OSF DOI: 10.31219/osf.io/8pcen

Reference: Hanssens, D., Van De Vijver, E., Waegeman, W., Everett, M.E., Moffat, I., Sarris, A., and De Smedt, P. 2021, Ambient temperature and relative humidity-based drift correction in frequency domain electromagnetics using machine learning, *Near Surface Geophysics*, 19: 541-556.

Title: Ambient temperature and relative humidity-based drift correction in frequency domain electromagnetics using machine learning

Authors: Daan Hanssens, Ellen Van De Vijver, Willem Waegeman, Mark E. Everett, Ian Moffat, Apostolos Sarris and Philippe De Smedt

Electromagnetic instrument responses suffer from signal drift that results in a variable response at a given location over time. If left uncorrected, spatiotemporal aliasing can manifest and global trends or abrupt changes might be observed in the data, which are independent of subsurface electromagnetic variations. By performing static ground measurements, we characterized drift patterns of different electromagnetic instruments. Next, we performed static measurements at an elevated height, approximately 4 metre above ground level, to collect a data set that forms the basis of a new absolute calibration methodology. By additionally logging ambient temperature variations, battery voltage and relative humidity, a relation between signal drift and these parameters was modelled using a machine learning (ML) approach. The results show that it was possible to mitigate the effects of signal drift; however, it was not possible to completely eliminate them. The reason is three-fold: (1) the ML algorithm is not yet sufficiently adapted for accurate prediction; (2) signal instability is not explained sufficiently by ambient temperature, relative humidity and battery voltage; and (3) the black-box internal (factory) calibration impeded direct access to raw data, which prevents accurate evaluation of the proposed methodology. However, the results suggest that these challenges are not insurmountable and that ML can form a viable approach in tackling the drift problem instrument specific in the near future.

Introduction

Frequency domain electromagnetic (FDEM) methods are commonly used to map the spatial distribution of the electrical conductivity and magnetic susceptibility of the subsurface (see Everett, 2012, and Pellerin, 2002 for an overview of applications). The phase-sensitive instrument responses – the quadrature phase and in-phase components of the FDEM signal – can be related to these physical properties (Keller and Frischknecht, 1996). In addition to responding to the physical properties of the subsurface, the phase-sensitive instrument responses are also impacted by two types of errors (Minsley *et al.*, 2012), which can be additive or multiplicative (Green and Lane, 2003) in nature:

1. Random or incoherent errors (Fig. 1) are that part of the signal that do not show any temporal nor spatial correlation. These types of errors are typically denoted as (random) noise and originate from cultural/anthropic or instrumental noise (Minsley *et al.*, 2012). A Gaussian distribution is often used to describe normally distributed noise (where white noise describes a zero-mean Gaussian distribution), a Cauchy distribution can express more heavy-tailed noise distributions and a mixture can be used to express other asymmetrical and/or heavy-tailed noise distributions in geophysics (Kitagawa *et al.*, 2004).

2. Systematic or coherent errors (Fig. 1) are that part of the signal that correlate spatially and/or temporally. Possible sources of systematic errors can be placed under four categories (Kumar and Ahmed, 2011): ambient induced errors; electromagnetic wave propagation related errors; data acquisition errors; and data processing artefacts.

Figure 1 Schematic example of a measured FDEM signal decomposition into true FDEM signal, systematic and random errors through time. Here, all presented errors are additive, whereas no multiplicative errors affect the measured signal. The true signal varies through time and suggests a movement.

There are numerous names and adjectives given to noise, which makes it impossible to formulate a definition to which everyone would agree, even within the domain of geophysics. Fortunately, such an agreement is not necessary if the readership knows what the authors have in mind. Here, we define noise as that part of the signal that does not carry any useful information (e.g., random errors) and cannot be explained or captured by any model (see, Scales and Snieder, 1998, for a more in-depth view on geophysical noise). Characteristics of (in this case, white) noise, however, can be described using a stochastic distribution, typically provided by the manufacturer of the FDEM instrumentation, or estimated using experimental static measurements (Gebbers *et al.*, 2009). For FDEM measurements, noise reduction techniques might be necessary to attenuate noise, especially when deploying the instrument in low signal-to-noise (S/N) conditions.

Note that it is near impossible to distinguish noise from more systematic errors such as signal drift (Al-Maskari *et al.*, 2014). However, only by accurately defining and attributing a part of the signal to the former can the latter be assessed.

FDEM instrument responses suffer from signal drift that results in a variable response at a given location over time. If left uncorrected, spatiotemporal aliasing can manifest, and global trends or abrupt changes can be observed in the data, which are independent of subsurface electromagnetic variations (Delefortrie *et al.*, 2014). Analogous to signal noise, signal drift becomes less significant when the S/N ratio is high (Abdu *et al.*, 2007). However, even then signal drift might persist given the accuracy of modern FDEM instrumentation (Delefortrie *et al.*, 2014). Several causes for signal drift have been mentioned, including hardware instability, ambient temperature variations, circuit design, exact placement of temperature compensation sensors and coil performance under heating (Sudduth *et al.*, 2001; Robinson *et al.*, 2004; Mester *et al.*, 2014). Given the influence of ambient temperature changes and solar radiation warming up on signal drift, instrument insulating has been proposed to mitigate temperature effects (Huang *et al.*, 2017). In addition, commercially available FDEM instrumentation generally incorporate internal calibration protocols. However, signal drift remains possible despite instrument insulation, calibration protocols or technical advances (Delefortrie *et al.*, 2014) and should be accounted for to ensure accurate FDEM data. Moreover, the black-box nature of the internal calibration protocols prevents direct access to the raw output data, the straightforward assessment of signal drift and the development of adaptive correction possibilities.

Signal drift can be broken down into two parts:

- Signal offset: a first additive component that is constant through time; and
- Signal variation: a second additive component that shows structured variation through time (e.g. signal drift correlated to ambient climatological conditions). Note that there might be a multiplicative component to signal drift, which shows variation that scales based on instrument response amplitude (Green and Lane, 2003). However, as we did not observe such variation in our study, we did not go into detail on this component.

Different calibration methods to tackle signal drift have been presented in the past and can be categorized as either relative or absolute calibration methods. A relative calibration corrects signal variation, yet shifts the data towards an arbitrarily chosen signal offset reference as opposed to an (absolute) zero or global reference. As a result, the relatively corrected FDEM data only allow qualitative interpretation. An absolute calibration, on the other hand, corrects both signal offset and variation, which allows quantitative interpretation of FDEM data and derived subsurface models.

Different relative correction procedures tackling baseline (Simpson *et al.*, 2009; Delefortrie *et al.*, 2014) and remnant drift (De Smedt *et al.*, 2016) have been developed. These methods principally make use of a tie line or calibration line that crosses the survey lines during the survey. The method assumes no signal drift during tie line collection. Relative correction procedures often require an additional signal offset correction (i.e., absolute correction), impeding direct quantitative interpretation. The advantage, on the other hand, is that these methods are time efficient and inexpensive.

Different absolute calibration procedures have been developed as well. Most of these procedures require additional geophysical data such as electrical resistivity tomography (Lavoué *et al.*, 2010; Minsley *et al.*, 2012; von Hebel *et al.*, 2019), signature responses of a metal sphere (Thiesson *et al.*, 2014), multi-elevation data (Tan *et al.*, 2019; Bobe and Van De Vijver, 2019), cone penetration resistance electrical conductivity (Delefortrie *et al.*, 2019), or borehole magnetic susceptibility (Delefortrie *et al.*, 2018) data. However, such ancillary data are often not readily available as these are both time-consuming and expensive to acquire. Another method of (absolute) calibration is to increase the height of the FDEM instrument substantially above ground level (Hossain *et al.*, 2010; Grellier *et al.*, 2013). Although this results in an increased depth of investigation (Huang, 2005), the S/N ratio is degraded significantly due to the relatively larger contribution of air to the measured signal. As a result, the influence of the subsurface on the measured signal becomes negligible. However, this method requires elevation of the FDEM instrument during survey, which is highly impractical and time-consuming. In addition, all these methods require additional forward modelling, inversions and/or (subjective) prior knowledge in the form of (model) assumptions.

The objective is two-fold: (1) signal drift characterization of different FDEM instruments and (2) signal drift prediction for one (DUALEM-1S) FDEM instrument. (1) By performing static ground measurements, we first characterized the drift patterns of different FDEM instruments as a function of ambient climatological parameters. (2) Next, we performed static measurements at an elevated height, approximately 4 m above ground level, to minimize the influence of subsurface properties (assuming the induced subsurface response is negligible) and to collect a data set that forms the basis of a new absolute calibration methodology. A noise reduction technique was applied to account for the influence of (random) noise such that all the variations that were measured could be attributed to signal drift. By additionally logging ambient temperature variations, battery voltage and relative humidity, a relation between signal drift and these parameters was established. By taking these parameters into account in a direct way, additional assumptions become obsolete, preventing the introduction of additional subjectivity (e.g., data inversion) in the methodology, making it more robust. A machine learning approach was used, as it does not require complex physical insight in the processes at hand and can be tailored to a specific instrument and incorporate its specific sensitivity – related to its internal design – to signal drift. Such a correction can be considered as a(n) (near-real-time) absolute calibration that can be performed in a highly time-efficient manner.

Static Ground Measurements

Data collection

Static ground measurements were made with four different types of frequency domain electromagnetic (FDEM) instrumentation (Table 1). The FDEM instruments were placed on the ground surface, away from potential sources of EM noise and metallic objects. The data and temperature logger were positioned at a separation distance larger than double the coil spacing to avoid influence on the measured signal. Two types of temperature loggers were used, recording at a sampling rate of 0.1 Hz:

1. a BSIDE BHT01 temperature and relative humidity logger (BSIDE, China); and
2. a Lascar EL-GFX-1 temperature logger (Lascar Electronics Ltd., United Kingdom).

Data processing

The temperature data were interpolated using the nearestneighbour technique to match the sampling frequency (see, Table 1) of the FDEM measurement. A second-order Butterworth tapered filter (Butterworth, 1930) was applied to the FDEM ground measurements to mitigate the effects of instrumental noise (see Table 1 for the relative cutoff frequency that was used). Following, the mean was subtracted from the signal to remove possible response offsets and facilitate time-series comparison.

Table 1 FDEM instruments used during static measurements, their specifications and measurement parameters. The grey background indicates the instrument that was used for static height measurements rather than static ground measurements. Both temperature loggers had a sampling rate of 0.1 Hz. Static ground measurements of the DUALEM-1 would lead to a similar discussion as for the static height measurements; hence, only the latter is presented

	Frequency	Tx-Rx spacing	Tx-Rx orientation Units	Sampling frequency	Relative cutoff frequency		
Instrument	Hz	m	-	Hz	Hz	Temp logger	Date
CMD-MiniExplorer	30,000	0.32 0.71 1.18	HCP	10	0.008	BHT01	28 October 2015
GEM-2	4950 10,230 21,030 43,350 89,430	1.66	HCP	2	0.01	BHT01	9 March 2016
Profiler EMP-400	1000 8000 15,000	1.21	HCP	3	0.05	EL-GFX-1	6 June 2017
EM31-MK2	9800	3.66	HCP	1	0.05	BHT01	16 April 2016
DUALEM-1S	9000	1 1.1	HCP PRP	8	0.001	BHT01	22 May 2019

Static Height Measurement Data collection

A static height measurement experiment was carried out using a DUALEM-1S instrument (Table 1). In addition to the instrument responses, internal temperature and battery voltage instrument output were logged using the standard instrument output. The instrument was mounted on a wooden frame that was constructed without any metal parts (wooden planks were attached using zip ties), approximately 4 m above ground level (Fig. 2), to avoid subsurface interference. Wooden sticks were used to fix the position of the frame onto the surface to avoid further noise generation through instrument motion. The temperature logger was attached to the wooden frame at a distance of approximately 3.5 m underneath the frequency domain electromagnetic (FDEM) instrument. The apparatus was set up in a shade-free area away from any buildings or other sources of noise and facing the south to avoid the influence of shadow.

Figure 2 Schematic sketch of wooden frame with mounted FDEM instrument and temperature logger. The battery and data logger for the FDEM instrument is illustrated alongside and should be moved as far as possible from the static set-up.

Lagged features construction

Internal and external temperature, battery voltage and relative humidity were selected as features x , while the FDEM instrument responses were selected as targets y for the machine learning algorithm (detailed below). All features were interpolated, when necessary, via the nearest-neighbour technique to match the sampling frequency of the FDEM measurement and mitigate the effect of non synchronous measurements between both sensors. Following, a noise reduction on the targets y was applied by using a second-order Butterworth tapered filter (Butterworth, 1930) with a relative cutoff frequency of 0.001 Hz. A data set x was constructed by systematic subsampling of the features using a 12.5-s interval to reduce data volume and by adding 30 prior time steps or lag observations per feature to each target (determined via trial-and-error). This approach accounted for a total lag between temperature measurement and its potential effect on FDEM response instability of 375 seconds.

Signal instability prediction

The first 80% (or 5+ hours) of the complete data set Ex were used as training data set $\mathbf{x}_{\text{train}}$, while the remaining data acted as test data. Following, a Ridge regression model f (Bishop, 2006) was used to fit $\mathbf{x}_{\text{train}}$ to y_{train} :

$$y_{\text{train}} = f(\mathbf{x}_{\text{train}}), \quad (1)$$

where y_{train} is the measured FDEM instrument response of the training data set. The implementation of the model f in scikit-learn (Pedregosa *et al.*, 2011) was used with regularization parameter α equal to 9.5, determined after cross-validation on the training data set. Following, all FDEM responses $\hat{y}$ were predicted for the whole time series.

$$\hat{y} = f(\mathbf{x}). \quad (2)$$

The mean squared error (MSE), root mean squared error (RMSE) and R^2 were used to evaluate the prediction error on the train and test data sets and were defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (3)$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (5)$$

where y_i is the i th observation, $\hat{y}_i$ is the i th prediction, $\bar{y}$ is the mean of the observations and N is the total number of observations. Ultimately, the RMSE was calculated for different subsets of features in the data set $\mathbf{x}$, with 100 permutations per feature and its associated decrease in RMSE. As a result, feature importance on the prediction can be determined.

Figure 3 Static ground measurements of the CMD-MiniExplorer, with (a) the external temperature ($^{\circ}\text{C}$), (b) relative humidity (%), (c) 0.32 m HCP, (d) 0.71 m HCP and (e) 1.18 m HCP (quadrature) QP and (in-phase) IP signal combinations. Additional instrument specifications can be found in Table 1.

Results and Discussion

Signal instability patterns

The logged data of the static ground measurements of the CMD-Mini-Explorer are shown in Figure 3 and their associated boxplots in Figure 4. The external temperature varies roughly between 20°C and 30°C and is inversely correlated with relative humidity. This can be explained by the increased ability of air to hold moisture when the temperature rises (absolute humidity), which results in a decrease in relative humidity (i.e., inverse correlation to temperature). It might be possible that an increased relative humidity has an influence of the quadrature phase (QP) signal due to an increased EC value of the air induced by the presence

of moisture. However, this effect is minimal and probably not detectable by frequency domain electromagnetic (FDEM) instrumentation under most conditions. The measurements confirm that the internal calibration of the FDEM instrument does not provide sufficient correction to output stable QP nor in-phase (IP) values and additional calibration is required. The following trends are apparent (Fig. 3):

- All responses are highly unstable and a second-order drift pattern can be distinguished; © 2021 European Association of Geoscientists & Engineers, *Near Surface Geophysics*, **19**, 541–556 *Drift correction in frequency domain electromagnetics* 547
- Erratic behaviour is observed as a sudden drop in QP, not related to ambient conditions, for both the HCP 0.71 m and 1.18 m between the fourth and fifth hours (indicated with a grey background in Fig. 3);
- A direct relationship between coil spacing and variation of drift can be observed for the QP response (Fig. 4). A similar trend seems present for the IP response; however, this trend is not monotonic (Fig. 4);
- A visible correlation between the FDEM responses and temperature can be distinguished, with a time lag between the occurring temperature and influence on the response. This is because heat requires a certain amount of time to ‘conduct’ through the instrument before having an observable effect on the response;
- The relative humidity seems to have an influence on the QP response in particular. However, this trend is less apparent and difficult to assess given the time delay of the effect and its direct correlation with temperature.

Figure 4 Boxplots of the static measurements of the CMD-MiniExplorer, GEM-2, Profiler EMP-400, EM31-MK2 and DUALEM-1S. The orange stripe represents the median of the data. The left and right columns represent the QP (ppt) and IP (ppt) measurements, respectively. Additional instrument specifications can be found next to the boxplots and in Table 1. Note that the IP signal was not recorded for the EM31-MK2 instrument.

The logged data of the static ground measurements of the GEM-2 are shown in Figure 5 and their associated boxplots in Figure 4. The external temperature varies roughly between 20°C and 40°C. Again, the internal calibration of the FDEM instrument does not provide sufficient correction to output stable QP nor IP values. Here, larger though similar signal fluctuations can be observed as for the CMD-Mini- Explorer. These larger fluctuations are induced by the combination of a higher operating frequency, instrument specifications and a wider external temperature range. A direct relationship between the variation of QP signal drift and operating frequency can be observed. This relationship is less obvious for the IP signal (i.e., a proportional increase with increasing operating frequency), although the highest signal variations are observed for the highest operating frequency (Fig. 4).

The signal variations of the Profiler EMP-400 (Fig. 6) are similar to the GEM-2. One important difference, however, is the particularly large influence of signal drift at instrument startup (i.e., during the very first minutes), which can be induced by changes in battery voltage, user proximity or a difference in start temperature between the external temperature and internal instrument temperature. A small rapid signal change can be observed during sensor shutdown as well, with a higher amplitude for higher frequencies.

The static ground measurements of the EM31-MK2 are shown in Figure 7 and the associated boxplots in Figure 4. The QP response is inversely related to the temperature and directly related to the relative humidity. Within a small time frame (i.e., approximately the first 18 minutes), a large signal drift can be observed. Again, the instrument requires an accurate correction to provide the user with reliable results.

Note that the static ground (as opposed to height) measurements are equally influenced by radiant heat of the soil or substrate. Moreover, clouds passing by, preventing direct sunlight for a certain timeframe, or collection of static data sets in shaded conditions might result in different signal variations. A general rule of thumb is that higher overall temperatures and larger temperature fluctuations will result in larger signal variations.

The static height measurements of the DUALEM-1S are shown in Figure 8 and their associated boxplots in Figure 4. The time series' variation pattern of both the external and internal temperature (and to a smaller degree the relative humidity) can be observed within the measured instrument responses, i.e. either in a direct or inverse relationship to each other. The battery voltage is complicated, as it drops immediately after instrument startup, stabilizes again and shows an exponentially decreasing trend afterwards. It is not possible to observe a relationship between the FDEM responses and the internal voltage in the logged time series. A small lag is present between both external and internal temperature, and FDEM responses and in/external temperature. Such a time lag can be expected, as heat requires a certain amount of time to conduct through the instrument's housing and influence the FDEM signal. Lagged observations, surpassing the lag between aforementioned the time series were added as features during construction of the lagged data set, to relate the measured temperature variations to their influence on signal drift that is only observed later in time. Similar temporal variations in the FDEM instrument response, as in the static ground measurements in Delefortrie *et al.* (2014), can be observed:

- The PRP QP response is rather stable (Fig. 4), showing no or limited linear drift;
- The PRP IP response is highly unstable (Fig. 4). A quadratic drift pattern can be distinguished, and erratic behaviour was observed for some discrete shorter time intervals;

- The HCP QP and IP responses show a quadratic drift pattern in time.

The relatively high stability of the QP signal as opposed to the IP signal (Fig. 8) is remarkable, suggesting that the factory internal correction strategy clearly has been designed focusing on the QP rather than IP signal. The effort involved in stabilizing the QP (as opposed to IP) response comes as no surprise as the majority of the end-users mainly use the QP-derived apparent electrical conductivity (ECa) assuming low induction number (LIN) conditions (see, McNeill, 1980).

This research demonstrates that, while manufacturers often claim that their FDEM instrument outputs stable signals, in practice this is mostly not the case. While efforts have been focused on outputting stable QP/ECa data, the IP signal remains largely unstable. Apart from the significant difference in signal stability for the QP and IP signals, the main instrumental factors that generally result in an increasing signal drift are (Fig. 4):

- higher and stronger variations in external temperature;
- higher operating frequencies;
- larger coil spacings.

Equally important is the influence of signal amplitude. While it is true that the absolute drift increases with larger coil spacings and higher operating frequencies, it is also true that the signal amplitude increases in these cases (analogous to S/N). Overall, this results in a higher signal to drift ratio – for additive drift – and therefore often a smaller relative influence of drift on the actual measured signal (see Fig. 9 for a boxplot in LIN ECa (mS/m) that supports these statements). As a result, drift errors can become particularly critical when signal amplitudes are low: especially under low S/N conditions,

Figure 5 Static ground measurements of the GEM-2, with (a) the external temperature (°C), (b) relative humidity (%), (c) 4950 Hz, (d) 10230 Hz, (e) 21030 Hz, (f) 43,350 and (g) 89,430 Hz HCP QP and IP signal combinations. Additional instrument specifications can be found in Table 1.

Figure 6 Static ground measurements of the Profiler EMP-400, with (a) the external temperature ($^{\circ}\text{C}$), (b) 1000 Hz, (c) 8000 Hz and (d) 15,000 Hz HCP QP and IP signal combinations. Note that the relative humidity was not logged here; additional instrument specifications can be found in Table 1.

Figure 7 Static ground measurements of the EM31-MK2, with (a) the external temperature (°C), (b) relative humidity (%) and (c) the HCP QP signal (ppt). Additional instrument specifications can be found in Table 1.

Signal instability prediction

While the FDEM instrument outputs an internally calibrated QP signal, an obvious drift pattern can still be distinguished in the HCP1 QP data (Fig. 8d). The internal calibration protocol appears to be insufficient to guarantee stable measurements independent of temperature fluctuations. The ML method is able to predict the measured signal instability trend for both the trainings and test data set (Fig. 10a). Despite our attempts at drift prediction, the ML method still introduces anomalous residuals (i.e., measurement – prediction) as it did not capture all the variability and is vulnerable to underprediction of sudden large signal variations. The HCP1 QP signal instability prediction is characterized by an RMSE_{test} value of 0.0026 ppt² (Table 2) for the test data set. Accounting for this drift prediction will effectively lower the influence of signal drift, yet will not entirely eliminate it.

The PRP1 QP signal appears to be either the least susceptible to signal instability effects, or the best internally calibrated (Fig. 10c). Although performing a noise filtering method, the largest part of the variation in PRP1 QP signal can be attributed to white noise, hence the combination of low RMSE and low R^2 (Table 2). As white noise is typically entirely random, it is not possible to predict in a structured way as a function of time. As a result, the prediction of the drift over time is approximately equal to the global mean of the recorded signal,

remediating signal instability relatively accurately, yet having a low R^2 . Here, signal instability correction is primarily a signal offset correction rather than a signal variation correction.

While it was possible to reduce the influence of signal drift for both the HCP1 and PRP1 IP signal (Fig. 10), it was not possible to correct it accurately. Both IP signals (and the PRP1 IP signal in particular, which suffers from the largest signal instability) are related to severe under- or overpredictions of the measured data, resulting in relatively high (R)MSE values (Table 2) and leaving large residuals (Fig. 10). Clearly, the logged features do not possess enough information to explain signal drift in the IP signal or the algorithm is not sufficiently adapted, given the high $RMSE_{train}$. As stated earlier, the IP signal is less accurately internally calibrated as opposed to the QP signal. A combination of both internal and additional calibration might lead to lower MSE values, conforming to QP signal predictions, but we cannot conclude this at this moment.

Figure 8 Static height measurements of the DUALEM-1S its (a) internal and external temperature, (b) relative humidity, (c) battery voltage, (d) HCP and (e) PRP, QP and IP signal combinations. Additional instrument specifications can be found in Table 1.

Feature permutation is the decrease in a prediction score when a single feature value is dropped. The drop in the prediction score is indicative of how much the prediction depends on that feature. Based on the feature permutation on the RMSE, the features that most effectively lower the RMSE are mainly internal and external temperature (Fig. 11). Relative humidity plays a relatively larger role in the prediction of QP instabilities as opposed to IP instabilities. This could be attributed to the influence of relative humidity on QP directly (although this effect is small and unlikely) and/or on the internal calibration protocol of the instrument. The influence

of battery voltage is not significant for any predictions. It may be possible that the relative importance of the different factors evaluated depends on ambient temperatures and other environmental conditions. However, this cannot be concluded at this point and requires further research.

Note that a different selection of training and test subsets could lead to other training and test prediction errors (Table*)

Figure 9 Boxplots of the LIN ECa (mS/m) of the static measurements of the CMD MiniExplorer, GEM-2, Profiler EMP-400, EM31-MK2 and DUALEM-1S. The orange line represents the median of the data. Additional instrument specifications can be found next to the boxplots and in Table 1.

Figure 10 Time series of (a) measured and predicted HCP1 QP signal (ppt) for the train and test data set and (b) associated residuals (ppt). The same order applies to the plots below but for the PRP1 QP (c–d), HCP1 IP (e–f) and PRP1 IP (g–h) configurations.

Figure 11 Effect of feature permutation on decrease in RMSE of the signal instability prediction for (a) HCP1 QP (ppt), (b) PRP1 QP (ppt), (c) HCP1 IP (ppt) and (d) PRP1 IP (ppt). The grey shade represents increasing lagged features, where the darkest bar represents the feature at t and the lightest bar at time stamp $t-30$.

Table 2 MSE (ppt²), RMSE (ppt) and R^2 (dimensionless) related to the train and test prediction of signal instability of each phase and orientation

Phase	Tx-Rx orientation	MSE		RMSE		R^2	
		Train	Test	Train	Test	Train	Test
QP	HCP	1.8e-5	6.6e-6	0.0042	0.0026	0.82	0.87
	PRP	3.4e-8	4.6e-8	0.0002	0.0002	0.58	0.60
IP	HCP	0.0001	0.0004	0.0139	0.0205	0.90	0.78
	PRP	0.0073	0.0063	0.0855	0.0793	0.90	0.81

2) or residuals as those in aforementioned figures. However, doing so would not lead to a structurally different outcome; i.e., selecting different training and test subsets would not alter the overall discussion nor conclusion. On the basis of the results presented above, one could argue that it is not possible for the ML algorithm to accurately predict signal drift, based on ambient temperature, relative humidity and battery voltage only, particularly for the IP signal. However, what does seem possible, and requires further research, is the combination of a relative calibration strategy (e.g. , Delefortrie *et al.*, 2014; De Smedt *et al.*, 2016) to tackle signal variation and prediction of the signal offset (correlated to ambient conditions), using the ML algorithm for the latter. As a result, the combination could lead to a nearly absolute correction, where the residuals are shifted towards the RMSE values (Table 2), yet no

additional (invasive) data need to be collected. The relative correction protocol would then be complementary to the method described here, and used to explain the variance that is not accurately captured by the ML algorithm, possible because it is not meaningfully related to the studied ambient parameters. It should be noted that for such a combined method to work, auto-calibration of the instrument on startup should not be performed and offset drift of the instrument over time should either be negligible or only vary slowly over time (such that only a new training data set needs to be collected).

A data-driven approach may be proven useful in the calibration of FDEM instrumentation as this can be tailored to a specific instrument, taking into account its technical specification. The most important requirement for a data-driven approach to work adequately is acquiring raw instrument output data and ancillary data that can be easily logged alongside and explain the variations in signal drift. Ideally, data are logged across multiple weeks, continuously and covering relevant parameters over a wide range of values. High performing recurrent neural networks, using long short-term memory or similar approaches can be used to train the model to fit the data (Hochreiter and Schmidhuber, 1997; Sherstinsky, 2018). While lagged observations were added manually to the data set here for the ML method, deep learning methods make their own features and relevant lags, effectively improving the signal instability prediction. Promising results were obtained with these types of deep learning (DL) architectures, implemented via Keras and TensorFlow (Abadi *et al.*, 2015). However, the data set in this study is not sufficient to robustly validate this approach.

Conclusion

Manufacturers often claim that their frequency domain electromagnetic (FDEM) instrument outputs stable signals, yet practical experiments often prove otherwise. The results of this study demonstrate that the in-phase signal remains particularly unstable and additional signal corrections are necessary to provide reliable FDEM data. By implementing a simple machine learning (ML) algorithm, signal drift was modelled based on (in order of significance) lagged internal and external temperature, relative humidity and battery voltage variations. While it was possible to reduce the effect of signal drift, it was not possible to completely eliminate it. The reason is three-fold:

1. The ML algorithm is not sufficiently adapted for accurate prediction. Other, more advanced DL algorithms and more data are necessary to further advance drift prediction;
2. Signal instability could not be explained completely by ambient temperature, relative humidity and battery voltage. Meaning that, unidentified features, which are or cannot be logged, have an additional significant influence on signal instability that is not accounted for in the here performed drift prediction;
3. The black-box calibration protocols that are incorporated into the manufacturers' instruments impede further exploitation of this methodology.

Logging relevant parameters and combining existing calibration strategies especially with ML (and possibly DL) algorithms that are instrument specific can help further improving the correction of signal drift. A combination with the ML approach (to correct signal offset) and a relative calibration method (to correct signal variation), on the other hand, could lead to a nearly absolute calibration strategy, if autocalibration of the instrument on startup is not performed and signal drift of the instrument over time is negligible. This suggests a new approach to tackle

the caration problem within FDEM as the data collection is simple and straightforward and could be repeated quite easily. However, it is required to span a wide range of different ambient conditions to provide a stable and robust training data set.

Acknowledgements

This research was conducted in the framework of the Ghent University BOF project ‘High resolution electromagnetic landscape mapping with multi-channel FDEM instruments’ (BOF/STA/202002/010). We would like to thank Valentijn Van Parys for his practical help and Marianne De Temmerman for granting access to her field. Willem Waegeman received funding from the Flemish Government under the ‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’ Programme. We also want to acknowledge the Australian Research Council grant #DE160100703.

Data Availability Statement

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

ORCID

Daan Hanssens <https://orcid.org/0000-0003-2633-5469>

References

- Abadi, M., Agarwal, A., Barham, A., Brevdo, E., Chen, Z., Citro, C., *et al.* (2015) *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. TensorFlow White Papers.
- Abdu, H., Robinson, D.A. and Jones, S.B. (2007) Comparing bulk soil electrical conductivity determination using the DUALEM-1S and EM38-DD electromagnetic induction instruments. *Soil Science Society of America Journal*, 71, 189.
- Al-Maskari, S., Li, X. and Liu, Q. (2014) *An Effective Approach to Handling Noise and Drift in Electronic Noses*. Cham: Springer International Publishing, pp. 223–230.
- Bishop, C. (2006) *Pattern Recognition and Machine Learning*, New York: Springer.
- Bobbe, C. and Van De Vijver, E. (2019) Offset errors in probabilistic inversion of small-loop frequency-domain electromagnetic data: A synthetic study on their influence on magnetic susceptibility estimation.
- GEM 2019 Xi'an: International Workshop and Gravity, Electrical & Magnetic Methods and Their Applications, Chenghu, China.
- Butterworth, S. (1930) On the theory of filter amplifiers. *Wireless Engineer*, 7, 536–541.
- De Smedt, P., Delefortrie, S. and Wyffels, F. (2016) Identifying and removing micro-drift in ground-based electromagnetic induction data. *Journal of Applied Geophysics*, 131, 14–22.

Delefortrie, S., De Smedt, P., Saey, T., Van De Vijver, E. and Van Meirvenne, M. (2014) An efficient calibration procedure for correction of drift in EMI survey data. *Journal of Applied Geophysics*, 110, 115–125.

Delefortrie, S., Hanssens, D. and De Smedt, P. (2018) Low signal-to-noise FDEM in-phase data: potential for magnetic susceptibility modelling. *Journal of Applied Geophysics*, 152, 17–25.

Delefortrie, S., Hanssens, D., Saey, T., Van De Vijver, E., Smetryns, M., Bobe, C. and De Smedt, P. (2019) Validating land-based FDEM data and derived conductivity maps: assessment of signal calibration, signal attenuation and the impact of heterogeneity. *Journal of Applied Geophysics*, 164, 197–190.

Everett, M.E. (2012) Theoretical developments in electromagnetic induction geophysics with selected applications in the near surface. *Surveys in Geophysics*, 33, 29–63.

Gebbers, R., Lück, E., Dabas, M. and Domsch, H. (2009) Comparison of instruments for geoelectrical soil mapping at the field scale. *Near Surface Geophysics*, 7, 179–190.

Green, A. and Lane, R. (2003) Estimating noise levels in AEM data. *ASEG Extended Abstracts*. Adelaide, Australia.

Grellier, S., Florsch, N., Camerlynck, C., Janeau, J.L., Podwojewski, P. and Lorentz, S. (2013) The use of Slingram EM38 data for topsoil and subsoil geoelectrical characterization with a Bayesian inversion. *Geoderma*, 200–201, 140–155.

Hochreiter, S. and Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, 9, 1735–1780.

Hossain, M.B., Lamb, D.W., Lockwood, P.V. and Frazier, P. (2010) EM38 for volumetric soil water content estimation in the root-zone of deep vertisol soils. *Computers and Electronics in Agriculture*, 74, 100–109.

Huang, H. (2005) Depth of investigation for small broadband electromagnetic sensors. *Geophysics*, 70, 135–142.

Huang, J., Minasny, B., Whelan, B.M., Mcbratney, A.B. and Triantafyllis, J. (2017) Temperature-dependent hysteresis effects on EM induction instruments: an example of single-frequency multi-coil array instruments. *Computers and Electronics in Agriculture*, 132, 76–85.

Keller, G.V. and Frischknecht, F.C. (1996) *Electrical Methods in Geophysical Prospecting*, London: Pergamon Press.

Kitagawa, G., Takanami, T. and Matsumoto, N. (2004) State space approach to signal extraction problems in seismology. In: Brillinger, D.R., Robinson, E.A. and Schoenberg, F. P. (Eds.) *Time Series Analysis and Applications to Geophysical Systems*. New York: Springer. Kumar, D. and Ahmed, I. (2011) Seismic noise. In: Gupta, H.K. (Ed.) *Encyclopedia of Solid Earth Geophysics*. Dordrecht: Springer.

Lavoué, F., Van Der Kruk, J., Rings, J., André, F., Moghadas, D., Huisman, J.A., *et al.* (2010) Electromagnetic induction calibration using apparent electrical conductivity modelling based on electrical resistivity tomography. *Near Surface Geophysics*, 8, 553–561.

McNeill, J.D. (1980) Electromagnetic terrain conductivity measurement at low induction numbers. Geonics Technical Note TN-6.

Mester, A., Zimmermann, E., Van Der Kruk, J., Vereecken, H. and Van Waasen, S. (2014) Development and drift-analysis of a modular electromagnetic induction system for shallow ground conductivity measurements. *Measurement Science and Technology*, 25, 055801.

Minsley, B.J., Smith, B.D., Hammack, R., Sams, J.I. and Veloski, G. (2012) Calibration and filtering strategies for frequency domain electromagnetic data. *Journal of Applied Geophysics*, 80, 56–66.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., *et al.* (2011) Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pellerin, L. (2002) Applications of electrical and electromagnetic methods for environmental and geotechnical investigations. *Surveys in Geophysics*, 23, 101–132.

Robinson, D.A., Lebron, I., Lesch, S.M. and Shouse, P. (2004) Minimizing drift in electrical conductivity measurements in high temperature environments using the EM-38. *Soil Science Society of America Journal*, 68, 339. Scales, J.A. and Snieder, R. (1998) What is noise? *Geophysics*, 63, 1122–1124.

Sherstinsky, A. (2018) Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. Elsevier.

Simpson, D., Van Meirvenne, M., Saey, T., Vermeersch, H., Bourgeois, J., Lehouck, A., *et al.* (2009) Evaluating the multiple coil configurations of the EM38DD and DUALEM-21S sensors to detect archaeological anomalies. *Archaeological Prospection*, 16, 91–102.

Sudduth, K.A., Drummond, S.T. and Kitchen, N.R. (2001) Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. *Computers and Electronics in Agriculture*, 31, 239–264.

Tan, X., Mester, A., Von Hebel, C., Zimmermann, E., Vereecken, H., Van Waasen, S. and Van Der Kruk, J. (2019) Simultaneous calibration and inversion algorithm for multiconfiguration electromagnetic induction data acquired at multiple elevations. *Geophysics*, 84, EN1–EN14.

Thiesson, J., Kessouri, P., Schamper, C. and Tabbagh, A. (2014) Calibration of frequency-domain electromagnetic devices used in nearsurface surveying. *Near Surface Geophysics*, 12, 481–491.

Von Hebel, C., Van Der Kruk, J., Huisman, J.A., Mester, A., Altdorff, D., Endres, A.L., *et al.* (2019) Calibration, Conversion, and quantitative multi-layer inversion of multi-coil rigid-boom electromagnetic induction data. *Sensors*, 19, 4753.