

モデル駆動型システムズエンジニアリングに基づくモデル管理およびデータ解析 ー Nano-JASMINE データ解析への応用 ー

初鳥 陽一^{*1} 宮下 尚^{*1} 清水 淳也^{*1} 山田 良透^{*2}

A Model Management and Data Analysis based on Model Driven Systems Engineering - Application for Nano-JASMINE data analysis -

Yoichi HATSUTORI^{*1}, Hisashi MIYASHITA^{*1}, Junya SHIMIZU^{*1} and Yoshiyuki YAMADA^{*2}

Abstract

Since data analysis for modern astronomy is getting more complicated than ever, effective information sharing among scientists and software engineers, and a distributed computing are important to compute large amount of data. In this study, a model construction and data analysis framework based on Model Driven Systems Engineering (MDSE) which makes it possible to effectively share information among project members are presented. Especially, we propose analytic framework based on a grey box modeling and least-square method used widely in data analysis. In particular, scientific theories and equations can be implemented as models of MDSE. In addition, proposed framework enables us to automatically generate effective machine codes from the models, which significantly reduces analysis efforts and development costs. Generated machine codes are computed by distributed processing technology based on MapReduce and GPGPU. As an example, proposed framework is applied to Japanese space astrometry project named Nano-JASMINE.

Keywords: Nano-JASMINE, Systems engineering, Data analysis framework, Distributed computing

概 要

近年、天文学におけるデータ解析は複雑化しつつあり、科学者と解析担当者間の効率的な情報共有および大規模なデータの分散処理が重要な課題となっている。本研究では、プロジェクトメンバー間での効果的な情報共有を可能とする、モデル駆動型システムズエンジニアリング (Model Driven Systems Engineering, MDSE) に基づくモデル構築およびデータ解析のフレームワークの提案を行う。特に、データ解析で多く用いられているグレーボックスモデリングと最小二乗法に基づく解析フレームワークを提案する。具体的には、科学理論や方程式は MDSE のモデルとして実装し、そのモデルから解析コードを生成することで、開発のコストや労力を低減する。生成されたコードは、MapReduce および GPGPU と呼ばれる分散処理技術に基づいて処理を行う。一例として、提案したフレームワークを Nano-JASMINE と呼ばれる位置天文観測衛星のデータ解析へ適用する。

記 号

i	星の ID	Au	1 天文単位
l	観測の ID	$bG(t)$	時刻 t における衛星の位置
$ui(t)$	時刻 t における星 i の位置	$q(t)$	時刻 t における衛星のクオータニオン
α_i	星 i の赤経	an	B-Spline 係数
δ_i	星 i の赤緯	$Bn(t)$	B-Spline 基底関数
ω_i	星 i の年周視差	η	CCD 座標における星の位置 (経度)
$\mu\alpha_i$	星 i の赤経方向の固有運動	ζ	CCD 座標における星の位置 (緯度)
$\mu\delta_i$	星 i の赤緯方向の固有運動		

^{*1} 日本アイ・ビー・エム株式会社 東京基礎研究所 (IBM Research Tokyo, IBM Japan)

^{*2} 京都大学大学院理学研究科 (Graduate School of Science, Kyoto University)

1. はじめに

近年の天文学におけるデータ解析は高精度化、複雑化が進んでいる。たとえば国立天文台、東京大学、京都大学が中心となって進めている JASMINE 計画¹⁾では、人工衛星を用いて銀河系内の星の観測を行い、得られた観測データを基に星の位置や年周視差をミリ秒角 (1 ミリ秒角は 1/3,600,000 度) やマイクロ秒角の精度で求めることが計画されている²⁾。JASMINE 計画では、主鏡口径 5cm の光学系を搭載する超小型衛星 Nano-JASMINE、主鏡口径 30cm の小型衛星 Small-JASMINE、主鏡口径 80cm の中型衛星 JASMINE という 3 機の衛星が検討されており、目標精度や観測範囲を改善しながら、最終的には銀河系中心のバルジと呼ばれる領域の位置天文観測を実施することを目標としている。表 1 に 3 機の衛星の目標精度、観測領域、打ち上げ年度について示す。このような高精度な位置天文観測を実現するためには、様々な解析手法を適用する必要や、人工衛星の姿勢パラメータや機器の劣化など観測結果に影響を及ぼす要素を洗い出し、解析の際に考慮に入れる必要がある。そのため、解析担当者がコードを実装する際には、複雑なモデルを理解し、大規模なデータを処理するための効果的なコードの実装を行うことが必要不可欠である。また、生成した解析コードを次のプロジェクトへ応用したいという科学者たちからの要求にこたえるためには、汎用的なコードの生成も必要となる。これらの要求を実現するためには、科学者と解析担当者との間における効果的な情報共有と大量のデータ処理を実現するための分散処理が重要な課題となる。そこで、本研究では、プロジェクトメンバー間での効果的な情報共有を可能とする MDSE の技術^{3), 4)}と、MapReduce^{5), 6)}および GPGPU^{7), 8)}の分散処理技術に基づく、モデルの実装およびデータ解析を行うフレームワークを提案し、すでにフライトモデルの開発が行われている Nano-JASMINE のデータ解析に対して適用する。

表 1 JASMINE 計画で計画されている 3 機の衛星の比較¹⁾

	Nano-JASMINE	Small-JASMINE	JASMINE
主鏡口径	5cm	30cm	80cm
衛星質量	35kg	400kg	1500kg
目標精度	3mas	10 μ as	10 μ as
観測領域	全天	バルジ数平方度	バルジ 200 平方度
運用予定	2013-2015	2017-2019	2020 年代

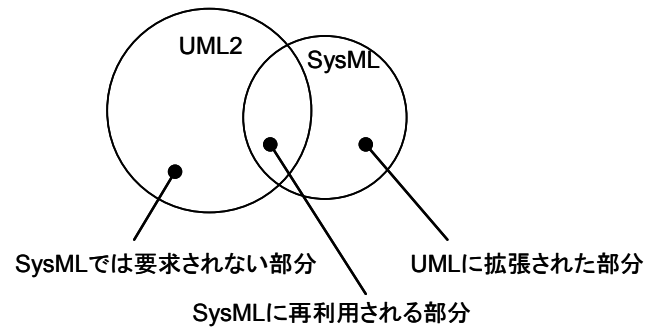
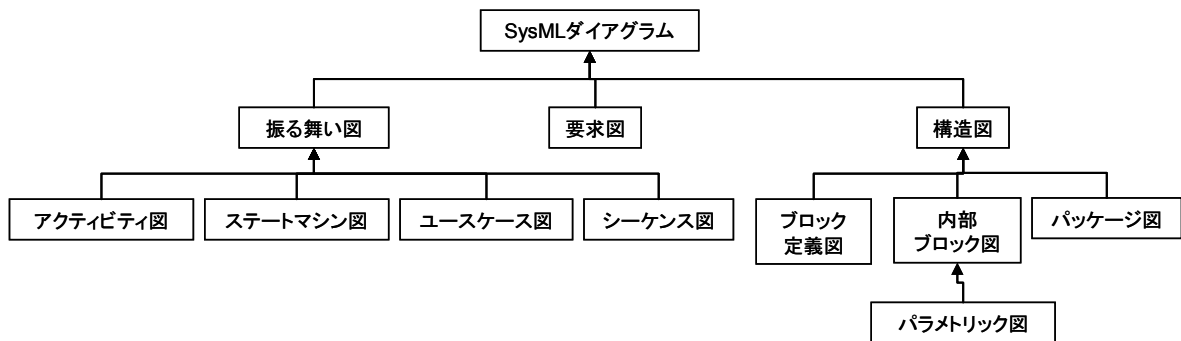
(mas: milli arc-sec, μ as: micro arc-sec)

2. MDSE

MDSE とは OMG (Object Management Group²⁾) と呼ばれるモデリング言語の標準化団体によって仕様が策定された SysML^{3), 4)} (Systems Modeling Language) を用いて、自動車・航空機などの複雑なシステムを形式的に記述し設計するための新しい手法のひとつである。SysML はソフトウェアモデリング言語の一つである UML2 (Unified Modeling Language) の拡張付きサブセットであり、ハードウェア、ソフトウェア、データ、人、手順、設備を含むシステムの仕様化、分析、設計、確認、検証をサポートするグラフィカルなモデリング言語である²⁾。図 1 に SysML と UML との関係を示し、図 2 に SysML で用いられるダイアグラムを示す。MDSE の技術は従来自動車や航空機、組み込みシステムの開発に用いられてきたが、本論文では、衛星データ解析に対して MDSE によるモデリング手法を適用する。具体的には、ブロック定義図 (Block Definition Diagram, BDD) と内部ブロック図 (Internal Block Diagram, IBD) と呼ばれるダイアグラムを用いて科学者から提供された数式を SysML に基づいて表現することで、複雑化が進んでいる天文学におけるデータ解析に対して科学者と解析担当者との間における効果的な情報共有を実現する。衛星データ解析に対して MDSE を適用する潜在的なメリットとしては、科学者と解析担当者双方が共通理解できる形式で記述することによるコミュニケーションの向上、システムの妥当性確認と検証、開発効率の向上などが挙げられる。

1. <http://www.jasmine-galaxy.org/index-j.html>

2. <http://www.omg.org>

図 1 SysML と UML の関係³図 2 SysML ダイアグラムの分類³

SysML によるモデル化では、システムの要素を記述する単位としてブロックが用いられ、値や制約、要求といった性質を定義することができる。BDD はブロックの特性と他のブロックとの構造的関係性を定義するためのダイアグラムであり、IBD は BDD で定義されたブロックの内部構造としてブロック間の接続方法を定義するためのダイアグラムである。BDD と IBD の関係性の例を図 3 に示す。図 3 左の BDD では A の構成要素として B,C,D を定義し、右の IBD では A の内部で B,C,D がどのように接続されるかを定義する。本論文では、数式やパラメータといった最小の構成要素をブロックで表現し、構造と流れを定義するために BDD と IBD を用いる。

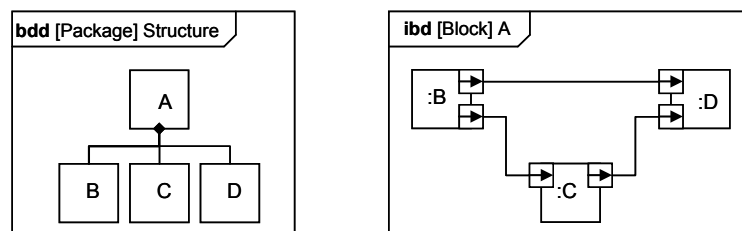


図 3 BDD と IBD の例（左：BDD，右：IBD）

3. JASMINE 計画におけるデータ解析

1 節で紹介した JASMINE 計画のデータ解析について述べる。JASMINE 計画では、人工衛星を用いて銀河系内の星の位置・距離・固有運動を高精度に測定することを計画しており、地球近傍を周回する人工衛星から星を観測した場合、星の運動は星自身のもつ固有運動と地球の公転運動に起因するみかけの楕円運動により螺旋運動に見える。Nano-JASMINE では 2 年間の観測を通して、同じ星を複数回観測することによって螺旋運動のパラメータを高精度に求めるこ

3. <http://www.omg.sysml.org>

とを目的としている⁹⁾。図4にJASMINE計画における観測のイメージ図として、地球の公転運動に起因するみかけの楕円運動と、螺旋運動と観測の想像図を示す。

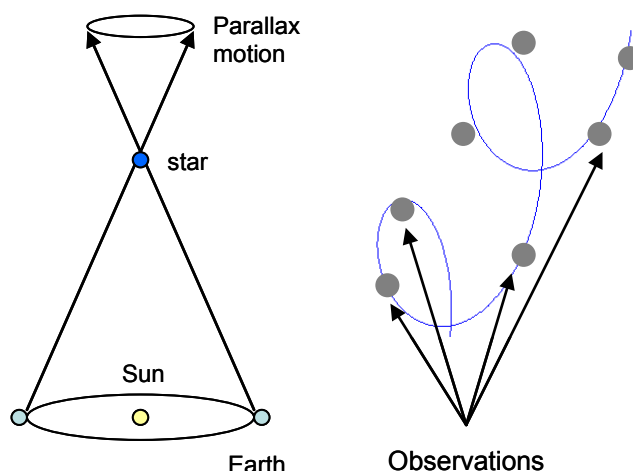


図4 地球の公転運動に起因するみかけの楕円運動（左）と、螺旋運動と観測の想像図（右）

図4に示すように、JASMINE計画におけるデータ解析とは、星の動きは螺旋運動であることを仮定し、そのパラメータを求める問題である。推定すべき星のパラメータは、星の位置（2個）・距離（1個）・固有運動（2個）の5つのパラメータであり⁹⁾、Nano-JASMINEでは地球近傍の約50万個の星それぞれに対して、5個のパラメータを同定する。しかし、Nano-JASMINEデータ解析の特徴として、星のパラメータだけでなく姿勢のパラメータも同時に推定する必要があるという点が挙げられる。たとえば、Nano-JASMINEの目標とする要求精度は3masと非常に高い一方で、露光時間中の指向安定度要求は740masとなっている¹¹⁾。これは得られた画像から星の位置を3masの精度で求めたいにもかかわらず、衛星の姿勢運動によって画像が740mas変動している可能性があることを示している。最小二乗法を用いて星のパラメータと姿勢のパラメータを同時に解くことで要求される精度を達成することが考えられているが、そのためには望遠鏡の収差など観測機器のパラメータ、CCDの放射線劣化など機器の劣化のパラメータといったミッション機器にかかわるパラメータだけでなく、人工衛星の姿勢のパラメータなど観測結果に影響を及ぼすバス機器のパラメータも同時に推定しなければならない。このように、ミッション系、バス系のパラメータを同時に解くという複雑なデータ解析のための情報共有の一つの手段としてMDSEを適用した。ただし、前述のように星の動きは螺旋運動であることが既知で、そのパラメータを求める問題を仮定している。このように、背景にある数式は既知とし、そのパラメータなど一部が未知であるとするモデル化手法はグレーボックスモデルと呼ばれており¹²⁾、本研究ではこのグレーボックスモデルで記述された衛星データ解析を前提としている。

4. 本研究で用いる解析フレームワーク

4.1 全体像

位置天文データ解析では、星のパラメータの他に、観測に影響を及ぼすパラメータを同時に推定する必要がある。そのため、星の螺旋運動や重力レンズ効果による補正、光行差による補正など物理学の知見以外にも、衛星の姿勢運動を表現する数式、搭載機器、機器の劣化、環境外乱や熱、衛星軌道など周辺環境を表現する数式など工学的な知見も解析に考慮する必要がある。位置天文観測に影響する様々な要素の例を図5に示す。そのため、解析においては複数の学問領域にまたがる数式群が必要になることから、科学者たちと解析担当者の間の効率的な情報共有が高精度な解析を行う上で重要な課題となる。そこで、MDSEに基づくモデルを構築し、解析コードを生成することで開発のコストや労力を低減する。また、生成されたコードを、分散処理を用いて大規模なデータを効率よく処理することを目指す。提案するフレームワークの全体像を図6に示す。図6において、科学者の提供する数式は解析担当者によってMDSEのモデルとして表現される。解析コードはMDSEによるモデルから生成し、MapReduceおよびGPGPUを用いて分散処理させることで大

規模観測データの効率的な処理を実施する．求められたパラメータを用いてモデルを更新し，また必要があればモデルを更新する．このモデルの一部もしくは全体を次のプロジェクトの解析の際に応用する．

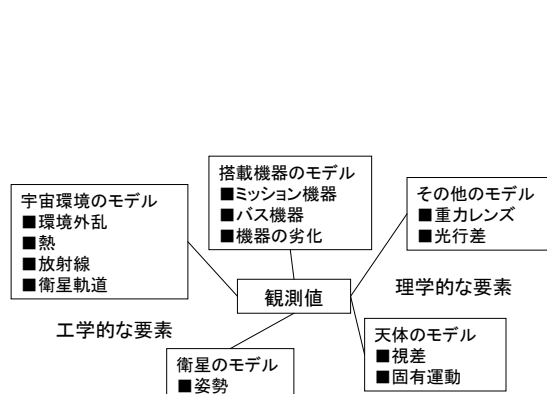


図5 観測値に影響を与える要素の例

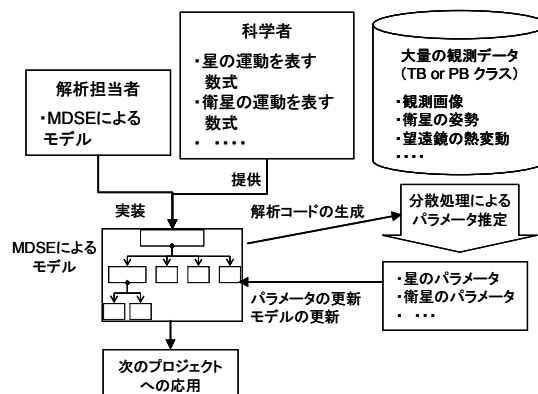


図6 本研究で用いるフレームワークの全体像

4.2 分散処理によるパラメータ推定

本研究では大規模なデータから最小二乗法を用いてパラメータを同定する問題を前提しており，正規方程式と呼ばれる

$$y = Ax \quad (1)$$

の形式の連立方程式から x の推定値を計算する．ここで， A は x の要素数を一辺とする正方疎行列で，Nano-JASMINE の問題では5節で述べるように A の一辺は1000万を超えるサイズの行列となるため，非常に膨大な数の要素を計算する必要がある．そこで，大量のデータを複数のプロセッサで計算させることが可能なフレームワークである MapReduce^{5), 6)} を適用する．MapReduce は MapPhase と ReducePhase に分かれており，MapPhase では計算を独立に実行可能な複数のタスクに分割し，結果を Key / Value ペアと呼ばれる形式で出力する．ReducePhase では MapPhase の出力結果を一つにまとめて最終結果へ変換する．今回の問題では星の観測は一つ一つが独立しており，1回の観測で行列 A の一部が埋まるため，観測ごとに処理を分割することができる．そこで，観測ごとの行列計算を MapPhase で実施し，行列のどの要素に対応するかを Key として出力し，計算された要素の値を Value として出力する．ReducePhase では，結果を全て統合することで行列 A を生成する．行列 A を計算した後，反復計算による x の推定値の計算には，行列演算に対して特に性能の高い GPGPU^{7), 8)} を用いる．GPGPU は General Purpose computation on Graphic Processing Units の略で画像処理用のハードウェアとして用いられてきた GPU を数値計算などの用途へ応用する技術であり，CPU に比べて複数のコアを持つことから並列計算の効果が高いことが知られており⁷⁾，天文学への利用に対する研究も進められている⁸⁾．

5. Nano-JASMINE データ解析への適用

5.1 Nano-JASMINE データ解析の BDD

Nano-JASMINE のデータ解析に必要な要素を定義し，BDD を用いたモデル化を行う．科学者から提供される数式群は Lindegren らの論文⁹⁾ を参考にした．Lindegren らの論文には ESA で進めている Gaia プロジェクトにおけるデータ解析のための数式群が定義されており，観測手法が同じ Nano-JASMINE データ解析にも応用ができる．そこで，Lindegren らの論文の数式群の一部を利用して BDD と IBD の構築を行った．位置天文観測において考慮すべき要素は星のモデル，姿勢のモデル，較正のモデルに分類され，それぞれが数式と推定すべきパラメータを持つ．たとえば，星の螺旋運動は

$$\begin{aligned} \mathbf{u}_i(t) &= \langle \mathbf{r}_i + t \cdot (\mathbf{p}_i \mu_{\alpha_i} + \mathbf{q}_i \mu_{\delta_i}) - \sigma \mathbf{b}_G(t) / A_u \rangle \\ \mathbf{p}_i &= [-\sin \alpha_i \quad \cos \alpha_i \quad 0]^T \\ \mathbf{q}_i &= [-\sin \delta_i \cos \alpha_i \quad -\sin \delta_i \sin \alpha_i \quad \cos \delta_i]^T \\ \mathbf{r}_i &= [\cos \delta_i \cos \alpha_i \quad \cos \delta_i \sin \alpha_i \quad \sin \delta_i]^T \end{aligned} \quad (2)$$

と記述される．ただし， $\langle \rangle$ は規格化を表し， $[\]^T$ は転置を表すものとする．衛星の姿勢はクォータニオンを 3 次の B-Spline 関数として，

$$\mathbf{q}(t) = [q_1 \quad q_2 \quad q_3 \quad q_4]^T = \left\langle \sum_{n=l-3}^l \mathbf{a}_n B_n(t) \right\rangle \quad (3)$$

と記述される．よって，星の方向ベクトル $\mathbf{v}(t)$ は

$$\mathbf{v}(t) = [\cos \zeta \cos \eta \quad \cos \zeta \sin \eta \quad \sin \zeta]^T = f(\mathbf{q}) \mathbf{u}(t) \quad (4)$$

$$f(\mathbf{q}) = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1 q_2 + q_3 q_4) & 2(q_1 q_3 - q_2 q_4) \\ 2(q_1 q_2 - q_3 q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2 q_3 + q_1 q_4) \\ 2(q_1 q_3 + q_2 q_4) & 2(q_2 q_3 - q_1 q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix}$$

から計算される．ここで式 (4) は非線形関数であるため，Lindegren らは非線形最小二乗法を適用することにより式 (1) の形式に変換している．具体的には，式 (4) の η に関する非線形方程式を星のパラメータ $\mathbf{s}_i = (\alpha_i, \delta_i, \omega_i, \mu\alpha_i, \mu\delta_i)$ とクォータニオン $\mathbf{q}(t)$ を用いて

$$\eta = h(\mathbf{s}_i, \mathbf{q}(t))$$

と表現し，この非線形の方程式を参照点 $(\mathbf{s}_i^*, \mathbf{q}(t)^*)$ 近傍で線形化する．観測値 η と参照点における値との差分 R_l に対してパラメータの偏微分を取ることで

$$R_l = \eta - h(\mathbf{s}_i^*, \mathbf{q}(t)^*) - \sum_i \frac{\partial R_l}{\partial p} R_i = \left[\sum_i \frac{\partial R_l}{\partial p} \frac{\partial R_i}{\partial p^T} \right] p \quad (5)$$

が得られる⁹⁾．ただし， p はパラメータの参照点からの差分を示す．式 (5) の左辺は式 (1) の左辺 y の要素に対応し，式 (5) 右辺の第一項は式 (1) の右辺の A の要素に対応し，第二項の p は x の要素に対応する．ここで，パラメータの数は，50 万個の星それぞれが 5 個のパラメータを持つため星のパラメータが 250 万個，クォータニオンは式 (3) の通り 3 次の B-Spline で表現するため，2 年間のミッション期間を 30 秒間隔のスプライン関数で記述することにより 840 万個のパラメータで表現される．式 (4) より星のパラメータとクォータニオンのパラメータには関係性があるため独立に解くことができず，式 (1) のそれぞれの要素のサイズは，星と姿勢のパラメータ数を足して，式 (1) 左辺の y が 1090 万 \times 1，右辺の A が 1090 万 \times 1090 万，右辺の x が 1090 万 \times 1 となる．また，Nano-JASMINE は一つの星辺りの観測回数が約 100 回であるため，50 万個の星全ての観測回数は 5000 万回になり，5000 万個の観測値から 1090 万個のパラメータを推定する問題となる．

上記の計算に必要なコンポーネントを BDD で定義すると，図 7 が得られる．図 7 では，データ解析として定義されたブロックは観測値とモデルから構成され，モデルは星のモデル，姿勢のモデル，較正のモデルに分類される．また，それぞれのモデルは，与えられた数式と，数式を計算するために必要なパラメータに分類される．図 8 には，BDD における星のモデルの概念図を示す．星のモデルは，パラメータと数式から構成され，パラメータには同定すべき 5 つのパラメータが定義される．また，数式には星の螺旋運動として提供されたモデルを定義する．このように，BDD ではデータ解析に必要な要素を，階層構造を用いて定義する．

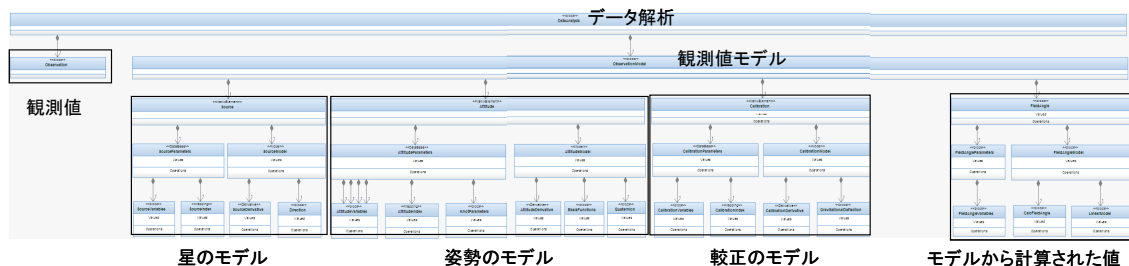


図 7 データ解析の BDD

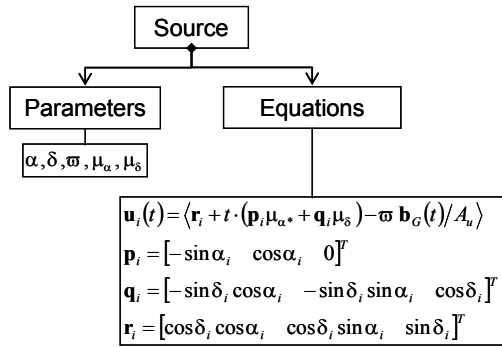


図 8 BDD における星のモデルの概念図

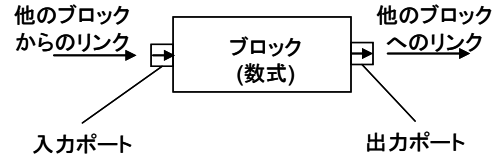


図 9 IBD におけるつながりの表現

5.2 Nano-JASMINE データ解析の IBD

前節で解析に必要な要素とその構造を定義した。IBD では BDD で定義されたブロック間のつながりを記述することで、データ解析におけるデータフローを表現する。図 9 のように数式での処理はブロックで記述し、必要な入力および出力はブロックのポートとして定義する。ブロック間のパラメータの受け渡しは、ポート間をリンクでつなげることで表現する。

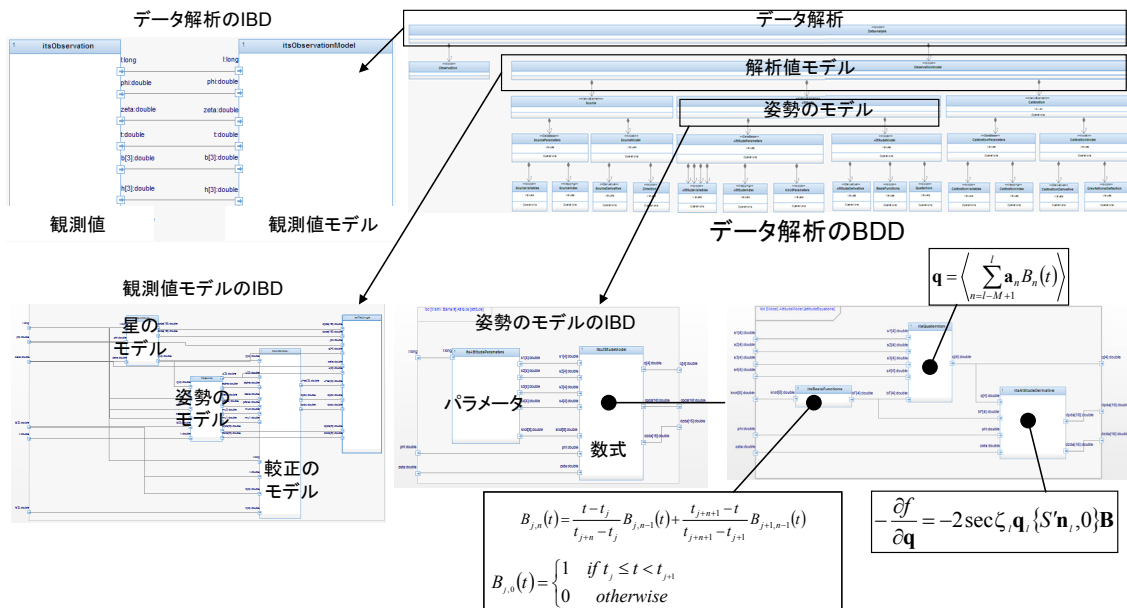


図 10 Nano-JASMINE データ解析における IBD

Nano-JASMINE のデータ解析は、式 (5) を用いて、式 (1) の各要素が計算される。この要素の計算の一連の流れを図 7 で定義した BDD の階層ごとに IBD を記述すると、図 10 が得られる。たとえば図 7 の BDD では、データ解析のブロックは観測値とモデルから構成されると定義したが、IBD では観測値からモデルへどのようなデータを受け渡すかを記述する。これにより観測値が与えられた際に、式 (5) の正規方程式の要素が計算されるまでのフローを、科学者と解析担当者双方が共通理解できる形式として記述でき、数式間のつながりを容易に把握することが可能になる。

5.3 連立方程式の生成と反復法による解法

5.2 節で正規方程式の要素を計算するためのモデルを MDSE に基づいて構築した。Nano-JASMINE におけるデータ解析では、式 (1) に基づいて一辺が 1090 万の正規方程式を生成し、パラメータを同定する必要がある。まず、Map Phase では観測値ごとに式 (5) を計算する。Key には、観測が式 (1) のどの要素であるかを示す値を設定し、Reduce Phase で全ての観測を集計することで連立方程式を生成する。この連立方程式は 1090 万個のパラメータを持つため、直接法で解くことが困難であり、反復法を用いて解く必要がある。本研究では反復法として一般的な共役勾配法を用いて計算を行う。このとき、5.1 節、5.2 節で述べたように、解析のモデルは星、姿勢、較正の 3 つの要素からなるが、その疎行列の特徴は大きく異なる。たとえば、星の要素はブロック対角に配置され、姿勢の要素は対角を中心とする帯行列に配置される。そこで、図 11 に示すように式 (1) の正方疎行列 A を要素ごとに分割し、それぞれ最適な格納形式を選択することで計算効率の向上を行った。分割の様子と、選択した格納形式、分割したことによる効果を図 11 に示す。ただし効果の評価には、星は α , δ , ω の 3 つのパラメータを持つと仮定した簡易モデルを用いて、1 か月分の姿勢データに対して星の数を 3000, 7000, 15000, 21000 と変化させたデータを用いた。全てを CSR で格納した場合を従来处理、分割した要素ごとに格納形式を選択したものを分割処理として比較を行った。その結果、分割した要素ごとに格納形式を選択した分割処理のほうが、全体を同一の形式で格納するよりも効率よく計算が可能であることが確認された。以降の計算では図 11 の格納形式を用いて計算を行うものとする。

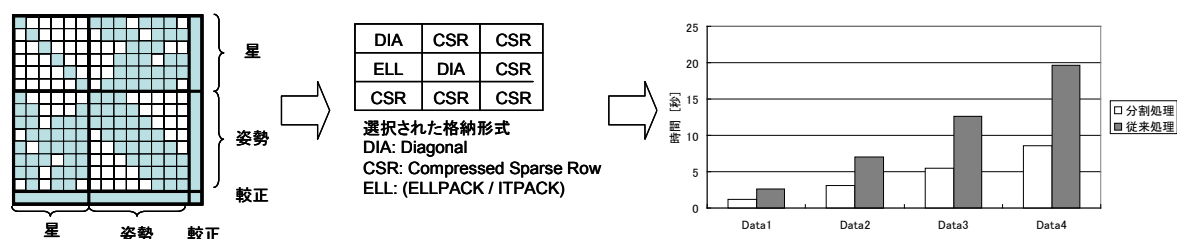


図 11 疎行列の分割と格納方法（左）と簡易モデルを用いた効果の検証（右）

5.4 位置天文データ解析の例

前節までの手法と数値解析による模擬データを用いて位置天文データ解析を行った。解析する星の数を 10312 個とし、それぞれが α , δ , ω の 3 つのパラメータを持つと仮定した。姿勢は、1 年間の姿勢データを 45 秒の knot で表現することで、666792 個のパラメータを持ち、較正の項は考慮しない。初期条件として、真の位置天文パラメータ α , δ , ω に対して 100mas, 100mas, 30mas (1σ) の標準偏差を持つ白色ノイズを付加し、姿勢には 50mas (1σ) の白色ノイズ、観測値には望遠鏡の収差などの影響を想定して 20mas (1σ) の白色ノイズを付加した。1 つの星は約 100 回観測される。ここで、5.1 節で述べたように、対象とする問題は非線形の方程式を線形化し、式 (1) の正規方程式を用いてパラメータを求める非線形最小二乗問題である。したがって、式 (1) から得られた解を用いて参照点を更新し、更新された参照点を用いて再度式 (1) の正規方程式を生成して解を求めていく必要がある。そこで、線形化・式 (1) の生成・式 (1) を用いた求解・参照点の更新という一連の流れをイタレーションの 1 回とカウントし、イタレーションが進むにつれてパラメータの誤差がどのように変化するかを確認した。ただし、Lindgren らの論文⁹⁾に基づき、評価指標として RSE (Robust Scatter Estimate) と呼ばれる値を用いた。これは分散の上から 10% と 90% の値の差分に 0.390152 を掛けた値で、ガウス分布であれば標準偏差と等しくなる値として定義されている。そこで、1 回イタレーションを行うごとに全ての α , δ , ω の誤差、更新の量の RSE を求めた。結果の一例として α と δ の様子を図 12 に示す。図 12 の横軸にはイタレーションの回数、縦軸には RSE の値を示す。 α と δ は初期値として 100mas (1σ) の白色ノイズが付加されているが、10 回のイタレーションにより 40mas (1σ) まで減少することが確認できた。また、更新の量は 5 回のイタレーションで 1mas 以下の値となり、これ以降はイタレーションをしても結果は変化しないことがわかる。実際の観測では真の値は得ることができないため、Lindgren らの論文においてもこの更新の量が十分に小さくなることが収束の条件であると定義されていることから、収束したと判断した。

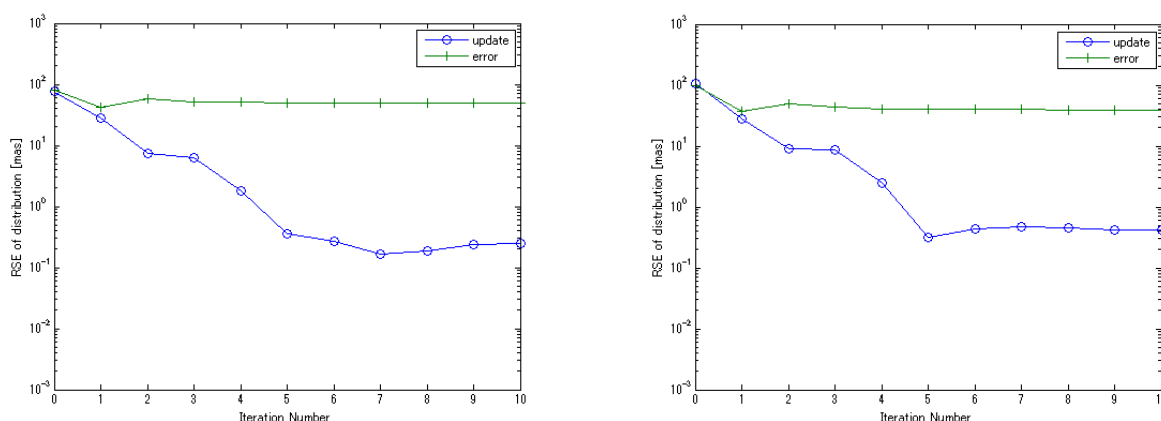


図 12 位置天文観測データ解析の例 (RSEと更新ベクトルのノルム, 左: α , 右: δ)

6. まとめ

大規模かつ複雑なデータ解析に対する科学者たちの要求を実現するため、解析担当者との効率的な情報共有および分散処理による効果的なデータ解析を可能にするモデル駆動型システムズエンジニアリングに基づくモデル管理およびデータ解析のフレームワークを実装した。一例として、位置天文観測衛星 Nano-JASMINE のデータ解析に適用し、解析モデルの生成、MapReduce による正規方程式の生成、GPGPU を用いた反復法によるパラメータ推定を実施した。現在は数値解析により生成した擬似観測データを用いて検証を行っているが、ノイズには白色性、正規性を仮定している。この仮定が成立しない場合、最小二乗法の解は真値に収束しないため¹¹⁾、今後は実機を用いた試験データからノイズの特性を分析することでモデル化誤差の有無を確認する予定である。ただし、地上で試験ができない姿勢制御の実データなどに関しては、打ち上げ後に得られるデータからモデル化誤差の有無を判断する必要がある。

参考文献

- 1) Gouda, N., Series of JASMINE missions, EAS Publications Series, Vol. 45, 2011, pp.393-396
- 2) Kobayashi, Y. et al., Nano-JASMINE: A 10-kilogram satellite for space astrometry, Proc. SPIE, 6265, 2006, pp.626544 1-10
- 3) Bock, C., Systems engineering in the product lifecycle, International Journal of Product Development, Vol. 2, 2005, pp.123-137
- 4) 西村他, システムズモデリング言語 SysML, 東京電機大学出版局, 2012, pp.2-14, 101-114
- 5) Jeffrey, D. et al., MapReduce: Simplified Data Processing on Large Clusters, Communications of the ACM, Vol. 51, No. 1, 2008, pp. 107-113
- 6) 白崎他, JVO 開発における大規模天文データ処理: 全天対応天文データ分散検索・解析機構の試験構築, 宇宙科学情報解析論文誌, 第一号, 2012, pp. 57-66
- 7) Narayanan, S. et al, A framework for efficient and scalable execution of domain-specific templates on GPUs, Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Processing, 2009, pp. 1-12
- 8) Christopher, J. et al, Astrophysical Supercomputing with GPUs: Critical Decisions for Early Adopters, publications of Astronomical Society of Australia, 2011, Vol. 28, pp.16-27
- 9) Lindegren, L. et al, The astrometric core solution for the Gaia mission Overview of models, algorithms and software implementation, Astronomy & Astrophysics, Vol. 538, id.A78, 2012, pp.1-48
- 10) 初鳥他, JASMINE 計画シリーズ: Nano-JASMINE から小型 JASMINE へ, 第 53 回宇宙科学技術連合講演会公演集, 2009, pp. 1101-1104
- 11) Inamori, T. et al, Compensation of time-variable magnetic moments for a precise attitude control in nano- and micro- satellite missions, Advances in Space Research, Vol. 48, Issue 3, 2011, pp. 432-440
- 12) 足立修一, システム同定の基礎, 東京電機大学出版局, 2009, pp. 4-5