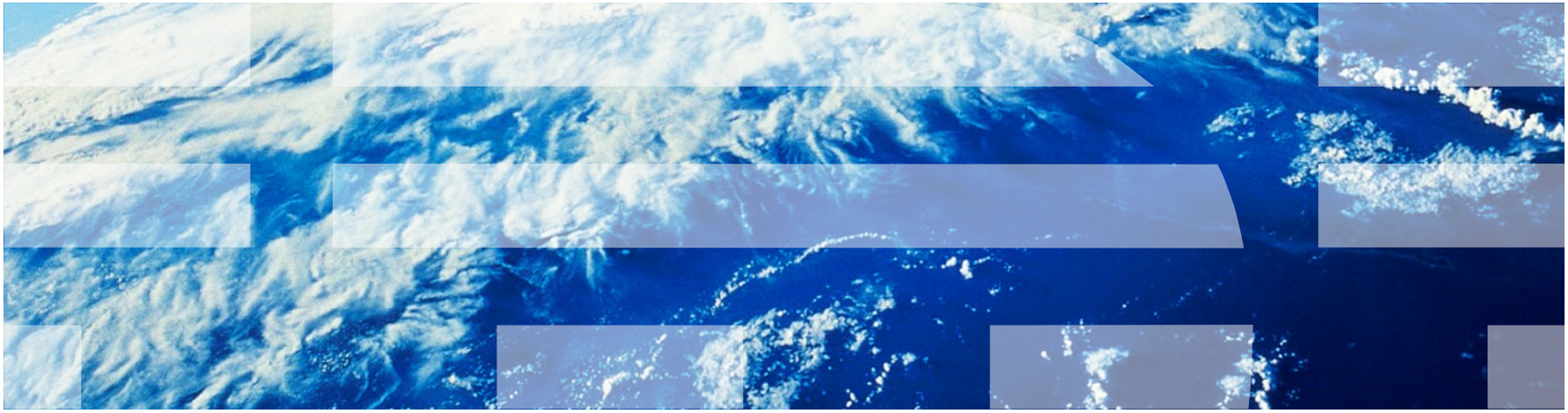
IBM Research — Tokyo
Yoichi Hatsutori
Feb. 15, 2012

平成23年度宇宙科学情報解析シンポジウム

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　－Nano-JASMINEデータ解析への適用－

○初鳥陽一、宮下尚、清水淳也（日本IBM）、山田良透（京都大学）



研究背景
大規模・複雑化するデータ解析における問題点と、そのソリューション

科学者（user）の要求

- ■膨大なデータの効率的な解析
- ■複数の学問領域にまたがる解析
- ■様々な解析手法の適用
- ■他のプロジェクトへの応用

解析担当者（vendor）の作業

- ■効果的なコードの実装
- ■複雑なモデルの実装
- ■モデルの管理
- ■汎用的なコードの生成

問題の大規模・複雑化に伴い
情報共有・分散処理が重要な課題

ソリューション

モデル駆動型システムズエンジニアリングによる管理および解析
Model-based System Identification Cloud (MbSIC) の提案

大量のデータ

【ソリューション】
システムズモデルによる
モデル管理

コード生成

【ソリューション】
MapReduce & GPGPUによる
分散処理系の構築
特徴ごとに行列を分割・格納形式
の選択

大規模・複雑なデータ解析の例
JASMINE計画（Japan Astrometry Satellite Mission for Infrared Exploration）

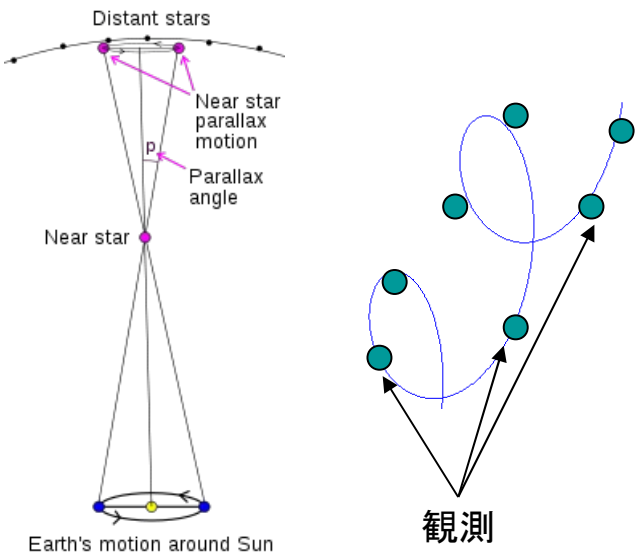
銀河系内の、特に銀河面、バルジなどのサーベイを行ない、数億個の星の位置・距離・固有運動を高精度に測定（位置天文学）

JASMINE計画では、3種類の衛星が検討・開発中

■Nano-JASMINE　口径5cm　zw-band　全天を観測　フライトモデル完成　2013年打ち上げ

■Small-JASMINE　口径30cm　Hw-band　バルジ方向を観測　検討中

■JASMINE　口径80cm　Kw-band　バルジ方向を観測　検討中



JASMINE計画での観測のイメージ図
複数の観測結果から、螺旋のパラメータを推定

JASMINE計画でのデータ解析

星の固有運動と、地球の公転運動に起因する楕円運動により、天球上での星の動きは螺旋運動にみえる。

星を複数回観測することにより、螺旋のパラメータを推定

※Nano-JASMINEでは約14万個の星に対して推定を行う

推定すべきパラメータ

人工衛星に搭載された望遠鏡を用いて星を観測するため、星のパラメータ以外に観測機器、機器の劣化、衛星の姿勢パラメータなども同時に推定する必要がある

Nano-JASMINEで推定するパラメータの数

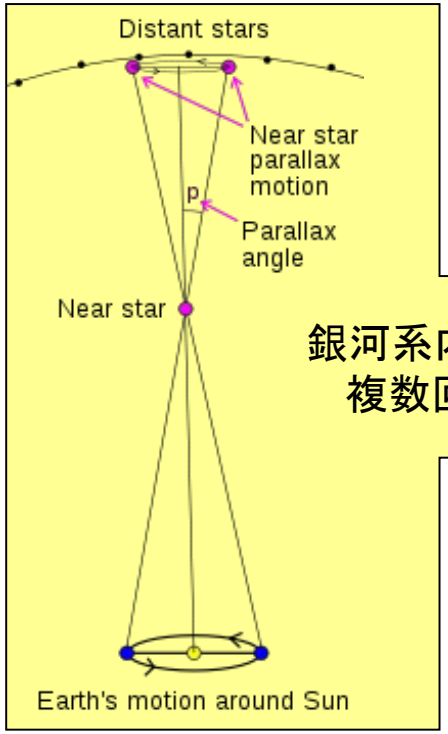
- 約14万個の星を観測。それぞれの星は5つの位置天文パラメータを持つ。
  - **星のパラメータ数　　70万個**（7×10^5）
- 3次のスプライン関数で衛星姿勢クオータニオンを近似。ミッション期間を2年間、スプラインのknotの間隔を30秒とする
  - **姿勢のパラメータ数　　840万個**（8.4×10^6）　　← 星のパラメータより多い！

- 星と姿勢のパラメータはカップリング項を持つため独立に解くことができない。そのため、約10^7のパラメータで構成される疎行列を解く必要がある。
  - 実際はカップリング項が最も情報量が多く、非対角項優位の大規模疎行列となる

参考

- ESAが打ち上げ予定のGAIAプロジェクトではおよそ1億個（10^8）の星に対して位置天文パラメータを求めることを計画
  - パラメータ数はおよそ10億個（10^9）

Model-based System Identification Cloudの全体像



銀河系内の星を
複数回観測

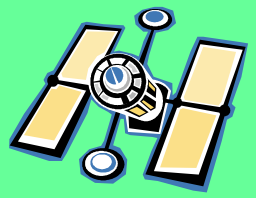
大量の観測データ
（TB or PB クラス）
■観測画像
■センサのノイズ
■衛星の姿勢
■熱の変化
■・・・

Model-based System Identification Cloud

分散処理による
パラメータ推定

得られる結果

星のパラメータ
（銀河系内の星の
三次元地図）



衛星のパラメータ
（搭載機器等）

解析コードの
生成

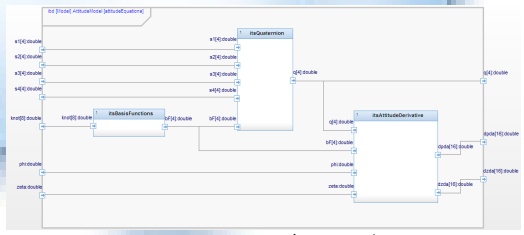
パラメータの更新
モデルの更新

科学者
■星の数学モデル
■衛星の数学モデル
■・・・

数式の提供

解析担当者
■実装

モデルの実装

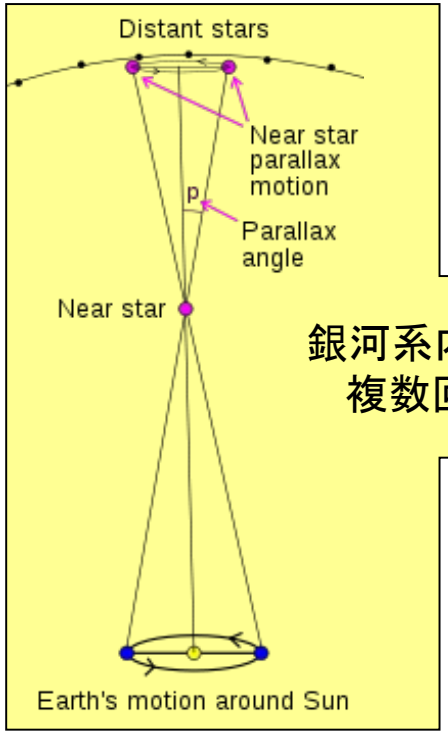


システムズモデル
（モデルベースの情報共有）

次のプロジェクト
への応用

Small-JASMINE
JASMINE

Model-based System Identification Cloudの全体像



銀河系内の星を
複数回観測

大量の観測データ

（TB or PB クラス）

■観測画像

■センサのノイズ

■衛星の姿勢

■熱の変化

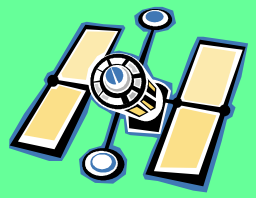
■・・・

Model-based System Identification Cloud

分散処理による
パラメータ推定

得られる結果

星のパラメータ
（銀河系内の星の
三次元地図）



衛星のパラメータ
（搭載機器等）

解析コードの
生成

パラメータの更新

モデルの更新

科学者

■星の数学モデル

■衛星の数学モデル

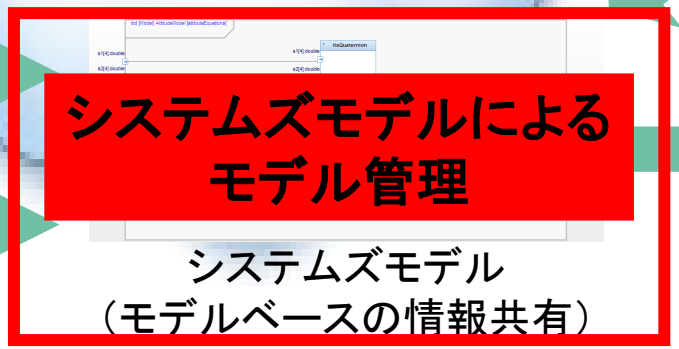
■・・・

数式の提供

解析担当者

■実装

モデルの実装

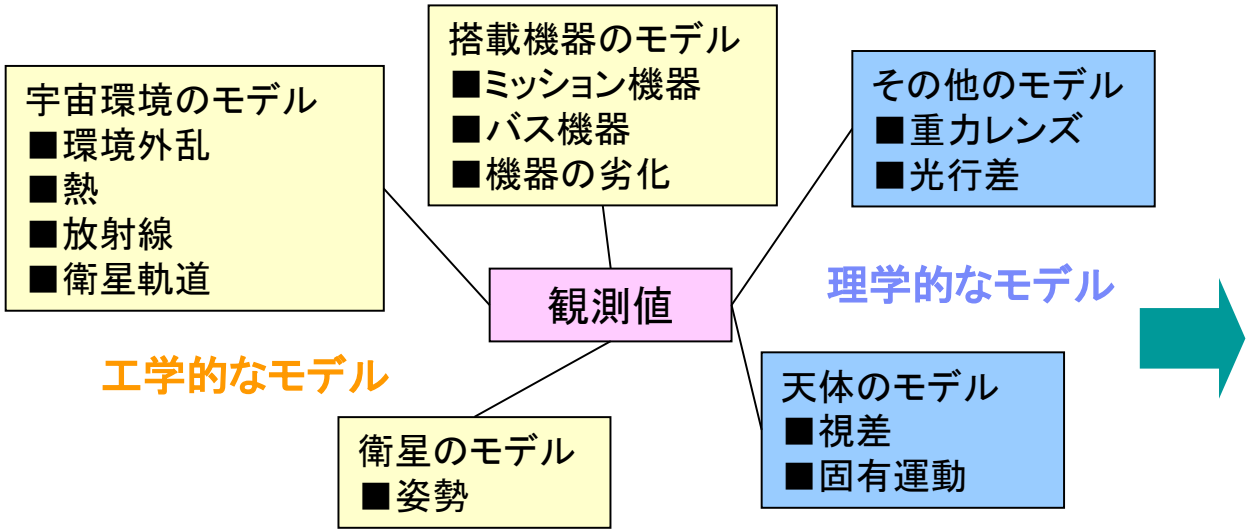


次のプロジェクト
への応用

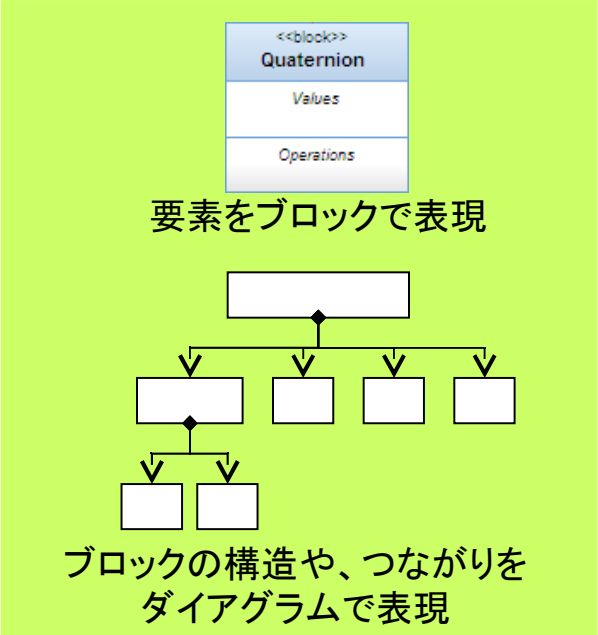
Small-JASMINE
JASMINE

モデル駆動型システムズエンジニアリング
Model Driven Systems Engineering (MDSE)

要求から設計まで、システム全体を形式的に記述し、表現するための新しい手法のひとつ



観測値に影響するさまざまな要素の例
（JASIMINEプロジェクトの例）

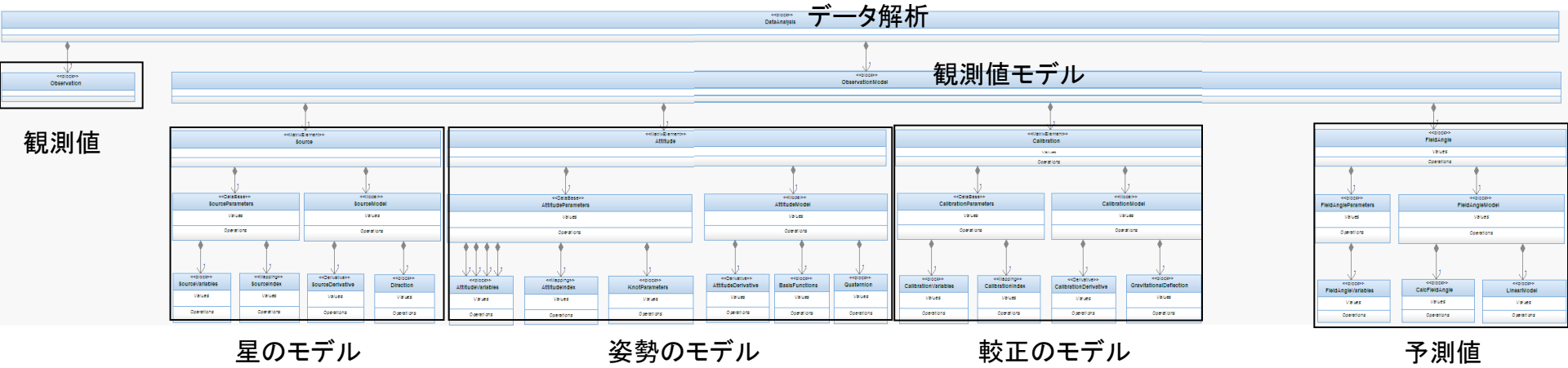


システムズモデルの記述の例

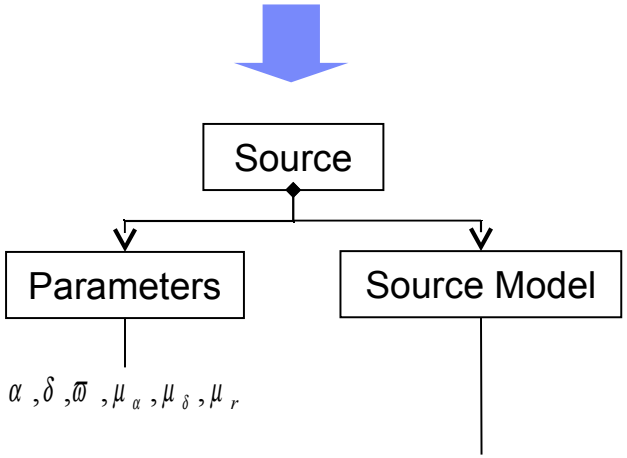
システムズモデル

理学、工学など様々な分野の研究者・開発者の間で、機能・制約・要求・振る舞いなどを共通理解しコミュニケーションするためのモデル

システムズモデル – ブロック定義図（Block Definition Diagram, BDD）



データ解析に用いるブロック定義図（全体像）



$$\mathbf{u}_i(t) = \langle \mathbf{r}_i + t \cdot (\mathbf{p}_i \mu_{\alpha *} + \mathbf{q}_i \mu_\delta + \mathbf{r}_i \mu_{ri}) - \varpi \ \mathbf{b}_G(t) / A_u \rangle$$

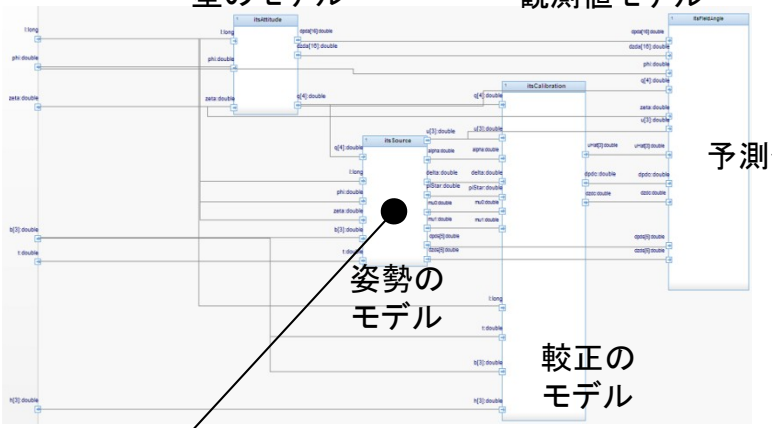
$$\mathbf{r}_i = [\cos\delta \ \cos\alpha \quad \cos\delta \ \sin\alpha \quad \sin\delta \ ]^T$$

データ解析に必要な数式や同定すべきパラメータ、入力するパラメータなどの要素を、それぞれブロックとして定義し、全体の構成を記述する。

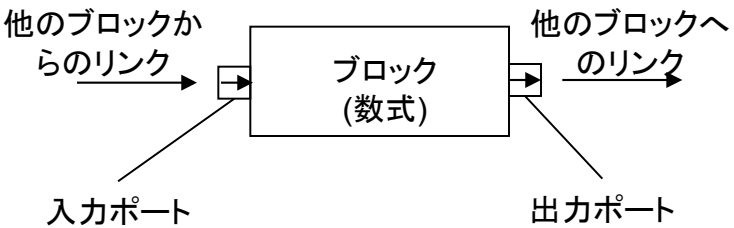
左図は星のモデルの一部で、星のモデルは星のもつ6個の位置天文パラメータと、星の位置を計算する数式から構成されていることがわかる

システムズモデル – 内部ブロック図（Internal Block Diagram, IBD）

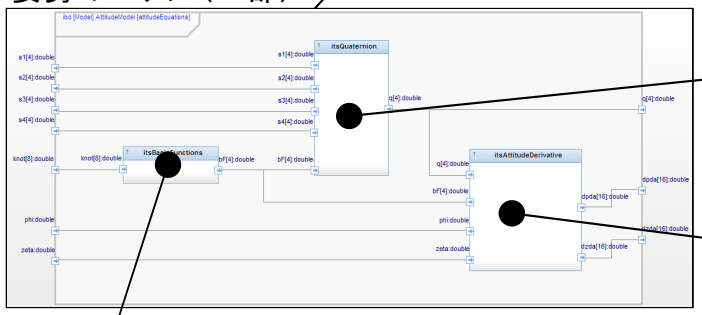
観測値
- 衛星姿勢
- 位置
- 速度
- 時間
- 星の位置



前頁のブロック定義図で定義されたブロック間のつながりをダイアグラムに記述



姿勢のモデル（一部）



$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \left\langle \sum_{n=l-M+1}^{l} \mathbf{a}_n B_n(t) \right\rangle$$

$$-\frac{\partial f}{\partial \mathbf{q}} = -2\sec\zeta_l \mathbf{q}_l \{S'\mathbf{n}_l, 0\}\mathbf{B}$$

入力 → 出力

内部の信号フロー

$$B_{j,n}(t) = \frac{t - t_j}{t_{j+n} - t_j} B_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} B_{j+1,n-1}(t)$$

$$B_{j,0}(t) = \begin{cases} 1 & if\ t_j \le t < t_{j+1} \\ 0 & otherwise \end{cases}$$

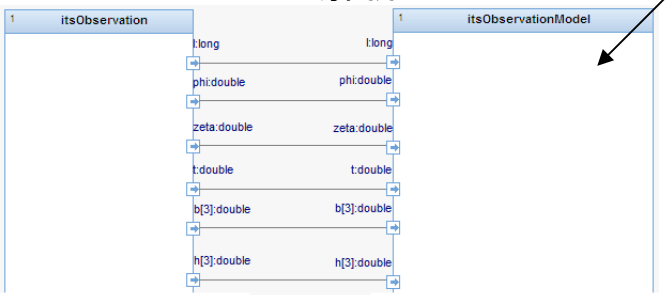
数式を計算するためのパラメータは、ブロックのポートとして定義

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

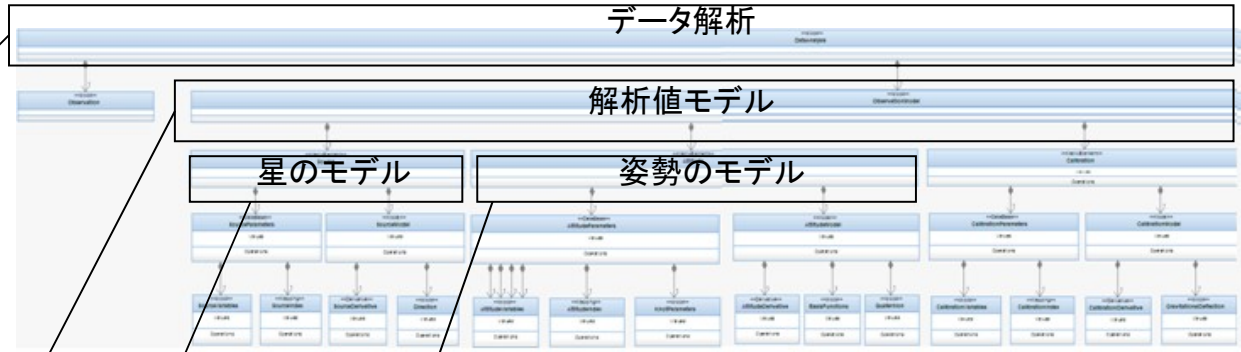
数式を提供する科学者（User）と解析担当者（Vendor）の双方が共通理解することのできる形でモデルを管理

システムズモデル － Nano-JASMINEの解析モデル

データ解析

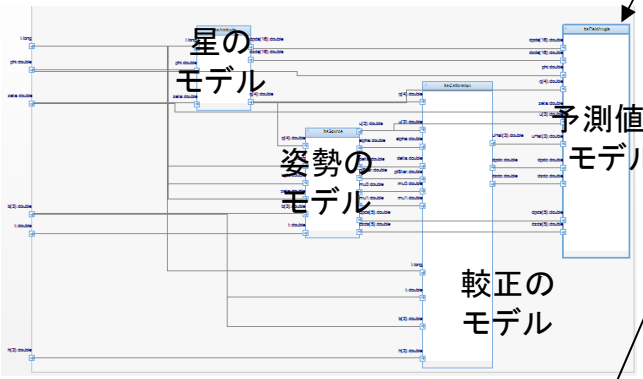


観測値　　観測値モデル



データ解析のBDD

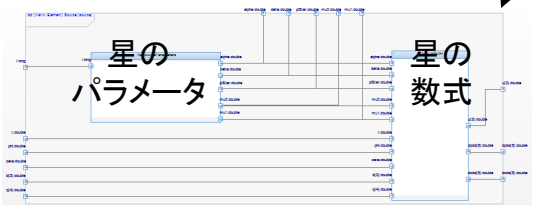
観測値モデル



BDD（上図）により定義されたブロック同士の関係性をIBDにて記述する
階層ごとにIBDを作成することで、系統的に解析モデルを取り扱う

姿勢のモデル

星のモデル



パラメータ推定のための連立方程式

$$B_{j,0}(t) = \begin{cases} 1 & \text{if } t_j \le t < t_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{j,n}(t) = \frac{t - t_j}{t_{j+n} - t_j} B_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} B_{j+1,n-1}(t)$$

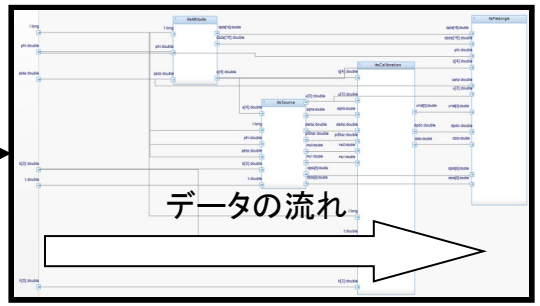
$$\left[\sum \frac{\partial R_l}{\partial p} \frac{\partial R_l}{\partial p'} W_l \right] d_i = -\sum \frac{\partial R_l}{\partial p} R_l W_l$$

$$R_l(\mathbf{s}, \mathbf{a}, \mathbf{c}) = \eta_{obs} - \eta_{predict}$$

$$\begin{bmatrix} \cos\eta \cos\delta \\ \sin\eta \cos\delta \\ \sin\delta \end{bmatrix} = f(\mathbf{q}) \cdot \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}$$

システムズモデル
による管理

観測値 η_{obs}　　予測値 $\eta_{predict}$



システムズモデルの一部
（観測値が入力されると、観測値と
予測値との差分が計算される）

モデルさえ記述すれば、
解析コードは自動的に生成できる

最小二乗法による定式化
（User側から提供）

$$N \cdot \Delta x = b$$

Nはn×nの対称行列
nはパラメータ数

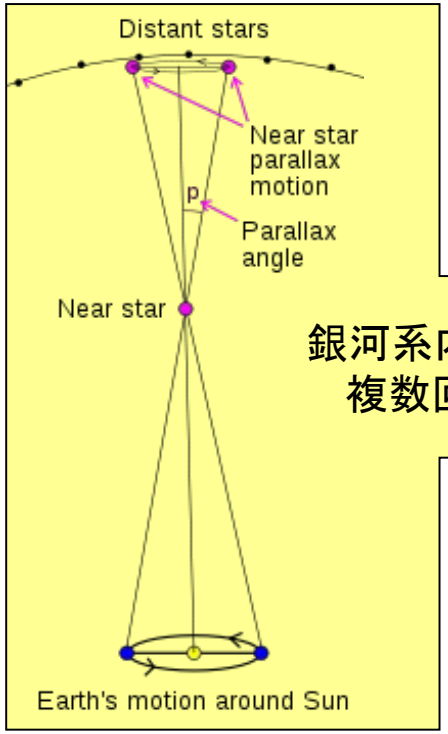
⊿xを求めたいが、パラメータ数が多いため反復法で計算

<u>パラメータ推定の流れ</u>

- ■行列N（$10^7 \times 10^7$）の生成
- ■N⊿x=bの式より⊿xを反復法により求める
- ■⊿xによりパラメータを更新
- ■行列N（$10^7 \times 10^7$）の生成
- ■・・・・

大規模なデータに対応するため
クラウドコンピューティングによる
分散処理を実装

Model-based System Identification Cloudの全体像



銀河系内の星を
複数回観測

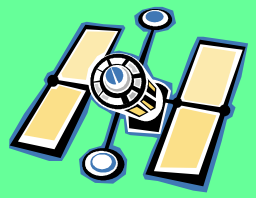
大量の観測データ
（TB or PB クラス）
■観測画像
■センサのノイズ
■衛星の姿勢
■熱の変化
■・・・

Model-based System Identification Cloud

分散処理による
パラメータ推定

得られる結果

星のパラメータ
（銀河系内の星の
三次元地図）



衛星のパラメータ
（搭載機器等）

解析コードの
生成

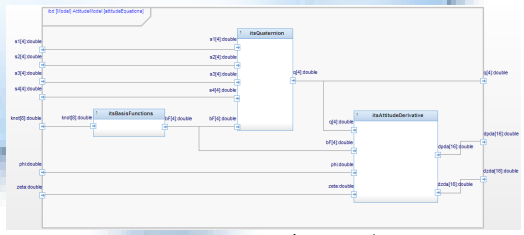
パラメータの更新
モデルの更新

科学者
■星の数学モデル
■衛星の数学モデル
■・・・

数式の提供

解析担当者
■実装

モデルの実装



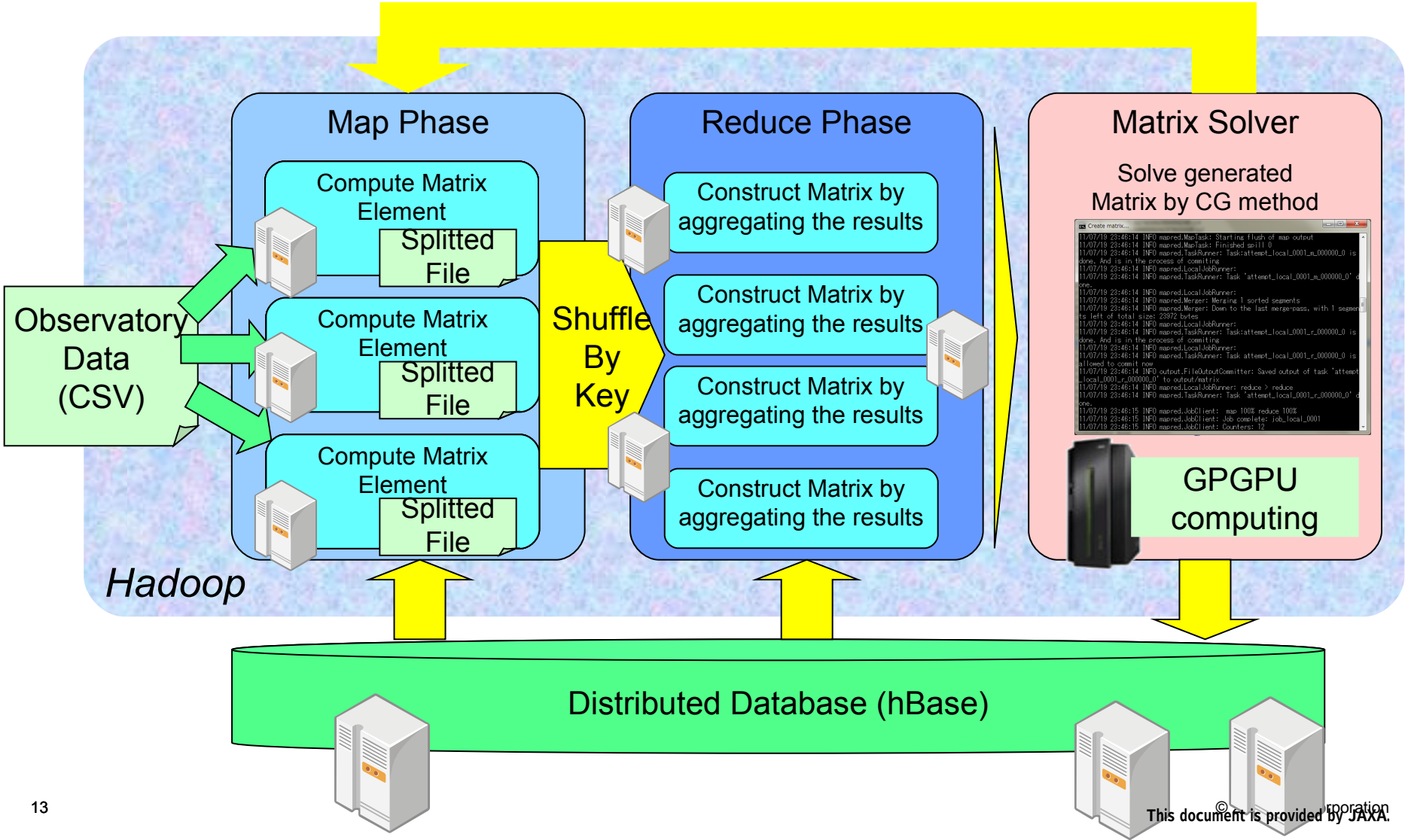
システムズモデル
（モデルベースの情報共有）

次のプロジェクト
への応用

Small-JASMINE
JASMINE

現在の構成

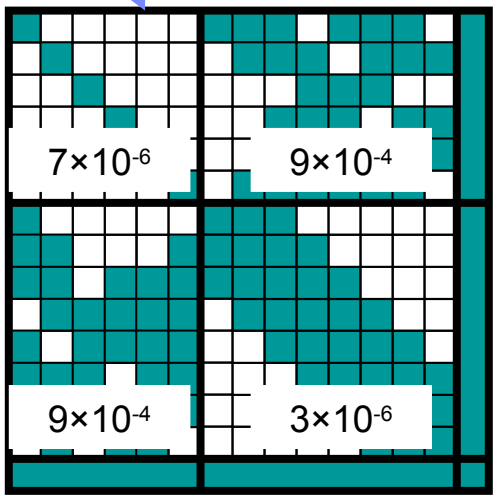
大規模疎行列の生成部分にはHadoop MapReduceを利用
反復法にてパラメータを解く部分にはGPGPUを利用



疎行列の分割と格納形式の選択

$$N \cdot \Delta x = b$$

星・姿勢・較正それぞれで疎行列の特性が大きく異なる



星の
パラメータ

姿勢の
パラメータ

較正のパラメータ

星の
パラメータ

姿勢の
パラメータ

較正の
パラメータ

大規模疎行列を効率よく解くために

■疎行列の分割

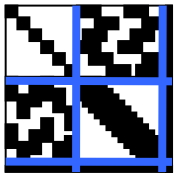
■それぞれに対して格納形式を選択

システムズモデルからは以下の情報が得られる

■パラメータの数

■パラメータ同士の関係性

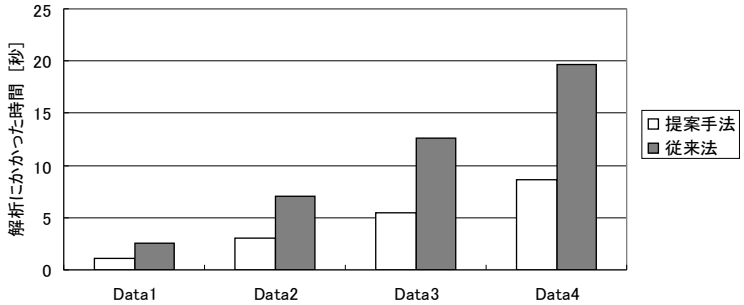
疎行列の非ゼロ要素位置を導出



分割生成された疎行列

DIA	CSR	CSR
ELL	DIA	CSR
CSR	CSR	CSR

選択された格納形式
（比較したものは全てCSRで格納）



星の個数を3000,7000,15000,21000させた
サンプルデータ解析の結果（toy model）

格納形式の判別に利用することで約2倍の効率化を実現

まとめ

- 科学者（User）の様々な要求に対応するため、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、データ解析のフレームワークを提案
  - Model-based Systems Identification Cloud

- 大規模複雑なデータ解析の一例として、位置天文観測衛星Nano-JASMINEのデータ解析に適用
  - 解析モデルの生成
  - Hadoop MapReduceによる疎行列生成
  - GPGPUを用いた反復法によるパラメータ推定

- 現状は、別途数値解析により用意したサンプルデータを入力データとして、検証中
- 今後は実機から得られるデータの利用、Small-JASMINE等他のプロジェクトの検討への利用を検討
  - HILS（Hardware In the Loop Simulator）
  - 軌道上データ（2013年打ち上げ後）