Hypothesis Testing on a Continental Scale: GPU Based Time Series Classification

Contact: adriaan(at)fwrite.org
keuris.earth.observation(at)proton.me

Adriaan van Natijne[1)], Lars Keuris[2)]

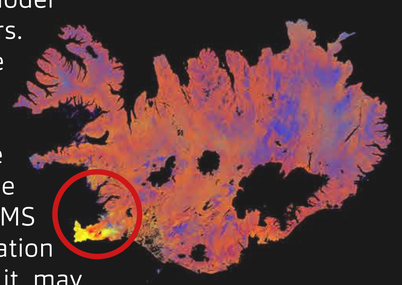
1) personal initiative, the Netherlands
2) Keuris Earth Observation, the Netherlands

Background

In recent years, pre-processed deformation datasets have become available to the general public, particularly due to the arrival of the European Ground Motion Service (EMGS). Thanks to this recent development, access to InSAR measurements has significantly improved. The continental scale of the EGMS deformation dataset enables users to assess potentially hazardous regions in a uniform and robust manner. However, to extract behavioral statistics from time series, such as trends and accelerations, of natural and anthropogenic deformation processes, a predefined model must be imposed on EGMS deformation data.

An Appropriate Model?

The EMGS comes with a predefined model and its corresponding model parameters. Unfortunately, this model is the same for all 15.5 billion locations. It is unlikely that the current EGMS model selection is optimal for all time series, and the large region over which these time series are distributed. The current EGMS model estimates velocity, acceleration and seasonality at once , therefore it may be needlessly complex, while in other cases it is insufficient to describe the deformation, as exemplified by jumps or trend breaks. However, the selection of an appropriate model requires iterative testing of many different models at each location, incurring a great computational cost.



Multiple Model Method

We extended the EGMS model of offset, velocity, acceleration and seasonality, to over 5000 model variations, including step functions and trend breaks. Discontinuous functions are of particular interest because they point out erratic processing (e.g. phase unwrapping errors) or hazardous physical behavior (e.g. sinkhole precursor). The variations include polynomials (up to the 3rd degree) and a seasonal signal combined with Heaviside functions to accommodate changes of deformation dynamics during the time series.

Parallelized GPU processing allows us to simultaneously impose many models on multiple time series and extract model parameters at unprecedented scales. We demonstrate this idea with a consumer laptop at ~15 million models per second. Subsequently, a best model for each time series is selected using the Bayesian Information Criterion (BIC). The proposed parallelized methodology was applied on our consumer laptop to all time series within the most recent EGMS dataset.

Results

With each measurement location having a new, tailored model, it becomes possible to distinguish the time series and ultimately the underlying deformation dynamics by the type of model that was selected. In the European map on the right, the blue color indicates that the majority of the time series are best modelled with polynomials. In the same fashion, green indicates seasonal behaviour and red demonstrates discontinuous behaviour. Color combinations are possible, because each grid cell in the example on the right is aggregated from all models in that cell.

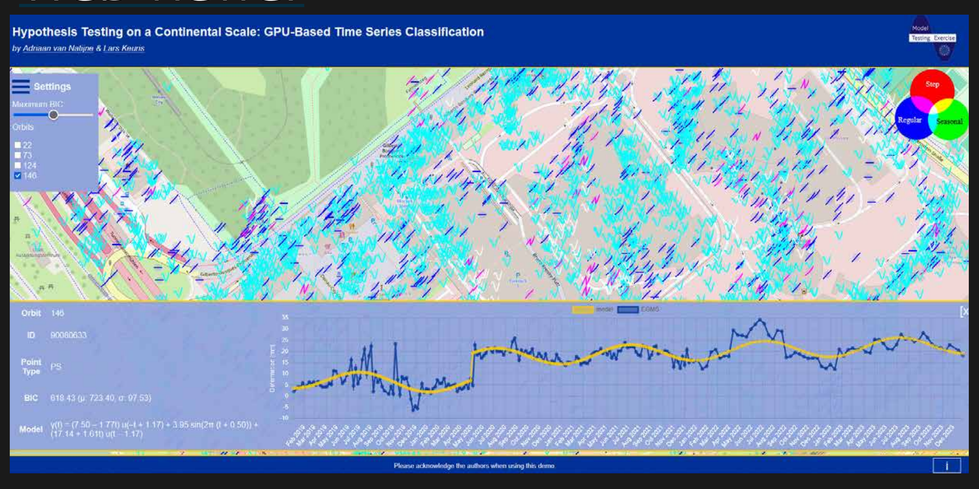
This map and the underlying individual time series are freely accessible and available in our novel web viewer. Each time series is clickable, displaying the original data and the best model fit. Aggregated results of all model fits show that the RMSE is comparable between the EGMS model fit and our tailored model fit (both ~4 mm). Thus a more realistic model is applied at no cost to the quality of the model fit.

Discussion

Our method helps to analyse the local heterogeneity in deformation processes, and provides a starting point for large scale analysis of the EGMS data. However, to link deformation anomalies to natural behavior it is necessary to interpret the deformation signal. Especially with more complex models, multiple signals are superimposed that may counteract each other and complicate the translation to geomorphological or mechanical (building) processes. To avoid this, the model could be rerun using only models with a clear physical interpretation. The requirement of the BIC that the true model is among the tested models, and that they will hold under the assumption that the deformation behavior is rooted in a geomorphological/mechanical process.

As a side effect, major differences in the model classification behavior between bursts and orbits helps identify processing differences and anomalies.

Web viewer

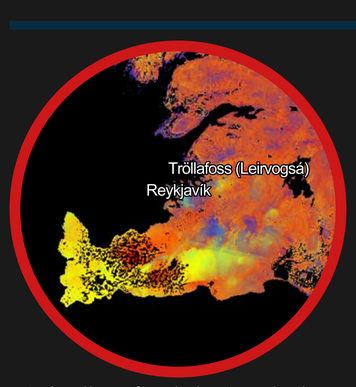


Remodelled webviewer (egms-new.fwrite.org) for the improved interpretation of the latest EGMS dataset.

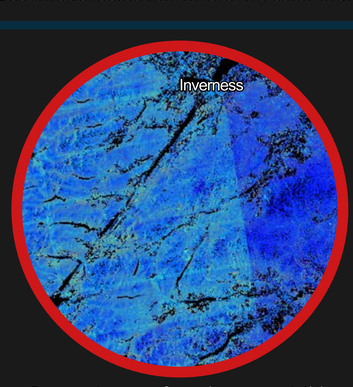
Conclusion

Our results show thatvin 94% of the cases the standard EGMS model fit can be improved with another model. This improves both the quantification of the deformation as well as the classification of the deformation type. Our results support a wide range of end users in the stability assessment at various scales and are more accessible and easy to interpret even with limited knowledge of the InSAR methodology.

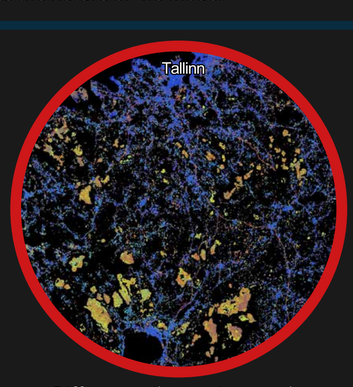
Best model selection where the color indicates the nature of the model.



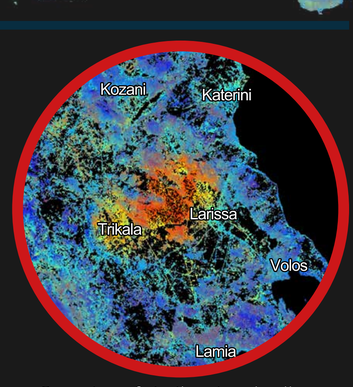
Iceland's surface behaviour displays many discontinuities.



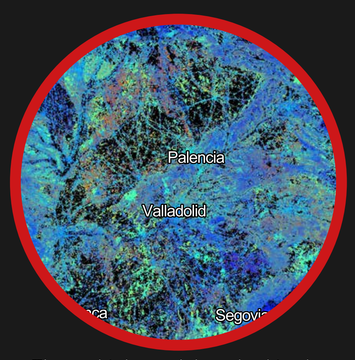
Processing artefacts become visible and dataset biases can be detected.



Different soil types impact the deformation signal distinctively.

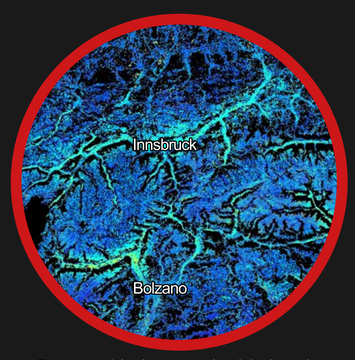


Grouping of similar signals allows them to be investigated together.



The multiple model method is also relevant at local scales.

Realisitc classifications allow for guided assessments.



Seasonal behaviour highlights the valleys in the European Alps.