Identification of deformation pattern changes caused by enhanced oil recovery (EOR) using InSAR

Ling Chang\textsuperscript{a}, Ou Ku\textsuperscript{b} and Ramon F. Hanssen\textsuperscript{c}

\textsuperscript{a}Department of Earth Observation Science, University of Twente, Enschede, The Netherlands; \textsuperscript{b}SkyGeo Netherlands B.V, Delft, The Netherlands; \textsuperscript{c}Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands

**ABSTRACT**
Continuous hydrocarbon production and steam/water injection cause compaction and expansion of the reservoir rock, leading to irregular downward and upward ground movements. Detecting such anthropogenic ground movements is of importance, as they may significantly influence the safety and sustainability of hydrocarbon production activities, in particular, enhanced oil recovery (EOR) and even lead to local hazards, e.g. earthquakes and sinkholes. As InSAR (Interferometric Synthetic Aperture Radar) can routinely deliver global ground deformation observations on a weekly basis, with millimetre-level precision, it can be a cost-effective, and less labour intensive tool to monitor surface deformation changes due to hydrocarbon production activities. Aimed at identifying the associated deformation pattern changes, this study focuses on InSAR deformation model optimization, in order to automatically detect irregularities, both spatially and temporally. We apply multiple hypothesis testing to determine the best model based on a library of physically realistic canonical deformation models. We develop a cluster-wise constrained least-squares estimation method for parameter estimation, in order to directly introduce contextual information, such as spatio-temporal correlation, into the mathematical model. Here a cluster represents a group of spatially correlated InSAR measurement points. Our approach is demonstrated over an enhanced oil recovery site using a stack of TerraSAR-X images.

1. Introduction
Petroleum reservoir compaction and expansion – due to steam/water injection – may unveil itself as ground surface subsidence and uplift, which has significant and often negative effects on wells, production efficiency, and safety. Consecutively monitoring ground deformations and early identification of anomalies in the expected deformation patterns can prevent damage or even disasters. For this goal, many in-situ and satellite-based techniques can be applied. For instance, in-situ instruments such as GPS (Global Positioning System) and tilt meters can be deployed at predefined locations where the changes in the behaviour would be most harmful, and deliver valuable observations at those specific locations.
However, anomalies may unexpectedly occur at locations where no in-situ instruments are installed. In addition, it is costly and laborious to collect continuous observations with in-situ instruments. To overcome those limitations, satellite Interferometric Synthetic Aperture Radar (InSAR) (Bamler and Hartl 1998; Ferretti, Prati, and Rocca 2001; Hanssen 2001) is a viable complementary or even alternative technique. InSAR is able to monitor ground movements and the associated deformation time series with competitive millimetre-level precision over wide areas. Many historical SAR data can be retrieved back to the early 1990s. With a relatively short revisit cycle, e.g. up to 3 day for Sentinel-1a/b (Torres et al. 2012), InSAR can deliver the relevant observations in a near-real time manner.

In principle, InSAR deformation time series can be parameterized in order to detect spatial and temporal pattern changes. However, the parameterization of the deformation time series is not straightforward, when the temporal (and spatial) deformation behaviour is not very systematic. For example, a temporal model may suffice for many years of continuous oil production, but fail when other activities, such as steam injection, start to play a role. In that case, the parameterization should be modified. A complicating factor in this requirement for reparameterization is the intrinsic 2 phase ambiguity (Bamler and Hartl 1998). Consequently, a sudden change in the deformation behaviour over an EOR site may be (erroneously) accommodated by a 2 phase jump, or by a change in the functional model parameterization. There is no unique solution for this problem unless additional constraints are introduced to the problem. Therefore, in this case, i) we analyse the variogram to determine the spatial and temporal correlation of adjacent and subsequent points, respectively, and categorize spatio-temporally correlated adjacent points as a cluster; ii) we apply cluster-wise multiple hypothesis testing to select the best deformation model, cf. Chang and Hanssen (2015); and iii) we develop a cluster-wise constrained least-squares estimation method to estimate the unknowns, thereby identifying changes in the deformation pattern. Compared to the point-wise method, which focused on fitting a temporal model per point, the cluster-wise method has the advantage to include spatial correlation in the functional and stochastic model, and use contextual information to constrain the model estimation.

This paper is organized as follows. In section 2, we briefly review the concept of the hypothesis testing theory, and present the principle of constrained least squares estimation. Section 3 describes the geological situation of our area of interest (AoI) and the SAR data used, followed by the results and discussion in Section 4. The conclusions are drawn in Section 5.

2. Method

The key to identify the deformation pattern changes is first to properly select the best deformation model. Given several potential best deformation models, we use the multiple hypothesis testing to determine the most probable one, see detail in Chang and Hanssen (2015). We briefly review this method in section 2.1. Then the parameters for the best model can be estimated by a constrained least squares estimation method, in which the contextual knowledge (indicating e.g. the signal smoothness either in space
or time) is treated as pseudo observations to constrain the mathematical model, see section 2.2.

2.1. Multiple hypothesis testing

Chang and Hanssen (2015) proposed a multiple hypothesis testing method for optimal deformation model selection per point. In this study, we improve this method towards deformation modelling per cluster. A cluster is a group of spatio-temporally correlated InSAR measurement points which is defined by variogram analysis. Suppose there are \( b \) points \([p_1, p_2, \cdots, p_b]\) within a cluster and \( m \) epochs in the time series, the mathematical model for the null hypothesis \( H_0 \) is expressed as

\[
E\{y\} = \begin{bmatrix}
A^{p_1} & \cdots & A^{p_m}
\end{bmatrix}
\begin{bmatrix}
x^{p_1} \\
\vdots \\
x^{p_b}
\end{bmatrix}, \quad D\{y\} = Q_{yy},
\]

where \( E\{.\} \) and \( D\{.\} \) represent the expectation and dispersion operator, respectively. The \( bm \times 1 \) observation vector is denoted by \( y = [y_{p_1}, y_{p_1}^1, \cdots, y_{p_m}^1, y_{p_1}^2, \cdots, y_{p_m}^2, \cdots, y_{p_1}^b, y_{p_2}^b, \cdots, y_{p_m}^b]^T \). Here the underscore indicates that the observations are random variables. The \( bm \times bm \) covariance matrix \( Q_{yy} \) describes the observation noise of all points within the cluster. The diagonal elements of \( Q_{yy} \) are the variances of all points, while the off-diagonal elements represent the spatio-temporal correlation of the points. For instance, \( Q_{y_{p_1}, y_{p_2}} \) represents the variance of point \( p_1 \) at the \( m \) th epoch of the deformation time series, and \( Q_{y_{p_i}, y_{p_j}} \) represents the spatial correlation of point \( p_1 \) and \( p_2 \) at the \( m \) th epoch of the deformation time series. The design matrix \( A \) is composed of the design matrices for \( b \) points, \( A^{p_1}, A^{p_2}, \ldots, A^{p_b} \). The number of unknowns \( x = [x^{p_1}, x^{p_2}, \ldots, x^{p_b}]^T \) is \( n \), where \( n = n^{p_1} + n^{p_2} + \cdots + n^{p_b}, \ (n^{p_i}, i \in [1, b], \) is the number of unknown parameters for point \( i \). In case of identical unknowns of all point within a cluster, that is \( x^{p_1} = x^{p_2} = \cdots = x^{p_b} \), the number of unknowns \( x \) reduces to \( n/b \).

Against the null hypothesis \( H_0 \), several alternative models should be investigated. To determine the most probable model per cluster, we test the null hypothesis \( H_0 \) against many alternative hypotheses \( H_j \), \( \forall j \), expressed as

\[
H_0 : \quad E\{y\} = A x , \quad D\{y\} = Q_{yy},
\]

\[
H_j : \quad E\{y\} = A x + C_j \nabla_j, \quad \nabla_j \neq 0; \quad D\{y\} = Q_{yy}.
\]

Here a specification matrix \( C_j \) and an additional new vector of unknown parameters \( \nabla_j \) with length of \( q \) are introduced in \( H_j \).

An Overall Model Test (OMT) (Teunissen, Simons, and Tiberius 2005) is applied to test for the correctness of the null hypothesis \( H_0 \). The test statistic for the OMT is
\[ T_{q=bm-n} = \hat{e}_0^T Q_{yy}^{-1} \hat{e}_0, \]  

(3)

where \( \hat{e}_0 \) is the vector of residuals between the observations and the functional model under \( H_0 \), that is \( \hat{e}_0 = y - A\hat{x} \) in which \( \hat{x} \) is estimated by Equation (8). If \( T_{q=bm-n} \) is greater than the predefined critical value \( \chi^2_{\alpha}(q_j, 0) \) (given a level of significance \( \alpha \)), \( H_0 \) will be rejected. In favour of the rejection of \( H_0 \), test ratio values of all alternatives \( H_j \)s are calculated to find the most probable model. The test ratio is expressed as

\[ T^j_{q_j} = \frac{\hat{e}_0^T Q_{yy}^{-1} \hat{e}_0 - \hat{e}_j^T Q_{yy}^{-1} \hat{e}_j}{\chi^2_{\alpha}(q_j)}, \]  

(4)

where \( q_j \) is the degree of freedom for the alternative hypothesis \( H_j \), \( \hat{e}_j \) is the residual between the observations and the functional model under \( H_j \). The most probable model \( H_B \) among all hypotheses is assumed to be

\[ \hat{T}_{q_B}^B = \max_j \{ T^j_{q_j} \}, \quad (\hat{T}_{q_B}^B > 1). \]  

(5)

### 2.2. Constrained least squares estimation

Information on the signal smoothness, e.g. by deciding to estimate identical unknowns for all points within a cluster, or by defining a specific change pattern in space (e.g. circular Gaussian-distributed pattern), can be introduced by defining pseudo observations for the parameter estimation. This can be generally expressed as

\[ E(d) = G x ; \quad D(d) = Q_{dd}, \]  

(6)

where the \( s \times 1 \) pseudo observation vector is denoted by \( d \). The noise of the pseudo observation vector is expressed in \( Q_{dd} \). Constrained by Equation (6), the unknowns for \( H_0 \) in Equation (2) are estimated by

\[ \hat{x} = (A^T Q_{yy}^{-1} A + G^T Q_{dd}^{-1} G)^{-1}(A^T Q_{yy}^{-1} y + G^T Q_{dd}^{-1} d), \]  

(7)

following the criterion \( \| A \hat{x} - y \|^2_{Q_{yy}^{-1}} + \| G \hat{x} - d \|^2_{Q_{dd}^{-1}} = \min \) (Teunissen, Simons, and Tiberius 2005). As such an estimation method requires that both the norm of the residuals of the InSAR observations and the residuals of the pseudo signal observations is minimal, we refer to it as constrained least squares estimation (CLSE). Likewise, the CLSE solution for \( H_j \) in Equation (2) is calculated by

\[ \left[ \begin{array}{c} \hat{x} \\ \hat{y} \end{array} \right] = \left( \begin{array}{c} A^T \\ C^T \end{array} \right) Q_{yy}^{-1} \left[ \begin{array}{c} A C \\ + G^T Q_{dd}^{-1} G \end{array} \right]^{-1} \left( \begin{array}{c} A^T \\ C^T \end{array} \right) Q_{yy}^{-1} y + G^T Q_{dd}^{-1} d. \]  

(8)

### 3. Geological situation of the AoI and data description

To investigate the performance of the cluster-wise multiple hypothesis testing and constrained least squares estimation, a \( 500 \times 500 \) m area of interest is chosen. The AoI, indicated by red in Figure 1, is located in a desert area in this hydrocarbon...
field there are more than 12,000 active oil/gas wells (DOGGER 2016). For instance, a row of pumpjacks for oil wells at this site is depicted in the upper-right inset in Figure 1. Continuous large-volume oil/gas extraction from shallow depths leads to severe reduction in pore pressure and soil compaction. The surface settlement has been reported by, e.g. Chase Jr, and Dietrich (1989); de Rouffignac et al. (1995); Dale, Narahara, and Stevens (1996); Fredrich et al. (2000); van der Kooij and Mayer (2002). Moreover, due to the local injection activities for the underground reservoir (Patzek 1992), the surface presents upward movements, as it has been shown in the literature such as Fielding, Blom, and Goldstein (1998); Wolhart et al. (2005).

We use 22 TerraSAR-X SAR data (Stripmap mode) acquired between day 0 and day 253 from an ascending orbit to demonstrate the proposed method. The radar sensor operates in X-band with 31 mm wavelength. The spatial resolution is 3 m in range and azimuth direction. The repeat cycle of TerraSAR-X satellite is 11 days.

4. Results and discussion

Due to the very irregular alternation between times of uplift (due to steam injection) and times of subsidence (due to oil production), the displacement time series of scatterers will be very non-stationary. As a consequence, assuming a single parametric model for the entire time series will yield two problems related to the high residues between the model and the observations. First, in estimating the coherence of the scatterer, model

Figure 1. The $500 \times 500$ m AoI indicated by the red square in an optical Landsat image. This site is located in an arid area. The upper-right photo shows an example of a row of pumpjacks for oil wells.
imperfections can be misinterpreted as loss of coherence, leading to the erroneous rejection of coherent points, and missing the signal of interest. Second, the model imperfections may lead to phase unwrapping errors, and effectively bias the results to conservative (smooth) models.

Using information from local domain experts, reporting that steam injection started day 190 we performed multi-epoch InSAR processing for the first 15 SAR images acquired between day 0 and day 165, with a crop size of $30 \times 50$ km. Within the $500 \times 500$ m AoI, 6321 points are detected for which the values of the estimated temporal coherence are larger than 0.3 (Touzi et al. 1999; Hanssen 2001). The deformation time series for those coherent points are estimated accordingly. Using the remaining 7 SAR images acquired between day 198 and day 253 the subsequent deformation time series of those 6321 points are estimated. It is suspected that some of the detected points may show abrupt deformation changes in relation to the reported injection activities. Hence, we actively reduce the likelihood for unwrapping errors by relaxing the deformation gradient threshold. During the processing, all 22 image data are aligned, and the 30 m resolution SRTM DEM (van Zyl 2001) is used to remove the topographic phase contribution. The first SAR acquisition is assigned to be the equivalent master image, which serves as the null-survey for the consecutive cumulative displacement estimates. Atmospheric influence is assumed to be marginal within the $500 \times 500$ m AoI.

Figure 2 shows the cumulative deformation maps over this area of interest, as well as the velocity map, shown in the bottom-right. Every sub-map is draped over the

![Figure 2](image_url)

**Figure 2.** Cumulative deformation maps over the $500 \times 500$ m AoI and the linear velocity map. Every sub-map is draped over the incoherently average amplitude image. The first 22 maps show the cumulative deformation, and the bottom-right sub-figure shows the corresponding linear velocity map (between day 0 and day 253). The sudden uplift on day 198 is indicated by the black circle. Note that the colour bar ranges are different between the cumulative maps and the linear velocity map. No motion is assumed at the first (reference) acquisition.
incoherently averaged amplitude image. It shows that the area indicated by the black circle experienced a sudden upward movement on day 198. The velocity map (bottom-right) is clearly not capable to detect such an atypical change in the deformation behaviour.

In order to identify the specific nonlinear spatial and temporal patterns over this AoI, we apply the method discussed in section 2. Acknowledging the expected temporal jumps caused by hydrocarbon extraction, the library of canonical deformation models includes the linear model, the outlier model, the step model, and the breakpoint model. We then determine the most probable model combination for every cluster of points using cluster-wise multiple hypothesis testing. The linear deformation model is considered as the null hypothesis, while all other models are tested as alternative models, i.e. linear plus single/bi-/tri-steps, linear plus outliers, and single/bi-/tri-breakpoints. To find the times of the steps, outliers, or breakpoints, in an alternative hypothesis, we use an epoch-by-epoch progressive search method. All time options are sequentially tested for each epoch except for the first and last acquisition. Hence, the number of the alternative hypotheses can be easily in the order of hundreds. Benefiting from computing only the test ratio as in Equation. (5), rather than estimating all parameters for each evaluated model, the computational load is significantly reduced. The level of significance is initially set to $1/(2 \times 22)$, or 2%, and the discriminative power of test is assumed to be 50%, c.f. de Heus et al. (1994); Chang and Hanssen (2015).

The spatio-temporal variogram of the deformations, see Figure 3(a) and 3(b), derived from all 22 acquisitions, shows increasing values for increasing spatial distances and temporal intervals, due to the spatial and temporal variability of the deformation signal. Fitting a Gaussian function to the empirical variogram yields a maximum spatial correlation distance of 305 m for day 209, see the red dashed line in Figure 3(c), where the variogram value is 11.6 mm$^2$. It has a nugget of 8.8 mm$^2$. By using a maximum allowable variability of 9 mm$^2$, we find that the corresponding spatial distance is 17 m, indicated by the black cross in Figure 3(c). Hence, points within this maximum distance are

![Figure 3.](image)
assumed to behave identically. We uniformly divide the AoI into a regular grid of 30 × 30 cells, to arrive at a cell size of 16.7 m.

Using the information that the points within a cluster behave in concert, the unknowns of each point are therefore identical. We formulate this as pseudo observations to constrain the parameter estimation. Applying our method, the results of cluster-wise multiple hypothesis testing show that 324 out of 900 clusters sustain the null hypothesis, as shown in a transparent colour, while 546 clusters follow the linear plus single/bi-/tri-steps deformation model, indicated in purple in Figure 4(a). In yellow, the 17 clusters are indicated that follow a linear plus single/bi-/tri-steps plus single/bi-/tri-breakpoints model, and in green, the 13 clusters that follow a linear plus single/bi-/tri-breakpoints model. The number of steps per cluster is depicted in Figure 4(b), which shows that the number of steps increases towards the centre of the AoI. The step size and step location in time and space for the clusters with linear + single/bi-/tri-step(s) model are shown in Figure 5. There are 503 clusters showing an instantaneous uplift on day 198, most likely due to steam/water injection activities. The sequential subsiding for 306 clusters on day 209 and 196 clusters on day 220 responded to the shut down of the well around day 200.

As an example, the parameter estimation results of a cluster are shown in Figure 6. The location of this cluster is indicated by the white star in Figure 4. There are 6 points in this cluster. The deformation time series of these 6 points are shown in Figure 6(a) in the different colours. For this cluster, the estimated linear velocity under the null hypothesis is $-84.1 \text{ mm year}^{-1}$. As the test statistic of the OMT in Equation (3) is larger than the critical value, the null hypothesis is rejected and cluster-wise multiple hypothesis testing is performed. The alternative hypothesis, with a linear plus bi-steps model, is found to be the best model for this cluster, which is shown in black. According to the CLSE, the estimated parameters, the velocity $\hat{v}$, the first step $\hat{\Delta}_1$ on day 198, and the second step $\hat{\Delta}_2$ on day 209 are $-114.7 \text{ mm year}^{-1}$, $+40.2 \text{ mm}$, and $-25.4 \text{ mm}$, respectively. Figure 6(b) gives the temporal evolution of these 6 points. It can show that the spatial pattern of this cluster, in which all points were smoothly subsiding till day 165, then they had a sudden upward movement on day 198 followed by downward movement. Comparison with the velocity estimators under the null hypothesis $H_0$ and the best alternative hypothesis $H_B$ shows that we may overestimate or underestimate the linear deformation if we stick to the null hypothesis, and an erroneous interpretation of the

**Figure 4.** (a) Deformation model classification. (b) The number of steps per cluster. The white five star indicates the location of the cluster, which will be discussed in Figure 6.
Figure 5. Step size and step location in time and space for the clusters following linear + single/bi-/tri-step(s) model.

Figure 6. (a) Deformation time series of 6 points within a cluster. (b) Temporal evolution of these 6 points. The location of this cluster is indicated by the white five star in Figure 4. The estimated velocity \( \dot{\nu} \) is \(-114.7 \text{ mm year}^{-1}\), the first step \( \Delta_1 \) on day 198, and the second step \( \Delta_2 \) on day 209 are \(+40.2 \text{ mm}\) and \(-25.4 \text{ mm}\), respectively.

deformation time series may occur. Finding the best alternative hypothesis improves the reliability of the estimators and unveils the deformation pattern anomalies.
5. Conclusions

Spatial and temporal deformation pattern anomalies in relation to hydrocarbon production can be detected by cluster-wise multiple hypothesis testing and constrained least squares estimation (CLSE). Particularly, using cluster-wise multiple hypothesis testing, we can determine the most probable time series model per cluster in a probabilistic and efficient way, and the CLSE system has the advantage to directly involve contextual knowledge such as spatio-temporal correlation in both a functional and stochastic structure, and considers such contextual knowledge as pseudo observations to constrain the model. The study on a hydrocarbon extraction area demonstrates that our method can better model the InSAR deformation time series, and identify sudden uplift or subsidence, at most about 4 cm, due to hydrocarbon production activities.

Acknowledgments

The authors would like to thank SkyGeo Netherlands B.V. for their collaboration and the German Space Agency (DLR) for providing the archived TerraSAR-X data.

Disclosure statement

No potential conflict of interest was reported by the authors.

References


