

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 4 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2011～2014

課題番号：23700278

研究課題名(和文)クラウドコンピューティングを利用したデータ同化システムの開発

研究課題名(英文)Development of data assimilation system using cloud computing technology

## 研究代表者

長尾 大道(Nagao, Hiromichi)

東京大学・地震研究所・准教授

研究者番号：80435833

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、ベイズ統計学の枠組みで数値シミュレーションモデルと観測データを融合するための計算基盤技術であるデータ同化を、クラウドコンピューティングを利用して実行するためのデータ同化システムを開発した。大地震に伴う地震音波伝播を題材に双子実験を実施したところ、各震源パラメータに関する周辺化事後分布は正しく真の値の周辺に分布し、かつ事後分布の最大値を与えるパラメータから再現された理論波形と擬似観測波形は、明らかに一致することが確認された。これにより、本研究で開発したデータ同化システムが適切に動作していることが実証された。

研究成果の概要(英文)：Data assimilation (DA) is fundamental computational technique to integrate numerical simulation models and observation data based on Bayesian statistics. The purpose of this study is to develop DA system using cloud computing technology. The prototype has been verified by a twin experiment that attempted to reproduce true hypocentric parameters from synthetic infrasound variations due to seismoacoustic wave propagation excited by a large earthquake. The estimated marginal posterior distributions related to the hypocentric parameters were clearly consistent with the true values, and the theoretical infrasound variations reproduced from the optimum parameters successfully explained the synthetic observation data. This result demonstrates that the proposed DA system works properly.

研究分野：データ同化

キーワード：データ同化 クラウドコンピューティング 地震音波 南海トラフ地震 粒子フィルタ マルコフ連鎖  
モンテカルロ 事後分布 微気圧

### 1 . 研究開始当初の背景

近年のハイパフォーマンスコンピューティング (HPC) 環境の大幅な飛躍により、様々な科学分野において大規模な数値シミュレーションが実施されるようになってきた。最近では非常に複雑なモデルに対する高時空間分解能シミュレーションへの期待が高まり、地球シミュレータや「京」等の理論性能が数十ペタ flops にも及ぶ超大型の単体スパコンが開発されたが、それでもなお、様々な自然現象をより現実的な精度で再現するためには、まだ計算機能力が不足していることがほとんどである。しかしながら、このような超大型単体スパコンは国家予算級の開発費を必要とするため、これまでのような単純な大型化指向路線からの変更を迫られている。そのため、今後ますます大規模化していくことが確実なシミュレーションを実施していくための次世代 HPC 環境を整備する必要性が日に日に高まっている。次世代 HPC 環境の候補の一つとしては、「クラウドコンピューティングシステム」(以下、CCS) が有力な方法として提案されており、情報検索サイト等、既に CCS の利用が実現されている分野も存在する。一方、CCS の科学利用についてはまだ議論が始まったばかりであり、日本国内においてもその仕様はまだ固まっていないのが現状である。この時点で、CCS に適した科学計算のためのアプリケーションソフトを今から開発しておくことは、まさに時代を先取りするものであり、逆にアプリケーションソフト側から必要な CCS 仕様を逆提案することも可能であると考えられる。

### 2 . 研究の目的

これまでに応募者は、CCS 上での利用を前提とした、粒子フィルタ法による時系列解析システム「CloCK-TiME」(Cloud Computing Kernel for Time-series Modeling Engine) を実験的に構築し、統計数理研究所データ同化研究開発センターのウェブサーバを通じて、一般ユーザへの公開を開始した (<http://sheep.ism.ac.jp/CloCK-TiME/>)。CloCK-TiME は、ウェブサーバ上に配置されたユーザインターフェース (UIF) を介してユーザから多変量時系列データおよび解析パラメータを受け取り、同センターが所有する PC クラスタを利用して解析した後、その結果を UIF 上に表示するシステムである (Nagao et al., 2010)。

本研究では、ベイズ統計学の枠組みでシミュレーションモデルと観測データをつなぎ、モデルパラメータを高精度に決定しつつ、将来予測が可能なシミュレーションモデルを構築するための計算基盤技術である「データ同化」を CCS 上で実現するためのデータ同化システムを開発する。データ同化は、シミュレーションによる将来の状態予測、観測データに基づくフィルタリングによる現在状態の推定を時間ステップごとに繰り返し、最終

的には状態およびモデルパラメータの確率密度関数を推定する手法であり (例えば、樋口 他, 2011)、身近なところでは気象予報において主要な役割を演じている。データ同化では、様々なモデルパラメータに対応したシミュレーションを反復実施する必要があるため、一般には大きな計算機資源を必要とする。そのため、シミュレーションやフィルタリングのための並列計算コーディングの手間が、当該分野の研究者にとっては大きな負担となっている。本研究で開発するデータ同化システムは、このようなコーディングの重い負担から研究者を解放し、かつ簡単にデータ同化研究ができる並列計算機環境を提供することが大きな目的である。

### 3 . 研究の方法

本研究で開発する CCS を利用したデータ同化システムでは、ノード間の通信コストを極力抑え、かつ各 PC の性能に最適な負荷バランスを配置するために、状態の逐次推定方法として粒子フィルタ法を採用する。開発後は、同システムの実際問題への応用として、地震関連現象のデータ同化研究に適用し、性能を評価する。また、シミュレーションモデルや観測データのアップロード、あるいはデータ同化計算を制御するためのパラメータを設定する UIF を備えた上、統計数理研究所のウェブサーバから公開することにより、データ同化法の様々な分野への普及を促進するとともに、CCS を科学利用するための技術に社会に還元する。

#### (1) 平成 23 年度の研究指針

初年度はまず、CCS 上で動作するデータ同化システムを開発する上で、克服すべき問題を十分に精査するところから始める。本システムは、前述した CloCK-TiME をベースに開発するが、(a) スパコン間でのデータ通信コストの低減、および (b) 性能の異なる PC クラスタの負荷バランスの 2 点については、今回新たに考慮すべき課題となる。解決の方向性としては、各時刻における逐次状態推定法として、モンテカルロ近似の一種である粒子フィルタ法を採用することが容易であろう。

粒子フィルタ法は遺伝的アルゴリズムとよく似た手法であり、「粒子」と呼ばれる実現値の集合によって状態ベクトルの分布関数を表現するが、時間発展をさせていく際に尤度の小さな粒子は、分布関数表現への貢献が小さいとして除外するという操作を実施する。観測時刻が進むにつれて粒子数が減り、分布関数の表現力が落ちることを避けるための方法としては、初期粒子数を極めて大きくしておく SIS 法と、各時刻において粒子リサンプリングを実施する SIR 法とがある。後者は PC に実装されているメモリを最大限に活用する場合に有効であるが、粒子リサンプリングによるノード間およびクラスタ間通信が頻発するため、本研究計画で開発する

システムには不向きである。したがって本システムでは、SIS法を採用することにより、(a)の問題を解決することが可能である。また粒子フィルタ法では、ノード間で大きな性能差がある場合には、粒子を各コアに均等に配分してしまうと、実行性能が小さな計算機によって決まってしまうという問題が生じる。これを解消するために、CloCK-TIMEでは計算実行前に各ノードの性能を調査し、それに合わせて粒子数(=負荷)を最適配分するように設計している。本研究で開発するシステムにもこの方法を導入することにより、(b)の問題を解決する。

平成23年度に実施予定のこれらの研究開発の成果は、統計関連学会連合大会をはじめとする国内学会、および国際学会において発表する。

## (2) 平成24年度以降の研究指針

前年度の準備を踏まえ、データ同化システムの本格的な開発に着手する。特に複数のPCクラスタ群によって構成されるCCSについては慎重な実験を繰り返し実施し、問題点を洗い出しておく。一通りの開発を終えた段階で、実際問題への応用例として地震関連現象のデータ同化研究に対して本システムを適用する。本システムの性能を十分評価した後、専用のユーザインターフェースを開発し、インターネット上を通じて一般に公開する。

これらの研究開発の成果は、統計学ならびに情報学に関連した国内学会および国際学会において発表し、数編の論文にまとめた上で国際誌へ投稿する。

## (3) 本研究の実施体制

データ同化の理論的側面ならびに本研究全般に関する指導および助言は、研究代表者の本研究計画申請時点での受入教員である統計数理研究所の樋口知之教授(現:同研究所長)から仰ぐ。

## 4. 研究成果

本研究で開発したデータ同化システムを検証するため、大地震によって大気中に励起される「地震音波」の伝播現象を応用テーマとして選んだ。以下では、地震音波について概説し、「地震音波データ同化システム」について検証した結果について述べる。

### (1) 大地震によって励起される地震音波

2011年3月11日の東北地方太平洋沖地震(東日本大震災)は、特に巨大津波による未曾有の大被害をもたらした。近い将来、確実に起こるとされる東海・東南海・南海連動地震(南海トラフ地震)に向けて、様々な科学分野の研究者が、それぞれの視点から努力を積み重ねているが、地震や津波を解析する上で、従来の地震波解析とは異なる切り口の一つとして、それらによって大気中に励起される「地震音波」がある。東日本大震災の際に

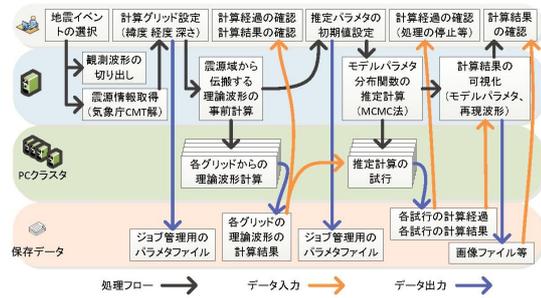


図1: 地震音波データ同化システムにおけるデータ同化プロセスの概要。

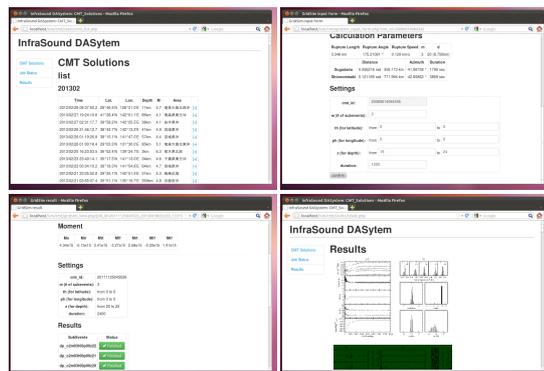


図2: データ同化システムのグラフィカルユーザインターフェース。

は、津波起源の音波によって励起されたと考えられる地震音波によって、地球の大気ならびに高度数十kmないし数百kmに存在する電離層が大規模に揺り動かされている様子が、日本およびその周辺国の微気圧計およびGNSS受信機によって観測された(Arai et al., 2011; Tsugawa et al., 2011; Nagao and Higuchi, 2012)。音波が津波よりずっと速く伝播するというこの性質を利用し、地震音波を詳しく解析することにより、津波が到来する前にその規模と到達時刻をいち早く知らせることのできる津波早期警戒システムの実現に期待が寄せられている。

### (2) 地震音波データ同化システム

津波早期警戒システムの実現のためには、地震音波伝播に関するシミュレーションと観測データの解析を同時に実施することが必要であることは自明であり、そのための基盤技術であるデータ同化の役割が非常に大きくなるであろうことは想像に難くない。

図1に、地震音波データ同化システム内で実施されているプロセスの概要を示す。本システムによって推定するモデルパラメータは、震源の位置、断層の破壊伝播速度、断層沿いに仮想的に配置した各サブイベントのモーメントマグニチュード、大気中を吹く背景風に伴う時間遅延、微気圧観測点における観測ノイズの分散である。本システムでは、まず予め用意された過去の地震リストから選択された地震イベントについて、気象庁

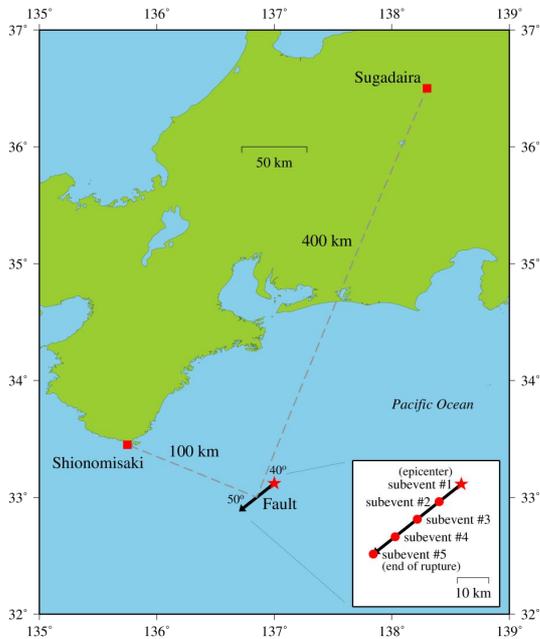


図 3: 地震音波データ同化システムの検証のための双子実験で仮定した南海トラフ地震。

が発表した震源に関するパラメータを得る。パラメータの事前分布を適当に定めることにより、ノーマルモード重合による数値シミュレーション (Kobayashi, 2007) によって得られる理論波形と観測波形との適合度を表す事後分布をマルコフ連鎖モンテカルロ法 (MCMC) によって推定する。本システムでは、図 2 に示すグラフィカル UIF より、データ同化計算を制御するための各種設定を柔軟に行うことが可能となっている。

### (3) 双子実験によるシステムの検証

本システムのアルゴリズムを検証するために、仮想的な震源モデルから数値シミュレーションによって生成した気圧変動を擬似観測データとし、本システムによって震源パラメータを再現する「双子実験」を実施する。図 3 に示す通り、この双子実験においては近い将来起こるとされる東南海地震を想定し、遠州灘沖においてマグニチュード 8 クラスの大地震が起こったと仮定する。ここでは真の震源の深さを 10km、断層の長さを 40km、断層の破壊伝播速度を 2km/秒、そして地震全体のモーメントマグニチュード  $M_0$  を  $10^{21}$ Nm、断層沿いに並べた 5 個のサブイベントのモーメントマグニチュードを、それぞれ  $0.15M_0$ 、 $0.25M_0$ 、 $0.35M_0$ 、 $0.20M_0$ 、 $0.05M_0$  とする。実際の南海トラフ地震においては、もっと長い断層が形成されることが想定されているが、あまりに長大な断層を仮定した場合には、サブイベントの個数も大幅に増やす必要が生じ、双子実験においてそのような大規模計算は検証実験の目的から不要であるため、ここでは比較的小規模な断層を仮定した。このように断層の長さを短くとると、観測点における気圧変動の振幅が増大するため、モーメント

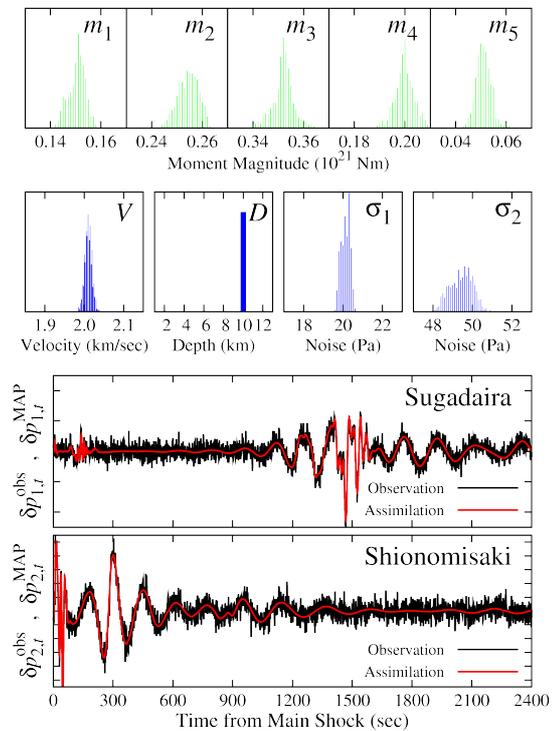


図 4: マルコフ連鎖モンテカルロによって推定された各震源パラメータについての周辺化事後分布、および事後分布の最大値を与えるパラメータから得られる地震音波の理論波形と擬似観測波形の比較。

マグニチュードを小さくした方がより現実的な振幅に近づくが、これは単純にスケールアップだけの問題であり、データ同化による推定には本質的に何ら影響を及ぼさないため、そのまましておく。震源域から励起された地震音波による大気変動は、実際に微気圧観測を実施している、断層の中心から 400km 離れた菅平観測点、ならびに 100km 離れた潮岬観測点において気圧変動として観測されるものとする。なお、この仮想地震のようないわゆる海洋プレート境界型地震の場合、震源の水平位置に関しては、地震波解析の段階で大きな推定誤差が生じることは考えにくい。本実験においては推定パラメータから除外する。また、大気中の背景風による地震音波伝播への影響をノーマルモード重合計算によって再現することは難しいため、遅延パラメータについても推定パラメータから除外する。したがって、本実験において推定すべきパラメータは、震源の深さ  $D$ 、破壊伝播速度  $V$ 、各サブイベントのモーメントマグニチュード  $m_i$  ( $i=1, \dots$ )、および各観測点における観測ノイズの分散  $\sigma_k^2$  ( $k=1, \dots$ ) となる。

図 4 上に、MCMC によって得られたサンプリングから推定した各パラメータについての周辺化事後分布を示す。データ同化を行う時系列の長さとしては、地震音波に伴う気圧変動が両観測点で顕著となる 2,400 秒間とした。なお、ここでは震源パラメータに関する事前分布として一様分布を仮定している。MCMC の

アルゴリズムとして Metropolis 法を採用したところ、約 25,000 ステップ目あたりから定常状態に達したため、30,005 ステップ目から 80,000 ステップ目まで 5 ステップごとにサンプリングを収集することによって得られた 10,000 個のサンプリングを用いて事後分布を推定した。本実験においては、MCMC のためのパラメータの初期値として、あえて最悪の値を与えた上で実施したものの、各パラメータの周辺化事後分布は、真の値の周辺に正しく分布していることが分かる。なお、ノーマルモードの計算アルゴリズムの制約上、震源の深さについては 1.25km おきの離散値に対してのみ計算可能であり、また他のパラメータと比較して非常に早く収束したため、得られた周辺化事後分布は Dirac のデルタ関数の形状となっている。気圧変動は、震源の深さに対して非常に敏感であるという特徴を持つため(Nagao et al., 2012)、このような早い収束をもたらしたと考えられる。

最後に、MCMC によって得られた事後分布の最大値を与えるパラメータ (MAP 解) から再現された理論波形と、擬似観測波形を比較したものを、図 4 下に示す。理論波形は観測波形の細部に至るまで説明できていることは明らかであり、このことから地震音波データ同化システムのアルゴリズムが適切に動作していることが分かる。時系列の長さをもっと短くした場合においても、地震音波に伴う気圧変動が起こっている途中であるにも関わらず、震源パラメータの推定および波形の再現がきちんとなされることが確認されており、本システムが早期警戒システムへと発展する可能性を十分に秘めていることを示唆している。

#### (4) まとめと今後の展望

本課題で開発した CCS 上で稼働するデータ同化システムの応用の一例として、地震の際に励起される地震音波に伴う気圧変動について、数値シミュレーションと微気圧観測データを融合することにより、震源パラメータを推定することを目的とした地震音波データ同化について述べた。近い将来起こるとされる東南海地震を想定した仮想地震を用いた双子実験を実施することにより、本システムのデータ同化アルゴリズムを検証した。双子実験においては、仮定した真の震源パラメータを MCMC によって正しく推定することに成功し、本システムの妥当性を示すことができた。

本課題で開発したデータ同化システムは、将来的には、様々なタイプの数値シミュレーションモデルの取り扱いや、データ同化計算の高速化等、数多くのハードルを乗り越える必要がある。そのためには、数値シミュレーションモデルの微分方程式群を記述するためのマークアップ言語の開発や、レプリカ交換 MCMC 法等の採用による分布関数の推定の効率化を図る必要があるものの、十分に達成

可能な目標である。本システムの開発によって、その礎を築くことはできたものの、これより先はあらゆる分野の研究者が知恵を出し合ってプロジェクトを大規模化し、共通の目標に向かって動いていく必要がある。

#### <引用文献>

Nagao, H. and T. Higuchi, Web application for time-series analysis based on particle filter available on cloud computing system, The Proceedings of 13th International Conference on Information Fusion, 1-6, 2010.

樋口知之, 上野玄太, 中野慎也, 中村和幸, 吉田亮, データ同化入門 朝倉書店 2011.  
Arai, N., M. Iwakuni, S. Watada, Y. Imanishi, T. Murayama, and M. Nogami, Atmospheric boundary waves excited by the tsunami generation related to the 2011 great Tohoku-Oki earthquake, Geophys. Res. Lett., 38, doi:10.1029/2011GL049146, 2011.

Tsugawa, T., A. Saito, Y. Otsuka, M. Nishioka, T. Maruyama, H. Kato, T. Nagatsuma, and T. Murata, Ionospheric disturbances detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake, Earth Planets Space, 63, 875-879, 2011.

Nagao, H. and T. Higuchi, Data assimilation of the Earth's atmospheric and ionospheric oscillations excited by large earthquakes, The Proceedings of 15th International Conference on Information Fusion, 2472-2479, 2012.

Kobayashi, N, A new method to calculate normal modes, Geophys. J. Int., 168, 315-331, 2007.

Nagao, H., N. Kobayashi, S. Nakano, and T. Higuchi, Fault parameter estimation with data assimilation on infrasound variations due to big earthquakes, The Proceedings of 14th International Conference on Information Fusion, 1-6, 2011.

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計7件)

長尾大道, 樋口知之, 地震音波データ同化システムの開発 双子実験による検証, 統計数理, Vol. 61, No. 2, 257-270, 2013, 査読有.

Nagao, H., T. Higuchi, S. Miura, and D. Inazu, Time-series modeling of tide gauge records for monitoring of the

crustal activities related to oceanic trench earthquakes around Japan, The Computer Journal, Vol. 56, No. 3, 355-364, doi:10.1093/comjnl/bxs139, 2013, 査読有.

長尾大道, 佐藤光三, 樋口知之, マルコフ連鎖モンテカルロ法を利用したトレーサー試験からフラクチャーの物理パラメータを推定する方法, 石油技術協会誌, Vol. 78, No. 2, 197-209, 2013, 査読有.

長尾大道, クラウドコンピューティングを利用した統計計算サービス, システム制御情報学会誌, Vol. 57, No. 4, 160-165, 2013, 査読無.

長尾大道, 樋口知之, データ同化 モデルパラメータの分布推定, 日本機械学会誌, Vol. 116, No. 1131, 89-92, 2013, 査読無.

長尾大道, 樋口知之, 三浦哲, 稲津大祐, 地球地殻の活動監視を目的とした粒子フィルタ法による長期潮位変動解析, 日本統計学会誌, Vol. 42, 119-133, 2012, 査読有.

本橋永至, 磯崎直樹, 長尾大道, 樋口知之, 状態空間モデルによるインターネット広告のクリック率予測, オペレーションズ・リサーチ, Vol. 57, 574-583, 2012, 査読有.

#### [学会発表] (計 11 件)

Nagao, H. and T. Higuchi, Cloud services to release techniques of data assimilation, 11th Annual Meeting of Asia Oceania Geosciences Society, 口イ トン札幌(北海道札幌市), 2014 年 7 月 29 日.

長尾大道, 樋口知之, クラウドサービスによるデータ同化技術の公開, 日本地球惑星科学連合 2014 年大会, パシフィコ横浜(神奈川県横浜市), 2014 年 4 月 30 日.

長尾大道, 樋口知之, 地震音波データ同化システムの開発, 2013 年度統計関連学会連合大会, 大阪大学豊中キャンパス(大阪府豊中市), 2013 年 9 月 10 日.

Nagao, H. and T. Higuchi, Data assimilation system for seismoacoustic waves, The 16th International Conference on Information Fusion, イスタンブール(トルコ), 2013 年 7 月 11 日.

長尾大道, 富澤一郎, 家森俊彦, 金尾政紀, 樋口知之, 固体地球起源の大気変動現象を解明するためのデータ同化システムの開発, 日本地球惑星科学連合 2013 年大会, 幕張メッセ(千葉県千葉市), 2013 年 5 月 24 日.

長尾大道, 樋口知之, クラウドコンピューティングシステムを利用した地球物理データの多変量時系列解析, 日本地球惑星科学連合 2013 年大会, 幕張メッセ(千葉県千葉市), 2013 年 5 月 20 日.

Nagao, H., Our attempts to broaden the application of data assimilation into many fields empowered by massively parallel computing, Data Assimilation, The 3rd workshop on numerical methods for solving the filtering problem and high order methods for solving parabolic PDEs, オックスフォード(イギリス), 2012 年 9 月 25 日.

Nagao, H. and T. Higuchi, Data assimilation of the Earth's atmospheric and ionospheric oscillations excited by large earthquakes, The 15th International Conference on Information Fusion, シンガポール, 2012 年 7 月 12 日.

Nagao, H. and T. Higuchi, Fault parameter estimation from coseismic infrasound variation excited by a large earthquake, European Geosciences Union General Assembly 2012, オーストリア・ウィーン, 2012 年 4 月 26 日.

長尾大道, 樋口知之, クラウドコンピューティングを利用したデータ同化システムの開発(序), 2011 年度統計関連学会連合大会, 九州大学伊都キャンパス(福岡県福岡市), 2011 年 9 月 6 日.

Nagao, H., R. Yoshida, and T. Higuchi, Hybrid Bayesian filter algorithm for multivariate time-series modellings on cloud computing systems, Yeditepe International Research Conference on Bayesian Learning, イスタンブール(トルコ), 2011 年 6 月 15 日.

#### [その他]

##### 報道

日本経済新聞 2012 年 1 月 26 日夕刊第一面

##### ハイライト論文選出

日本地球惑星科学連合 2012 年大会

##### Web サービス「CLOCK-TIME」

<http://sheep.ism.ac.jp/CLOCK-TIME/>

#### 6. 研究組織

##### (1) 研究代表者

長尾 大道 (NAGAO, Hiromichi)

東京大学・地震研究所・准教授

研究者番号: 80435833