Probabilistic Estimation of InSAR Displacement Phase Guided by Contextual Information and Artificial Intelligence

Philip Conroy , Simon A. N. van Diepen , Sanneke van Asselen , Gilles Erkens ,
Freek J. van Leijen , *Member, IEEE*, and Ramon F. Hanssen , *Senior Member, IEEE*

Abstract— Phase unwrapping, also known as ambiguity resolution, is an underdetermined problem in which assumptions must be made to obtain a result in SAR interferometry (InSAR) time series analysis. This problem is particularly acute for distributed scatterer InSAR, in which noise levels can be so large that they are comparable in magnitude to the signal of investigation. In addition, deformation rates can be highly nonlinear and orders of magnitude larger than neighboring point scatterers, which may be part of a more stable object. The combination of these factors has often proven too challenging for the conventional InSAR processing methods to successfully monitor these regions. We present a methodology which allows for additional environmental information to be integrated into the phase unwrapping procedure, thereby alleviating the problems described above. We show how problematic epochs that cause errors in the temporal phase unwrapping process can be anticipated by the machine learning algorithms which can create categorical predictions about the relative ambiguity level based on the readily available meteorological data. These predictions significantly assist in the interpretation of large changes in the wrapped interferometric phase and enable the monitoring of environments not previously possible using standard minimum gradient phase unwrapping techniques.

Index Terms— Artificial intelligence, peatland, phase unwrapping, recurrent neural network (RNN), SAR interferometry (InSAR), subsidence.

I. INTRODUCTION

ACTIVELY monitoring ground motion is of paramount importance in The Netherlands, a country in which many of its regions lie below sea level. There is a strong link between soil height and phreatic groundwater level, and indeed

Philip Conroy, Simon A. N. van Diepen, Freek J. van Leijen, and Ramon F. Hanssen are with the Department of Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands (e-mail: p.n.conroy@tudelft.nl; s.a.n.vandiepen@tudelft.nl; f.j.vanleijen@tudelft.nl; r.f.hanssen@tudelft.nl).

Sanneke van Asselen is with the Deltares Research Institute, 3584 BK Utrecht, The Netherlands (e-mail: sanneke.vanasselen@deltares.nl).

Gilles Erkens is with the Deltares Research Institute, 3584 BK Utrecht, The Netherlands, and also with the Faculty of Geosciences, Utrecht University, 3584 CB Utrecht, The Netherlands (e-mail: gilles.erkens@deltares.nl).

Digital Object Identifier 10.1109/TGRS.2022.3203872

the Dutch have actively managed the water tables in the country for centuries using a system of dams and canals through which excess water can be pumped away. Thus, better knowledge of subsidence processes is needed for flood protection, building damage risk assessments, and understanding greenhouse gas emissions caused by peat oxidation [1], [2]. The Dutch government has committed to a 1-Mt ($\sim$25%) reduction in CO_2 equivalents per year originating from its peatlands before 2030, a decision with significant impact on future land use, agriculture, and the economy, which requires adequate assessment of subsidence rates [3].

Yet, there is currently a large gap in our monitoring capabilities of these low-lying peatland regions. While SAR interferometry (InSAR) techniques using point scatterers (PS) have been successfully used to monitor subsidence in The Netherlands [4], [5], [6], these PS points are usually founded at greater depths and the movement of the surrounding landscape has had to be indirectly inferred. Direct monitoring of the land has proven to be much more challenging. The soft soils which comprise the majority of the country's agricultural lands are prone to rapid deformation and are very difficult or even impossible to directly monitor using standard distributed scatterer (DS) InSAR techniques [7], [8] which use smoothness constraints to perform spatial phase unwrapping [9], [10], [11]. Recently, a significant amount of research has been conducted to improve 2-D phase unwrapping [12], and there is great interest in the applicability of the machine learning techniques in solving the problem [13]; however, this research has largely been limited to the spatial case and does not consider the temporal effects of a rapidly deforming region. Rapid soil motion and nonstationary coherence cause large fluctuations affecting both the functional and stochastic models, which rules out the use of time-domain phase unwrapping techniques such as integer least-squares (ILS) [14], as the stochastic model will become too large reliably to evaluate an assumed deformation model.

Using newly available ground-truth measurements, we find that in the case of Sentinel-1 observations, rapid soil uplift due to increases in the ground water level between satellite overpasses can cause phase displacements larger than half a cycle ($\lambda/4$) at the C-band. This renders time series analysis of the region using Sentinel-1 imagery impossible using standard techniques, because the phase unwrapping algorithms

will typically assume the smaller phase displacement in the opposite direction to be the correct solution. This introduces a systematic error in all subsequent points of the time series corresponding to one 2π ambiguity level. While this problem may potentially be avoided using the L-band SAR data, there is currently no operational L-band radar mission observing The Netherlands at the temporal frequencies required for monitoring this environment.

It is therefore necessary to augment the standard unwrapping procedure with additional information to prevent errors in the unwrapping direction. This is done by considering the *direction of ground motion* (DOGM), and therefore the correct phase unwrapping direction, as stated in a hidden Markov model (HMM). Using a modification of the widely used Viterbi algorithm [15], we can integrate additional contextual information about the system into a generalized probabilistic framework which can be used to guide the unwrapping of the interferometric phase in the time domain. In our case, this additional information comes from a recurrent neural network (RNN) which takes meteorological data as its input to predict the DOGM. This was chosen because meteorological data are easily accessible and interpretable for our areas of study, but in general, any model which predicts the state of ground motion can be used. We show that this framework is able to reliably anticipate rapid soil deformation events and correctly unwrap the interferometric phases.

The remaining sections of the article are organized as follows: Section II presents the results of the extensometer measurements taken in two regions of the Dutch polderlands. These data are used to simulate a InSAR phase time series which is used to demonstrate that unwrapping errors are inevitable in this region when using standard approaches. Section III motivates the use of additional contextual information which can be used to aid in this problem. An RNN is used to process sequences of meteorological data and estimate the DOGM to aid the unwrapping process. Section IV describes how the outputs of RNN are used in tandem with the wrapped phases to perform phase unwrapping. Section V shows the results of this RNN-aided phase unwrapping versus a standard minimum gradient algorithm. A general discussion is presented in Section VI, and Section VII concludes this article.

II. EXTENSOMETER DATA AND SIGNAL SIMULATION

A. Extensometer Data

Extensometers placed in various locations across the Netherlands have been continuously monitoring the vertical movement of the surface and shallow subsurface of peaty soft soils for several years [16]. The longest running of these stations is in Rouveen, which has been collecting data since October 2018 and exhibits relatively modest levels of movement, within a range of 4 cm; see Fig. 1(a). On the other hand, some of the largest measured soil variations are observed in Zegveld, within a range of 9 cm; see Fig. 1(b). Therefore, these two locations show a representative range of ground movements which can be reasonably expected to be encountered in Dutch agricultural areas. Inspection of the extensometer measurements (blue trace) shows a very high degree of
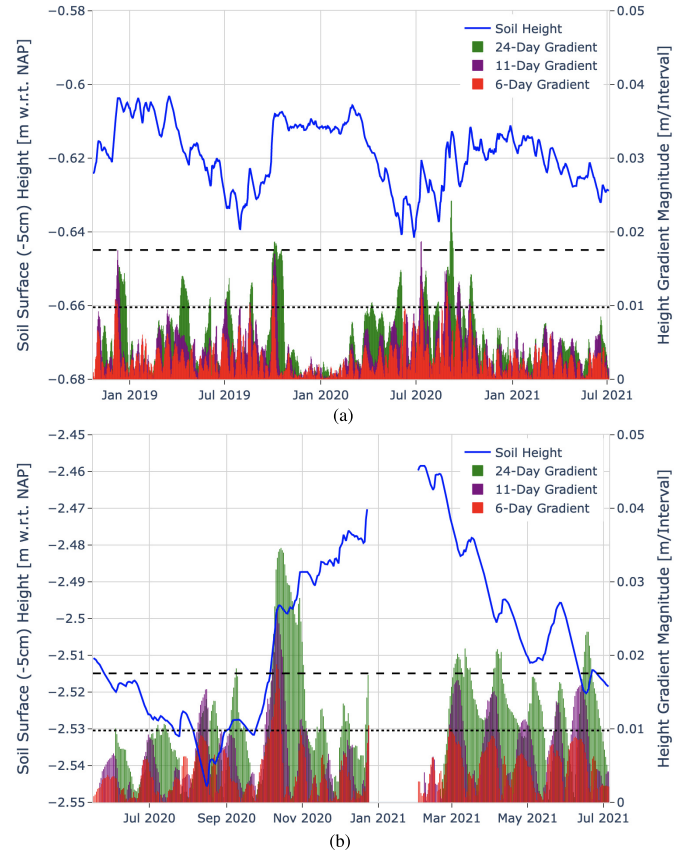

Fig. 1. (a) Rouveen and (b) Zegveld daily extensometer soil height measurements with respect to The Netherlands vertical data (NAP) (blue trace, left-hand *y*-axis) and the corresponding height gradient magnitudes computed over 6-, 11-, and 24-day intervals (red, purple, green bars, right-hand *y*-axis). Representative C- and X-band unwrapping thresholds projected from the slant range to the vertical are indicated by the horizontal dashed and dotted lines, respectively (right-hand *y*-axis). Not visible: L-band unwrapping threshold projected to the vertical (0.0726 m). The gap in data around January 2021 in Zegveld is caused by uplift which exceeded the range of the extensometer scale and could not be recorded.

reversible deformation following a strong seasonal trend. The most stable periods are during the winter months, when cold, rainy conditions in The Netherlands keep the soils saturated with moisture. As temperatures rise and precipitation levels decrease, the soils dry out and become more unstable. A significant amount of movement over short time scales can also be observed. This creates an additional complication for InSAR observations, as significant movement can occur even between the relatively frequent six-day Sentinel-1 overpass cycle, which is shown by the six-day height gradient magnitude (red bars). In the case of Zegveld, shifts in soil height can cause strong gradients in the measured signal which exceed the unwrapping threshold (projected from the slant range to the vertical using a typical Sentinel-1 incidence angle of 37°, and shown by the dashed horizontal line), even in an ideal, noiseless case. We can also infer that we do not need to expect the deformation to exceed multiple ambiguity levels at the C-band within one overpass. For other missions, the outlook is significantly worse. In the case of Radarsat2 (corresponding to the 24-day gradient shown by green bars), which also operates at the C-band, the gradients routinely

exceed the unwrapping threshold in both the locations. With TerraSAR-X (corresponding to the 11-day gradient shown by purple bars, and the dotted line representing the X-band unwrapping threshold height projected from the slant range to the vertical), the signal shifts are consistently larger than the unwrapping threshold, rendering any time series interpretation extremely challenging. On a more hopeful note, the future Radar Observation System for Europe in L-band (ROSE-L) mission operating at the L-band with a six-day revisit time should be able to comfortably follow the observed signal without significant risk of phase unwrapping errors. (The L-band unwrapping threshold is significantly larger than the dynamic range of the signals plotted in Fig. 1 and is therefore not shown.) While this mission will undoubtedly be a very useful asset for subsidence monitoring in The Netherlands, it is expected to be launched in the year 2027 and will then require an additional several years to build up a dataset of observations to allow for the time series analysis. This means that scientists and policymakers will need to wait for about another ten years before the benefits of the ROSE-L mission become truly available to them, and many decisions until that time will need to be based on the available Sentinel-1 data. While other L-band missions have been flown, notably advanced land observing satellite (ALOS)-1, ALOS-2, and argentine microwaves observation satellite (SAOCOM), there have not been enough acquisitions to provide the necessary coverage.

Note that the gap in the data from end December 2020 to mid January 2021 in Fig. 1 in Zegveld is caused by the uplift which exceeded the range of the extensometer scale and could not be recorded. We conservatively do not use the data from this time in our subsequent analysis; however, it could be interpreted that an additional spike in the observed deformation gradients would be present here as well.

B. Simulated InSAR Signal

To assess the ability of an algorithm to correctly unwrap the observed phase, we simulate an InSAR signal based on the ground truth provided by the extensometers. This allows us to control the level of noise in the data and reference the obtained solutions to a known true value. We produce this simulated signal, ϕ_{sim}, by downsampling the extensometer data to one observation per six days, converting the vertical displacement into phase, projecting from the vertical axis onto the slant range direction, adding noise, and wrapping the resultant phase on the interval $[-\pi, \pi)$. For simplicity, we use the shorthand $f(t)$ to denote a time series which has been sampled with a period of six days. The simulated signal is given by

$$\phi_{\text{sim}}(t_i, t_j) = W\left\{ \Delta z_{\text{ext}} \cdot \frac{4\pi \cos \theta}{\lambda} + n(t_i, t_j) \right\} \quad (1)$$

where $W\{\cdot\}$ is the phase wrapping operator [17], Δz_{ext} is the change in the extensometer signal between acquisition epochs t_i and t_j, λ is the radar wavelength of 0.0556 m, corresponding to the wavelength of the Sentinel-1 radar, θ is the radar incidence angle, and $n(t_i, t_j)$ is the additive heteroscedastic noise. The variance of $n(t_i, t_j)$ differs for each interferogram and is governed by the coherence $\gamma(t_i, t_j)$, and equivalent
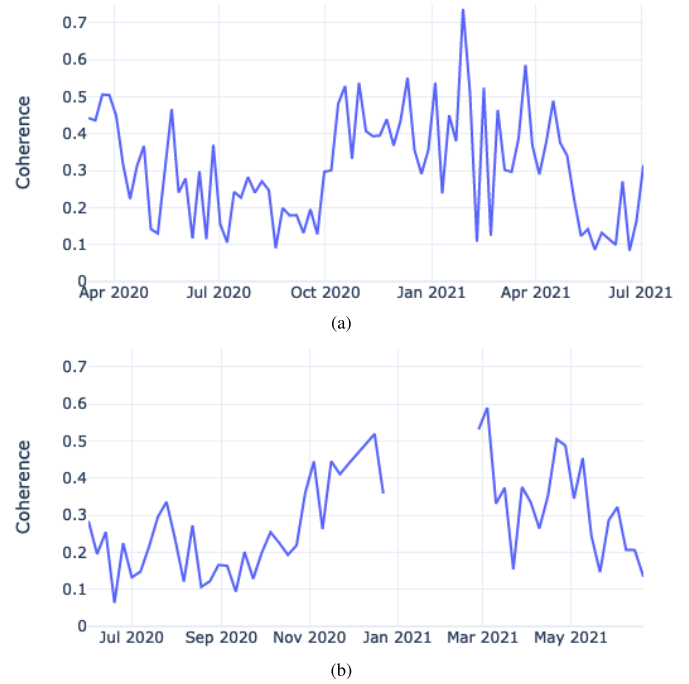

Fig. 2. Daisy-chain coherence values used to generate additive noise in (a) Rouveen and (b) Zegveld. The term "daisy-chain" refers to the off-diagonal elements of the coherence matrix showing the relative coherence between consecutive epochs.

number of looks, L [17]. The coherence of a multilooked set of pixels Ω in an interferogram is estimated from the magnitude of the complex sample coherence given in [17] by

$$\gamma(t_i, t_j) = \frac{\left| \sum_{n \in \Omega} S_{i_n} S_{j_n}^* \right|}{\sqrt{\left(\sum_{n \in \Omega} |S_{i_n}|^2 \right) \left(\sum_{n \in \Omega} |S_{j_n}|^2 \right)}} \quad (2)$$

where S_{i,j_n} is the nth pixel in SAR images acquired at times t_i and t_j. We focus on the daisy-chain coherence, which refers to the coherence observed in interferograms of consecutive acquisitions ($j = i - 1$), i.e., the magnitudes of the first off-diagonal of the full coherence matrix, as shown in Fig. 2. For simplicity, we denote the daisy-chain phase $\phi(t_i, t_{i-1})$ as $\Delta \phi$.

The additive phase noise is modeled as a random process which follows a circular Gaussian distribution. The probability density function (PDF) of the phase component is [17], [18], [19]:

$$f(\phi | \gamma, L, \phi_0) = \frac{\Gamma(L + 1/2)(1 - \gamma^2)^L \beta}{2\sqrt{\pi} \Gamma(L)(1 - \beta^2)^{L+1/2}}$$
$$+ \frac{(1 - \gamma^2)^L}{2\pi} \cdot {}_2F_1\left(L, 1; \frac{1}{2}; \beta^2 \right) \quad (3)$$

where $\beta = \gamma \cos(\phi - \phi_0)$, $\Gamma(\cdot)$ is the gamma function, and $F(\cdot)$ is the hypergeometric function. The PDF is evaluated over the interval $\phi = [-\pi, \pi)$, and the mean, ϕ_0, is taken as 0 to center the additive noise around the signal. An equivalent alternative formulation is possible in which one sets $\phi_0 = W\{\Delta z_{\text{ext}} \cdot 4\pi \cos \theta / \lambda\}$, due to the fact that $f(\phi | \gamma, L, \phi_0)$ produces random samples of wrapped phases. In this case, the distribution would describe the entire signal, ϕ_{sim}, and not only the noise component, and the phase wrapping would be stated
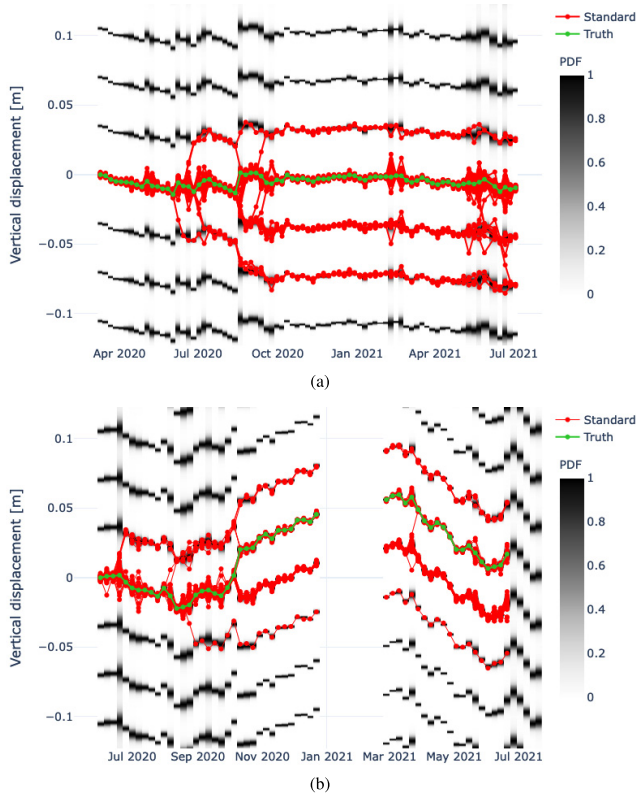
Fig. 3. Normalized multimodal PDF (greyscale) based on (1) and (3) from downsampled extensometer soil height measurements (green trace) with additive noise, corresponding to the Sentinel-1 revisit time for (a) Rouveen and (b) Zegveld. During times of high coherence, the PDF is clearly defined and a clear peak is visible at each ambiguity level. During low-coherence epochs, the PDF becomes "smeared-out" and distinguishing ambiguity levels becomes challenging. In all, 50 realizations of random noise are generated and a solution obtained using the standard minimum gradient approach (red traces).

implicitly. We choose to use $\phi_0 = 0$ formulation because it allows the reader to more clearly see all the steps which are taken to produce the simulated phases.

Thus, every epoch has a different noise variance $\sigma_n^2(\gamma, L)$, which is based on the coherence value at that epoch and determines the shape of the distribution $f(\gamma(t_i, t_j), L)$ which is randomly sampled. The simulated coherence values used to generate these noise values are shown in Fig. 2, which are the representative values for Sentinel-1 observations over Rouveen and Zegveld during the periods in which the extensometers were active. Epochs with coherences of less than 0.05 can occasionally occur, but are omitted from this study because this results in almost complete information loss in the interferogram, and phase unwrapping becomes redundant because only noise remains in the observation. In a fully realized InSAR processor, these interferograms must be omitted or bridged using some form of assumption or pseudoobservation. We use a multilooking factor $L = 100$, as strong multilooking is an important part of a strategy for noise reduction to enable InSAR over grasslands on peaty soil. Without such high levels of multilooking, noise levels in the summer months of low coherence can become so strong that it is impossible to interpret the interferometric phase.

To visualize the effect of noise, Fig. 3 shows 50 unwrapped time series (red traces) obtained by 50 realizations of random

noise generated by the PDF (greyscale). The PDF is mul- timodal, which reflects that the observed phase is wrapped and the true ambiguity level is unknown. The epochs of low coherence widen the PDF such that it is no longer a set of clearly defined peaks, which can cause the observed phase change between epochs to exceed the unwrapping threshold.

The divergence of the red traces illustrates how noise can have a significant effect on unwrapping decisions, and therefore on the inferred vertical displacement. Strong noise can either introduce new unwrapping errors, as in the case of Rouveen, or exacerbate the existing problems already present, as in the Zegveld case, cf. Fig. 1(a) and (b), respectively.

C. Implications of Rapid Ground Motion and Noise

Figs. 1–3 demonstrate the difficulty of directly producing InSAR time series estimates of the Dutch polderlands. First, from a signal perspective, Fig. 1 shows that the ground motion which we attempt to measure is highly rapid and nonlinear. Large shifts can occur even within the six-day Sentinel-1 revisit interval, creating problems for phase unwrapping even in a noise-free scenario. The second additional challenge is that of noise, which is related to that of temporal decorrelation, as shown in Figs. 2 and 3. Additive noise can occasionally become very large during periods of low coherence and can cause phase wrapping to occur when there would be none in the hypothetical noise-free scenario. This is of particular importance during the late summer and autumn months, when the soil height gradients are at their largest. The observer, and by extension, the algorithm, needs more information to correctly unwrap the observed phase in this environment. Section III will describe how this may be accomplished using the readily available meteorological data.

III. MAKING USE OF CONTEXTUAL INFORMATION

A. Motivation for a RNN

Studies with the extensometer data [16] demonstrate a very strong relationship between the shallow ground displacements and the phreatic groundwater level. It is therefore critical for us to understand the dynamics of the phreatic groundwater to anticipate large deformation events in the shallow subsurface. The available *in situ* measurements have shown that the groundwater follows a clear seasonal pattern with short-term dynamics. Modeling the phreatic groundwater of the Dutch peatlands is an ongoing scientific endeavor and has proven challenging in the past, and most models are only able to estimate an approximate annual maximum and minimum [20]. Numerical models that are capable of simulating phreatic groundwater dynamics are also available, but are not accurate or granular enough for our application [21].

Despite not having direct access to the groundwater level, several inferences can still be made. Based on *in situ* mea- surements, we assert that the critical factors which determine the short-term changes in the phreatic groundwater level, and therefore the surface height, are precipitation and temperature. With three parameters [precipitation, temperature, and day of year (DOY)], we can train a neural network to anticipate shifts in the surface height, which captures both the seasonality and
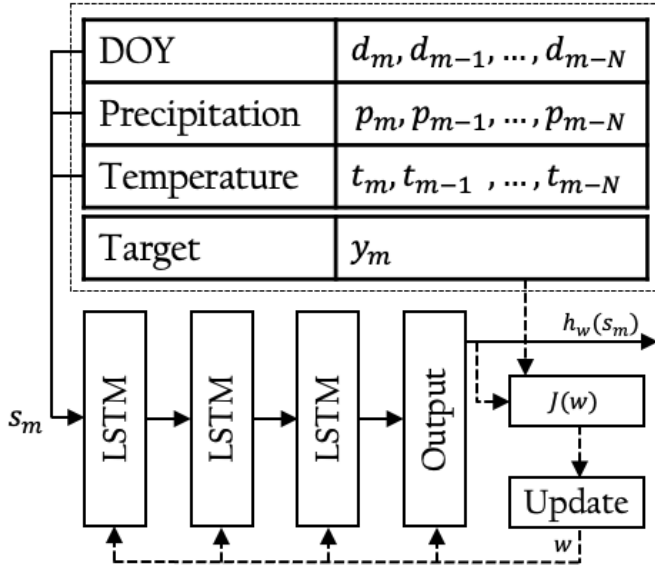
Fig. 4. RNN flow diagram with an instance of a training data sequence s_m at epoch m. The feedforward prediction path is shown with solid lines, and the feedback path used for training the network is shown with dashed lines.

the short time scales we require. Synoptic precipitation and temperature values are commonly measured at weather stations and are freely available throughout The Netherlands and many other parts of the world. Furthermore, we do not need to know the exact surface height at any given time; what is important for us is simply to know what correct phase unwrapping *direction* is. That is, we only need to predict whether or not the ground level will move significantly upward or downward with respect to the previous epoch. This allows us to turn the time series prediction problem into a classification problem, which is conceptually much easier for the machine learning algorithms to solve.

B. RNN Architecture

An RNN is optimized for the task of processing sequences of input data in time and predicting an output classification and is therefore a suitable choice for our application [22]. The RNN comprises an input layer, three long short-term memory (LSTM) hidden layers, and a three-node output layer, corresponding to the probabilities of the three output labels. LSTM layers are chosen because of their ability to use both the current and past values of a time series in computation. Multiple layers are used to capture the nonlinear relationships between the input sequences and the target value. The outputs of each layer are renormalized, and a 10% dropout is added to help prevent overfitting [23]. The RNN is created using the TensorFlow Python library.

C. Training Data

The RNN is designed to take in a set of sequential precipitation and temperature data linked to a target value for training. An instance of the training set, pictured in Fig. 4, is a sequence of DOY values, and daily precipitation and temperature values

starting from the day in question (the current epoch, m) going back to the Nth previous day. These are coupled with one integer target value. The target value is extracted from the extensometer data by downsampling to one observation every six days to match the Sentinel-1 overpass cycle and assigning an integer label of 0, 1, or 2 corresponding to insignificant, upward, or downward motion, respectively.

A set of M training sequences are used to train the network, which takes the precipitation and temperature inputs and attempts to predict the target value at the output. Error in the predictions is quantified through the categorical cross-entropy loss function J, as [24]

$$J(w) = -\frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{C} y_{m,i} \cdot \log\left[h_{w,i}(s_m)\right] \qquad (4)$$

where w is the vector of weights in the neural network model (also known as the model parameters), y_m is the one-hot vector representation of the true target at epoch m, and $h_w(s_m)$ is the estimated target. C is the length of the one-hot vectors, i.e., the number of categories. A one-hot vector is a mapping of an integer category to a vector with a length equal to the number of categories, in which all the elements are zero apart from the element corresponding to the given category. For example, in the set of categories {0, 1, 2}, the category "2" maps to the vector [0, 0, 1]. The model parameters are updated in each training iteration so as to minimize J. The problem statement can then be formulated as finding the optimal set of weights that minimize J

$$w_{opt} = \arg\min_{w} J(w). \qquad (5)$$

The training dataset is split between 80% training and 20% validation. The network is trained over many iterations as the algorithm attempts to reduce the training loss. The final model which is saved is the iteration in which the validation loss is minimized, as shown in Fig. 5. The model iteration used is given by the vertical black dashed line.

D. RNN Model Testing

The RNN is tested using an independent extensometer dataset obtained from a separate location 4.5 km away. The training and testing locations are shown in Fig. 6. This extensometer is installed on a site with a significantly thinner soft soil layer in the shallow subsurface. This site was chosen so that a large variation in the soil uplift and subsidence rates may be observed. What this means for our model test is that the soil properties of this location are substantially different enough to represent the natural variations in soil encountered throughout an observed scene, but are still geographically close enough to encounter the same weather conditions.

The model is evaluated using 977 predictions on data gathered from October 15, 2018, to July 4, 2021. The performance of the model is determined by creating a confusion matrix of the real versus predicted movement classes, as shown in Table I. It can be seen that the network can very accurately differentiate between significant upward and downward motion (UP versus DOWN) with little to no error (bottom-right
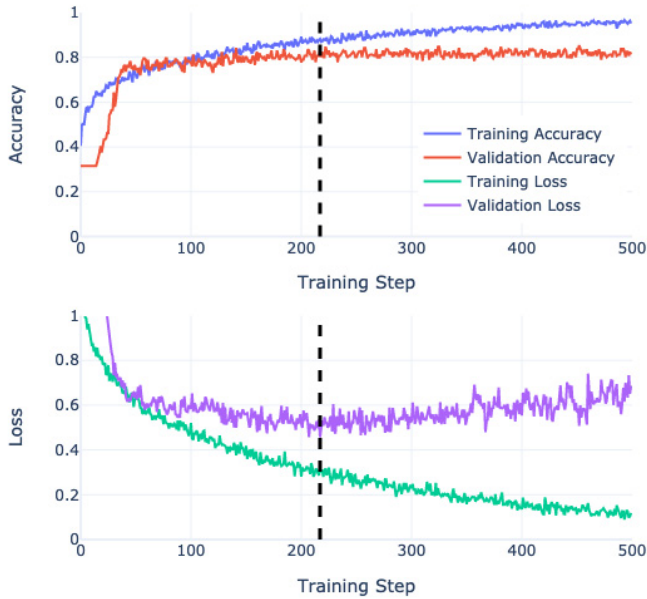
Fig. 5. RNN training curves. The black dashed line indicates the training step in which the optimum model weights are achieved. Accuracy is defined as the ratio of the number of correct predictions to the total number of predictions, and loss is defined by (4).

Fig. 6. Aerial view of extensometer locations used for training and testing of RNN in Rouveen, NL.

TABLE I
RNN TEST CONFUSION MATRIX

	True STAY	True UP	True DOWN
Predicted STAY	0.61	0.12	0.22
Predicted UP	0.14	0.88	0.02
Predicted DOWN	0.24	0	0.76

quadrant of Table I. It has greater difficulty in distinguishing between UP/DOWN versus STAY. This is due to the fact that the choice of threshold which defines what amount of movement is considered large enough to be an UP/DOWN
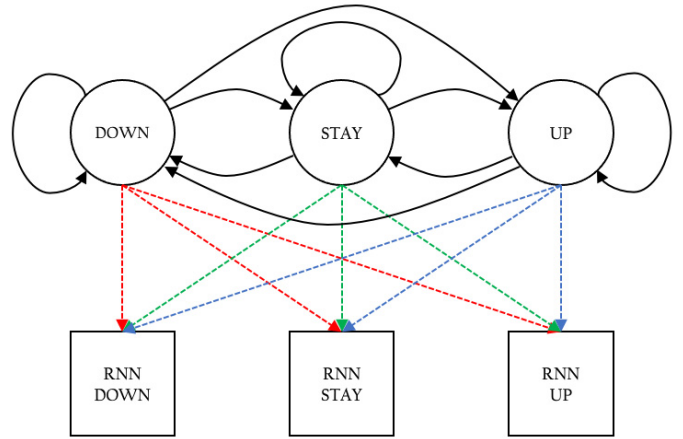

Fig. 7. HMM used to integrate wrapped phase information with RNN output. Solid lines represent transition probabilities, and dashed lines represent emission probabilities.

state versus what is not large enough and is considered a STAY is a somewhat arbitrary distinction. This result is, however, not very concerning, because the important distinction to make is between significant upward and downward motion, i.e., correctly predicting UP versus DOWN.

IV. RNN-AIDED PHASE UNWRAPPING

A. Introduction

The RNN model described in Section III is able to estimate the DOGM; however, these predictions are not 100% accurate (as shown in Table I), and all this information must still be integrated into a phase unwrapping routine. In addition, the wrapped phases and coherences are also important information which should be used. To integrate all this information together, the problem can be posed as an HMM, which is a general formulation describing how a system may transition between various unobservable "hidden" states (the unwrapped phase transitions) given a set of noisy observations (the RNN estimation).

B. Hidden Markov Model

The system of wrapped interferometric phase observations and estimated RNN classes is considered to act together as an HMM, as pictured in Fig. 7. For simplicity, we consider three hidden states corresponding to the average ground movement: uplift (UP), subsidence (DOWN), or no significant movement (STAY), which correspond closely to (but are distinct from) the RNN output classes. These states are hidden, because the nature of wrapped phases is such that one cannot discern the DOGM from them directly. The additional observations in this HMM therefore come from the RNN output classes.

To complete the model, a set of *transition* and *emission* probabilities are required. The transition probabilities represent the probability of moving to a given state from the current state. In our implementation, these are estimated using the wrapped phases, $\Delta\phi$. The probabilities are estimated in two steps: 1) estimating the probability of downward versus upward motion and 2) estimating the significance of the

motion, by comparing the magnitude of the phase change with the distribution given in (3). The up/down probabilities are determined by considering the two nearest ambiguity levels as branches

$$b_1 = \Delta\phi$$
$$b_2 = \Delta\phi - \text{sign}(\Delta\phi) \cdot 2\pi \qquad (6)$$

where $\Delta\phi$ is the observed change in the wrapped phases between epochs, which is equal to the equivalent daisy-chain phase. These branches correspond to the UP or DOWN state, depending on the sign of the observed phase change. The initial probabilities of the two complementary branches are estimated based on the magnitude of $\Delta\phi$ and are given by

$$p_{b_1} = 1 - \frac{1}{2}[\text{erf}(|\Delta\phi| - \pi) + 1]$$
$$p_{b_2} = 1 - p_{b_1} \qquad (7)$$

where $\text{erf}(\cdot)$ denotes the error function. As $|\Delta\phi|$ increases, the probability of remaining on branch b_1 smoothly decreases from 1. When $|\Delta\phi| = \pi$, both the branches are equally likely. In this initial estimation, the branch corresponding to the minimum phase gradient solution is favored. In the absence of any additional information, this branch would always be selected by the algorithm, corresponding to the standard unwrapping solution. Which branch corresponds to UP and which to DOWN depends on the sign of $\Delta\phi$.

While the probabilities estimated by (7) provide the relative likelihoods of transitioning to the UP versus DOWN state, to complete the set of transition probabilities, the likelihood of the STAY state must also be estimated. This is done using the temporal coherence to compare the magnitude of the phase to three times the estimated phase noise standard deviation, σ_{ϕ_n}, of that epoch, which is determined using the circular Gaussian distribution given in (3). This provides an estimate as to whether or not an observed phase change contains significant movement with respect to the estimated level of noise in the measurement. This probability of significance is estimated by comparing the magnitude of the phase change with the *n*-sigma noise level

$$p_{\text{sig}} = \text{erf}\left(\frac{1}{\sqrt{2}} \cdot \frac{|\Delta\phi|}{n\sigma_{\phi_n}}\right). \qquad (8)$$

The best value of n is empirically found to be 1.5. Finally, to obtain the set of transition probabilities, the probabilities of (7) are conditioned on p_{sig}

$$T(\text{UP}) = p_{b_{1,2}} \cdot p_{\text{sig}}$$
$$T(\text{DOWN}) = p_{b_{2,1}} \cdot p_{\text{sig}}$$
$$T(\text{STAY}) = 1 - p_{\text{sig}} \qquad (9)$$

where the subscript $b_{1,2}$ refers to the branch corresponding to the upward motion, and $b_{2,1}$ is the opposite branch corresponding to the downward motion depending on the signs of the terms in (6). A graphical example of how an observed phase change relates to the transition probabilities is shown in Fig. 8.

The *emission* probabilities describe the likelihood of an external observation (in our case the RNN output) while being
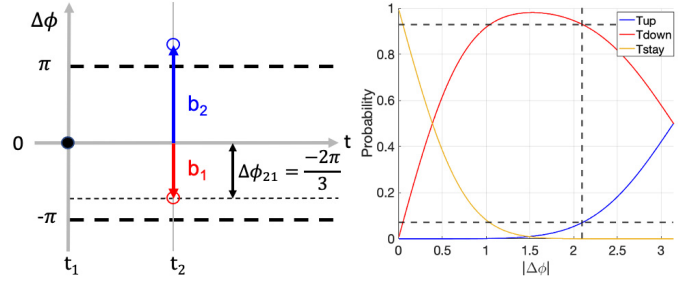

Fig. 8. Example estimation of the transition probabilities based on an observed phase change $\Delta\phi = -2\pi/3$. Left: blue indicates the upward branch and red indicates the downward branch. Right: blue indicates T(UP), red T(DOWN), and yellow T(STAY). The black dashed lines correspond to the UP and DOWN probabilities estimated for $\Delta\phi$ (not shown: T(STAY) = 0).

in a given state. These probabilities are estimated by taking the values of the appropriate column of the confusion matrix obtained from testing the RNN (Table I).

There are two key differences which differentiate our implementation and a typical HMM example: 1) the transition probabilities are usually assumed to be static, but in our case they change with every epoch and 2) in our case, the transition probabilities are independent of the current state, while usually they are not. The effect this has on the calculation is described below.

C. Modified Viterbi Algorithm

The Viterbi algorithm [15] describes how to predict the most likely sequence of states in an HMM. It does this by predicting the most likely current state and subsequently using that state's transition probabilities in combination with the external observations to predict the next state. Thus, the sequence of most likely states, called the Viterbi path, is determined. The output of this calculation is a "*trellis*" of probabilities P which describe the likelihood of being in a given state s at epoch t (corresponding to t_i in (1)), which can be written as

$$P(t, s) = P(t - 1, k) \cdot T(k, s) \cdot E(t, s) \qquad (10)$$

where k is the most likely state at time $t - 1$, T is the matrix of state transition probabilities, and E is the matrix of emission probabilities. This calculation is simplified in our implementation by the fact that the next state does not depend on the current state, that is, while the set of state transition probabilities change from epoch to epoch, they are the same for each state. Since $P(t - 1, k)$ is the same for each choice of k, it becomes a scalar which no longer affects the outcome and is dropped. We therefore calculate the trellis of probabilities as

$$P(t, s) = T(t, s) \cdot E(t, s) \qquad (11)$$

which shows how RNN estimation is combined with the wrapped phases to inform the algorithm how the phase should be unwrapped. T represents the information present in the wrapped phases, and E is the information from the RNN. If one were to imagine the case in which E were not present, that is, $P(t, s) = T(t, s)$, phase unwrapping would be completely informed by the wrapped phases and the algorithm

would behave identically to the standard minimum gradient implementations. Inversely, if one were to ignore the wrapped phases and take $P(t, s) = E(t, s)$, the phase unwrapping would be driven completely by the RNN. Therefore (11) shows how the final probabilities are assessed based on the balance of the relative levels of confidence in the two sources of information at that epoch.

D. Phase Unwrapping

The phase unwrapping works by stepping through the trellis to find the most likely state at each epoch and integrating the corresponding phase. The two branches in (6) are considered as the two possible *paths* the algorithm can take to unwrap the phase. The algorithm steps through each epoch and selects the appropriate ambiguity level, i.e., -2π, 0, or $+2\pi$, to add to the phase at that epoch according to the value of the most likely state. In the case of an UP state, the upper branch is chosen, and in the DOWN state, the lower branch is chosen. In the STAY case, the branch corresponding to the smallest phase change is selected. This way, the algorithm attempts to find the correct ambiguity level without modifying the observed phases. When the observed phase change is low, or the estimated level of noise is high, the algorithm becomes more conservative and the STAY state becomes more likely, causing the algorithm to default back to the standard minimum gradient solution.

V. RESULTS

A. Optimal Path Through Noisy Data

The RNN-aided algorithm is used to unwrap 50 instances of noisy phases originally shown in Fig. 3, shown again with the RNN-aided solution in Fig. 9.

While the RNN-aided algorithm solutions contain significantly less error than their standard counterparts, it can be observed that noise still has a significant effect on the solution. We therefore include the option to retain multiple solutions which branch out in the problematic epochs where the algorithm cannot confidently choose a certain path. Fig. 10,

The effect of introducing the RNN-aided algorithm is clearly shown, which shows the histograms of the error in the final epoch of the unwrapped time series with respect to the true value in 1000 realizations of noise. Here, a downward skew can be seen in the standard algorithm's results at both the locations, which is particularly pronounced in the Zegveld case. This downward skew is strongly mitigated by the RNN-aided algorithm, and the correct solution is more consistently obtained.

B. Unwrapping Success Rates Versus Coherence

The RNN-aided phase unwrapping algorithm is tested against a standard minimum gradient unwrapping algorithm to test its ability to correctly interpret the simulated deformation phase as given in (1). Several values of temporal coherence ranging from 0.05 to 0.95 are swept through to generate the test cases; see Fig. 11 for Rouveen and Zegveld. For each coherence level, 1000 independent noise simulations are
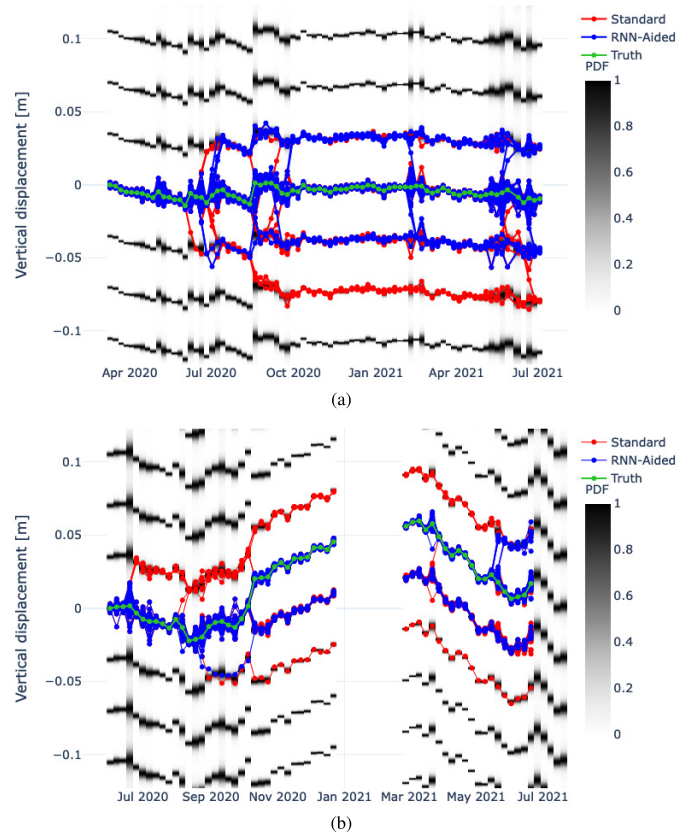

Fig. 9. Normalized multimodal PDF (greyscale) based on (1) and (3) from downsampled extensometer soil height measurements (green trace) with additive noise, corresponding to Sentinel-1 revisit time for (a) Rouveen and (b) Zegveld. In all, 50 realizations of random noise are generated and solutions obtained using the standard minimum gradient (red traces) and the RNN-aided (blue traces) approaches.

generated according to the circular Gaussian distribution in (3). Note that for this particular test, while the noise is random from epoch to epoch, the coherence is held constant for the entire run through the time series. This simplification is made to generate the test statistics and accurately compare the two algorithms in a standardized manner. For each time series run, the integer number of unwrapping errors is tallied and the unwrapping error rate is determined by dividing the total number of errors by the number of unwrapping operations performed for each coherence level ($N_{errors}/N_{epochs} \cdot N_{iterations}$). The position of the unwrapping error in the time series is not considered in this tally.

The RNN-aided unwrapping algorithm consistently outperforms the standard algorithm in both the cases. In the case of Rouveen, both are able to correctly unwrap the time series and retrieve the correct deformation phase in high-coherence scenarios, because the deformation phase gradient never exceeds the unwrapping threshold. But as coherence drops and noise levels increase, errors begin to appear in the solution. The RNN-aided algorithm is more resistant to this noise, as it does not solely rely on the wrapped phases to estimate the correct ambiguity level. In the Zegveld case, the deformation gradients are stronger and do exceed the unwrapping threshold even in a noiseless scenario. Thus, we
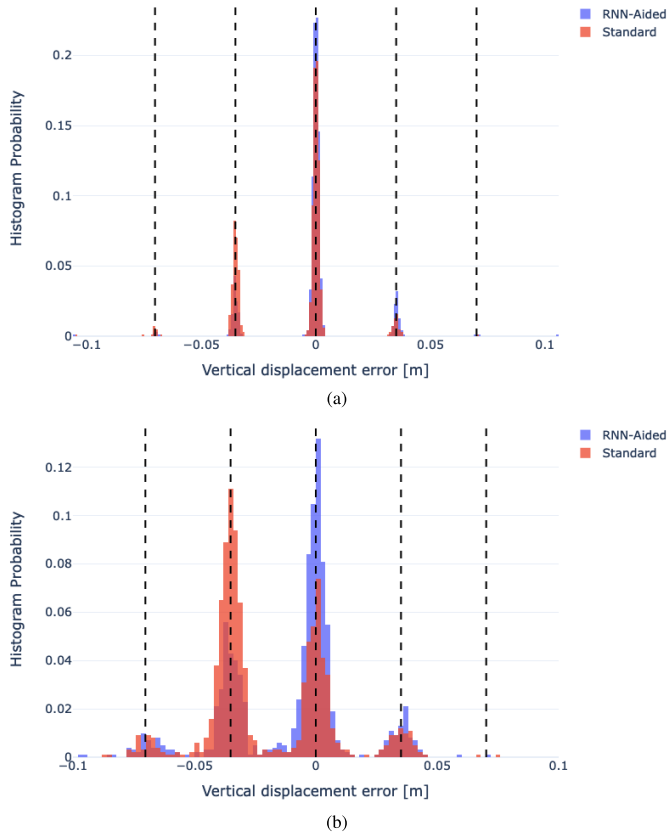
(a)

(b)

Fig. 10. Histograms of the error in the final epoch of the obtained time series in (a) Rouveen and (b) Zegveld of 1000 instances of simulated noise. (Red) Standard algorithm. (Blue) RNN-aided algorithm. The nearest ambiguity levels are indicated by the black dashed lines. The standard algorithm produces results which are downwardly skewed. This skew is corrected by the RNN-aided algorithm.
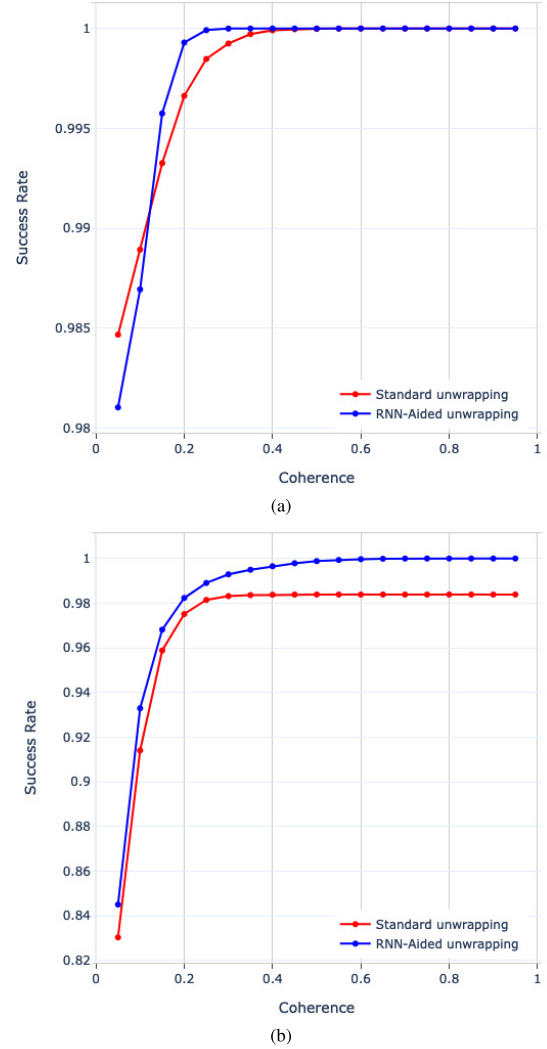

(a)

(b)

Fig. 11. Algorithm performance curves of unwrapping SR versus coherence in (a) Rouveen and (b) Zegveld. (Red) Standard algorithm. (Blue) RNN-aided algorithm. In the Zegveld case, the standard algorithm cannot exceed an SR of approx. 98%. This means that at least one phase unwrapping error can be expected for a time series containing 50 epochs (just under one year of Sentinel-1 data).

conclude that the standard algorithm will intrinsically never be able to estimate the correct ambiguity level. However, the RNN-aided algorithm is able to consistently unwrap the time series correctly. As coherence decreases, the likelihood of making an error increases, as spikes of noise can affect the algorithm's ability to interpret the phase. Nevertheless, the RNN-aided algorithm is still able to reject some of the effects of noise and perform better, or as well as, the standard algorithm.

VI. DISCUSSION

Extensometer measurements gathered throughout the Dutch polderlands have shown that there is a very close relationship between the phreatic groundwater level and the soil height [16]. While ground water measurements are not directly available, we are able to anticipate large shifts in the soil height which disrupt standard phase unwrapping by means of a motion classifier RNN. This model can very reliably anticipate the large changes in the ground motion of nearby independent locations, as shown by the lower right quadrant of the confusion matrix in Table I. The model has greater difficulty in differentiating between significant and insignificant levels of motion in both the directions, which is a somewhat arbitrary distinction. This can be avoided by removing the STAY class

altogether; however, in that case the model predictions can become unstable, as even tiny or zero deformations must be given an UP or DOWN label by the model. We find that including a STAY class corresponding to small deformations (<3 mm) provides more accuracy in making the distinction between large UP/DOWN events, which we choose to prioritize.

Previous InSAR studies of the Dutch peatlands which have attempted to directly monitor the ground via DS processing [7] have failed to match the observations of other geodetic techniques such as leveling, lidar surveys, and extensometer measurements, and often strongly overestimate the true linear subsidence rate. The reason for this (as discussed in Section I) is due in large part to the conventional phase unwrapping techniques being unable to cope with the rapid movements which occur within the shortest temporal baselines, as shown in Fig. 1, which results in estimates skewed downward, as shown

in Fig. 10. By removing this downward skew, we hope to enable accurate monitoring of this region with DS InSAR for the first time.

The results of Fig. 11 show that the RNN can provide a significant performance gain by supplying the unwrapping procedure with more information. The deformation test scenario chosen based on the Zegveld extensometer measurements as shown in Fig. 1 shows a period of strong uplift in the autumn of 2020. When downsampling these measurements to six-day intervals corresponding to the Sentinel-1 overpass cycle, there is at least one epoch in which the deformation phase exceeds half a cycle ($\lambda/4$) and the standard unwrapping procedures fail, even in the presence of little or no noise. On the other hand, the RNN-aided algorithm is able to consistently retrieve the correct time series. In the Rouveen case, both the algorithms are able to achieve a 100% success rate (SR) given a high enough coherence level. However, the RNN-aided algorithm is able to achieve 100% SR at lower levels of coherence, which shows how integrating more information into the processing workflow will enable greater performance. Coherence itself can be viewed as a measure of the information content in an interferogram (or lack of coherence as a measure of entropy). Thus, introducing new information to the algorithm creates an increase in the "*effective coherence*" of the interferogram. For example, in the Rouveen case, Fig. 11 shows an effective coherence gain of $\Delta\gamma = 0.175$ at SR = 100% when moving from the standard unwrapping algorithm ($\gamma = 0.4$) to the RNN-aided algorithm ($\gamma = 0.225$).

One can also observe that the RNN-aided algorithm performs differently between locations, with better performance in Rouveen than in Zegveld. This is likely due to the fact that the available dataset in Rouveen is much longer than in Zegveld. The Rouveen dataset begins 1.5 years before the Zegveld one, which means that the RNN has more than twice as much training data available to it. Therefore, one may also infer that the potential for improvement as available datasets expand is very large, and that the performance of these models will continue to improve as more data become available.

VII. Conclusion

Phase unwrapping is an underdetermined problem which is conventionally sidestepped by assuming that the correct solution is the one which corresponds to the smallest phase change between epochs. While this is often a reasonable approach, ground-based measurements have shown that this is not a good enough solution in certain regions, such as the Dutch peatlands. If InSAR is to be used to monitor ground deformation in these regions, a different approach is required. This necessitates the integration of additional information to help solve the phase unwrapping problem.

We demonstrated how ground-based measurements can be used to create a model that anticipates large shifts in the ground level based on the readily available environmental information such as precipitation and temperature. Predictions from this model are integrated into the phase unwrapping process by considering the system of wrapped phases and model predictions as an HMM. The relative probabilities of

which direction the observed phases should be integrated into the time series can then be estimated using the Viterbi algorithm. While in our scenario we chose to focus on a model which can anticipate ground movement caused by changes in precipitation and temperature, this HMM framework can also be used in other cases in which a researcher would wish to integrate additional information into a temporal phase unwrapping application.

Our results show that using this technique can alleviate some of the worst effects which rapid soil movements have on DS InSAR estimates when combined with an effective multilooking strategy. For the future, we plan to integrate this phase unwrapping procedure into a processing framework such as the Delft persistent scatterer interferometry (DePSI) processor [25], which will allow for mixed PS-DS monitoring of the Dutch peatlands.

References

[1] G. Erkens, M. J. van der Meulen, and H. Middelkoop, "Double trouble: Subsidence and CO_2 respiration due to 1,000 years of Dutch coastal peatlands cultivation," *Hydrogeol. J.*, vol. 24, no. 3, pp. 551–568, May 2016.

[2] E. Stouthamer *et al.*, "Dutch national scientific research program on land subsidence: Living on soft soils—Subsidence and society," *Proc. Int. Assoc. Hydrolog. Sci.*, vol. 382, pp. 815–819, Apr. 2020.

[3] C. Schouten. (2020). *Letter to Parliament by the Dutch Minister of Agriculture on the Reduction of Greenhouse Gas Emissions From Peatlands (in Dutch)*. [Online]. Available: https://www.rijksoverheid.nl/documenten/publicaties/2020/07/13/verplicht-format-bijlage-onderbouwing-en-evaluatie-veenplan-1e-fase

[4] M. C. Cuenca and R. F. Hanssen, "Subsidence due to peat decomposition in The Netherlands, kinematic observations from radar interferometry," in *Proc. ESA Fringe Workshop*, Frascati, Italy, 2008, pp. 1–6.

[5] M. C. Cuenca, R. F. Hanssen, A. Hooper, and M. Arikan, "Surface deformation of the whole Netherlands after PSI analysis," in *Proc. ESA Fringe Workshop*, Frascati, Italy, 2011, pp. 19–23.

[6] R. F. Hanssen *et al.* (2018). *Land Motion Service of The Netherlands*. [Online]. Available: http://bodemdalingskaart.nl

[7] Y. Morishita and R. F. Hanssen, "Deformation parameter estimation in low coherence areas using a multisatellite InSAR approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4275–4283, Aug. 2015.

[8] F. M. G. Heuff and R. F. Hanssen, "InSAR phase reduction using the remove-compute-restore method," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 786–789.

[9] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, no. 4, pp. 713–720, Aug. 1988.

[10] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, no. 4, pp. R1–R54, Aug. 1998.

[11] C. W. Chen and H. A. Zebker, "Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 18, no. 2, pp. 338–351, Feb. 2001.

[12] H. Yu, Y. Lan, Z. Yuan, J. Xu, and H. Lee, "Phase unwrapping in InSAR: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 40–58, Mar. 2019.

[13] L. Zhou, H. Yu, Y. Lan, and M. Xing, "Artificial intelligence in interferometric synthetic aperture radar phase unwrapping: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 10–28, Jun. 2021.

[14] B. M. Kampes and R. F. Hanssen, "Ambiguity resolution for permanent scatterer interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2446–2453, Nov. 2004.

[15] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

[16] S. van Asselen, G. Erkens, and F. de Graaf, "Monitoring shallow subsidence in cultivated peatlands," *Proc. Int. Assoc. Hydrolog. Sci.*, vol. 382, pp. 189–194, Apr. 2020.

[17] R. F. Hanssen, *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht, The Netherlands: Kluwer Publishers, 2001.

[18] B. C. Barber, "The phase statistics of a multichannel radar interferometer," *Waves Random Media*, vol. 3, no. 4, pp. 257–266, Oct. 1993.

[19] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, Sep. 1994.

[20] J. van der Gaast, H. Vroon, and H. Massop, *Groundwater Level Based on Mappable Features*, vol. 41. Engelsbrand, Germany: STOWA, 2010.

[21] W. J. De Lange et al., "An operational, multi-scale, multi-model system for consensus-based, integrated water management and policy analysis: The Netherlands hydrological instrument," *Environ. Model. Softw.*, vol. 59, pp. 98–108, Sep. 2014.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[23] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.

[24] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, 2018, pp. 1–11.

[25] F. van Leijen, "Persistent scatterer interferometry based on geodetic estimation theory," Ph.D. dissertation, Dept. Geosci. Remote Sens., TU Delft, Delft, The Netherlands, 2014.

Philip Conroy received the B.Sc. degree *(cum laude)* in electrical engineering from the University of Calgary, Calgary, Canada, in 2014, and the M.Sc. degree in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2017. He is currently pursuing the Ph.D. degree in satellite radar interferometry with the Delft University of Technology, Delft, The Netherlands.

From 2017 to 2019, he worked with the European Space Agency under the Young Graduate Traineeship Program, where he was involved in digital signal processing and ionospheric studies in support of Earth observation and planetary science radar missions.

Simon A. N. van Diepen received the B.Sc. *(cum laude)* and master's degrees in aerospace engineering from the Delft University of Technology, Delft, The Netherlands, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in satellite radar interferometry. His thesis at the Space Research Organization Netherlands (SRON) focused on developing an algorithm for estimating methane emissions based on imagery from the TROPOspheric Monitoring Instrument (TROPOMI) instrument.

Sanneke van Asselen received the master's degree *(cum laude)* in physical geography from the University of Amsterdam, Amsterdam, The Netherlands, in 2003, and the Ph.D. degree *(cum laude)* in physical geography from Utrecht University, Utrecht, The Netherlands, in 2010. Her thesis concerned the effects of peat compaction on Holocene delta evolution, including quantification and prediction of subsidence due to peat compaction.

From 2003 to 2005, she worked as a Project Officer and the Researcher for the European Union (EU) concerted action Soil Conservation and Protection in Europe. From 2010 to 2017, she was a Post-Doctoral Researcher on global land use change with the Institute for Environmental Research, Free University of Amsterdam, and as a Post-Doctoral Researcher and the Coordinator of the Future Deltas Research Program at Utrecht University. Since 2017, she has been a Researcher with the Deltares Research Institute, focused on measuring and monitoring movement in peat soils and understanding the related driving geomechanical and biogeochemical processes.

Gilles Erkens received the M.Sc. degree in quaternary geology from the Free University of Amsterdam, Amsterdam, The Netherlands, in 2003, and the Ph.D. degree from Utrecht University, Utrecht, The Netherlands, in 2009, where his studies focused on sediment dynamics in the Rhine catchment over the past 10 000 years.

From 2008 to 2010, he worked as a Post-Doctoral Researcher with the Utrecht University studying peat deposits of the coastal and deltaic area of The Netherlands, determining the rate and magnitude of land subsidence and related greenhouse gas emissions. After working at the Dutch Geological Survey, he joined Deltares Research Institute in 2011, to work on land subsidence in coastal lowlands around the world. He is currently chairing the Land Subsidence Team, Deltares Research Institute, and is a part-time Senior Researcher with Utrecht University. In this team he continues his work on measuring, modeling, and understanding different components of land subsidence. He is a member of the United Nations Educational, Scientific and Cultural Organization (UNESCO) Land Subsidence International Initiative (LASII) working group.

Freek J. van Leijen (Member, IEEE) received the master's degree in geodetic engineering from the Delft University of Technology, Delft, The Netherlands, in 2002. His thesis concerned stochastic modeling of the tropospheric variability using Global Positioning System (GPS) and InSAR observations.

In 2003, he rejoined the Delft University of Technology as a Ph.D. Student, working on InSAR research. He has developed a range of software algorithms for persistent scatterer interferometry. After working for five years at SkyGeo, Delft, a commercial company delivering radar remote sensing services, he returned to the Delft University of Technology in 2013 to continue his research in the field of satellite radar interferometry.

Ramon F. Hanssen (Senior Member, IEEE) received the M.Sc. degree in geodetic engineering and the Ph.D. degree *(cum laude)* from the Delft University of Technology, Delft, The Netherlands, in 1993 and 2001, respectively.

He was with the International Institute for Aerospace Survey and Earth Science (ITC), Stuttgart University, Stuttgart, Germany; the German Aerospace Center (DLR), Weßling, Germany; Stanford University, Stanford, CA, USA, as a Fulbright Fellow; and the Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA, involved in microwave remote sensing, radar interferometry, signal processing, and geophysical application development. Since 2008, he has been an Antoni van Leeuwenhoek Professor in Earth observation with the Delft University of Technology, where he has been leading the research group on mathematical geodesy and positioning since 2009. He has authored a textbook on radar interferometry.