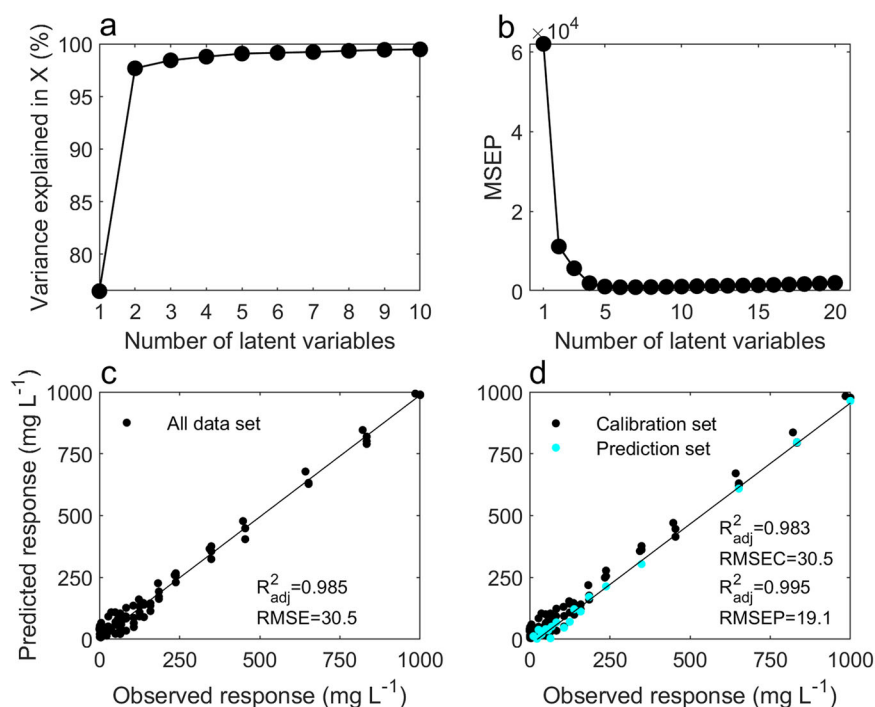

Fig. 3 Characterization of the PLSR model. a Variance explained in the spectra, **b** mean MSEP values for each latent variable in the model, **c** regression of all model-predicted concentrations on known (observed) values, and **d** validating the model with 80% training (calibration) set and 20% testing (prediction) set.

accurate prediction model. It is important to note that using a training set with a larger sample size would yield a more accurate predictive model. However, since this work was aimed at developing a method that could be further used for the development of a dedicated sensor for in situ measurement of ammonium in the field, minimizing pre-preparation and sample number was also an important consideration.

The first necessary step was to find the optimal number of LVs using the explained variance and mean squared error of prediction (MSEP) (Fig. 3a, b). Overfitting caused by choosing a large number of LVs might lead to a PLSR model that includes noise in the data, resulting in a biased training model, whereas a number of LVs that is too small could result in information loss. The results showed that with 2 LVs, the explained variance in the

spectra is >98%. This was not surprising, since we used a broad wavelength with strong ammonium absorption (Fig. 1). The explained variance term only supports the estimation of the minimum number of LVs, and not the optimal number. The minimum MSEP value was obtained with 6 LVs; however, the change in the error dropped dramatically after 4 LVs, suggesting this as a more optimal number of LVs (Fig. 3b). Note that in the column-transport experiment, we also checked the PLSR model with 6 LVs and found that our estimation of 4 LVs gives better results.

The 4-LV PLSR model was evaluated with adjusted correlation coefficient (R^2_{adj}) and root mean square error (RMSE) by regressing the observed on the predicted responses (concentrations). The results were $R^2_{adj} = 0.985$, root mean squared error of calibration
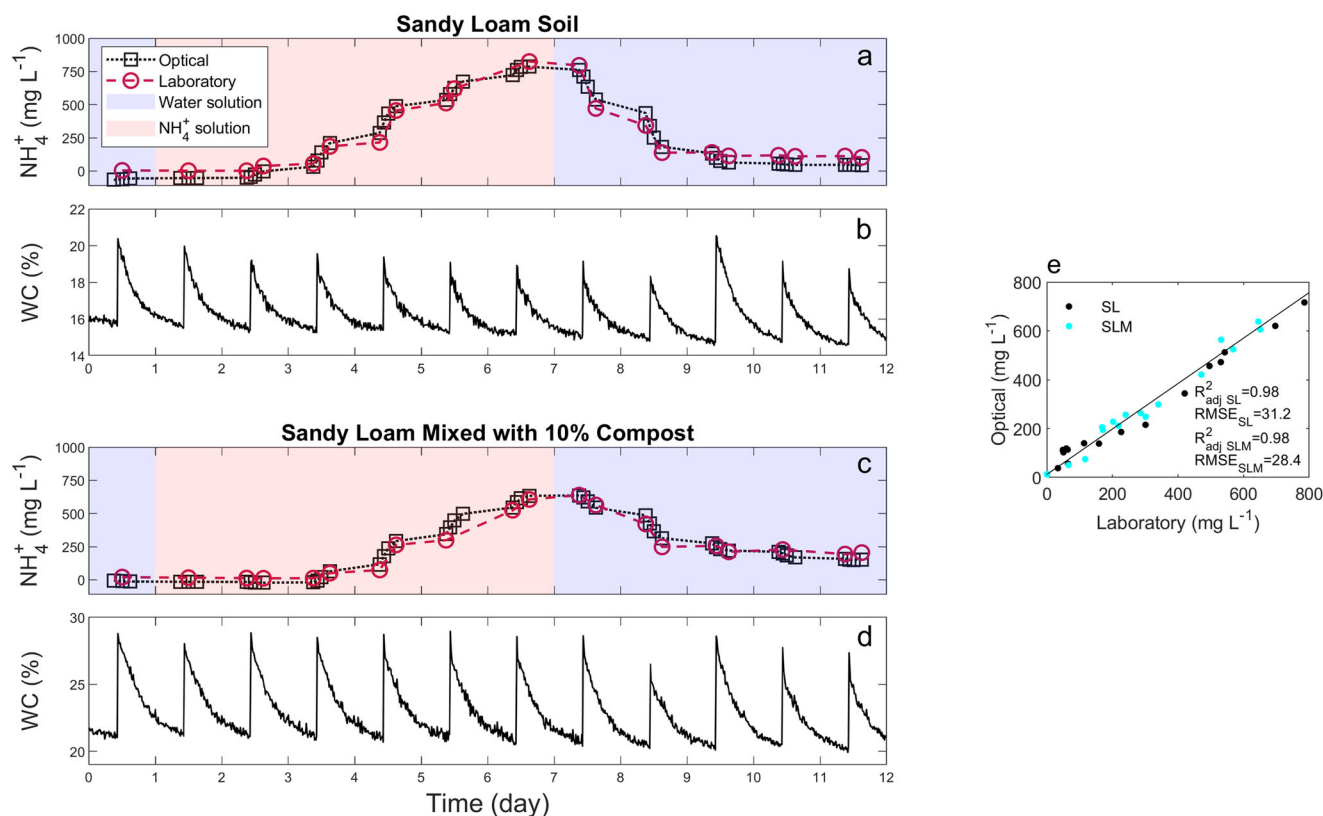
Fig. 4 Ammonium breakthrough curves in soil columns. a, c Breakthrough curves of ammonium in the sandy loam soil (SL) and SL mixed with 10% compost (SLM) columns, respectively, obtained from the optical technique and standard laboratory procedure. **b, d** Daily WC fluctuations in both columns. **e** Regression plot of the optical vs. laboratory measurements.

(RMSEC)=30.5 (mg L^{-1}) and $p < 0.01$ (Fig. 3c). For validation, the model was trained with 80% of the data (calibration set) and was tested on the remaining 20% (prediction set). The results from the calibration set were $R^2_{adj} = 0.983$, RMSEC = 30.5 (mg L^{-1}) and $p < 0.01$ (Fig. 3d, black points). The prediction set results were $R^2_{adj} = 0.995$, root mean squared error of prediction (RMSEP)=19.1 (mg L^{-1}) and $p < 0.01$ (Fig. 3d, cyan points). The R^2_{adj} values of the trained PLSR model were >0.98. However, The RMSE values were in the dozens of milligrams per liter. From an analytical perspective, the presented methodology may not demonstrate high accuracy. However, for fertilized agricultural soils and sites prone to ammonium contamination, it may be satisfactory. Nevertheless, this PLSR model for optical measurements of ammonium needs to be refined at a finer scale in order to achieve higher accuracy.

The motivation for assessing the limit of detection (LOD) was to define the lowest ammonium concentration that can be reliably detected in a solution using this approach. The LOD was tested using Allegrini and Olivieri's technique[47], which was adapted to the PLSR model and follows the International Union for Pure and Applied Chemistry (IUPAC) standards. The LOD results were 1.40–2.62 mg L^{-1} (LOD$_{min}$–LOD$_{max}$) (Supplementary Fig. 1). This range specifies that concentrations below the LOD$_{min}$ cannot be detected, while those above the LOD$_{max}$ may be, in a high confidence of level ($p < 0.05$). Concentrations between these values ($1.40 < y < 2.62$) need to be carefully tested. Use of a more sensitive device or a different spectral band might have resulted in a different LOD, but these were not tested here. The results emphasize how well the model may be modified to lower concentrations and so may be more adapted to shallow agricultural soils where these concentrations are more prevalent. All of the results and information on the LOD calculations can be found in Supplementary Methods.

Real-time ammonium breakthrough experiment

Real-time measurement of ammonium transport in two soil columns was tested with the trained PLSR model. One column was filled with sandy loamy soil (SL) and the other, SL mixed with compost (SLM) in a 10% weight ratio, to determine the impact of soil organic matter and other natural soil-water constituents on ammonium transport and spectral properties. The columns were irrigated daily with fresh water or ammonium solution, and the porewater samples were collected from the column 3–4 times a day using special customized low-dead-volume suction cups to track the ammonium breakthrough (Fig. 4a, c).

Water content (WC) and electrical conductivity (EC) in the soil columns were continuously monitored throughout the experiment with a time domain transmissometer (TDT) sensor. The hydraulic and chemical conditions in the two soils were very different, because the compost in the soil is expected to increase the water retention, salinity and dissolved organic carbon. As expected, the mean volumetric WC in the SL was 15% and 19.3% before and after irrigation events, respectively (Fig. 4b, d), whereas in the SLM column, the mean WC was much higher, reaching 20.75% and 28.15% before and after the irrigation cycles, respectively. The EC values of the water samples from the SL column were almost insignificant and ranged between 0.01 and 0.5 dS m^{-1}, whereas in the SLM column, the values were much higher and ranged between 0.2 and 1.1 dS m^{-1} throughout the experiment, signifying salinity release from the compost (Supplementary Fig. 2). The mean pH values in both columns were relatively similar, 7.42 ± 0.05 and 7.49 ± 0.05 in the SL and SLM columns, respectively. However, the initial total organic carbon (TOC) values were drastically different, 7 and 147 mg L^{-1} for the SL and SLM soils, respectively.

Along with the standard laboratory analysis, the ammonium concentration in the soil pore water, as predicted by the PLSR-

trained model for the NIR spectroscopic analysis, presented ammonium breakthrough curves in the soil columns (Fig. 4). A clear appearance of ammonium breakthrough appeared after 2.5 and 3.5 days in the SL and SLM soil columns, respectively. The ammonium breakthrough in the SLM was significantly retarded compared to the breakthrough in the SL soil due to the high organic content of the former. Six days after initiation of the ammonium application, the concentration in the SL column reached a maximum of ~900 mg L^{-1} and the irrigation water was replaced with tap water. Note that the maximum concentration in the SLM column reached only 705 mg L^{-1}. Immediately after switching back to fresh water, a reduction in ammonium concentration was observed, with an expected delay in the SLM column. Ammonium retardation in both soils is related to fixation and absorption to both minerals and organic matter in the soil. Therefore, ammonium mobility in the soil is dictated by the degree of absorption saturation. Although ammonium transport in soil is of great interest, in this study, we focused on the optical analytical procedure and not on the transport process.

Through these long-term measurements, we got the comparative results ($R^2 = 0.98$, RMSE $= 31.2$ mg L^{-1} and $p < 0.01$ in the SL column; and $R^2 = 0.98$, RMSE $= 28.4$ mg L^{-1} and $p < 0.01$ in the SLM column; Fig. 4e). The validation data points for the ammonium measurements were based on the total N (TOC/TN). At 900 mg L^{-1} of ammonium, a few ppm of nitrate is not a significant concern. However, since ammonium can be transformed to nitrate through nitrification in soil and the ammonium concentration was lower at other stages of the experiment, it was important to determine the actual concentrations of all N forms. An ion chromatograph was used to measure the concentration of nitrate and nitrite in five samples from each column. In the SL column, nitrate concentrations ranged from 0.5 to 1 mg L^{-1} (0.2–0.8% of TN) and nitrite concentrations ranged from 0.5 to 2.6 mg L^{-1} (0.4–3% of TN). In the SLM column, which contains more organic matter, nitrate concentrations were higher, ranging from 0.7 to 18 mg L^{-1} (2.8–8.6% of TN), and nitrite concentrations ranged from 1.1 to 12 mg L^{-1} (1.6–7.3% of TN).

The ammonium concentration range studied here was rather high, reaching 1000 mg L^{-1}. Although in aquatic environments, the concentration range that is considered highly polluted is in the range of a few to tens of milligrams per liter, in intensively fertigated soils, and under landfills and manure ponds, the concentration may reach hundreds of milligrams per liter[23,29,31]. Nevertheless, a low-concentration PLSR model can be built to adapt the approach to aquatic environments and less intensively fertigated soils.

METHODS

Optimizing waveband for the PLSR model

To optimize the optical band for our PLSR model, we performed a correlation coefficient (R^2) test using standard solutions of ammonium chloride salt (NH$_4$Cl; Merck EMSURE®) in double-distilled water. The dilutions were 31, 62, 125, 250, 500, 1000 mg L^{-1} to cover a wide range of concentrations. The samples were measured under the whole NIR spectrum (1000–2500 nm) using a Cary® 5000 UV–Vis–NIR (Agilent) spectrometer. The correlation coefficient was calculated for each wavelength separately. R^2 values were calculated using the Microsoft Excel *RSQ* function.

Soil solution collection and spiking procedure for the training model

Ammonium measurements in soil pore water combining NIR absorption spectroscopy with the PLSR algorithm requires the development of a training model based on a collection of samples with known ammonium concentrations and varying chemical composition. Four agricultural soils from the southern Israel coastline were used: (i) organic greenhouse soil based on compost fertilization, (ii) conventional greenhouse soil based on industrial fertilizers, (iii) soil from an open cultivated field of mixed crops, and (iv) sandy soil mixed with commercial compost (Table 1). The soils were packed into a 45-cm long and 29-cm diameter column. A customized small-volume ceramic suction cup was placed at 25 cm depth. The columns were irrigated daily with tap water, and soil pore water was collected using the suction cups. The soil pore water collected from each of the columns was separated into 27 vials (a total of 106 samples), 4 mL per sample. Each vial was spiked with a different amount of standard solution made up of Merck EMSURE ammonium chloride (NH$_4$Cl) salt dissolved in tap water (EC of 270 μS). Spikes consisted of 1–1000 mg L^{-1} to cover a wide range of concentrations, from a minimum reflecting natural conditions to a maximum matching that following fertilization events[48]. The ammonium concentrations in the pore water were corrected based on the ratio of ammonium (NH$_4^+$) to ammonia (NH$_3$) and a pH–temperature table.

NIR measurements and data processing

All NIR spectra were measured in the Cary 5000 UV–Vis–NIR spectrometer equipped with a PbS detector. Water solutions can be measured at NIR wavelengths by increasing the signal-to-noise ratio; this is done by reducing the light pathlength in a narrow-width cuvette, and by selecting the best wavelength region[39]. Therefore, a thin 1-mm pathlength fused quartz cuvette with 0.4 mL volume was used to measure the spiked solutions. The first absorption scan for optimizing the wavelength was between 1000 and 2500 nm with a spectral resolution of 1 nm. After the optimization, we only scanned the 2000–2500 nm range, at a scan rate of 300 nm min^{-1} with spectral resolution of 0.5 nm. Baseline correction was performed with double-distilled water as a reference.

To average the noise produced by the spectrometer, each sample was scanned three times and the mean spectrum was calculated. Then, the spectral data were smoothed with a Gaussian-weighted moving average filter with 21-point window size. Smoothing was performed using the MATLAB R2019a *smoothdata* function.

Table 1. Initial total organic carbon (TOC) and ammonium concentrations in the collected soil pore water.

Location	Soil	Site type	Ammonium (mg L^{-1})	TOC ± SD (mg L^{-1})	pH
Zikim	Loam	Organic greenhouse	2.00	257.4 ± 0.6	6.84
Zikim	Sandy loam	Conventional greenhouse	0.44	63.2 ± 0.5	6.55
Afek	Clay soil	Open cultivated field	ND	136.2 ± 0.3	6.88
Sde Boker	Sandy loam	Mixed sandy soil and compost	ND	14.0 ± 0.2	7.41

ND not detected.

PLSR model

A multivariant statistical PLSR algorithm was developed to predict the ammonium concentration in soil pore water from the spectral absorption data in the NIR wavelength range. The algorithm projects the original spectral dataset (x_i also denoted as "predictors") and the chemical concentrations (y also denoted as "responses") of the solution, to smaller orthogonal and uncorrelated dimensions, also known as LVs, and finds the maximum covariance between those parameters[49].

This chemometric technique is widely used for soil analysis, because it can quantify several properties or chemical values, and can find underlying patterns or relationships from overlapping spectral peaks[50]. Moreover, this method has several advantages over other regression techniques such as principal component regression (PCR), because the algorithm considers both the predictors and the responses and calculates them simultaneously. This creates a smaller and more compact model than PCR; it also handles missing data better and takes into consideration errors in the predictors[51]. Other techniques, such as deep learning convolutional neural network, have no significant advantages over PLSR when using particularly small-size training models ($n < 1400$) and the computation costs are higher[52].

In practice, the LVs can be used to predict the concentration of ammonium in new samples based on their spectral profiles using the PLSR. This can be done using Eq. 1:

$$y = x_1b_1 + x_2b_2 \ldots x_ib_i + \varepsilon \qquad (1)$$

where the predictors (x_1, x_2, etc.) are combined with their corresponding coefficients (b_1, b_2, etc.) to predict the responses (y). The error term (ε) represents the difference between the predicted response and the actual response. To calculate the LVs and the regression coefficients Eqs. 2 and 3 can be used:

$$T = XW \qquad (2)$$

$$B = YT \qquad (3)$$

Here the LVs (T) are calculated from the predictors (X – spectral dataset) and the weights (W). The regression coefficients (B) are calculated from the responses (Y) and the LVs. These LVs and coefficients iteratively estimated through a process of weighted least squares regression models until they maximize the explained variance in the responses. The number of LVs is a user-defined parameter. In this work the PLSR model was built using the MATLAB 2019a *plsregress* function.

PLSR training and validation

The optimal number of PLSR LVs was evaluated with the MSEP values. We calculated these values for each number of LVs using five fold cross-validation and 100 Monte-Carlo repetitions. After the number of LVs was set, we validated our training model with a test set. The spectral data observations were divided into 80% training and 20% testing. Then we built our PLSR model with 80% of the spectral profiles and tested it on the remaining 20% of the observations. The training and testing datasets were evaluated with R^2_{adj}, RMSEC for the training, and RMSEP for the testing.

Trained PLSR model testing with real-time data

The PLSR model was tested using a new set of data derived from spectral measurements of porewater samples collected during a column-transport experiment. These samples contained unknown concentrations of ammonium and therefore were optimal for our model validation. We used the 2100–2300 nm spectra of each sample, and the regression coefficients (b_i) computed during the training phase to predict the unknown ammonium concentrations. Then the ammonium concentrations from our model were compared to those from standard laboratory analysis and were evaluated with R^2_{adj} and RMSE.

Ammonium-transport experiment

A column experiment was conducted to test the applicability of the new methodology for continuous measurements of ammonium in soil. Two columns, 45 cm long and 29 cm diameter, were packed with two soils: SL and SLM (SL with 10% weighted commercial compost; Humus Factories Ltd.). The columns were drained by creating a continuous hydraulic gradient with mineral wool that was placed at the bottom of the column, into a 0.5-m long and 2.54-cm diameter drainage pipe (Fig. 5). A customized ceramic suction cup and WC sensor (Digital TDT® SDI-12, Acclima) were placed at a depth of 25 cm depth from the top of the soil
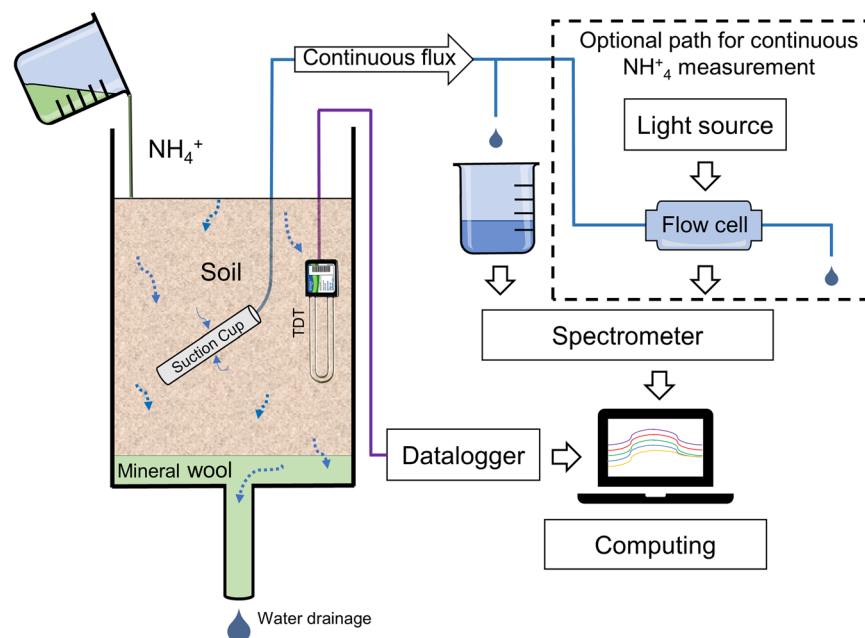

Fig. 5 Column setup. Illustration of the experimental column setup for measurement of ammonium transport in soil. The 'optional path' shows how to apply the approach with a real-time field sensor for measuring ammonium in the soil pore water.

column (Fig. 5). The columns were irrigated daily with 1 L (equivalent to 15.8 mm) of tap water. Once the daily oscillations in soil WC stabilized, the irrigation water was replaced with 1000 mg L^{-1} ammonium solution (EMSURE NH_4Cl salt mixed with tap water). Soil pore water was collected from the suction cups 3–4 times per day and the NIR absorption spectrum was optically measured in the Cary 5000 UV–Vis–NIR spectrometer. Note that soil pore-water sampling depends on soil water potential which vary between the irrigation cycles. Therefore, porewater sampling sequence is not evenly distributed. The Acclima TDT soil-moisture sensor was used to measure the soil bulk EC and WC at 15-min time resolution intervals using a Campbell Scientific CR300 datalogger. The column's wetting and drainage cycles were designed to mimic the natural unsaturated soil conditions of irrigated soil with daily volumetric WC fluctuation ranging between 15 and 30%. In this experiment, we deliberately used sandy loam soils with low clay content, to reduce the ammonium ion adsorption mechanism and obtain faster breakthrough curves.

Chemical analysis

All of the pH values in this work were measured using a pH meter. TOC and TN were estimated with a multi N/C® 2100 s (Analytic Jena AG) TOC/TN analyzer. Ammonium concentrations in the pore water from the four agricultural sites were measured using Nessler's reagent and absorbance at 425 nm in a TECAN Infinite® M200 spectrometer. A Thermo Scientific Dionex™ ICS-5000 ion chromatograph was used to measure nitrate and nitrite in the ammonium-transport experiment.

DATA AVAILABILITY

All data generated during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. UNESCO. *The United Nations world water development report 2018: Nature-Based Solutions for Water* (2018).
2. Smith, K. A., Jackson, D. R. & Pepper, T. J. Nutrient losses by surface run-off following the application of organic manures to arable land. 1. Nitrogen. *Environ. Pollut.* **112**, 41–51 (2001).
3. Kurtzman, D., Shapira, R. H., Bar-tal, A., Fine, P. & Russo, D. Nitrate fl uxes to groundwater under citrus orchards in a Mediterranean climate: Observations, calibrated models, simulations and agro-hydrological conclusions. *J. Contam. Hydrol.* **151**, 93–104 (2013).
4. Turkeltaub, T., Kurtzman, D. & Dahan, O. Real-time monitoring of nitrate transport in the deep vadose zone under a crop field-implications for groundwater protection. *Hydrol. Earth Syst. Sci.* **20**, 3099–3108 (2016).
5. Weissman, G. et al. Increased irrigation water salinity enhances nitrate transport to deep unsaturated soil. *Vadose Zo. J.* **19**, 1–16 (2020).
6. Driscoll, C. T. et al. Nitrogen pollution in the northeastern United States: sources, effects, and management options. *Bioscience* **53**, 357–374 (2003).
7. Carey, R. O. & Migliaccio, K. W. Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems. *Environ. Manag.* **44**, 205–217 (2009).
8. Baram, S., Kurtzman, D., Ronen, Z., Peeters, A. & Dahan, O. Assessing the impact of dairy waste lagoons on groundwater quality using a spatial analysis of vadose zone and groundwater information in a coastal phreatic aquifer. *J. Environ. Manag.* **132**, 135–144 (2014).
9. Aharoni, I., Siebner, H., Yogev, U. & Dahan, O. Holistic approach for evaluation of landfill leachate pollution potential – From the waste to the aquifer. *Sci. Total Environ.* **741**, 140367 (2020).
10. The European Parliament and the Council of the European Union. Directive (EU) 2020/2184 of the European Parliament and of the Council. *Off. J. Eur. Union* **2019**, 1–62 (2020).
11. Wang, X. L., Lu, Y. L., Han, J. Y., He, G. Z. & Wang, T. Y. Identification of anthropogenic influences on water quality of rivers in Taihu watershed. *J. Environ. Sci.* **19**, 475–481 (2007).
12. Singh, B. & Craswell, E. Fertilizers and nitrate pollution of surface and ground water: an increasingly pervasive global problem. *SN Appl. Sci.* **3**, 1–24 (2021).
13. Ding, J. et al. Identifying diffused nitrate sources in a stream in an agricultural field using a dual isotopic approach. *Sci. Total Environ.* **484**, 10–18 (2014).
14. Howarth, R. W. Coastal nitrogen pollution: a review of sources and trends globally and regionally. *Harmful Algae* **8**, 14–20 (2008).
15. Rabalais, N. N. et al. Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences* **7**, 585–619 (2010).
16. UNEP and WHRC. Reactive Nitrogen in the Environment: Too Much or too Little of a Good Thing. *United Nations Environ. Program.* 1–56 (2007).
17. Beman, J. M., Arrigo, K. R. & Matson, P. A. Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* **434**, 211–214 (2005).
18. Masclaux-Daubresse, C. et al. Nitrogen uptake, assimilation and remobilization in plants: Challenges for sustainable and productive agriculture. *Ann. Bot.* **105**, 1141–1157 (2010).
19. Wen, Q.-X. & Cheng, L.-L. Fixation and release of ammonium. in *Nitrogen in Soils of China.* 67–86 (Springer Dordrecht, 1997).
20. Nieder, R., Benbi, D. K. & Scherer, H. W. Fixation and defixation of ammonium in soils: a review. *Biol. Fertil. Soils* **47**, 1–14 (2011).
21. Li, J., Zhang, J. & Ren, L. Water and nitrogen distribution as affected by fertigation of ammonium nitrate from a point source. *Irrig. Sci.* **22**, 19–30 (2003).
22. Baethgen, W. E. & Alley, M. M. Nonexchangeable ammonium nitrogen contribution to plant available nitrogen. *Soil Sci. Soc. Am. J.* **51**, 110–115 (1987).
23. Briones, A. M. Jr et al. Ammonia-oxidizing bacteria on root biofilms and their possible contribution to N use efficiency of different rice cultivars. *Plant Soil* **250**, 335–348 (2003).
24. Bryla, D. R. & Machado, R. M. A. Comparative effects of nitrogen fertigation and granular fertilizer application on growth and availability of soil nitrogen during establishment of highbush blueberry. *Front. Plant Sci.* **2**, 46 (2011).
25. Kabala, C., Karczewska, A., Gałka, B., Cuske, M. & Sowiński, J. Seasonal dynamics of nitrate and ammonium ion concentrations in soil solutions collected using MacroRhizon suction cups. *Environ. Monit. Assess.* **189**, 304 (2017).
26. Wells, N. S., Hakoun, V., Brouyère, S. & Knöller, K. Multi-species measurements of nitrogen isotopic composition reveal the spatial constraints and biological drivers of ammonium attenuation across a highly contaminated groundwater system. *Water Res.* **98**, 363–375 (2016).
27. Du, Y., Ma, T., Deng, Y., Shen, S. & Lu, Z. Sources and fate of high levels of ammonium in surface water and shallow groundwater of the Jianghan Plain, Central China. *Environ. Sci. Process. Impacts* **19**, 161–172 (2017).
28. Jiao, J. J. et al. Abnormally high ammonium of natural origin in a coastal aquifer-aquitard system in the pearl river delta, China. *Environ. Sci. Technol.* **44**, 7470–7475 (2010).
29. Buss, S. R., Herbert, A. W., Morgan, P., Thornton, S. F. & Smith, J. W. N. A review of ammonium attenuation in soil and groundwater. *Q. J. Eng. Geol. Hydrogeol.* **37**, 347–359 (2004).
30. Aharoni, I., Siebner, H. & Dahan, O. Application of vadose-zone monitoring system for real-time characterization of leachate percolation in and under a municipal landfill. *Waste Manag.* **67**, 203–213 (2017).
31. Ham, J. M. Seepage losses from animal waste lagoons: a summary of a four-year investigation in kansas. *Am. Soc. Agric. Eng.* **45**, 983–992 (2002).
32. Baram, S., Arnon, S., Ronen, Z., Kurtzman, D. & Dahan, O. Infiltration mechanism controls nitrification and denitrification processes under dairy waste lagoon. *J. Environ. Qual.* **41**, 1623–1632 (2012).
33. Mienis, O. & Arye, G. Long-term nitrogen behavior under treated wastewater infiltration basins in a soil-aquifer treatment (SAT) system. *Water Res.* **134**, 192–199 (2018).
34. Dahan, O. Vadose zone monitoring as a key to groundwater protection. *Front. Water* **2**, 1–9 (2020).
35. 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).
36. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M. & Wetterlind, J. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* **107**, 163–215 (2010).
37. Tuli, A., Wei, J.-B., Shaw, B. D. & Hopmans, J. W. In situ monitoring of soil solution nitrate: proof of concept. *Soil Sci. Soc. Am. J.* **73**, 501–509 (2009).
38. Yeshno, E., Arnon, S. & Dahan, O. Real-time monitoring of nitrate in soils as a key for optimization of agricultural productivity and prevention of groundwater pollution. *Hydrol. Earth Syst. Sci.* **23**, 3997–4010 (2019).
39. Jensen, P. S. & Bak, J. Near-infrared transmission spectroscopy of aqueous solutions: Influence of optical pathlength on signal-to-noise ratio. *Appl. Spectrosc.* **56**, 1600–1606 (2002).

40. Büning-Pfaue, H. Analysis of water in food by near infrared spectroscopy. *Food Chem.* **82**, 107–115 (2003).

41. Hall, J. W. et al. Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation. *Appl. Spectrosc.* **50**, 102–108 (1996).

42. McShane, M. J. & Coté, G. L. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. *Appl. Spectrosc.* **52**, 1073–1078 (1998).

43. Arnolda, S. A. et al. At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. *Enzym. Microb. Technol.* **27**, 691–697 (2000).

44. Cervera, A. E., Petersen, N., Lantz, A. E., Larsen, A. & Gernaey, K. V. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. *Biotechnol. Prog.* **25**, 1561–1581 (2009).

45. Petersen, N. et al. In situ near infrared spectroscopy for analyte-specific monitoring of glucose and ammonium in Streptomyces coelicolor fermentations. *Biotechnol. Prog.* **26**, 263–271 (2010).

46. Fastelli, M., Comodi, P., Maturilli, A. & Zucchini, A. Reflectance spectroscopy of ammonium salts: implications for planetary surface composition. *Minerals* **10**, 1–22 (2020).

47. Allegrini, F. & Olivieri, A. C. IUPAC-consistent approach to the limit of detection in partial least-squares calibration. *Anal. Chem.* **86**, 7858–7866 (2014).

48. Sato, S., Morgan, K. T., Ozores-Hampton, M. & Simonne, E. H. Spatial and temporal distributions in sandy soils with seepage irrigation: l. Ammonium and nitrate. *Soil Sci. Soc. Am. J.* **73**, 1044–1052 (2009).

49. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).

50. P. Leone, A., A. Viscarra-Rossel, R., Amenta, P. & Buondonno, A. Prediction of soil properties with PLSR and vis-NIR spectroscopy: application to Mediterranean Soils from Southern Italy. *Curr. Anal. Chem.* **8**, 283–299 (2012).

51. Dunn, K. G. Latent variable modelling. In *Process Improvement using Data* 315–391 (2020).

52. Ng, W., Minasny, B., de Sousa Mendes, W. & Melo Demattê, J. A. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *Soil* **6**, 565–578 (2020).

AUTHOR CONTRIBUTIONS

Conceptualization – R.Y.; Experimental setup – R.Y., E.Y., O.D.; Analysis – R.Y., S.A., I.V.-F.; Writing – R.Y., S.A., O.D.; Project supervision – S.A., O.D.

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/s41545-023-00243-z.

Correspondence and requests for materials should be addressed to Rotem Yupiter.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.