Scalable Inference and Model Selection for Large-Scale Time Series with Complex Dependence Structures

Lionel Voirol

Ph.D. Public Defense

Prof. Olivier Scaillet (Chair)
Prof. Stéphane Guerrier (Thesis advisor)
Prof. Maria-Pia Victoria Feser
Dr. Chiara Montagna

Geneva School of Economics and Management
University of Geneva
Switzerland

UNIVERSITÉ DE GENÈVE
GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

Outline

Introduction

Modern sensing technologies and digital systems generate massive time series with complex dependence structures.

In many applications, the goal is to perform statistical inference.

Classical likelihood-based methods do not scale well, which limits the analysis of long signals and/or large sensor networks.

Practitioners often rely on approximations or ad hoc methods that reduce computational cost but provide limited or unclear statistical guarantees.

Thesis Objective

Develop scalable statistical methods with clear theoretical guarantees for reliable statistical inference on large dependent datasets.

Applications include inertial sensor calibration in engineering and the estimation of tectonic rates using collections of position time series from large-scale GNSS station networks.

Chapter I: Accounting for Vibration Noise in Stochastic Measurement Errors of Inertial Sensors

This chapter corresponds to the following published article:

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 72, 2024 2117

Accounting for Vibration Noise in Stochastic Measurement Errors of Inertial Sensors

Mucyo Karemera, Lionel Voirol, Davide A. Cucci, Wenfei Chu, Roberto Molinari, and Stéphane Guerrier

Karemera, M.*, Voirol, L.*, Cucci, D. A., Chu, W., Molinari, R., Guerrier, S., "*Accounting for Vibration Noise in Stochastic Measurement Errors of Inertial Sensors*", *IEEE Transactions on Signal Processing*, 72, 2117–2129, 2024.

* indicates co-first authors and authors are listed alphabetically.

Context

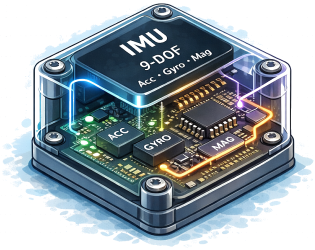
Inertial sensors are used in drones, smartphones and autonomous vehicles.

They measure acceleration and angular velocity at high frequency.

Their measurements are corrupted by errors that quickly accumulate, requiring the noise structure to be adequately modeled.

These signals and calibration data are then used in sensor-fusion algorithms, to estimate position, velocity, and attitude.

Measurement errors are generally modeled using composite stochastic models.



The GMWM estimator

Likelihood-based inference becomes computationally prohibitive for large and complex time series models.

The Generalized Method of Wavelet Moments[1] (GMWM) proposes to exploit the property of the Wavelet Variance (WV) to concisely capture the dependence structure of a time series in order to inverse the mapping between $\boldsymbol{\theta}_0$ and $\boldsymbol{\nu}(\boldsymbol{\theta}_0)$ to estimate $\boldsymbol{\theta}_0$:

Let $\varepsilon_1, \ldots, \varepsilon_n \sim F_{\boldsymbol{\theta}_0}$ be a realization of the data-generating process with true parameter $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2,$$

where:

- $\hat{\boldsymbol{\nu}}$ is the empirical wavelet variance[2] computed on $\varepsilon_1, \ldots, \varepsilon_n$
- $\boldsymbol{\nu}(\boldsymbol{\theta})$ is the model-implied wavelet variance
- $\boldsymbol{\Omega}$ is a positive definite weighting matrix

[1] Guerrier, Stéphane et al., 2013, *Journal of the American Statistical Association*

[2] Percival, Donald P., 1995, *Biometrika*

Context

The stochastic properties of device errors may vary with external conditions, such as temperature or motion[3].

The impact of these dynamics on measurement errors is studied in controlled environments, such as rotation tables.

Vibration noise arises from internal mechanics, motion dynamics, or external conditions, and contaminates the measurement errors of sensing devices.

This type of perturbation is not properly captured by standard stochastic error models, leading to inaccurate calibration and degraded navigation performance.

[3] Stebler, Yannick et al., 2015, *IEEE Transactions on Instrumentation and Measurement*; Radi, Ahmed et al., 2018, *Gyroscopy and Navigation*

Chapter Objective

1. Define a stochastic model that adequately captures vibration perturbation
2. Derive its wavelet variance for GMWM-based estimation
3. Establish the theoretical properties of the GMWM when incorporating this additional component

Sinusoidal processes

We consider a stochastic periodic process that we parameterize as follows:

$$S_t = \alpha \sin(\beta t + U) \tag{1}$$

with:

- amplitude $\alpha \in \mathbb{R}_+ = (0, +\infty)$,
- angular frequency $\beta \in (0, \pi]$ and
- phase $U \sim \mathcal{U}(0, 2\pi)$.

To make use of the GWMM, we need the theoretical WV of the processes defined in (1).

We derive the theoretical WV of S_t at scale j denoted as $\nu_j(\alpha, \beta)$:

$$\nu_j(\alpha, \beta) = \frac{\alpha^2 \left\{1 - \cos\left(\frac{\beta \tau_j}{2}\right)\right\}^2}{\tau_j^2 \left\{1 - \cos(\beta)\right\}}$$

where τ_j denotes the length of the wavelet filter at the j^{th} level of decomposition.

Model

We consider a general class of composite stochastic model:

$$\varepsilon_t = W_t + Q_t + \sum_{k=1}^{K} Y_t^{(k)} + D_t + R_t + \sum_{l=1}^{L} S_t^{(l)},$$

where

- W_t is a White Noise,
- Q_t is a Quantization Noise,
- $Y_t^{(k)}$ is the k^{th} AR(1) process,
- D_t is a Deterministic Drift,
- R_t is a Random Walk and
- $S_t^{(l)}$ is the l^{th} sinusoidal process.

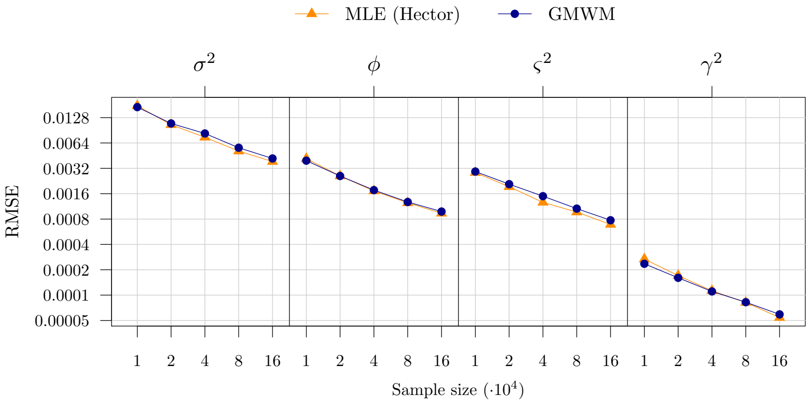


Figure: RMSE of the estimated parameters for the MLE (orange line) and the GMWM (blue line) for sample sizes $T = T_i \cdot 10^4$ with $\{T_1, \dots, T_5\} = \{1, 2, 4, 8, 16\}$

Simulation Studies

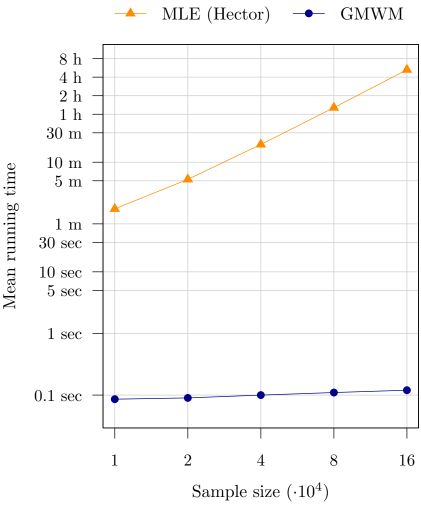


Figure: Mean running time of the MLE (orange line) and the GMWM (blue line) for sample sizes $T = T_i \cdot 10^4$ with $\{T_1, \ldots, T_5\} = \{1, 2, 4, 8, 16\}$

We consider two signals collected from a GMAT IMU under two settings.

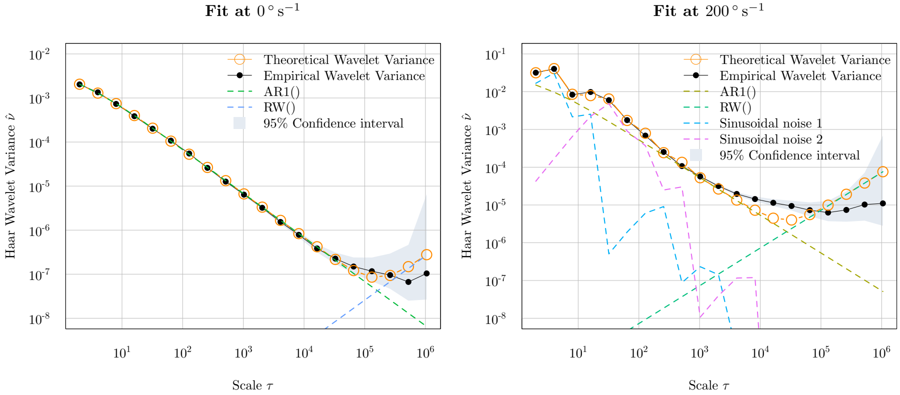


Figure: Empirical and model-implied wavelet variance of the estimated models for signals collected on a static IMU and when rotating at $200\,^\circ\,\mathrm{s}^{-1}$

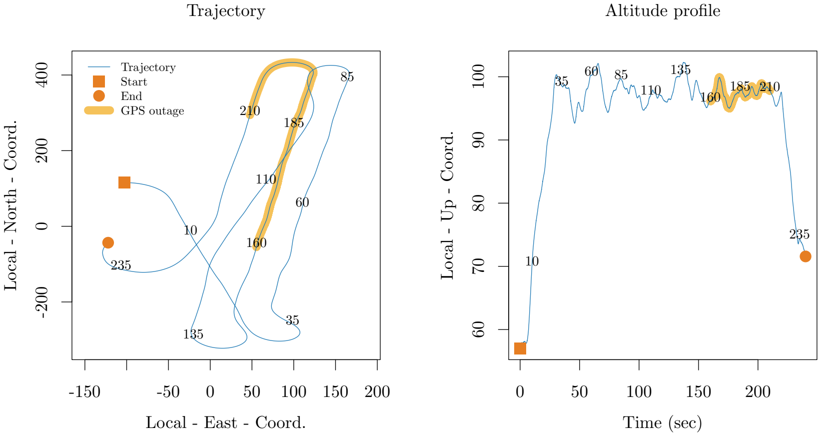


Figure: Trajectory in the emulation study

Emulation Study

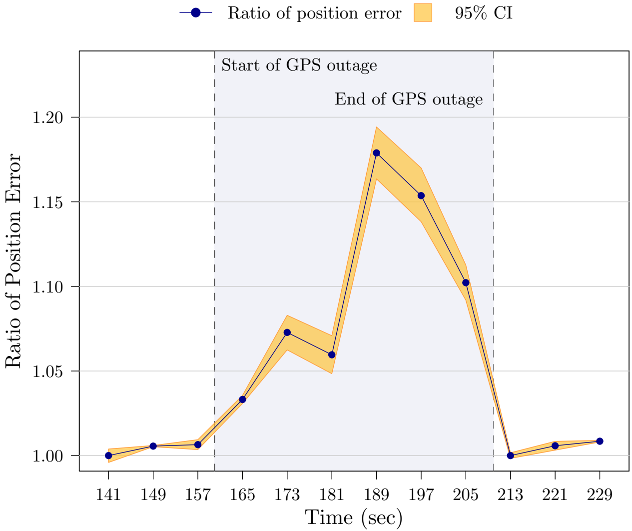


Figure: Mean ratio of position error

Chapter II: The Generalized Method of Wavelet Moments with eXogenous inputs

This chapter corresponds to the following published article:

Journal of Geodesy (2023) 97:14
https://doi.org/10.1007/s00190-023-01702-8

ORIGINAL ARTICLE

Check for updates

The Generalized Method of Wavelet Moments with eXogenous inputs: a fast approach for the analysis of GNSS position time series

Davide A. Cucci[1] · Lionel Voirol[1] · Gaël Kermarrec[2] · Jean-Philippe Montillet[3,4] · Stéphane Guerrier[1,5]

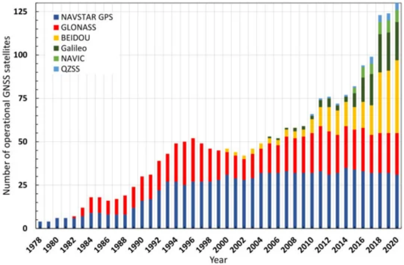
Received: 22 June 2022 / Accepted: 26 December 2022 / Published online: 6 February 2023

Cucci, D. A.*, Voirol, L.*, Kermarrec, G., Montillet, J.-P., Guerrier, S., "*The Generalized Method of Wavelet Moments with eXogenous inputs: A Fast Approach for the Analysis of GNSS Position Time Series*", *Journal of Geodesy*, 97(2), 2023.

* indicates co-first authors and authors are listed alphabetically.

The Global Navigation Satellite System (GNSS) is an important tool to study multiple geophysical phenomena[4] such as long-term tectonic movements[5], fault strain accumulation, post-glacial rebound[6], hydrological loading[7] and crustal deformations[8].

Networks of permanent stations record sequences of position measurements from various satellite constellations.



[4] Bock, Yehuda et al., 2016, *Reports on Progress in Physics*; Herring, T. A. et al., 2016, *Reviews of Geophysics*
[5] Bock, Yehuda et al., 2016, *Reports on Progress in Physics*
[6] Milne, Glenn A et al., 2001, *Science*
[7] Bevis, Michael et al., 2002, *Marine Geodesy*
[8] Williams, Simon, 2003, *Journal of Geophysical Research: Solid Earth*

Context

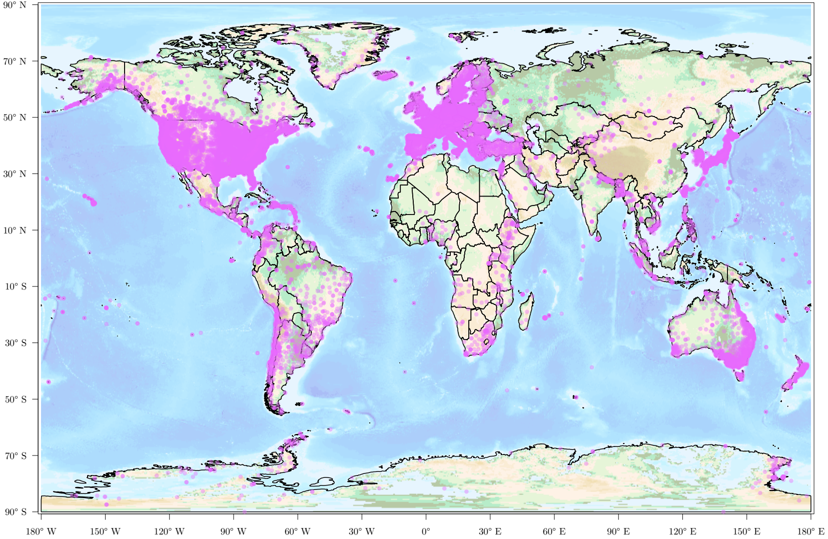


Figure: GNSS Stations processed by the Nevada Geodetic Laboratory

Context

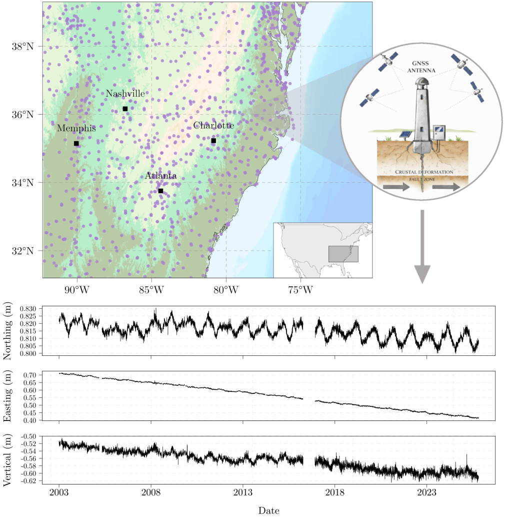


Figure: GNSS network along the U.S. east coast with example station and position time series

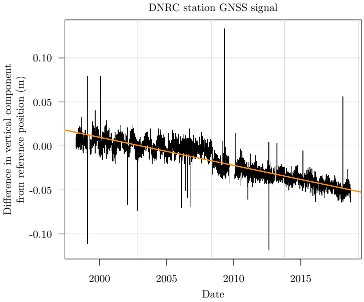


Figure: GNSS position time series and estimated trend

A general formulation of the models typically used to model GNSS signals can be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)\right\}, \tag{2}$$

where

- $\mathbf{Y} \in \mathbb{R}^n$ is the response vector,
- $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a fixed design matrix,
- $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is the vector of unknown parameters,
- $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is a zero-mean residual vector, with covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$, that depends on the unknown parameter vector $\boldsymbol{\gamma}_0 \subset \mathbb{R}^q$. This matrix does not have a block diagonal structure and can be non-Toeplitz.

Hence, we define

$$\boldsymbol{\theta}_0 = \begin{bmatrix} \boldsymbol{\beta}_0^{\mathrm{T}} & \boldsymbol{\gamma}_0^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \boldsymbol{\Theta} \subset \mathbb{R}^{p+q}.$$

Trajectory Model:

With t_i representing the i^{th} time point, the mean function capturing trend, seasonal variations, offsets, and post-seismic relaxations is written as[9]:

$$\begin{aligned}\mathbb{E}[Y_i] = a + b\left(t_i - t_0\right) + \sum_{h=1}^{m}\Big\{c_h \sin\left(2\pi f_h t_i\right) + d_h \cos\left(2\pi f_h t_i\right)\Big\} + \\ \sum_{j=1}^{r} e_j H\left(t_i - t_j\right) + \sum_{k=1}^{s} l_k \left\{1 - \exp\left(\frac{-(t_i - t_k)}{\tau_k}\right)\right\} H\left(t_i - t_k\right),\end{aligned} \tag{3}$$

for $i \in \{1, \ldots, n\}$, and with $\boldsymbol{\beta}_0 = \Big[a, b, c_1, \ldots, c_m, d_1, \ldots, d_m, e_1, \ldots, e_r, l_1, \ldots, l_s\Big]^{\mathrm{T}}$.

Stochastic Model:

The stochastic component is modeled as a combination of stochastic processes, typically coupling white noise with correlated processes such as Matérn or power-law processes.

[9] Langbein, John, 2008, *Journal of Geophysical Research: Solid Earth*; Bos, M. S. et al., 2008, *Journal of Geodesy*; Williams, Simon, 2003, *Journal of Geophysical Research: Solid Earth*

Maximum Likelihood Estimation: Computational Challenges

Inference is typically performed via Maximum Likelihood Estimation (MLE).

The MLE is implemented in softwares such as `CATS`, `Est_noise`, and `Hector`[10].

Each evaluation of the likelihood requires inversion of the $n \times n$ matrix $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$.

Computational complexity: $\mathcal{O}(n^\delta), \ \delta \in [2, 3]$.

In practice, the analysis of complex geodynamic processes requires to estimate signals from thousands of GNSS stations which record daily observations over decades and where different noise models must be tested[11].

This procedure which has to be performed routinely becomes impractical due to the large amount (e.g., weeks) of processing time required[12].

[10] Williams, Simon, 2008, *GPS Solutions*; Langbein, John, 2008, *Journal of Geophysical Research: Solid Earth*; Bos, M. S. et al., 2008, *Journal of Geodesy*

[11] He, Xiaoxing et al., 2021, *Remote Sensing*

[12] He, X et al., 2019, *Journal of Geodesy*; Bos, Machiel S. et al., 2020; Tunini, Lavinia et al., 2024, *Earth System Science Data*

Chapter Objective

Develop a computationally efficient alternative to the MLE for linear models with dependent errors, and apply it to large collections of GNSS position time series for estimating horizontal tectonic velocities and vertical crustal uplift rates.

The GMWMX

We propose a two-step estimation procedure combining a Generalized Least Squares (GLS) approach and the GMWM that we introduce as GMWMX for Generalized Method of Wavelet Moments with eXogenous variables.

Step 1: Estimation of $\boldsymbol{\beta}_0$ (GLS step)
Given a covariance matrix $\mathbf{\Sigma}$, define

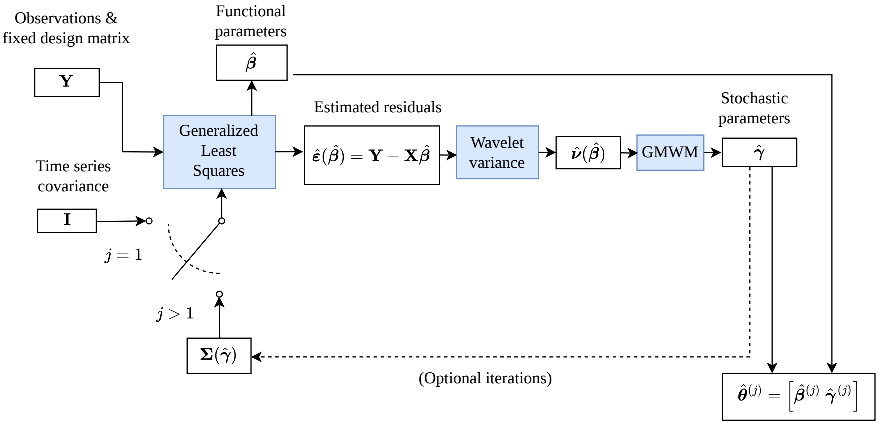
$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{\Sigma}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{\Sigma}^{-1} \mathbf{Y},$$

and the estimated residuals computed using $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\beta}}) = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Step 2: Estimation of γ_0 (GMWM step)
Consider $\hat{\boldsymbol{\nu}}(\hat{\boldsymbol{\beta}})$, the empirical WV computed on $\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\beta}})$ and, $\boldsymbol{\nu}(\gamma)$, the model-implied WV. We define:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{\hat{\boldsymbol{\nu}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\nu}(\gamma)\right\}^{\mathrm{T}} \mathbf{\Omega} \left\{\hat{\boldsymbol{\nu}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\nu}(\gamma)\right\}.$$



We denote the estimator with one or two iterations as the GMWMX-1 and GMWMX-2, respectively.

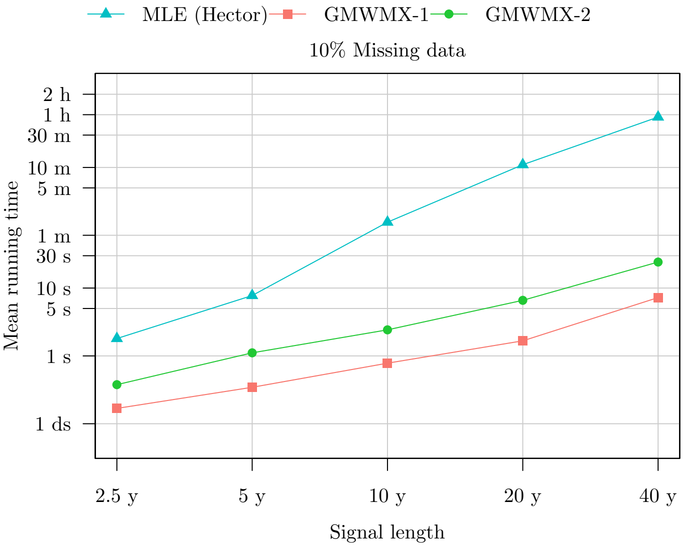


Figure: Mean running time of the MLE, the GMWMX-1 and the GMWMX-2 as a function of the sample size.

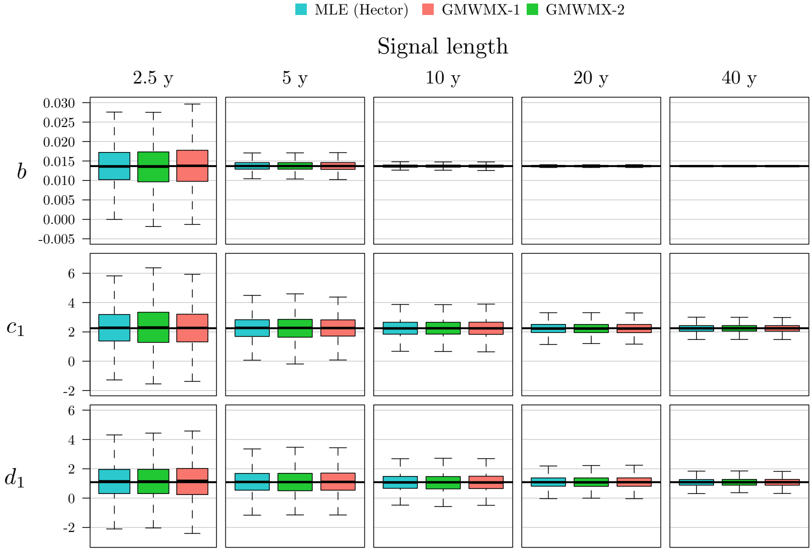


Figure: Boxplots of the estimated parameters of the model with the GMWMX-1, the GMWMX-2 and the MLE

Simulation Studies

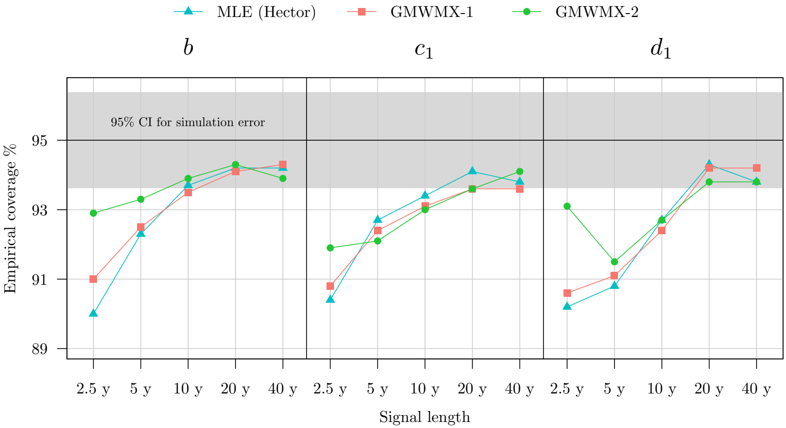


Figure: Empirical coverage of the confidence intervals at level $1-\alpha=0.95$ for the parameters b, c_1 and d_1 for GMWMX-1, GMWMX-2 and the MLE as a function of sample size. The grey area represents a 95% confidence interval of the simulation error.

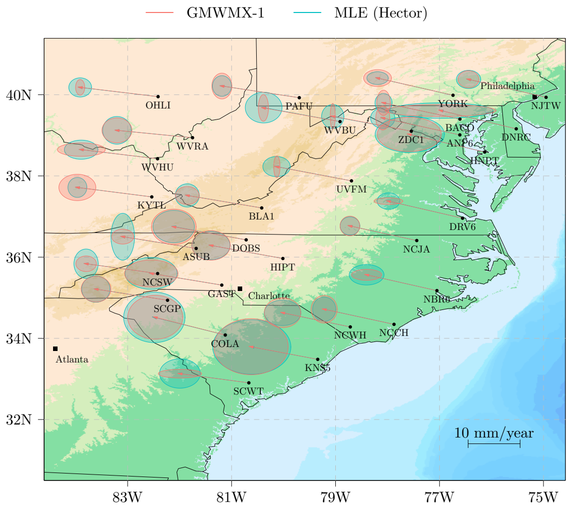


Figure: Estimated North-East velocity solutions for 33 GNSS receivers distributed over the East coast of the USA using the GMWMX-1 and Hector software (MLE).

This chapter is based on the following manuscript:

Towards Open Science: Monitoring Crustal Deformations in North America

Lionel Voirol[1], Haotian Xu[2], Yuming Zhang[3], Luca Insolia[4,5], Roberto Molinari[2] & Stéphane Guerrier[4,5]

[1]Geneva School of Economics and Management, University of Geneva, Switzerland

[2]Department of Mathematics and Statistics, Auburn University, United States

[3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, United States

[4]School of Pharmaceutical Sciences, University of Geneva, Switzerland

[5]Department of Earth Sciences, University of Geneva, Switzerland

Voirol, L., Xu, H., Zhang, Y., Insolia, L., Molinari, R., Guerrier, S., "*Towards Open Science: Monitoring Crustal Deformations in North America*".

GNSS position time series often contain missing observations because of receiver failures, maintenance, or poor observation conditions[13].

Valid inference depends on the choice of the stochastic model[14]. Misspecified noise structures can produce incorrect standard errors by factors ranging from 5 to 11[15].

Model selection becomes part of the inferential problem.

In practice, this is usually done by fitting several candidate stochastic models and selecting one with likelihood-based criteria such as AIC or BIC[16].

This adds a substantial computational burden to large-scale GNSS analyses.

[13] Bos, M. S. et al., 2008, *Journal of Geodesy*; Bos, MS et al., 2013, *Journal of Geodesy*

[14] Langbein, John et al., 1997, *Journal of Geophysical Research: Solid Earth*; Williams, Simon, 2003, *Journal of Geodesy*

[15] Mao, Ailin et al., 1999, *Journal of Geophysical Research: Solid Earth*; Wang, Wei et al., 2012, *Advances in space research*

[16] He, Xiaoxing et al., 2017, *Journal of Geodynamics*; He, X et al., 2019, *Journal of Geodesy*

Chapter Objective

1. Incorporate a missingness mechanism directly into the modeling framework
2. Develop a computationally efficient model selection criterion for the stochastic noise structure
3. Study in depth the theoretical properties of the resulting estimator

Model

Recall the model previously defined as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)\right\}.$$

A reasonable strategy to model missing data is to consider the variable $\tilde{\mathbf{Y}}$ related to the latent variable $\mathbf{Y}$ by the following mechanism:

$$\tilde{\mathbf{Y}} = \mathbf{Z} \odot \mathbf{Y},$$

where $\odot$ denotes the Hadamard product and $\mathbf{Z} = [Z_1, \ldots, Z_n]^{\mathrm{T}}$ is a binary-valued stationary process independent of $\mathbf{Y}$ which encodes observed entries by 1 and 0 otherwise, with $\mathbb{E}[Z_i] = \mu(\boldsymbol{\vartheta}_0) \in (0,\, 1]$ and covariance matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta}_0) \in \mathbb{R}^{n \times n}$, parametrized by $\boldsymbol{\vartheta}_0 \in \boldsymbol{\Upsilon} \subset \mathbb{R}^k$.

We consider a general class of models for $\mathbf{Z}$ as:

$$Z_i = \begin{cases} Z_{i-1}, & \text{with probability } \rho, \\ W_i \sim \mathrm{Bernoulli}(\mu(\boldsymbol{\vartheta}_0)), & \text{with probability } 1-\rho, \end{cases} \tag{4}$$

With this new parametrization, we define

$$\boldsymbol{\theta}_0 = \left[\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}}, \boldsymbol{\vartheta}_0^{\mathrm{T}} \right]^{\mathrm{T}}$$

and $\tilde{\mathbf{X}} = \left(\mathbf{Z} \otimes \mathbf{1}_p^{\mathrm{T}}\right) \odot \mathbf{X} \in \mathbb{R}^{n \times p}$ as the design matrix $\mathbf{X}$ with zero-valued rows for missing observations in $\mathbf{Y}$, where $\otimes$ denotes the Kronecker product and $\mathbf{1}_p$ is a vector of ones of dimension p.

We propose a three-steps statistical procedure that we introduce as WAMORE for Wavelet Moment Regression.

Step 1: Estimation of $\boldsymbol{\vartheta}_0$

Using $\mathbf{Z}$, we first estimate $\boldsymbol{\vartheta}_0$ by Maximum Likelihood. Let $\hat{\boldsymbol{\vartheta}}$ denote our estimator.

Step 2: Estimation of $\boldsymbol{\beta}_0$

We then consider a Least Squares approach to obtain an estimate of $\boldsymbol{\beta}_0$ and define:

$$\hat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{Y}}.$$

Step 3: Estimation of $\boldsymbol{\gamma}_0$

Let $\hat{\boldsymbol{\nu}}$ represent the empirical WV computed on the estimated residuals process $\hat{\boldsymbol{\varepsilon}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\nu}\left(\boldsymbol{\gamma}, \boldsymbol{\vartheta}\right)$ the theoretical WV of $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta} = \boldsymbol{\varepsilon} \odot \mathbf{Z}$. Given $\hat{\boldsymbol{\vartheta}}$, we estimate $\boldsymbol{\gamma}_0$ with:

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\left(\boldsymbol{\gamma}, \hat{\boldsymbol{\vartheta}}\right)\right\|_{\boldsymbol{\Omega}}^{2} \tag{5}$$

We aim to construct a computationally efficient model selection criterion for the stochastic model.

Recall the error term $\boldsymbol{\varepsilon} \sim \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_0\right)\right\}$ with $\boldsymbol{\gamma}_0 \subset \mathbb{R}^q$ and consider $\hat{\boldsymbol{\nu}}$, the empirical WV computed on $\boldsymbol{\varepsilon}$.

A reasonable fit metric is given by:

$$C = \mathbb{E}_0\left[\mathbb{E}\left\{\|\hat{\boldsymbol{\nu}}_0 - \boldsymbol{\nu}(\hat{\gamma})\|_{\boldsymbol{\Omega}}^2\right\}\right] = \mathbb{E}_0\left(\mathbb{E}\left[\left\{\hat{\boldsymbol{\nu}}_0 - \boldsymbol{\nu}(\hat{\gamma})\right\}^{\mathrm{T}} \boldsymbol{\Omega}\left\{\hat{\boldsymbol{\nu}}_0 - \boldsymbol{\nu}(\hat{\gamma})\right\}\right]\right),$$

where $\hat{\gamma}$ is computed on the sample $\boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\nu}}_0$ is computed on an independent copy, $\boldsymbol{\varepsilon}_0$ satisfying $\boldsymbol{\varepsilon} \stackrel{d}{=} \boldsymbol{\varepsilon}_0$.

Model Selection

It can be shown that:

$$C = \mathbb{E}\left[\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\gamma})\|_{\boldsymbol{\Omega}}^2\right] + 2\mathrm{tr}\left[\mathrm{cov}\left\{\hat{\boldsymbol{\nu}}, \boldsymbol{\nu}(\hat{\gamma})\right\}\boldsymbol{\Omega}\right]$$

Using a first-order approximation of $\boldsymbol{\nu}(\hat{\gamma})$ around $\boldsymbol{\gamma}_0$, an estimator of C is obtained. With $\boldsymbol{\Omega} = \mathbf{V}^{-1}$, the resulting complexity penalty simplifies to:

$$\hat{C} = \|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\gamma})\|_{\boldsymbol{\Omega}}^2 + 2q$$

The empirical WV as a quadratic form

In order to obtain $\boldsymbol{\nu}(\boldsymbol{\gamma}, \boldsymbol{\vartheta})$ to estimate $\boldsymbol{\gamma}_0$ in (5) and $\mathbf{V} = \text{var}(\hat{\boldsymbol{\nu}})$ to compute $\hat{C}$, we need to derive their explicit form.

Denoting the vector of Wavelet coefficients at scale j as $\mathbf{W}_j$ computed on a signal $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ with covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$, we can write:

$$\mathbf{W}_j = \begin{bmatrix} W_{j,1} \\ W_{j,2} \\ \vdots \\ W_{j,M_j} \end{bmatrix} = \mathbf{B}_j \boldsymbol{\varepsilon} = \begin{bmatrix} \mathbf{h}_j^{\mathrm{T}} & 0 & \dots & 0 \\ 0 & \mathbf{h}_j^{\mathrm{T}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{h}_j^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

where $\mathbf{B}_j$ is an $M_j \times n$ matrix and $\mathbf{h}_j^{\mathrm{T}}$ is a vector of length 2^j containing the wavelet filter. Hence, we can rewrite the empirical WV as a quadratic form:

$$\hat{\nu}_j = \frac{1}{M_j} \sum_{i=1}^{M_j} W_{j,i}^2 = \frac{1}{M_j} \mathbf{W}_j^{\mathrm{T}} \mathbf{W}_j = \frac{1}{M_j} (\mathbf{B}_j \boldsymbol{\varepsilon})^{\mathrm{T}} \mathbf{B}_j \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{A}_j \boldsymbol{\varepsilon}, \text{ where } \mathbf{A}_j = \frac{\mathbf{B}_j^{\mathrm{T}} \mathbf{B}_j}{M_j}$$

Using the quadratic form representation, we obtain the theoretical WV:

$$\mathbb{E}(\hat{\nu}_j) = \nu_j = \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{A}_j \boldsymbol{\varepsilon}\right] = \mathrm{tr}\left\{\mathbf{A}_j \boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_0\right)\right\},$$

Using the Gaussian quadratic-form result, we obtain $\mathbf{V}$, the covariance of the empirical WV $\hat{\boldsymbol{\nu}}$:

$$\mathrm{cov}\{\hat{\nu}_i, \hat{\nu}_j\} = \mathrm{cov}(\boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{A}_i \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{A}_j \boldsymbol{\varepsilon}) = 2\mathrm{tr}\{\mathbf{A}_i \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0) \mathbf{A}_j \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)\}.$$

First explicit form of $\mathbf{V}$[17].

We provide a computationally efficient algorithm to obtain $\mathrm{cov}\{\hat{\nu}_i, \hat{\nu}_j\}$ allowing to compute these quantities without passing by their matrix expression.

[17] Zhang, Nien Fan, 2008, *Metrologia*

Theoretical properties

Inference on the functional parameters

- Consistent estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$
- Asymptotic normality of $\hat{\boldsymbol{\beta}}$ under short- and long-memory dependence regime
- Valid confidence intervals under dependent errors and missing observations

Model selection

- Consistent estimator of the target criterion: $\hat{C} = C + o_p(1)$
- Same asymptotic overfitting and underfitting probabilities as the AIC

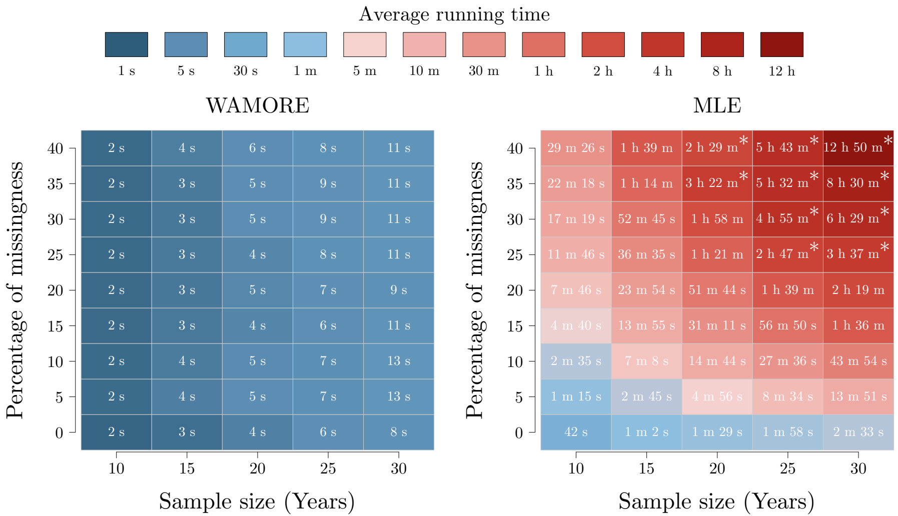


Figure: Average running time of the WAMORE (left panel) vs the MLE (right panel) for the setting with the WN + MAT + RW model. Settings annotated with a "*" were evaluated only on 5 Monte Carlo realizations due to computational constraints.

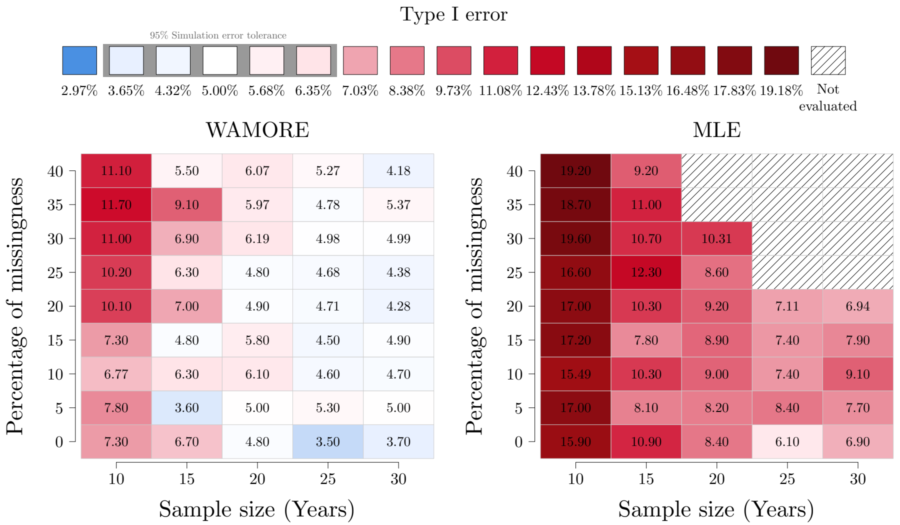


Figure: Empirical type I error (in %) of the WAMORE (left panel) vs the MLE (right panel) for the setting with the WN + MAT + RW model.

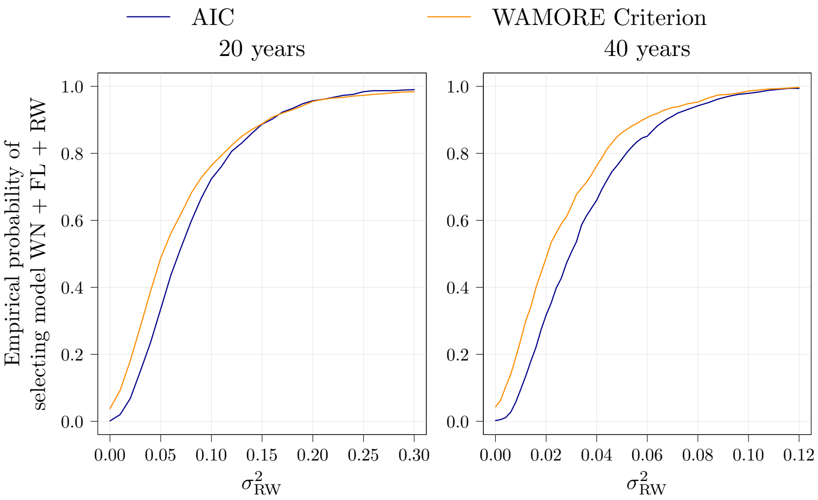


Figure: Selection rate of the AIC and the proposed WAMORE criterion for setting considering 20 years of daily data (left panel) and 40 years of daily data (right panel).

Case Study

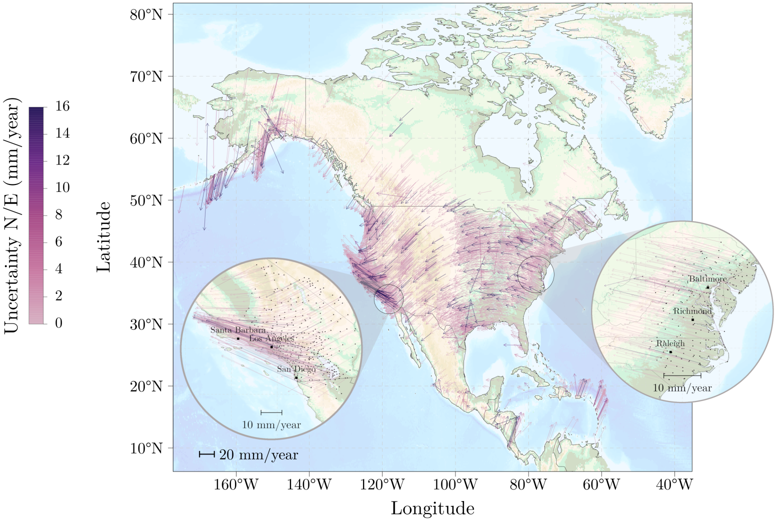


Figure: Estimated tectonic velocities and associated uncertainty using the WAMORE methodology. Arrow length encodes velocity magnitude, and color represents uncertainty, computed as the ℓ_2-norm of the north and east component standard errors.

https://data-analytics-lab.shinyapps.io/wamore-crustal-deformation-explorer

Figure: Interactive WAMORE crustal deformations explorer.

Other Works

Signal processing & engineering

Cucci, Davide A, Voirol, Lionel, Khaghani, Mehran, Guerrier, Stéphane, 2023, "On Performance Evaluation of Inertial Navigation Systems: The Case of Stochastic Calibration", *IEEE Transactions on Instrumentation and Measurement*, pp. 1–17

Voirol, Lionel, Guerrier, Stéphane, Zhang, Yuming, Karemera, Mucyo, Radi, Ahmed, 2020, "Optimally Weighted Wavelet variance-based Estimation for Inertial Sensor Stochastic Calibration", *2020 12th International Conference on Electrical Engineering (ICEENG)*, pp. 482–486

Health science and biostatistics

Wehrli, Lydia, Galdadas, Ioannis, Voirol, Lionel, Smievsko, Martin, Cambet, Yves, Jaquet, Vincent, Guerrier, Stéphane, Gervasio, Francesco Luigi, Nef, Serge, Rahban, Rita, 2023, "The Action of Physiological and Synthetic Steroids on the Calcium Channel CatSper in Human Sperm", *Frontiers in Cell and Developmental Biology*

Ramzy, George M, Norkin, Maxim, Koessler, Thibaud, Voirol, Lionel, Tihy, Mathieu, Hany, Dina, McKee, Thomas, Ris, Frédéric, Buchs, Nicolas, Docquier, Mylène, 2023, "Platform Combining Statistical Modeling and patient-derived Organoids to Facilitate Personalized Treatment of Colorectal Carcinoma", *Journal of Experimental & Clinical Cancer Research*, p. 79

Medical education

Pfarrwaller, Eva, Voirol, Lionel, Piumatti, Giovanni, Karemera, Mucyo, Sommer, Johanna, Gerbase, Margaret W, Guerrier, Stéphane, Baroffio, Anne, 2022, "Students' Intentions to Practice Primary Care are Associated with their Motives to Become Doctors: A Longitudinal Study", *BMC Medical Education*, p. 30

Pfarrwaller, Eva, Voirol, Lionel, Karemera, Mucyo, Guerrier, Stéphane, Baroffio, Anne, 2023, "Dynamics of Career Intentions in a Medical Student Cohort: A four-year Longitudinal Study", *BMC Medical Education*, pp. 1–11

Balbinot, Fernanda, Voirol, Lionel, Guerrier, Stéphane, Karemera, Mucyo, Feser, Raphaël, Baroffio, Anne, Gerbase, Margaret W, 2025, "Unveiling Empathy Determinants Across Borders: A Comparative Analysis of Medical Students from two geo-sociocultural Backgrounds", *BMC Medical Education*, p. 554

Thank You!

https://lionelvoirol.com
Lionel.Voirol@unige.ch

Balbinot, Fernanda, Lionel Voirol, Stéphane Guerrier, Mucyo Karemera, Raphaël Feser, Anne Baroffio, and Margaret W Gerbase (2025). "Unveiling Empathy Determinants Across Borders: A Comparative Analysis of Medical Students from two geo-sociocultural Backgrounds". In: *BMC Medical Education* 25.1, p. 554.

Bevis, Michael, Wolfgang Scherer, and Mark Merrifield (2002). "Technical Issues and Recommendations Related to the Installation of Continuous GPS Stations at Tide Gauges". In: *Marine Geodesy* 25.1-2, pp. 87–99.

Bock, Yehuda and Diego Melgar (2016). "Physical Applications of GPS Geodesy: A Review". In: *Reports on Progress in Physics* 79.10, p. 106801.

Bos, M. S., R. M. S. Fernandes, S. D. P. Williams, and L. Bastos (2008). "Fast Error Analysis of Continuous GPS Observations". In: *Journal of Geodesy* 82.3, pp. 157–166.

Bos, Machiel S., Jean-Philippe Montillet, Simon D. P. Williams, and Rui M. S. Fernandes (2020). "Introduction to Geodetic Time Series Analysis". In: *Geodetic Time Series Analysis in Earth Sciences*. Cham: Springer International Publishing, pp. 29–52.

Bos, MS, RMS Fernandes, SDP Williams, and L Bastos (2013). "Fast Error Analysis of Continuous GNSS Observations with Missing Data". In: *Journal of Geodesy* 87.4, pp. 351–360.

- Cucci, Davide A, Lionel Voirol, Mehran Khaghani, and Stéphane Guerrier (2023). "On Performance Evaluation of Inertial Navigation Systems: The Case of Stochastic Calibration". In: *IEEE Transactions on Instrumentation and Measurement* 72, pp. 1–17.
- Guerrier, Stéphane, Jan Skaloud, Yannick Stebler, and Maria-Pia Victoria-Feser (2013). "Wavelet-variance-based Estimation for Composite Stochastic Processes". In: *Journal of the American Statistical Association* 108.503, pp. 1021–1030.
- He, X, MS Bos, JP Montillet, and RMS Fernandes (2019). "Investigation of the Noise Properties at Low Frequencies in Long GNSS Time Series". In: *Journal of Geodesy* 93.9, pp. 1271–1282.
- He, Xiaoxing, Machiel Simon Bos, Jean-Philippe Montillet, Rui Fernandes, Tim Melbourne, Weiping Jiang, and Wudong Li (2021). "Spatial Variations of Stochastic Noise Properties in GPS Time Series". In: *Remote Sensing* 13.22, p. 4534.
- He, Xiaoxing, Jean-Philippe Montillet, Rui Fernandes, Machiel Bos, Kegen Yu, Xianghong Hua, and Weiping Jiang (2017). "Review of Current GPS Methodologies for Producing Accurate Time Series and their Error Sources". In: *Journal of Geodynamics* 106, pp. 12–29.

- Herring, T. A., R. W. King, S. C. McClusky, L. Floyd M.and Wang, M. Murray, and et al. (2016). "Plate Boundary Observatory and Related Networks: GPS Data Analysis Methods and Geodetic Products". In: *Reviews of Geophysics* 54, pp. 759–808. DOI: 10.1002/2016RG000529.
- Langbein, John (2008). "Noise in GPS Displacement Measurements from Southern California and Southern Nevada". In: *Journal of Geophysical Research: Solid Earth* 113.B5, pp. 1–12.
- Langbein, John and Hadley Johnson (1997). "Correlated Errors in Geodetic Time Series: Implications for time-dependent Deformation". In: *Journal of Geophysical Research: Solid Earth* 102.B1, pp. 591–603.
- Mao, Ailin, Christopher GA Harrison, and Timothy H Dixon (1999). "Noise in GPS Coordinate Time Series". In: *Journal of Geophysical Research: Solid Earth* 104.B2, pp. 2797–2816.
- Milne, Glenn A, James L Davis, Jerry X Mitrovica, H-G Scherneck, Jan M Johansson, Martin Vermeer, and Hannu Koivula (2001). "Space-geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia". In: *Science* 291.5512, pp. 2381–2385.
- Percival, Donald P. (1995). "On Estimation of the Wavelet Variance". In: *Biometrika* 82.3, pp. 619–631. ISSN: 00063444. URL: http://www.jstor.org/stable/2337538.

- Pfarrwaller, Eva, Lionel Voirol, Mucyo Karemera, Stéphane Guerrier, and Anne Baroffio (2023). "Dynamics of Career Intentions in a Medical Student Cohort: A four-year Longitudinal Study". In: *BMC Medical Education* 23.1, pp. 1–11.
- Pfarrwaller, Eva, Lionel Voirol, Giovanni Piumatti, Mucyo Karemera, Johanna Sommer, Margaret W Gerbase, Stéphane Guerrier, and Anne Baroffio (2022). "Students' Intentions to Practice Primary Care are Associated with their Motives to Become Doctors: A Longitudinal Study". In: *BMC Medical Education* 22.1, p. 30.
- Radi, Ahmed, Sameh Nassar, and Naser El-Sheimy (2018). "Stochastic Error Modeling of Smartphone Inertial Sensors for Navigation in Varying Dynamic Conditions". In: *Gyroscopy and Navigation* 9.1, pp. 76–95.
- Ramzy, George M, Maxim Norkin, Thibaud Koessler, Lionel Voirol, Mathieu Tihy, Dina Hany, Thomas McKee, Frédéric Ris, Nicolas Buchs, Mylène Docquier, et al. (2023). "Platform Combining Statistical Modeling and patient-derived Organoids to Facilitate Personalized Treatment of Colorectal Carcinoma". In: *Journal of Experimental & Clinical Cancer Research* 42.1, p. 79.
- Stebler, Yannick, Stéphane Guerrier, and Jan Skaloud (2015). "An Approach for Observing and Modeling Errors in MEMS-based Inertial Sensors under Vehicle Dynamic". In: *IEEE Transactions on Instrumentation and Measurement* 64.11, pp. 2926–2936.

References V

Tunini, Lavinia, Andrea Magrin, Giuliana Rossi, and David Zuliani (2024). "Global Navigation Satellite System (GNSS) Time Series and Velocities about a Slowly Convergent Margin Processed on High Performance Computing (HPC) Clusters: Products and Robustness Evaluation". In: *Earth System Science Data* 16.2, pp. 1083–1106.

Voirol, Lionel, Stéphane Guerrier, Yuming Zhang, Mucyo Karemera, and Ahmed Radi (2020). "Optimally Weighted Wavelet variance-based Estimation for Inertial Sensor Stochastic Calibration". In: *2020 12th International Conference on Electrical Engineering (ICEENG)*. IEEE, pp. 482–486.

Wang, Wei, Bin Zhao, Qi Wang, and Shaomin Yang (2012). "Noise Analysis of Continuous GPS Coordinate Time Series for CMONOC". In: *Advances in space research* 49.5, pp. 943–956.

Wehrli, Lydia, Ioannis Galdadas, Lionel Voirol, Martin Smievsko, Yves Cambet, Vincent Jaquet, Stéphane Guerrier, Francesco Luigi Gervasio, Serge Nef, and Rita Rahban (2023). "The Action of Physiological and Synthetic Steroids on the Calcium Channel CatSper in Human Sperm". In: *Frontiers in Cell and Developmental Biology* 11.

Williams, Simon (2003a). "Offsets in Global Positioning System Time Series". In: *Journal of Geophysical Research: Solid Earth* 108.B6, pp. 1–13.

Williams, Simon (2003b). "The Effect of Coloured Noise on the Uncertainties of Rates Estimated from Geodetic Time Series". In: *Journal of Geodesy* 76.9, pp. 483–494.

— (2008). "CATS: GPS Coordinate Time Series Analysis Software". In: *GPS Solutions* 12.2, pp. 147–153.

Zhang, Nien Fan (2008). "Allan Variance of Time Series Models for Measurement Data". In: *Metrologia* 45.5, p. 549.