Jemil Butt*, Andreas Wieser, and Stefan Conzett

Intrinsic random functions for mitigation of atmospheric effects in terrestrial radar interferometry

DOI 10.1515/jag-2016-0042
Received October 8, 2016; accepted February 14, 2017

Abstract: The benefits of terrestrial radar interferometry (TRI) for deformation monitoring are restricted by the influence of changing meteorological conditions contaminating the potentially highly precise measurements with spurious deformations. This is especially the case when the measurement setup includes long distances between instrument and objects of interest and the topography affecting atmospheric refraction is complex. These situations are typically encountered with geo-monitoring in mountainous regions, e.g. with glaciers, landslides or volcanoes.

We propose and explain an approach for the mitigation of atmospheric influences based on the theory of intrinsic random functions of order k (IRF-k) generalizing existing approaches based on ordinary least squares estimation of trend functions. This class of random functions retains convenient computational properties allowing for rigorous statistical inference while still permitting to model stochastic spatial phenomena which are non-stationary in mean and variance. We explore the correspondence between the properties of the IRF-k and the properties of the measurement process. In an exemplary case study, we find that our method reduces the time needed to obtain reliable estimates of glacial movements from 12 h down to 0.5 h compared to simple temporal averaging procedures.

Keywords: Atmospheric correction, intrinsic random functions, terrestrial radar interferometry, monitoring in alpine environment, Geostatistics

***Corresponding author: Jemil Butt,** ETH Zürich, Institute of Geodesy and Photogrammetry, Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland, e-mail: jemil.butt@geod.baug.ethz.ch
Andreas Wieser, ETH Zürich, Institute of Geodesy and Photogrammetry, Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland, e-mail: andreas.wieser@geod.baug.ethz.ch
Stefan Conzett, Terra Vermessungen AG, Obstgartenstrasse 7, CH-8006 Zürich, Switzerland, e-mail: conzett@terra.ch

1 Introduction

Terrestrial radar interferometry (TRI) is a technology providing spatiotemporally dense measurements for the quantification of geometric surface changes along the line-of-sight over distances up to a few km. This is not only of immediate practical interest in applications like structural health monitoring, operation and safeguarding of open pit mines or monitoring of rockfalls endangering critical infrastructure but could also facilitate better understanding of dynamic processes underlying large-scale geological natural hazards through scientific measurements.

TRI is therefore regularly deployed in mountainous regions, e.g. to survey glaciers [15], to assess the likelihood of geologically predisposed areas becoming active landslides [4] or to observe the flanks of volcanoes for deformation patterns indicating an increase in activity [14]. Even though TRI, as a remote sensing technology, has certain advantages over classical point-based geodetic techniques due to its high sensitivity to small surface displacements, inherently areal sampling and remote operability without need for any in-situ components, it shares with them some of their limitations.

Like for any other type of measurement, TRI data consist of a signal part and a noise part. The latter is further decomposable into a spatiotemporally highly autocorrelated component having its origins in the essentially unpredictable meteorological changes along the propagation path, and a second, spatiotemporally uncorrelated, component subsuming thermal noise, crosstalk between electronic circuits and movements of the surveyed object on very short length- or timescales or in manners otherwise inaccessible to systematic analysis.

This second component becomes directly visible in the interferograms at locations corresponding to regions with weak backscattering. Particularly, it appears in the form of unsystematic disturbances seemingly adhering to the restrictive probability laws of white noise and is therefore easily modeled stochastically and amenable to statistical standard treatments.

The noise component corresponding to atmospheric influences is more problematic. In addition to the atmospheric phase screen (APS) reaching magnitudes and spa-

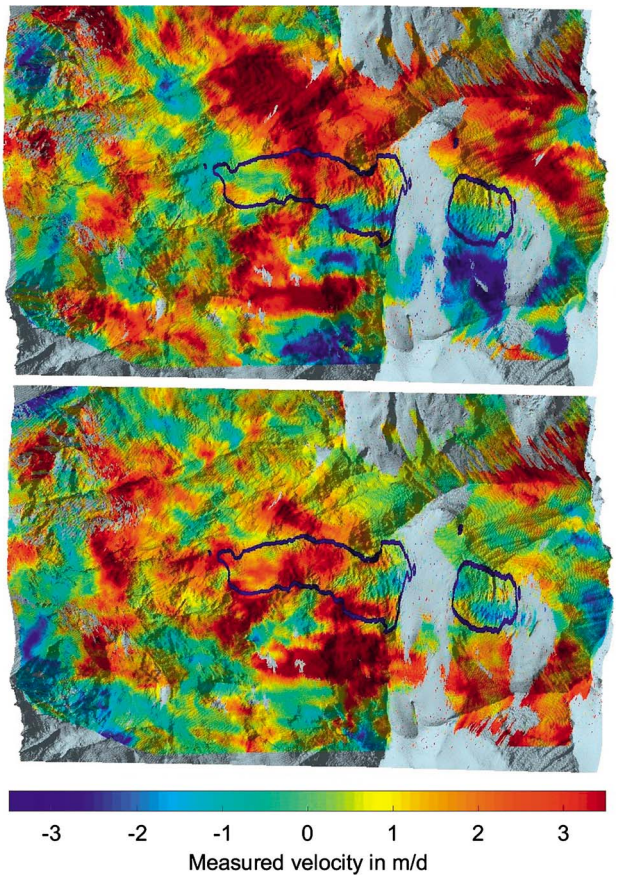


Figure 1: Two successive 2-minute interferograms, in which the APS masks the deformations that are actually limited to the area outlined in black.

tial extents allowing it to completely mask the underlying signal (see Fig. 1), its stochastic properties also depend on the topography and are consequently non-constant because the APS associated with a point $t \in T$ (e.g., $T = \mathbb{R}^3$) is the result of an integration of differential atmospheric effects along the line joining t and the instrument position. Accordingly, the APS cannot be second order stationary [5, p. 253], and thus widely known inference methods like simple Kriging hold no optimality properties [7, p. 352]. This raises the question of stochastically optimal inference of an instationary APS.

This inference problem has been approached from different perspectives. Most of them employ a blend of (i) deterministic relations between APS, refractive index and meteorological quantities and (ii) stochastic relations of the deterministically unmodelled residuals to deduce the APS from the measurements. A purely deterministic approach was investigated in [9], where the authors gathered meteorological data and predicted the atmospheric phase delay using tools from weather forecasting. An alternative to the strictly deterministic viewpoint consists in mixing a polynomial model for the APS with a multiple regression model meant to explicitly incorporate the information from altitude and phase measurements on known stable points into the coefficients of the polynomial APS [11]. Presupposing less regularity of the APS and placing more emphasis on the structure found in the data leads the authors of [4] to estimate the APS over the area potentially containing displacements with a spatial low-pass filter.

It is in recognition of the irregular movement patterns and highly variable meteorological properties of air in mountainous terrain that we adopt a data-driven viewpoint similar to the one proposed in [4]. In section 2 an interpretation of the APS as an intrinsic random function will be presented and followed in section 3 by a scheme allowing rigorous statistical inference for intrinsic random functions in form of what is called the BLIE (Best Linear Intrinsic Estimator) in the geostatistical literature (see [12]). Section 4 contains results validating the proposed correction method in the context of a monitoring campaign targeting an alpine glacier in southern Switzerland.

2 A mathematical model for the APS

The area potentially containing moving objects will be assumed known. A set of persistent scatterers (PS) lying outside of it has to be extracted from the data for later use as reference points with good backscattering characteristics and signal-to-noise ratio that provide reliable information regarding the APS over stable regions. To this end, the amplitude dispersion index (ADI) described for example in [13] may be used although care must be taken to eliminate any PS detected in unstable areas. The goal is then to derive for all locations $t \in T$ the least squares estimator minimizing the expected quadratic error $E[(\hat{A}(t) - A(t))^2]$ which represents the deviation between predicted and true APS, and to quantify the uncertainties associated with the resulting estimates. Acknowledging the effect of noise unaccounted for otherwise, we will refrain from employing a strict interpolation procedure and instead perform estimation and smoothing simultaneously using a coherence-based noise parameter.

The APS itself will be considered a random field

$$A(\cdot) : T \ni t \mapsto A(t) \in L^2(\Omega) \tag{1}$$

where $L^2(\Omega)$ denotes the Hilbert space of square integrable random variables on the probability space Ω (see for example [7, p. 25]). The above definition is to be interpreted

as $A(\cdot)$ being an assignment of a random variable to each location $t \in T$. Measurements $M(\cdot)$ on the PS will be demanded to satisfy

$$M(\cdot) = A(\cdot) + N(\cdot) \tag{2}$$

with noise $N(\cdot)$ being a zero-mean random field with diagonal covariance matrix Σ_N and $A(\cdot)$ and $N(\cdot)$ uncorrelated. To model drifts in variance and expected value of $A(\cdot)$, the usual regularity assumption of second order stationarity (s.o.s.) of $A(\cdot)$ is dropped and replaced by the weaker [12, p. 440] assumption of $A(\cdot)$ being an intrinsic random function of order k (IRF-k).

IRF-k's can be understood as random functions, for which a differential operator ∇^{k+1} of order $k+1$ exists, such that $\nabla^{k+1}A(\cdot)$ satisfies $E[\nabla^{k+1}A(t)] = 0 \ \forall t \in T$ and $E[\nabla^{k+1}A(t)\nabla^{k+1}A(s)] = C(t-s) \ \forall s,t \in T$, where $C : T \times T \to \mathbb{R}$ is a covariance function unaffected by any translation $\Gamma_{t_0} : T \ni s \mapsto s + t_0 \in T$ acting jointly on locations s and t:

$$C(\Gamma_{t_0}(s) - \Gamma_{t_0}(t)) = C(s + t_0 - t - t_0) = C(s - t). \tag{3}$$

As this implies $\nabla^{k+1}A(\cdot)$ to have translation invariant first and second moment functions, a more succinct way of stating $A(\cdot)$ to be an IRF-k consists of stating $\nabla^{k+1}A(\cdot)$ to be s.o.s.

Direct physical meaning can be attributed to the case $k = 0$ as it corresponds to the assumption of $\nabla A(\cdot)$ being s.o.s., even though $A(\cdot)$ itself, as a sum of atmospheric phase variations along wave propagation paths of different lengths, will exhibit an instationary variance depending on the distance between instrument and targeted area. As its variance is not translation invariant, $A(\cdot)$ can not be s.o.s., although the supposition of s.o.s. derivatives of $A(\cdot)$ seems justifiable due to a lack of visible drifts in the variance of velocities differentiated with respect to the range direction (see Fig. 2) and will make rigorous inference possible.

Arguments and definitions will be made more formal in the following section adhering closely to notations laid out in [5, pp. 238–270] and loosely to the theory presented in [12]. A small glossary listing mathematical symbols and their meaning is supplied at the end of the paper to further aid the reader.

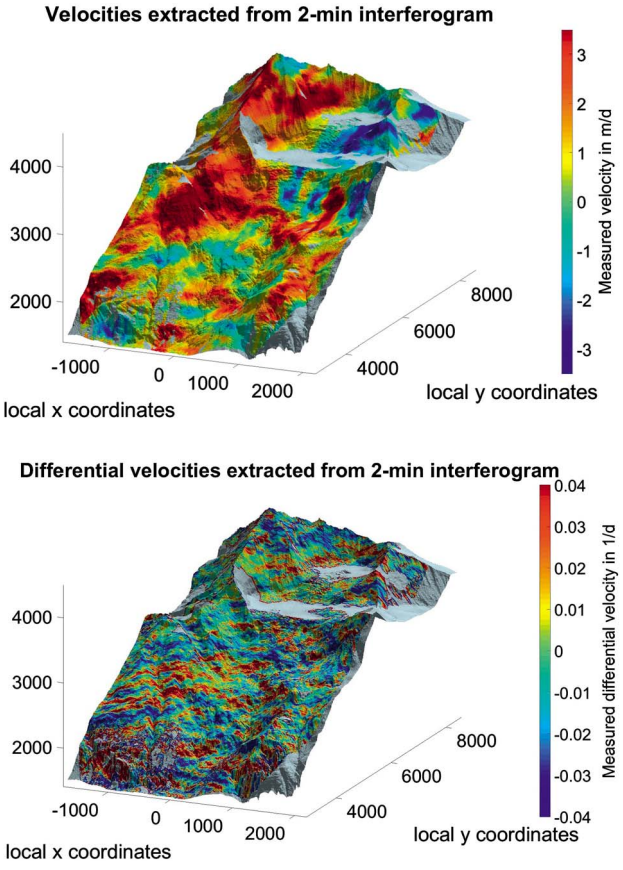


Figure 2: Measured displacement velocities and their derivative in range-direction. The latter seems to be zero-mean with a homogeneous correlation function.

3 Inference in the space of IRF-k

3.1 Intrinsic random functions

Let δ be the Dirac measure and $\delta_t, t \in T$ denote its translate, i.e., $\delta_t(T') = 1$ if $t \in T' \subset T$ and $\delta_t(T') = 0$ otherwise. Then the discrete measure $\lambda = \sum_i \lambda_i \delta_{t_i}$, $\lambda_i \in \mathbb{R}$ freely chosen, acts on a function $X : T \to U$ via

$$X_\lambda := \sum_i \int_T \lambda_i X(s) \delta_{t_i}(ds) = \sum_i \lambda_i X(t_i).$$

If U is either $L^2(\Omega)$ or $\mathbb{R}$ and X therefore a random field or a normal, scalar function, define the action of λ on X around t as

$$X_\lambda(t) := \sum_i \int_T \lambda_i X(s) \delta_{t+t_i}(ds) = \sum_i \lambda_i X(t + t_i). \tag{4}$$

The space Λ^k of discrete measures allowable at order k consists of those λ for which all polynomials $P^k : T \to \mathbb{R}$ of degree at most k vanish:

$$\lambda \in \Lambda^k \Leftrightarrow \sum_i \lambda_i P^k(t + t_i) = 0 \quad \forall \ t \in T$$

whereby the number of λ_i is almost arbitrary. This emulates closely polynomials of degree k being in the kernel of the differential operator of order $k+1$. Such X_λ are then

called allowable linear combinations of order k, or ALC-k for short. If $\lambda \in \Lambda^k$ then polynomial drifts of $X(\cdot)$ up to order k in mean and up to order $2k$ in variance are mapped to 0 by λ according to

$$\begin{aligned} X(t) &:= Y(t) + P^k(t) \\ X_\lambda(t) &= Y_\lambda(t) + \underbrace{P^k_\lambda(t)}_{0 \text{ since } \lambda \in \Lambda^k} = Y_\lambda(t) \end{aligned}$$

$$\begin{aligned} E[X(s)X(t)] :=& \underbrace{K(s,t)}_{\text{only function of } t-s} + \sum_{i\leqslant k} P^i(s) f_i^1(t) \\ &+ \sum_{j\leqslant k} P^j(t) f_j^2(s) + \sum_{i,j\leqslant k} P^i(s) P^j(t) \\ E[X_\lambda(s)X_\lambda(t)] =& K_{\lambda\lambda}(s,t) + \sum_{i\leqslant k} P^i{}_\lambda(s) f_i^1{}_\lambda(t) \\ &+ \sum_{j\leqslant k} P^j{}_\lambda(t) f_j^2{}_\lambda(s) + \sum_{i,j\leqslant k} P^i{}_\lambda(s) P^j{}_\lambda(t) \\ =& \underbrace{K_{\lambda\lambda}(s,t)}_{\text{only function of } t-s} \end{aligned}$$

where $K_{\lambda\lambda}(s,t)$ is understood as a notational extension of eq. (4) equivalent to

$$K_{\lambda\lambda}(s,t) = \sum_{i,j} \lambda_i \lambda_j K(t+t_i, s+s_j).$$

$X_\lambda(t)$ is therefore free of the influences of those polynomial drifts and would be stationary if these were the only terms inciting instationary behavior of $X(\cdot)$. If indeed this is the case and the random field $X_\lambda(\cdot) : T \ni t \mapsto X_\lambda(t) \in L^2(\Omega)$ is zero-mean and s.o.s. for $\lambda \in \Lambda^k$:

$$\begin{aligned} E[X_\lambda(t)] &= 0 && \forall t \in T \\ E[X_\lambda(s)X_\lambda(t)] &= C(t-s) && \forall s,t \in T \end{aligned} \tag{5}$$

then $X(\cdot)$ is called an IRF-k. The special case $k = 0$ is again particularly instructive. If X exhibits a linear trend in variance as typical for integrated s.o.s. processes [5, p. 253], the X_λ satisfies the equations $E[X_\lambda] = 0$ and $E[X_\lambda(s)X_\lambda(t)] = K_{\lambda\lambda}(t-s)$ and is s.o.s. for all $\lambda \in \Lambda^0$ (e.g. λ the discrete derivative). Consequently, X is an IRF-0.

For explanatory purposes we will showcase the necessary constructions and the derivation of the BLIE for the APS $A(\cdot)$ in the 1D case only. The extensions required to handle data in more than one dimension are straightforward (see [12]) and any non-obvious adjustments to the procedure – necessary for the application to 2D random fields in sec. 4 – will only originate from the polar geometrical nature of the measurements. To avoid unnecessary clutter of symbols, the linear map $L^\lambda : F \ni f(\cdot) \mapsto f_\lambda(\cdot) \in F$, with F some space of functions, will also be denoted by λ. No confusion should arise due to the different meanings of λ as a measure and the linear map that maps a function f to f_λ.

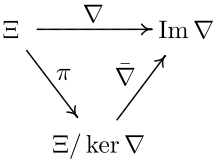
We will now define two specific measures ∇ and $\int$ that are inverse to each other in a certain sense to be made explicit later. In section 3.3 they will be the basis for a formalization of the idea that differencing in range direction can make interferograms s.o.s. Let T be a set of pixel indices now, Ξ the space of all random fields, and

$$\begin{aligned} \nabla &: \Xi \ni A(\cdot) \mapsto A_\nabla = A(\cdot) - A(\cdot - 1) \in \Xi \\ \textstyle\int &: \Xi \ni A(\cdot) \mapsto A_\int(\cdot) = \sum_{k=0}^{\cdot - a} A(a+k) \in \Xi \end{aligned}$$

where $A(\cdot) : T \ni t \mapsto A(t) \in L^2(\Omega)$, $T \subset \mathbb{Z}$ is a discrete (1D) random field and a is an arbitrary integer constant corresponding to the lower limit of integration. A short calculation shows

$$\begin{aligned} \nabla \textstyle\int A(\cdot) &= \nabla \left(\sum_{k=0}^{\cdot - a} A(a+k) \right) \\ &= \sum_{k=0}^{\cdot - a} A(a+k) - \sum_{k=0}^{\cdot - a - 1} A(a+k) \\ &= A(\cdot) \\ \textstyle\int \nabla A(\cdot) &= \textstyle\int (A(\cdot) - A(\cdot - 1)) \\ &= \sum_{k=0}^{\cdot - a} A(a+k) - A(a+(k-1)) \\ &= A(\cdot) - A(a-1) \end{aligned} \tag{6}$$

so that ∇ and $\int$ are not strictly inverses on Ξ. However, analogous to integration and differentiation in real analysis, $\int \nabla$ differs from the identity function id_Ξ only by $A(a-1)$ which is an element of $\ker \nabla := \{X \in \Xi : X_\nabla = 0\}$ and can be considered as giving rise to the identity on the quotient space $\Xi/\ker\nabla$. By the universal property of quotient spaces [1, p. 89] $\exists! \bar{\nabla}$ such that



$$\begin{aligned} \pi : \Xi \ni X &\mapsto \bar{X} \in \Xi/\ker\nabla \\ \nabla &= \bar{\nabla} \circ \pi \\ \bar{X} &= \{Y \in \Xi : Y - X \in \ker \nabla\} \end{aligned}$$

commutes, that is $\bar{\nabla} \circ \pi X = \nabla X \; \forall X \in \Xi$. Here, as in $\bar{X}$, the overbar is understood to indicate the equivalence class of X under the relation $X \sim Y \Leftrightarrow X - Y \in \ker\nabla$ and $\Xi/\ker\nabla$ is the space of equivalence classes with addition $\bar{X} + \bar{Y} = \overline{X+Y}$ and scalar multiplication $\alpha\bar{X} = \overline{\alpha X}, \alpha \in \mathbb{R}$. Since $\bar{\nabla}\bar{X} = \nabla X$ is well defined according to $X \sim Y \Rightarrow \nabla(X-Y) = 0 \Leftrightarrow X_\nabla = Y_\nabla$ and by the first isomorphism theorem for

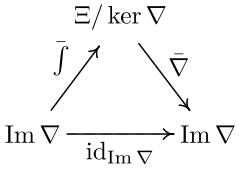
modules [1, p. 89] $\operatorname{Im}\nabla \cong \Xi/\ker\nabla$ we surmise that $\bar{\nabla}$ might be an isomorphism between $\operatorname{Im}\nabla$ and $\Xi/\ker\nabla$. Indeed it is

$$\begin{aligned} &\text{surjective by} && \bar{\nabla}(\Xi/\ker\nabla) = \bar{\nabla}\circ\pi(\Xi) = \operatorname{Im}\nabla \\ &\text{injective by} && \bar{X} \neq \bar{Y} \Rightarrow X - Y \notin \ker\nabla \\ & && \Rightarrow \nabla X \neq \nabla Y \Rightarrow \bar{\nabla}\bar{X} \neq \bar{\nabla}\bar{Y} \end{aligned}$$

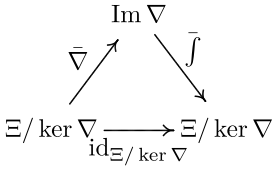
and lastly a homomorphism by

$$\begin{aligned} \bar{\nabla}(\bar{X}+\bar{Y}) &= \nabla(X+Y) = \bar{\nabla}\bar{X} + \bar{\nabla}\bar{Y} \\ \bar{\nabla}(\alpha\bar{X}) &= \nabla\alpha X = \alpha\bar{\nabla}\bar{X}. \end{aligned}$$

Furthermore, $\bar{\int} = \pi\circ\int : \operatorname{Im}\nabla \to \Xi/\ker\nabla$ serves as a two sided inverse homomorphism. This means that the following diagrams commute.



$$\begin{aligned} \bar{\nabla}\bar{\int}X(\cdot) &= \bar{\nabla}\pi\int X(\cdot) \\ &= X(\cdot) \end{aligned}$$



$$\begin{aligned} \bar{\int}\bar{\nabla}\bar{X}(\cdot) &= \pi\int\nabla X(\cdot) \\ &= \bar{X}(\cdot) \end{aligned}$$

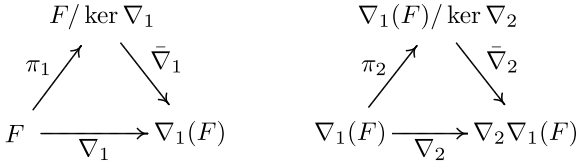
As of now $\bar{\nabla} : \Xi/\ker\nabla \to \operatorname{Im}\nabla$ is identified as an isomorphism with inverse $\bar{\int} : \operatorname{Im}\nabla \to \Xi/\ker\nabla$. This implies in fact the recapturability of the equivalence class of $A(\cdot)$ from $A_\nabla(\cdot)$, since $\bar{A}(\cdot) = \overline{\int}A_\nabla(\cdot)$ satisfies $\bar{\nabla}\bar{A}(\cdot) = A_\nabla(\cdot)$ meaning that $\bar{A}$ is the equivalence class of all solutions $A(\cdot)$ to $\nabla A(\cdot) = A_\nabla(\cdot)$ with $A_\nabla(\cdot)$ given. By exchanging the space of functions on $T\times T$ for Ξ in the quotient space constructions, it is possible to show that equivalence classes $\bar{\sigma}$ of covariance functions are recoverable from C in the same way equivalence classes $\bar{A}$ of random functions are from A_∇. They turn out to be not only the minimum requirement for best linear estimation, but also directly inferrable from the data by extending the inversion procedure on the quotient space $\Xi/\ker\nabla$ to $F/\ker\nabla\nabla$ and applying it to a parametric model of C.

3.2 Derivation of the BLIE

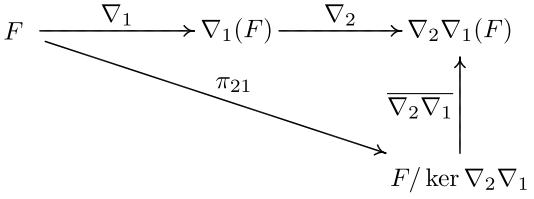
The main impediment complicating statistical inference for instationary random fields is the fact that the covariance function σ depends not only on the difference between pixels but also on the location, making it impossible to infer it from one realization of a random field. However, it will turn out that for optimal estimation the instationary covariance is in fact not strictly required; the equivalence class of its stationary part, termed generalized covariance (GC) in [12], will be sufficient. The GC in turn depends only on the equivalence class of the random field, enabling reliable estimation in presence of drift in mean and variance. Since $A_\nabla(\cdot)$ can be assumed stationary, the covariance function of the ALC-0 $A_\nabla(\cdot)$

$$E\left[A_\nabla(s)A_\nabla(t)\right] = \sigma_{\nabla\nabla}(t,s) = C(t-s)$$

is stationary and can be inferred from the data. Employing arguments from section 3.1 it can be shown that for $\nabla_1\sigma = \sigma_{\nabla\delta}$ and $\nabla_2\sigma = \sigma_{\delta\nabla}$ the inverse of $\overline{\nabla_2\nabla_1} : F/\ker\nabla_2\nabla_1 \to \nabla_2\nabla_1(F)$ is $\pi_{21}\int_1\int_2$ with F some function space containing σ and other quantities as defined below.



- π is the natural projection onto the quotient space
- $\bar{\nabla}\bar{\sigma} = \nabla\sigma$; $\bar{\sigma}^i$ equiv. class of σ in $F/\ker\nabla_i$
- $\nabla_i(F)$ is space of functions of type $\nabla_i\sigma$, $\sigma\in F$



Checking that they are inverses is simple:
For $C\in\nabla_2\nabla_1(F)$, $\nabla_1 e = d$, $\nabla_2 d = C$

$$\begin{aligned} i)\quad \overline{\nabla_2\nabla_1}\pi_{21}\int_1\int_2 C &= \overline{\nabla_2\nabla_1}\pi_{21}\int_1(d+q_2) \\ &= \overline{\nabla_2\nabla_1}\pi_{21}(e+\underbrace{q_1}_{\in\ker\nabla_1} + \int_1\underbrace{q_2}_{\in\ker\nabla_2}) \\ &= \overline{\nabla_2\nabla_1}\bar{e}^{21} = \nabla_2\nabla_1 e = C \end{aligned}$$

$$\begin{aligned} ii)\quad \pi_{21}\int_1\int_2\overline{\nabla_2\nabla_1}\bar{e}^{21} &= \pi_{21}\int_1\int_2\nabla_2(\nabla_1 e) \\ &= \pi_{21}\int_1\nabla_1 e + q_2 \\ &= \pi_{21}(e+q_1+\int_1 q_2) = \bar{e}^{21} \end{aligned}$$

This directly translates to the formula

$$\bar{\sigma}^{21} = \pi_{21}\int_1\int_2 C \tag{7}$$

since $\nabla_2\nabla_1\sigma = \sigma_{\nabla\nabla} = C$ is known allowing inference not of the instationary covariance function but only of its equivalence class $\pi_{21}\sigma$. For any $\lambda\lambda : \sigma \mapsto \sigma_{\lambda\lambda}$ with $\ker\lambda\lambda \supseteq$

$\ker \nabla_2\nabla_1 \overline{\lambda\lambda} : F/\ker \nabla_2\nabla_1 \ni \bar{\sigma}^{21} \mapsto \sigma_{\lambda\lambda} \in \nabla_2\nabla_1(F)$ is well defined:

$$\begin{aligned}\sigma_1 \sim \sigma_2 &\Leftrightarrow \sigma_1 - \sigma_2 \in \ker \nabla_2\nabla_1 \subseteq \ker \lambda\lambda \\ &\Rightarrow \lambda\lambda(\sigma_1 - \sigma_2) = 0 \Leftrightarrow \lambda\lambda\sigma_1 = \lambda\lambda\sigma_2\end{aligned}$$

Thus for any $\lambda \in \Lambda^0$, $\sigma_{\lambda\lambda}$ only depends on the equivalence class $\bar{\sigma}^{21}$ which can be calculated from the stationary covariance C of the first derivatives of $A(\cdot)$. We estimate the atmosphere at $t_0 \in T$ linearly from n measurements:

$$\hat{A}(t_0) = \sum_{j=1}^{n} \lambda_j M(t_0 + s_j) = \lambda^T M$$

The error $\epsilon(t_0) = \hat{A}(t_0) - A(t_0) = A_{\tilde{\epsilon}} + N_\lambda$ with $\tilde{\epsilon} = \sum_{j=1}^{n} \lambda_j \delta_{s_j} - \delta$ is an instationary random field. Forcing the residual term $A_{\tilde{\epsilon}}$ to be drift independent and stationary is equivalent to adding the constraint $\tilde{\epsilon} \in \Lambda^0$ effectively constraining $\tilde{\epsilon}$ to eliminate constant polynomials translating to the equation $\sum_{i=1}^{n} \lambda_i - 1 = 0$. Setting $\bar{K} = \pi_{21} \int_1 \int_2 C$ and $C = \sigma_{\nabla\nabla}$, the error variance can be written as

$$\begin{aligned}E[\epsilon^2] &= E[(A_{\tilde{\epsilon}} - N_\lambda)^2] \overset{A \perp\!\!\!\perp N}{=} E[A_{\tilde{\epsilon}}^2] + E[N_\lambda^2] \\ &= \sigma_{\tilde{\epsilon}\tilde{\epsilon}}(t_0, t_0) + \lambda^T \Sigma_N \lambda \\ &\overset{\tilde{\epsilon} \in \Lambda^0}{=} K_{\tilde{\epsilon}\tilde{\epsilon}}(t_0, t_0) + \lambda^T \Sigma_N \lambda \\ &= \lambda^T \underbrace{(K_{ij} + \Sigma_N)}_{K} \lambda - 2\lambda^T K_{t_0} + K(t_0, t_0) \\ \left\{K_{ij}\right\}_{kl} &= K(t_0 + s_k, t_0 + s_l) \\ \left\{K_{t_0}\right\}_k &= K(t_0 + s_k, t_0) \qquad k, l = 1, \ldots, n\end{aligned}$$

Minimizing $E[\epsilon^2]$ subject to $\sum_{j=1}^{n} \lambda_j - 1 = 0$ leads to the system of equations [2, pp. 84–86]

$$\begin{aligned}K\lambda + \underline{1}\mu &= K_{t_0} \\ \underline{1}^T\lambda &= 1\end{aligned} \tag{8}$$

where μ is a single Lagrange-multiplier and $\underline{1} \in \mathbb{R}^n$ is a vector of ones. Successive substitution ultimately yields

$$\begin{aligned}\hat{A}(t_0) &= \sum_{j=1}^{n} \lambda_j M(t_0 + s_j) \\ \lambda &= K^{-1}\left(K_{t_0} - \underline{1}(\underline{1}^T K^{-1} \underline{1})^{-1}(\underline{1}^T K^{-1} K_{t_0} - 1)\right) .\end{aligned} \tag{9}$$

This estimator can be seen to coincide with ordinary Kriging (see [6] for a similar formulation) but with the GC K instead of the stationary covariance C and an additional term accounting for the effects of noise. How this formula is to be adapted for practical application under consideration of coordinate dependent aspects will be discussed in the next section.

3.3 Geometrical considerations and practical implementation

Let $s \in T$ be a point and let $c_s = (x_s, y_s)^T$ and $p_s = (r_s, \varphi_s)^T$ denote Cartesian and polar coordinates of this point, respectively. The coordinate transform given by

$$\phi : [0, \infty) \times [0, 2\pi) \ni p_s \mapsto c_s \in \mathbb{R}^2$$

$$\phi(p_s) = \begin{bmatrix} r_s \cos(\varphi_s) \\ r_s \sin(\varphi_s) \end{bmatrix} \quad \phi^{-1}(c_s) = \begin{bmatrix} \sqrt{x_s^2 + y_s^2} \\ \operatorname{atan}(y_s/x_s) \end{bmatrix}$$

maps the polar coordinates of a point s to its Cartesian coordinates. If $f(\cdot)$ is a function from T to $\mathbb{R}$ then $f^c(\cdot)$ and $f^p(\cdot)$ will be the corresponding functions acting on the Cartesian and polar coordinates satisfying $f(s) = f^c(c_s)$ and $f(s) = f^p(p_s)$. The identity $f(s) = f^c(c_s) = f^c(\phi p_s)$ implies the transformation laws $f^c \circ \phi = f^p$ and $f^p \circ \phi^{-1} = f^c$. An analogous statement holds for functions with multiple inputs for which the coordinate transform is applied to each input separately, e.g.:

$$\begin{aligned}\sigma^c \circ (\phi, \phi) &= \sigma^p \\ \sigma^c &= \sigma^p \circ (\phi^{-1}, \phi^{-1})\end{aligned}$$

Following the line of argumentation outlined in section 2, the derivative of the APS $A(\cdot)$ in line-of-sight may be regarded as s.o.s. It will be numerically approximated by the discrete derivative of $A(\cdot)$ in range direction. Introducing the measure ∇ as below, $A_\nabla(\cdot) \approx (\partial/\partial r)A(\cdot)$ and a linear map $\int$ can be found such that the relation between ∇ and $\int$ closely resembles that of ∇ and $\int$ as defined in the 1D case in section 3.1.

Note, however, that apart from the errors introduced by discretization, further sources of uncertainty exist:

1. The derivative in range direction is used instead of the inaccessible derivative in line-of-sight.
2. Height information is not used although it might in reality have a significant role to play in the determination of atmospheric correlations.
3. The covariance of A_∇ is treated as though it were stationary even if it is likely to be height dependent and empirically known only at ground level.

Given the lack of reliable information regarding the stochastic properties of differential phase delay and the ill-posedness of the estimation problem, none of these simplifying neglections can be proven justified within the framework in which the BLIE is valid. For the time being we will just assume their validity and not try to find a more faithful stochastic model. However, ∇ and $\int$ still need to

be defined in terms of coordinates to provide a link between image geometry and stochastic aspects of the measurements:

$$\nabla : A(\cdot) \mapsto A_\nabla(\cdot) = \frac{1}{\Delta r}\left[A^p(p_\cdot) - A^p(p_\cdot - \Delta)\right]$$
$$\int : A(\cdot) \mapsto A_\int(\cdot) = \sum_{k=0}^{\frac{r_\cdot - r_a}{\Delta r}} A^p(p_\cdot + k\Delta)\Delta r$$

Here $\Delta = (\Delta r, 0)^T$, Δr is the range difference between two pixels and $a \in T$ is a point undetermined apart from its membership to the line joining $\cdot$ and the instrument. Closedness of p_T under translation by Δ is presumed only to avoid problems of the definitions in the vicinity of the boundary of p_T.

Calculations mirroring those in section 3.2 show that $\bar{\nabla}$ and $\overline{\int}$ are isomorphisms between $\Xi/\ker\nabla$ and $\operatorname{Im}\nabla$ and $\pi_{21}\int_1\int_2$ and $\overline{\nabla_2\nabla_1}$ are isomorphisms between $F/\ker\nabla_2\nabla_1$ and $\operatorname{Im}\nabla_2\nabla_1$ for $F \ni \sigma$ and $\nabla_2\nabla_1\sigma = \sigma_{\nabla_2\nabla_1}$. Thus one of the GC's K satisfying $K - \sigma \in \ker\nabla_2\nabla_1$, as required in the BLIE, can be seen to be

$$K(s,t) = \int_1\int_2 C^p(p_s, p_t) = \sum_{k=0}^{\frac{r_s - r_a}{\Delta r}} \sum_{j=0}^{\frac{r_t - r_b}{\Delta r}} C^p(p_a + k\Delta, p_b + j\Delta)\Delta r^2 \quad . \tag{10}$$

Here $a = (r_0, \varphi_s)^T$ and $b = (r_0, \varphi_t)^T$ are chosen to lie on the lower border of the interferogram. Equation (10) can be interpreted as assuring that the GC $K(s,t)$ can be computed as the double integral of the stationary covariance function C^c (C^p need not be stationary) along the two lines L_s, L_t joining points a, b to s, t (see Fig. 3).

We recognize equation (10) as a discretization of the integral representation of $K(s,t)$ in the continuous case:

$$\begin{aligned} K(s,t) &= \int_{L_s}\int_{L_t} C(u,v)dudv \\ &= \int_{r_0}^{r_s}\int_{r_0}^{r_t} C^p\left(\begin{bmatrix} r_1 \\ \varphi_s \end{bmatrix}, \begin{bmatrix} r_2 \\ \varphi_t \end{bmatrix}\right) dr_1 dr_2 \\ &= \int_{r_0}^{r_s}\int_{r_0}^{r_t} C^c\left(\phi\begin{bmatrix} r_1 \\ \varphi_s \end{bmatrix}, \phi\begin{bmatrix} r_2 \\ \varphi_t \end{bmatrix}\right) dr_1 dr_2 \end{aligned}$$

This can be understood intuitively. The total correlation between the APS associated to points s and t consists of the sum of all individual correlations between points in the atmosphere along the propagation paths L_s and L_t. Even though $K(s,t)$ is instationary, it differs from a stationary GC only by an element of $\ker\nabla_2\nabla_1$ as proven in [8, pp. 179–186].

To summarize our findings and tie them to a practically feasible algorithm, we propose the following scheme:

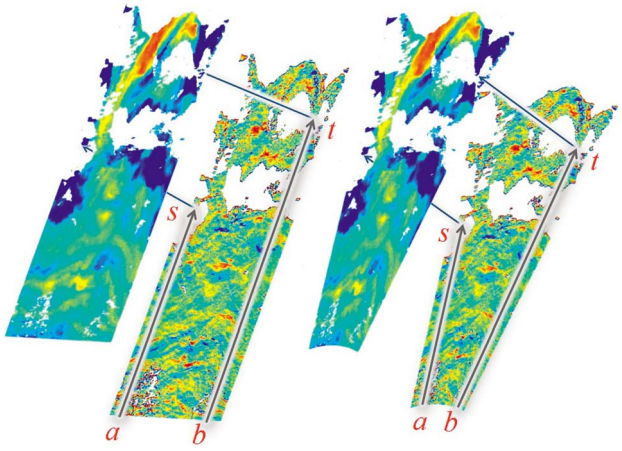


Figure 3: Geometric relations between A (top layer) and A_∇ (lower layer) are inherited by the covariance functions.

1. Choose PS lying outside the deformation area by thresholding on the ADI and using prior knowledge on stable areas.
2. Calculate A_∇ from a version of A previously smoothed with a median filter to reduce the impact of noise on the estimation of the derivative.
3. Infer $C^c(c_s, c_t) = E[A_\nabla(s)A_\nabla(t)]$ by minimizing the squared error between the empirical covariance and the parametric anisotropic stationary covariance model.
4. Calculate the GC $K(s,t)$ using eq. (10) and estimate Σ_N by using formulas linking coherence and signal-to-noise ratio (see e.g. [10, p. 98]) and a rough prior of the variance of the APS.
5. Estimate the APS at the unobserved locations using its noisy observations on the PS using eq. (9) and subtract it from the measurements to estimate the deformations.

We finally propose to perform these steps for a set of overlapping windows to lessen the computational burden and the impact of neglected but potentially existing instationarity in $C^c(c_u, c_v)$.

In our tests, step 3. proved to be problematic. In accordance to Fig. 2 the estimation of an anisotropic covariance function yielded very short correlation lengths in range-direction. This, however, affected unfavourably the robustness of the output and the runtime of the algorithm due to the demand for smaller steps in the numerical integration of C^c. To sidestep these challenges, isotropy of C^c was assumed during the production of the results presented in section 4. Additionally, r_0 was set to 0 and C^c chosen in a way to optimize the fit between K and the empirical covariance of $A(\cdot)$, circumventing occasionally unreliable inference of K from a single dataset.

4 Validation and conclusion

For testing and validation purposes we draw upon the data gathered during a 2014 measurement campaign in the alpine regions of southern Switzerland, see [3]. The dataset features rugged terrain with height differences exceeding 1 km, a spatiotemporally highly variable APS, and stable reference areas surrounding a rapidly moving glacier. Data were collected with a sampling interval of 2 minutes from a location across the valley, at a distance of up to 8 km from the monitored area. Even though movements locally reach 2 m/day, the influence of the APS can mask the signal to the point where atmospheric artifacts and real displacements become indistinguishable (see Fig. 4).

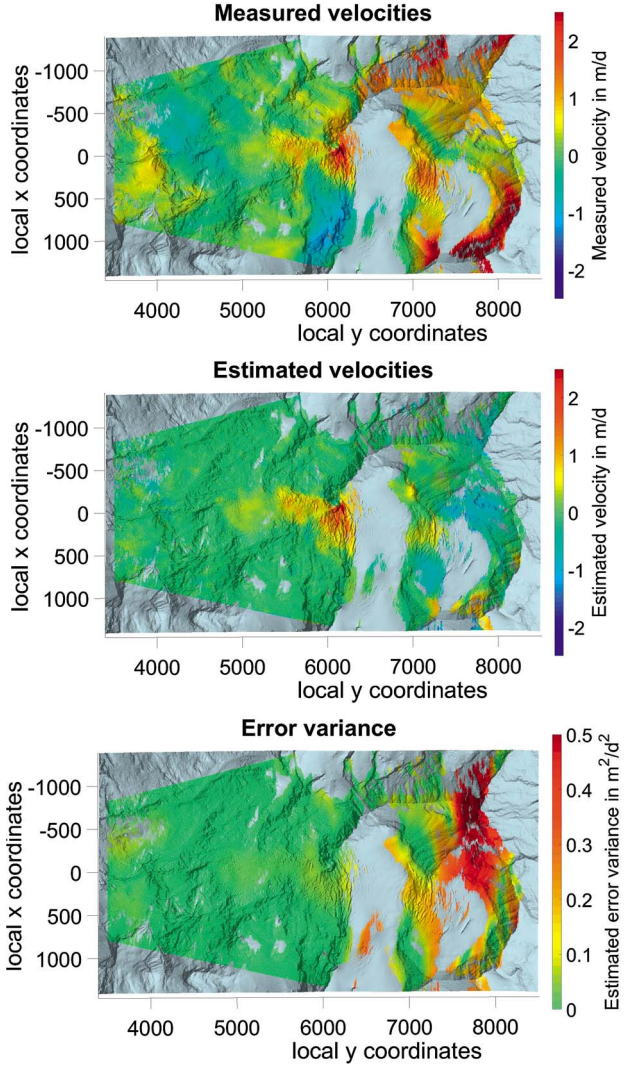


Figure 4: From top to bottom: Average of five two-minute interferograms, estimated velocities after subtraction of the APS and error variance.

For the estimation of the APS the phase values of approximately 3500 PS were used, the distribution of which was very sparse in the right fifth of the interferograms shown in Fig. 4 ($\approx$ 25PS/km^2) and increasingly dense towards their center ($\approx$ 350PS/km^2). As can be seen in Fig. 4, the correction scheme performs well in regions with a high amount of PS and produces poor results in regions with few PS (right) or bad coherence (left). This is also supported by the accompanying estimation of the error variance, which is helpful for quantifying the uncertainty associated with the APS correction procedure. It is worth noting that the left parts of the images contain the backscattering of objects in a distance of about 4 km to the instrument while for the right, topographically highly irregular parts, this distance is about 8 km. Consequently the error variances are significantly higher in the latter area.

Depending on the number of PS, the error variance for deformation estimation using 2-minute interferograms is in certain regions of a magnitude that rivals the amplitude of the signal. This can be explained with short correlation lengths induced by turbulent atmospheric behavior; averaging in time improves the stochastic properties of the residuals by mitigating temporal high-frequency components. Fig. 5 shows the average total errors derived from cross validation on the stable areas as a function of averaging time and size of the area without any PS in it.

Two obvious trends are clearly visible: larger averaging times lead to more reliable estimations in both cases and the larger the area without reliable APS-observations, the more uncertain the results. Less than 0.05 m/day average total error are reached within 20 minutes with our approach, whereas simple averaging in time needs 12 h, and the widely employed fitting of second order polynomials (not shown) around 2 h. Due to the stochastic nature of the estimator, we expect that high irregularity of the APS

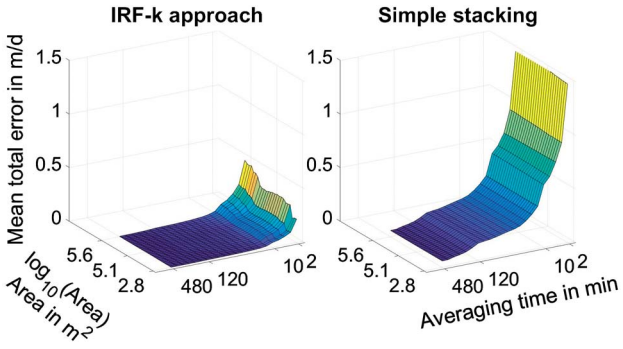


Figure 5: Average total errors of estimated deformations for the IRF-k approach and simple temporal averaging. Colouring added for better visual distinction between error levels.

and dense distribution of PS will further widen the gap between the method proposed here and comparable methods relying on deterministic assumptions.

We conclude the suitability of IRF-k's for mitigating the atmospheric phase screen and facilitating deformation monitoring based on interferograms heavily affected by instationary autocorrelated atmospheric noise as expected in mountaineous terrain. We will further pursue an approach generalizing IRF-k's to linear random functionals and reproducing kernel Hilbert spaces of distributions to better include geometric correlations induced by the topography and best represented via linear operators acting on stationary covariance functions in 3D space.

Glossary of mathematical terms and symbols

General notation

X^T	Transpose of X
$\hat{X}$	Estimator for X
$\bar{X}$	Equivalence class of X

Index sets and variables

T	Space, time or pixel indices
c_s	Cartesian coordinates of $s \in T$
p_s	Polar coordinates of $s \in T$
L_s	Line joining Radar and $s \in T$

Spaces

$L^2(\Omega)$	Space of square integrable random variables
Ξ	Space of random fields
Λ^k	Space of allowable measures
$\mathrm{Im}(f)$	Image of the map f
$\ker(f)$	Kernel of the map f
X/Y	Quotient space X by Y

Random fields

$M(\cdot)$	Measurements
$A(\cdot)$	Atmospheric phase screen
$N(\cdot)$	Noise
$\epsilon(\cdot)$	Estimation error

Measures and functions

δ	Dirac measure
λ	Arbitrary discrete measure
σ, K	Covariance functions
C	Translation invariant covariance function
P^k	Polynomial of up to order k
π	Canonical projection
ϕ	Coordinate transform
f^c	Function acting on cart. coord.
f^p	Function acting on pol. coord.

Matrices and operators

Σ_N	Covariance matrix of noise
E	Expectation operator
∇^{k+1}	Differential operator of order at most $k+1$
∇	Derivative in range direction
$\int$	Integration in range direction

Acknowledgment: The radar data used in this paper were collected in cooperation with Prof. Martin Funk, ETH Zürich, and Prof. Martin Truffer, University of Alaska, who also provided the instrument.

Funding: The Swiss Federal Office for the Environment (Bundesamt für Umwelt) has supported the data collection financially.

References

[1] Ash, R. B. (2013). Basic Abstract Algebra. Courier Corporation New York.

[2] Berlinet, A. and C. Thomas-Agnan (2011). Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer Science & Business Media Berlin Heidelberg.

[3] Butt, J., S. Conzett, M. Funk, and A. Wieser (2016). Terrestrial Radar Interferometry for Monitoring Dangerous Alpine Glaciers: Challenges and Solutions. In: *Proc. of GeoMonitoring*, Braunschweig, Germany.

[4] Caduff, R., A. Kos, F. Schlunegger, B. W. McArdell and A. Wiesmann (2014). Terrestrial radar interferometric measurement of hillslope deformation and atmospheric distrubances in the Illgraben debris-flow catchment, Switzerland. *IEEE Geoscience and Remote Sensing Letters*, Vol 11, No 2, pp. 434–438.

[5] Chilès, J. P. and P. Delfiner (2012). Geostatistics. Modeling Spatial Uncertainty. John Wiley & Sons New York.

[6] Cressie, N (1990). The origins of Kriging. *Mathematical Geology*, Vol 22, No 3, pp. 239–252.

[7] Christakos, G. (2013). Random Field Models in Earth Sciences. Elsevier Amsterdam.

[8] Gelfand, I. M. and N. Y. Vilenkin (1964). Generalized Functions: Applications of harmonic analysis. Academic Press Amsterdam Boston.

[9] Gong, W., F. Meyer, P. W. Webley, D. Morton, and S. Liu (2010). Performance analysis of atmospheric correction in InSAR data based on the weather Research and Forecasting Model (WRF). In: *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2900–2903.

[10] Hanssen, R. F. (2006). Radar Interferometry. Data Interpretation and Error Analysis. Springer Science & Business Media Berlin Heidelberg.

[11] Iglesias, R., X. Fabregas, A. Aquasca, J. J. Mallorqui, C. Lopez-Martinez, J. A. Gili, and J. Corominas (2014). Atmospheric phase screen compensation in ground-based SAR with a multiple-regression model over mountaineous regions. *IEEE Transactions on Geoscience and Remote Sensing*, Vol 52, No 5, pp. 2436–2449.

[12] Matheron, G. (1973). Intrinsic random functions and their applications. *Advances in Applied Probability*, Vol 5, No 3, pp. 439–468.

[13] Noferini, L., M. Pieraccini, D. Mecatti et al. (2005). Permanent scatterers analysis for atmospheric correction in ground-based DAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, Vol 43, No 7, pp. 1459–1470.

[14] Rödelsperger, S., M. Becker, C. Gerstenecker, G. Läufer, K. Schilling, and D. Steineck (2010). Digital elevation model with the ground-based SAR IBIS-L as basis for volcanic deformation monitoring. *Journal of Geodynamics*, Vol 49, Nos 3–4, pp. 241–246.

[15] Voytenko, D., T. H. Dixon, C. Werner et al. (2012). Monitoring a glacier in southeastern Iceland with the portable Terrestrial Radar Interferometer. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3230–3232.