

クエリーの高速化

- 今まで非常に長い時間がかかったクエリーの高速化
- 7億超のレコード数のテーブルに対する試験（Query1を除く）

計測Query

Spark環境でUDFを使用可にくださったことで元々のクエリに近い形で調査可能に

Query	使用テーブル (Sparkは、を_に読み替え)	特徴 ()内は元の環境で問題となった処理時間等	UDF +	count(*) 結果行数 PostgreSQL@SV76/ Spark@SV68(parquet)
1	pdr2_dud.forced, forced2, meas	天体の明るさや色の条件、および空の領域の範囲の条件を組み合わせる興味ある天体を選択する典型的なクエリ (S17A 12h00m06s 検索時間が12時間を超えてタイムアウト)	なし	9,575/9,493
2	pdr2_wide.forced (sparkはwith_coord使用)	多数の空の領域の範囲を一度に指定して天体を選ぶクエリ (S17A 30m26s)	coneSearch	191,788/191,788
3	pdr2_wide.photoz_demp †, forced	少数の条件からカタログの多くの部分を持って来るようなクエリ (S17A 2h2m48s - 結果サイズオーバーでエラーになっている)	なし	436,333,410/436,333,410
4	pdr2_wide.forced pdr2_wide.meas2	少数の条件からカタログの多くの部分を持って来るようなクエリ (S17A 3h46m12s)	boxSearch	33,088,312/33,140,174
17	pdr_wide.forced[,2-5]	特定のフィールドに存在する天体の多くの部分を区切って持ってくるようなクエリ (pdr2 2h59m58s: order by, limit, offset ありの数値)	search_w05	179,918,836/179,918,836

3つの実行クエリー

残りの2つは長すぎるので割愛しました。。

実行Query3 (for PostgreSQL@SV76)

```
SELECT count(*) FROM pdr2_wide.photoz_demp AS zp, pdr2_wide.forced AS pm where pm.object_id=zp.object_id AND zp.photoz_best is not null;
```

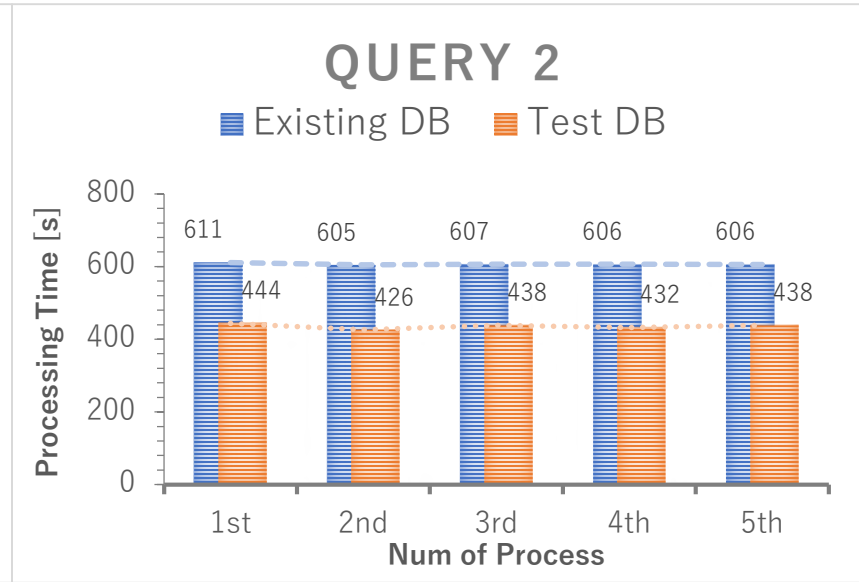
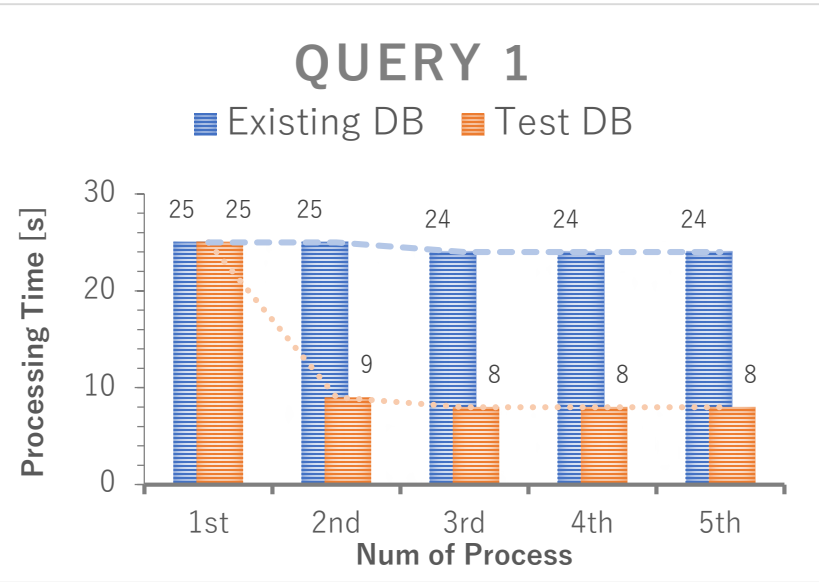
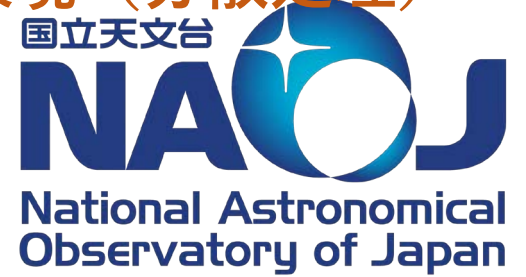
実行Query4 (for PostgreSQL@SV76)

```
SELECT count(*) FROM pdr2_wide.forced AS main LEFT JOIN pdr2_wide.meas2 AS m1 USING (object_id) WHERE boxSearch(coord, 330.0, 360.0, -2.0, 6.0) AND isprimary AND r_cmodel_mag<26 AND g_cmodel_mag<26 AND y_cmodel_mag<26;
```

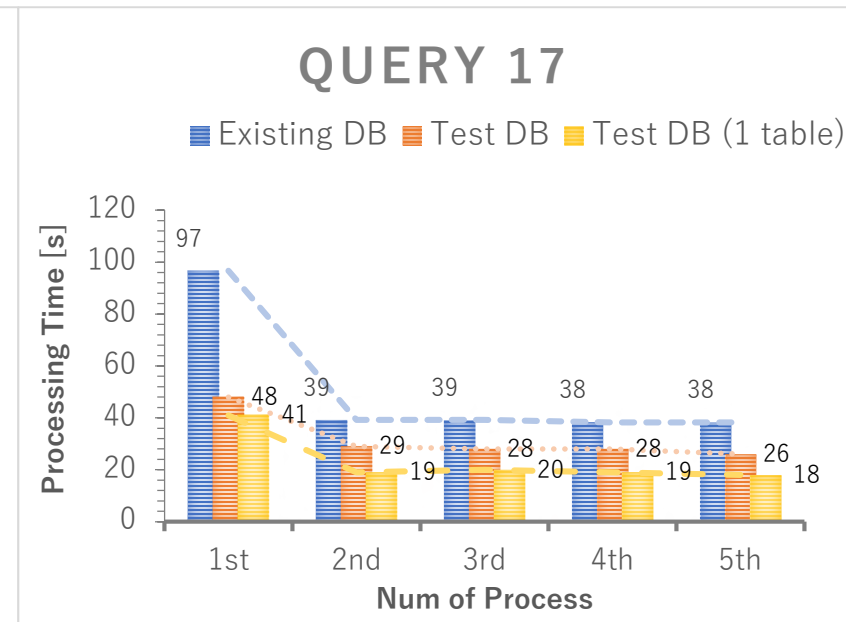
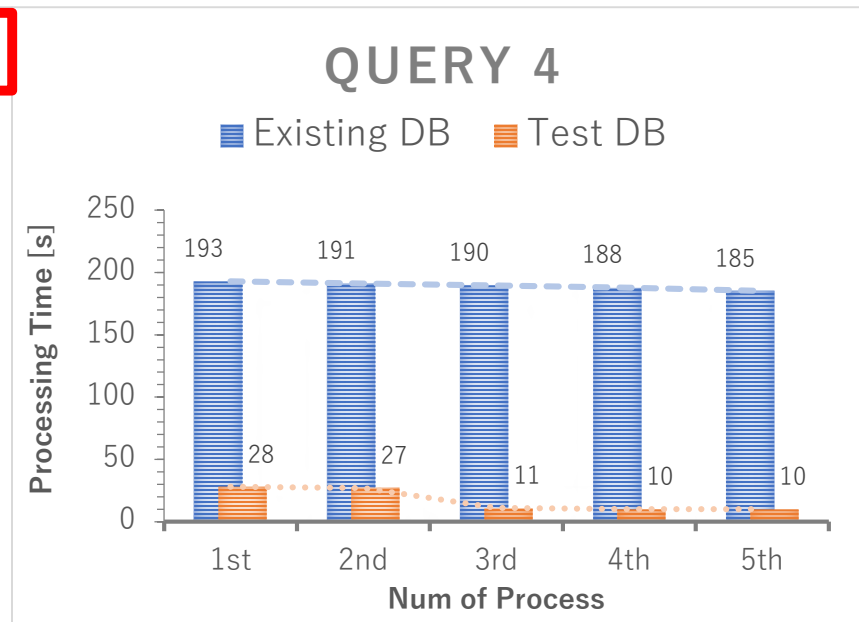
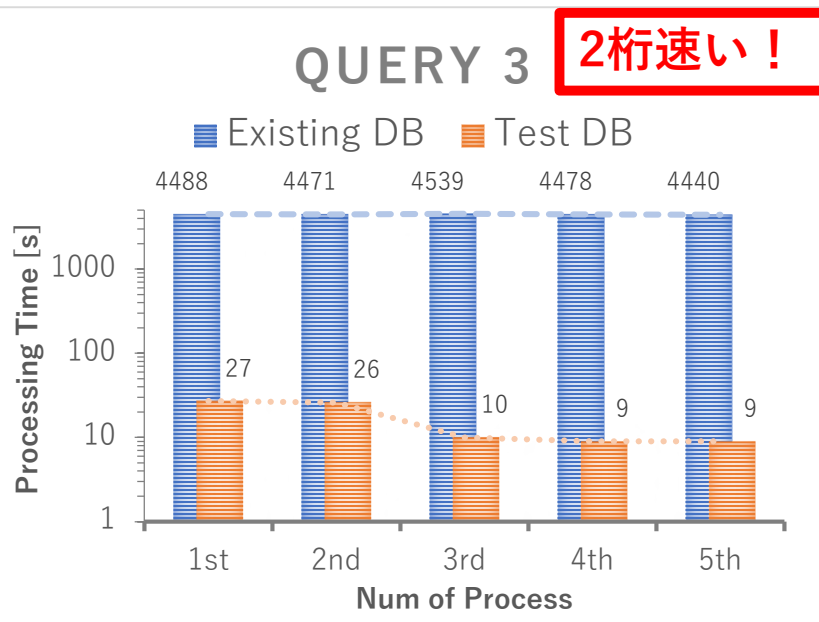
実行Query17 (for PostgreSQL@SV76)

```
SELECT count(*) FROM pdr2.pdr2_wide.forced LEFT JOIN pdr2.pdr2_wide.forced2 USING (object_id) LEFT JOIN pdr2.pdr2_wide.forced3 USING (object_id) LEFT JOIN pdr2.pdr2_wide.forced4 USING (object_id) LEFT JOIN pdr2.pdr2_wide.forced5 USING (object_id) WHERE pdr2.pdr2_wide.search_w05(object_id);
```

検索時間の比較 PostgreSQLクラスター（旧来处理） vs Spark環境（分散処理）



- 計測した全てのqueryにおいてSpark環境(分散処理)の方が処理が高速(query3は2桁の違い)
- 初回の処理時間のみに注目した場合、ほぼ差がないqueryもあり(query1)



変光天体の検出、分類のプロトタイプ

(小規模データによる試験実装：約234万天体)

- 3億5341万レコード（延べ数） × 164カラムのテーブルを使用

天体の形や色の特徴などにより、調べたい天体をあらかじめ絞り込む

~1万1000天体（点状天体の場合）

天体ごとに時間方向の統計値（平均、標準偏差、 χ^2 等の統計値）を計算

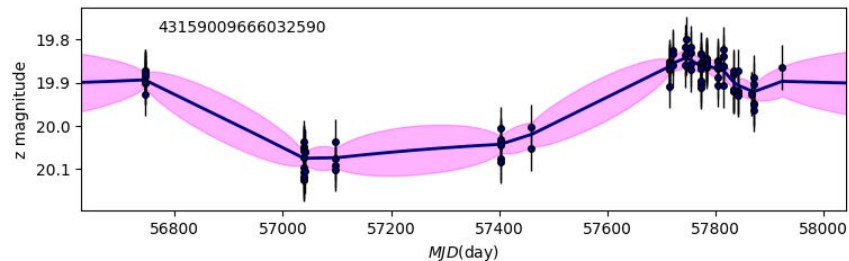
明るさの統計的特徴が一定の基準を満たすものを変光天体候補として同定

~400天体（点状天体の場合）

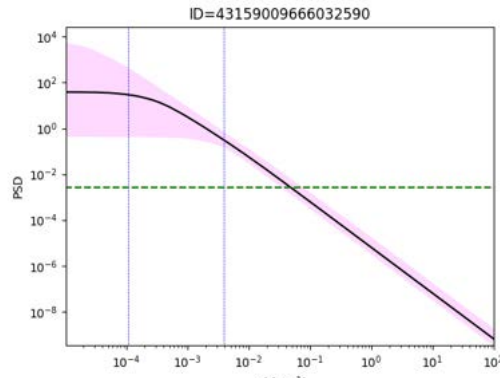
時系列解析によるモデルパラメータを使った（自動）分類

巨大BH天体、銀河系内のコンパクト天体、周期的変光星の候補を分離

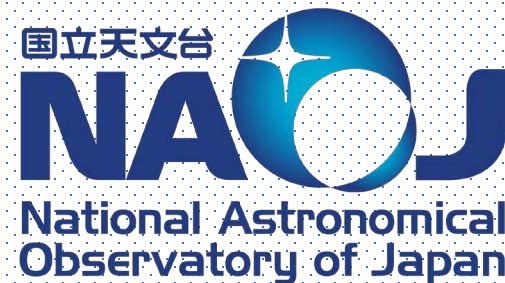
モデル推定したライトカーブ



3年間



モデル推定したパワースペクトル密度

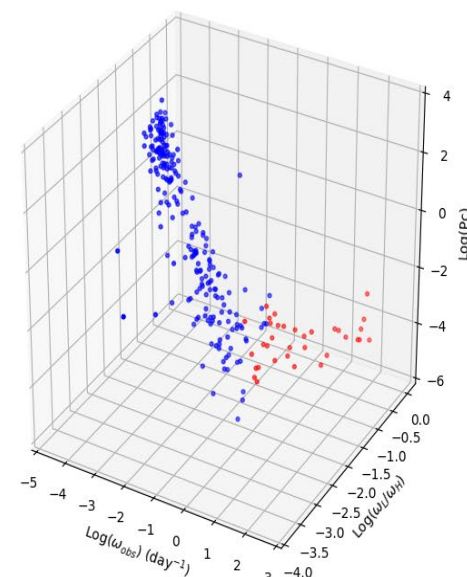
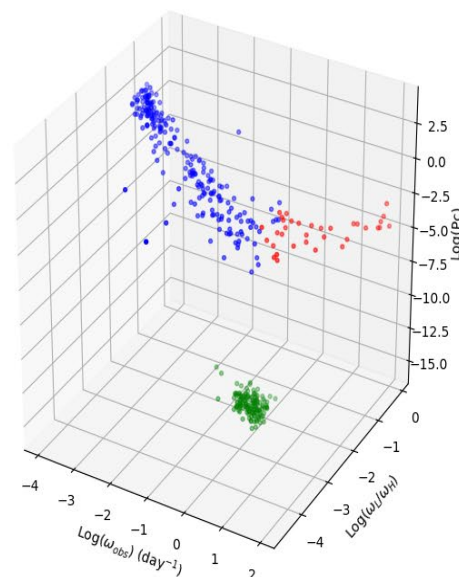
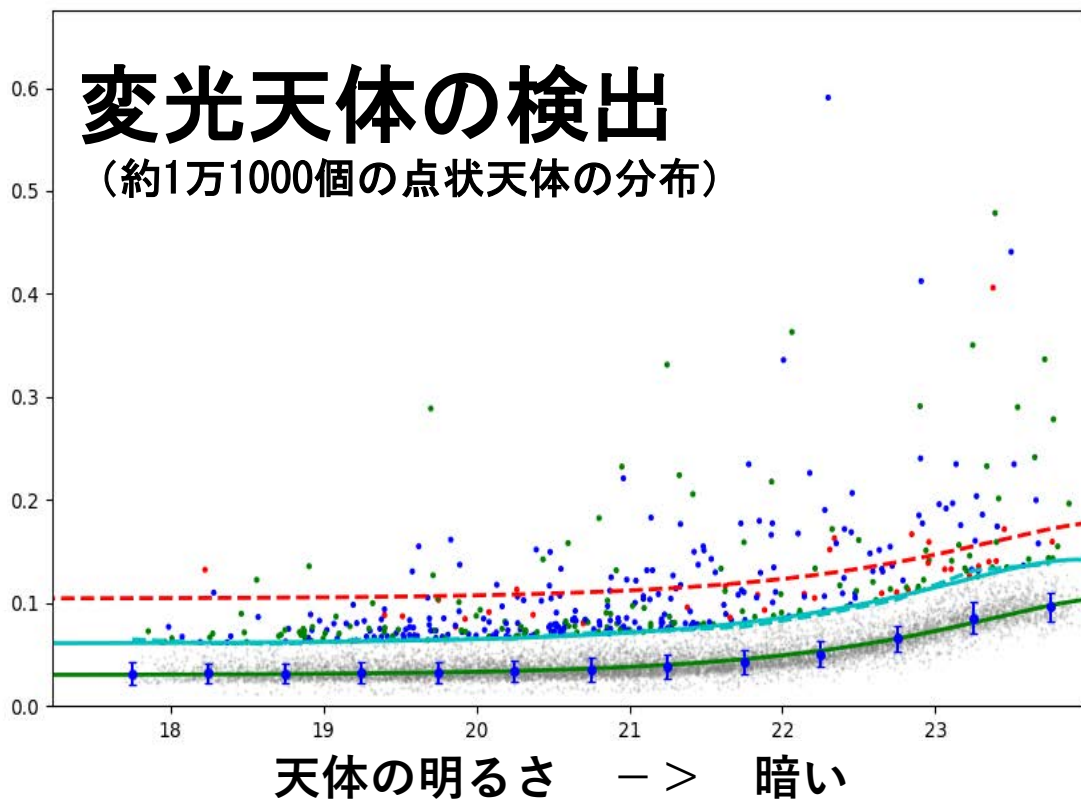


青：(1型) 巨大BH天体候補 赤：銀河系内天体候補 緑：ランダム項の係数が大きく変動パワーは小さい (星?)

天体の明るさの標準偏差

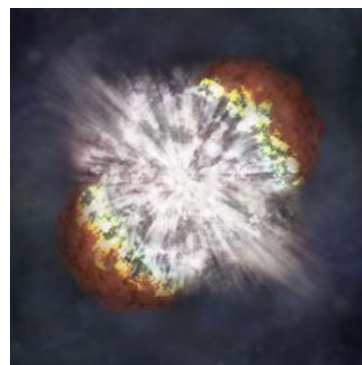
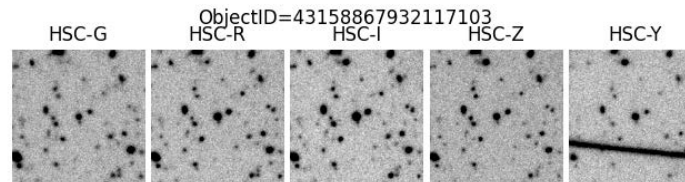
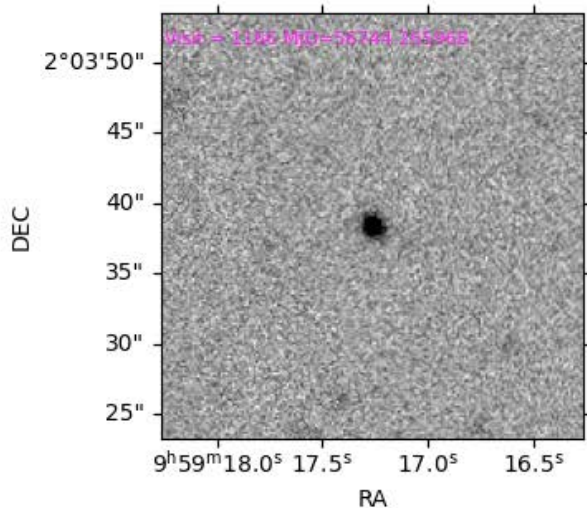
変光天体の検出

(約1万1000個の点状天体の分布)

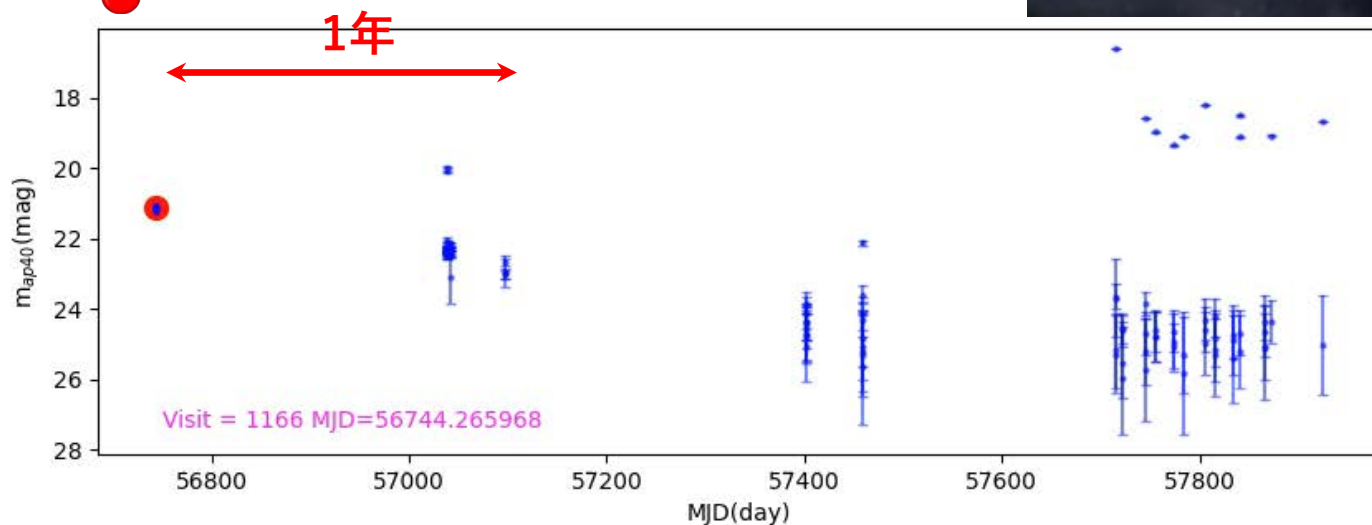


時系列解析による変光天体の自動分類

2年かけて25倍以上暗くなり、 3年後には見えなくなった天体



ゆっくりと減光しているが、いつがピークだったのかはわからない。1年後でも十分に見えている。2年後でもギリギリ見えている？変動幅の大きいクエーサーの可能性も完全には否定できないが、過去のデータを見ることで、さらに情報を足す必要がある。



その後、ここから3年前（2010～11年）、9年前（2005年）のデータから、クエーサーの可能性は低いと判断している。X線やガンマ線による検出の報告も今のところなさそう。

今後の展開（予定）

• クエリーの高速化

- ✓ より多くのパラメータを用いたデータ分析
- ✓ 例えば、天体までの距離推定の高度化に関する研究
 - 新規データに対する差分的解析 → 特異天体の即時追観測の可能性

• 変光天体探査

- ✓ 現在は1視野分 → 今年度中には20視野以上についての分析を行う（延べ数数十億天体 × 5フィルター分について）
- ✓ 最終的には1000視野について同様の解析を行う事が今後数年での目標の1つ

◆ 新規データの即時解析による突発天体検出と追観測アラート発令システム

- ✓ 本システムの発するアラートから世界中の天文台が追跡観測を開始
 - ➡ 夜を探して地球を巡る（マルチメッセンジャー天文学）

• 移動天体探査

- ✓ 既存天体（小惑星・彗星）のデータベース上での存在確認 → **太陽系の成り立ちを探る**
- ✓ 新移動天体の発見 → 地球（我々）への影響 ➡ 世界中の天文台による追観測

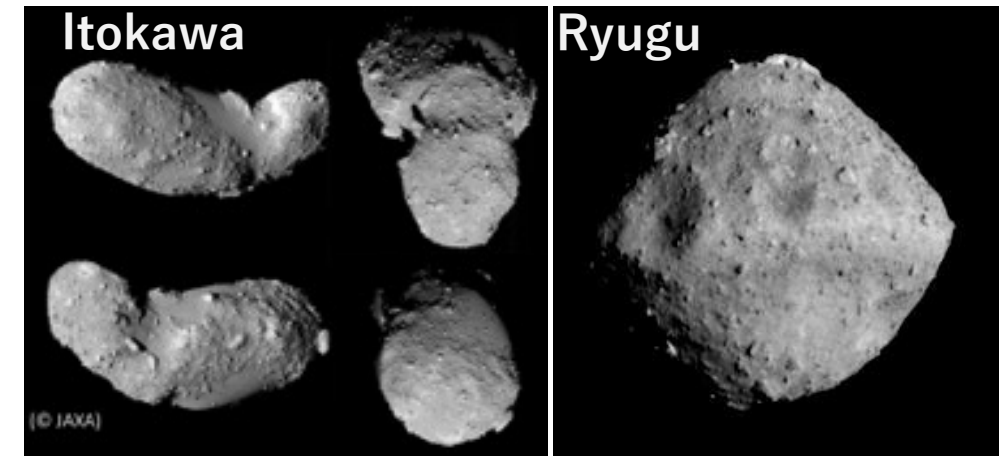
移動天体探査における高速検索データベースの活用

既知の太陽系小天体(2020-08-31現在)

- ・ 小惑星・外縁天体 963399
(小惑星番号あり 546077)
- ・ 地球近傍天体 (NEO) 23138
- ・ 彗星 4173

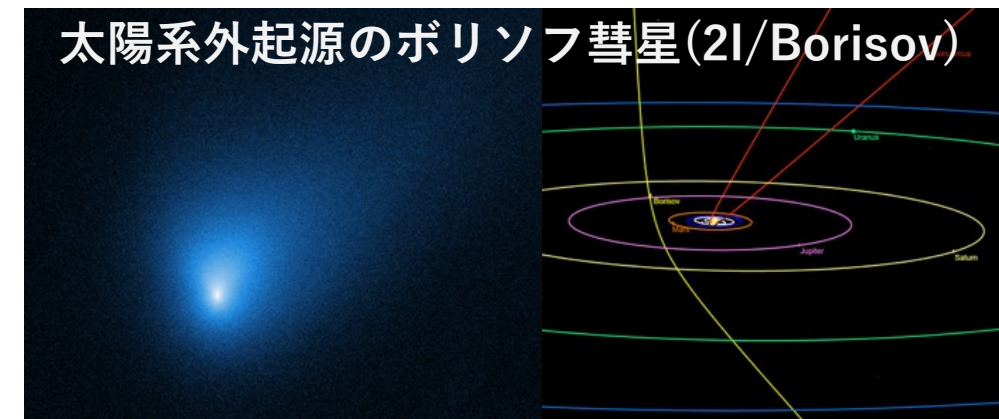
これまで多数の太陽系小天体が発見されているが、その多くは観測数が少なく天体の詳細までは分かっていない。広視野で多数の天体を検出できるHSCのデータから、移動天体である小惑星・彗星を検出しその性質を明らかにする。

- ・ 各観測時刻ごとに約100万天体の座標を計算し、すばるの検出天体DBとのマッチングをおこなう
- ・ 将来的には、すばるで新発見の移動天体探査も



(JAXA)

(JAXA, 東京大, 高知大,
立教大, 名古屋大, 千葉工大,
明治大, 会津大, 産総研)



(NASA, ESA, D. Jewitt)