

[EJ] Oral | M (Multidisciplinary and Interdisciplinary) | M-GI General Geosciences, Information Geosciences & Simulations

[M-GI29] [EJ] Data-driven analysis, modeling and prediction in geosciences

convener: Tatsu Kuwatani (Japan Agency for Marine-Earth Science and Technology), Dmitri Kondrashov (University of California, Los Angeles), Hiromichi Nagao (Earthquake Research Institute, The University of Tokyo), Sergey Kravtsov (University of Wisconsin Milwaukee), Chairperson: Hiromichi Nagao (Earthquake Research Institute, The University of Tokyo), Chairperson: Sergey Kravtsov (University of Wisconsin Milwaukee)

Sat. May 20, 2017 10:45 AM - 12:15 PM A01 (Tokyo Bay Makuhari Hall)

It is important to extract essential processes and structures from observed datasets in order to understand and predict the dynamic behavior of the earth and planetary systems. Recently, powerful data-driven methodologies have been proposed to extract, model and predict useful information contained in high-dimensional datasets that are ubiquitous in earth and space. This session aims to provide an opportunity to highlight recent advances in such data-driven techniques across disciplines and to have a productive discussion for interdisciplinary collaborations.

[MGI29-07] Data-driven science for solving problems in geosciences

*Akaho Shotaro¹ (1. The National Institute of Advanced Industrial Science and Technology)

10:45 AM - 11:00 AM

[MGI29-08] Source Term Estimation for atmospheric release In Nuclear Accidents Using ensemble Kalman filter: a validation with wind tunnel experiment

*sheng fang¹ (1. Tsinghua University)

11:00 AM - 11:15 AM

[MGI29-09] Bayesian inversion analysis of nonlinear spatiotemporal dynamics of heterogeneous reactions in rock-water interactions

*Toshiaki Omori¹, Ryota Morimoto¹, Tatsu Kuwatani², Atsushi Okamoto³, Koji Hukushima⁴ (1. Kobe University, 2. JAMSTEC, 3. Tohoku University, 4. University of Tokyo)

11:15 AM - 11:30 AM

[MGI29-10] Detection of principal dynamical modes of changing climate

*Dmitry Mukhin¹, Andrey Gavrilov¹, Evgeny M Loskutov¹, Alexander Feigin¹ (1. Institute of Applied Physics of RAS)

11:30 AM - 11:45 AM

[MGI29-11] Data-adaptive Harmonic Decomposition and Stuart-Landau closure modes

*Dmitri Kondrashov^{1,2} (1. University of California, Los Angeles, 2. Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russia)

11:45 AM - 12:00 PM

[MGI29-12] Conditional stochastic model chains in reduced space: Towards efficient simulation of non-stationary typhoon precipitation patterns

*Boyko Dodov¹ (1. AIR Worldwide Boston)

12:00 PM - 12:15 PM

Data-driven science for solving problems in geosciences

*Akaho Shotaro¹

1. The National Institute of Advanced Industrial Science and Technology

Recent technological development of measurement and observation enables us to obtain a large amount of high-dimensional data. Effective use of high-dimensional data requires a robust framework to make the tight connection of information science to the original purpose of data analysis derived from various scientific disciplines [1]. Since 2013, we have launched a big scientific project entitled as “Initiative for high-dimensional data-driven science through deepening sparse modelling (FY2013-FY2017)” funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. The aim of this project is to establish a novel framework of data analysis for natural sciences, namely, data-driven science. Its target fields are very wide including geosciences, astronomy, biology, and medical and brain sciences. In this presentation, we introduce the concept of data-driven science and some applications to geosciences [e.g. 2-5].

- [1] Igarashi Y, Nagata K, Kuwatani T, Omori T, Nakanishi-Ohno Y, Okada M (2016) Three levels of data-driven science. *J Phys Confere Ser* 699: 012001-1 - 012001-13.
- [2] Kuwatani T, Nagata K, Okada M, Toriumi M (2014) Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth Planet Space* 66: 5-1 - 5-9.
- [3] Kuwatani T, Nagata K, Okada M, Watanabe T, Ogawa Y, Komai T, Tsuchiya N (2014) Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits. *Sci Rep* 4: 7077-1—7077-6.
- [4] Nakata R, Kuwatani T, Okada M, Hori T (2016) Geodetic inversion for spatial distribution of slip under smoothness, discontinuity, and sparsity constraints. *Earth Planet Space* 68, 20-1 –20-10.
- [5] Araki T, Ochi T, Matsumoto N, Akaho S (2015) Robust estimation of spatio-temporal distribution of slow slip event by switching model, *Spatial Statistics*.

Keywords: data-driven science, sparse-modeling, machine learning

Source Term Estimation for atmospheric release In Nuclear Accidents Using ensemble Kalman filter: a validation with wind tunnel experiment

*sheng fang¹

1. Tsinghua University

In a nuclear accident, the source term which contains the release rate of atmospheric radionuclide leakage is a key issue of the nuclear emergency response. One effective way to obtain the source term is the inverse modelling method that is based on radionuclide transportation process and environment radiation monitoring data. However, the inverse modelling method may be sensitive to specific site conditions. Therefore, a case-by-case validation is important.

In this study, a source term estimation method based on the Ensemble Kalman filter (EnKF) data assimilation technique was proposed for source term estimation and a wind tunnel experiment that simulated a highly heterogeneous Chinese nuclear power plant site was performed to validate this method.

The EnKF method takes the concentrations measured at different positions as input parameters and iteratively refined the source term estimate by reducing the discrepancy between the experimental measurements and the concentration prediction that is obtained by certain atmospheric transport model based on the estimated source term. In order to improve the accuracy of the transport model, CALMET was used to generate the wind field that is necessary to drive RIMPUFF.

A 1:2000 wind tunnel was performed to comprehensively evaluate the performance of the proposed method, which simulated both ground and stack release scenarios with a dominant wind direction at a typical Chinese nuclear power plant. The incoming flow in the wind tunnel is adjusted according to the annual mean wind speed and vertical profile that has been measured in recent years near the NPP site. The experiment simulated a ground release scenario by using CO as the tracer gas, which the release position is in the center of the nuclear power plant site. Concentrations are measured at 264 positions in the downwind direction of the release, which equals to 6500m away from the source in a real world scale. The estimated release rate was compared with the true release rate that was used in the wind tunnel experiment, in order to assess the convergence and accuracy of the proposed method. Meanwhile, the concentration measured in the wind tunnel was both qualitatively and quantitatively compared with the simulation values that are calculated by CALMET-RIMPUFF using the estimated release rate.

The validation results demonstrate that the proposed method has a fast convergence rate. The estimated release rate matches the real one used in the experiment well for both release cases, which the bias is less than 50% for the worst case estimate. As for the concentrations predicted with the estimated release rate, they are not only qualitatively consistent with the spatial distribution of the measured concentrations, but also show satisfactory results with respect to statistical evaluation metrics for both release cases. The Pearson correlation coefficient is higher than 74~87%, the FAC2 is 42~54% and the FAC5 is 77~80% for all the experiment cases. The source term estimation of the stack release case slightly outperforms that of the ground release case, due to the better atmospheric transport modelling. The experiment also reveals that CALMET-RIMPUFF tends to underestimate the concentrations, which are the primary source of the bias in the estimates. Therefore, the atmospheric model is critical for the performance of the source term estimation, which shall be our future work focus.

Keywords: Source term estimation, Data assimilation, Atmospheric dispersion, Radioactive release

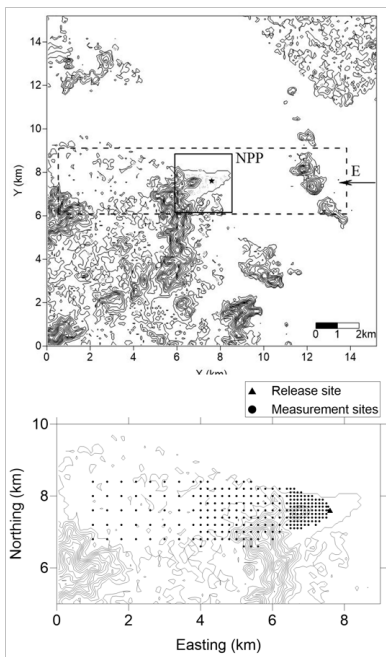


Fig. 1 Topography and measurement network of the wind tunnel experiment

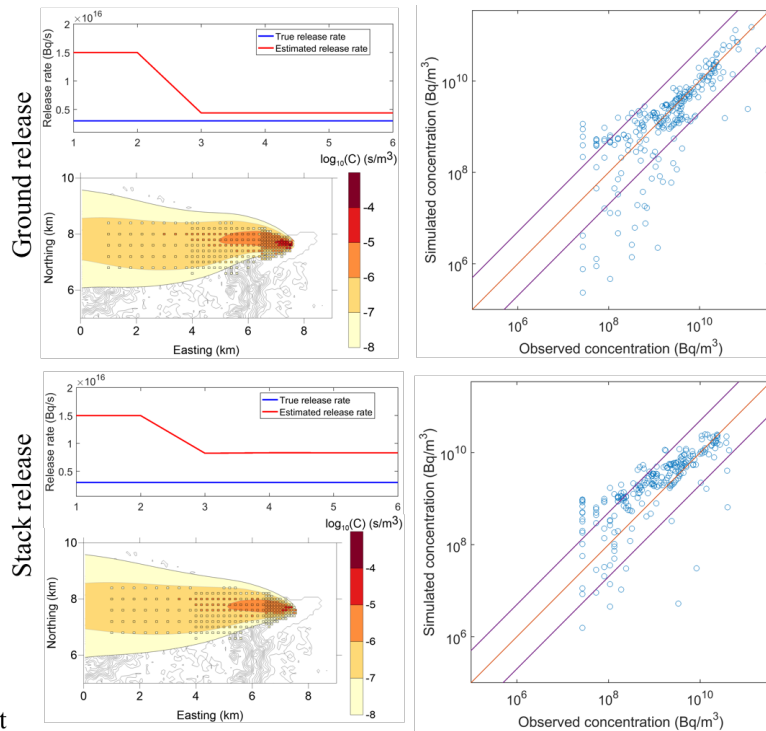


Fig. 2 source term estimation results. Convergence curve, simulated plume using the source term estimate and the corresponding scatter plot

Bayesian inversion analysis of nonlinear spatiotemporal dynamics of heterogeneous reactions in rock-water interactions

*Toshiaki Omori¹, Ryota Morimoto¹, Tatsu Kuwatani², Atsushi Okamoto³, Koji Hukushima⁴

1. Kobe University, 2. JAMSTEC, 3. Tohoku University, 4. University of Tokyo

It is essential to extract nonlinear dynamics from time-series data as an inverse problem in natural sciences. We propose a Bayesian statistical framework for extracting nonlinear spatiotemporal dynamics of surface heterogeneous reactions from sparse and noisy observable data. Surface heterogeneous reactions are chemical reactions with conjugation of multiple phases, and they have the intrinsic nonlinearity of their dynamics caused by the effect of surface-area between different phases. We employ sequential Monte Carlo algorithm and other statistical algorithm to partial observation problem, in order to simultaneously estimate the time course of hidden variables and the kinetic parameters underlying dynamics. Using our proposed method, we show that the rate constants of dissolution and precipitation reactions, which are typical examples of surface heterogeneous reactions, and the diffusion constants, as well as the spatiotemporal changes of solid reactants and products, were successfully estimated only from the observable temporal changes in the concentration of the dissolved intermediate product.

[1] Omori et al., Bayesian inversion analysis of nonlinear dynamics in surface heterogeneous reactions, Phys. Rev. E, 94, 033305 (2016)

Keywords: Data-driven science, Machine learning

Detection of principal dynamical modes of changing climate

*Dmitry Mukhin¹, Andrey Gavrilov¹, Evgeny M Loskutov¹, Alexander Feigin¹

1. Institute of Applied Physics of RAS

The success in empirical climate modeling strongly depends on selection of model variables used for reduced representation of system's dynamics. In fact, we can say that the states of the climate system are determined by a combination of external perturbations (forcings) and unknown internal variables. Thus, the detection of principal internal modes of changing climate is crucial point in modeling problem. In the presentation the method for extraction such modes from data is presented. The method is based on the Nonlinear Dynamical Mode (NDM) expansion [1,2], but takes into account forcing time series applied to the system: each NDM is represented by hidden time series governing the observed variability, which, together with external forcing signals, are mapped onto the data space.

In this work the method is used for reconstructing and studying the principal modes of global climate variability on inter-annual and decadal time scales, adjusted for the external forcings such as anthropogenic emissions, variations of the solar activity and volcanic activity. The structure of the obtained modes as well as their response to external factors, e.g. forecast their change in 21 century under different CO2 emission scenarios, are discussed.

[1] Mukhin, D., Gavrilov, A., Feigin, A., Loskutov, E., & Kurths, J. (2015). Principal nonlinear dynamical modes of climate variability. *Scientific Reports*, 5, 15510. <http://doi.org/10.1038/srep15510>

[2] Gavrilov, A., Mukhin, D., Loskutov, E., Volodin, E., Feigin, A., & Kurths, J. (2016). Method for reconstructing nonlinear modes with adaptive structure from multidimensional data. *CHAOS*, 26(12), 123101. <http://doi.org/10.1063/1.4968852>

Keywords: nonlinear dynamical modes, data expansion, empirical modeling

Data-adaptive Harmonic Decomposition and Stuart-Landau closure modes

*Dmitri Kondrashov^{1,2}

1. University of California, Los Angeles, 2. Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russia

Novel signal processing technique will be presented that estimates power and phase spectra of multivariate dataset via data-adaptive modes obtained in time-embedded phase space. The key feature of the Data adaptive Harmonic decomposition (DAH) method relies on the construction of covariance matrices that exploit cross correlations differently than in Principal Component Analysis and Multichannel Singular Spectrum Analysis. Eigenmodes associated with DAH covariance matrices form an orthogonal set of oscillating data-adaptive harmonic modes (DAHMs) that come in pairs and in exact phase quadrature for a given Fourier frequency, aka *sine* and *cosine*.

The recent Multilayer Stochastic Model (MSM) framework introduced in [Kondrashov, Chekroun and Ghil, 2015] emphasizes the ubiquitous role of nonlinear, stochastic as well as memory effects for the derivation of data-driven closure models with good skill in simulating and predicting main dynamical features of the targeted spatiotemporal field as an output of a high-end geophysical model, or as a set of observations. However, if the input data are not numerous enough and exhibit mixture of different spatiotemporal scales, the analysis may reveal multiple predictors and complex model structure. The DAH decomposition provides an attractive data-adaptive alternative via multilayer stochastic Stuart-Landau models (MSLM), which reduce the data driven modeling effort to elemental MSMs stacked per frequency with fixed and much smaller number of coefficients to estimate. In particular, the pairs of data-adaptive harmonic coefficients (DAHCs), obtained by projecting the input dataset onto DAHMs, can be effectively modeled within a universal parametric family of simple nonlinear stochastic models - coupled Stuart-Landau oscillators stacked per frequency, and driven at all frequencies by the same noise realization. DAH-MSLM results for climate modeling and prediction will be presented.

Keywords: stochastic inverse modeling, climate prediction, data-adaptive decomposition

Conditional stochastic model chains in reduced space: Towards efficient simulation of non-stationary typhoon precipitation patterns

*Boyko Dodov¹

1. AIR Worldwide Boston

Stochastic simulation of realistic and statistically robust patterns of Tropical Cyclone (TC) induced precipitation is a challenging task. It is even more challenging in a catastrophe modeling context, where tens of thousands of typhoon seasons need to be simulated in order to provide a complete view of flood risk. Ultimately, one could run a coupled global climate model and regional Numerical Weather Prediction (NWP) model, but this approach is not feasible in the catastrophe modeling context and, most importantly, may not provide TC track patterns consistent with observations. Rather, we propose to leverage NWP output for the observed TC precipitation patterns (in terms of downscaled reanalysis 1979-2015) collected on a Lagrangian frame along the historical TC tracks and reduced to the leading spatial principal components of the data. The reduced data from all TCs is then grouped according to timing, storm evolution stage (developing, mature, dissipating, ETC transitioning) and central pressure and used to build a dictionary of stationary (within a group) and non-stationary (for transitions between groups) covariance models. Provided that the stochastic storm tracks with all the parameters describing the TC evolution are already simulated, a sequence of conditional samples from the covariance models chosen according to the TC characteristics at a given moment in time are concatenated, producing a continuous non-stationary precipitation pattern in a Lagrangian framework. The simulated precipitation for each event is finally distributed along the stochastic TC track and blended with a non-TC background precipitation. The proposed framework provides means of efficient simulation (10000 seasons simulated in a couple of days) and robust typhoon precipitation patterns consistent with observed regional climate and visually undistinguishable from high resolution NWP output. The framework is used to simulate a catalog of 10000 typhoon seasons implemented in a flood risk model for Japan.

Keywords: Reduced models, Stochastic simulation, Typhoon precipitation