

バーチャル天文台を利用した天文学研究の成果と今後の期待

白崎裕治 小宮 悠
国立天文台

話の概要

✓ 白崎

- 天文データアーカイブの現状
- JVO を利用した科学成果について紹介

✓ 小宮

- 今年度から JVO グループに参加
- 北大出身の理論天文学者
- JVO をつかって科学研究を行うこと
- JVO を利用したサイエンスユースケースについて紹介

天文データアーカイブの現状

- ✓ 全世界の天文データは1年で倍増
 - CCDの普及 → 多数のCCDを利用した観測装置
- ✓ 観測装置の予算規模も増加
 - データの質や価値も増加
 - データを公開し広く使われることが予算を付ける条件となっている。→ アーカイブ化が必須
- ✓ 無償で利用可能なデータを利用した研究
 - アーカイブデータを利用した天文学研究が主要な研究手法の一つとなりつつある。
 - これらのデータをより利用しやすい形で配信する必要
 - バーチャル天文台

バーチャル天文台

- ✓ 個々の研究組織が管理・運用している天文データを効率的に利用するための仕組み
 - International Virtual Observatory Alliance (IVOA)
 - 標準化されたインターフェイスにより一様な方法でさまざまなデータベースを利用可能
 - ポータルサイト、VOアプリケーション、独自スクリプトによる自動化 …
 - 大量データを利用した天文学研究が容易になる。
- ✓ データコンテンツの充実が急務
 - アーカイブデータ(生データ)をパイプライン処理する仕組みも必要な要素
 - 最先端の使えるデータを流通させることが現在の最大の課題

VO を利用した科学論文

- ✓ 主要な査読付き天文ジャーナル紙 (A&A, ApJ, AJ, MNRAS)
- ✓ アブストラクトに Virtual Observatory を含む論文の内、科学成果についての論文のみピックアップ
 - 2006年 2本、2007年 5本、2008年 3本、2009年 2本
- ✓ VOがどのように使われたか
 - VOTool を利用した → データ収集、Workflow
 - 従来の方法でもできる研究？
- ✓ インパクトのある科学研究成果が必要
 - 今まで誰もできなかった最先端の研究ができることを実践して見せることが VO への理解を得るために現在必要
 - 「これまでより便利」にというだけでは推進する動機として弱い
 - 最先端研究の実践をしつつ、だれでも使える便利な道具として成長させる必要がある。

JVO を使った科学研究

- ✓ いまだかつてだれもやったことのない 最先端研究
- ✓ すばるの全アーカイブデータ を利用
 - すばるの全アーカイブデータを使って研究をしている人はいない。極めて困難。
 - 今回は Suprime-Cam にターゲットを絞る
- ✓ 遠方宇宙 の研究
 - 観測領域を遠方へ伸ばすのは常に最先端
 - 広視野撮像能力を生かして 銀河の大規模構造の進化を AGN と銀河のクラスタリング観測から探る。
 - $z > 0.6$ (50億年) の宇宙におけるこの種の研究は銀河の種類を限定しない観測としては世界初、かつ大規模。

PASJ submitted, arXiv:0907.5380

PASJ: Publ. Astron. Soc. Japan , 1-??,
© 2009. Astronomical Society of Japan.

Early Science Result from the Japanese Virtual Observatory: AGN and Galaxy Clustering at $z = 0.3$ to 3.0 *

Yuji SHIRASAKI¹ Masahiro TANAKA¹ Masatoshi OHISHI¹ Yoshihiko MIZUMOTO¹ Naoki YASUDA² Tadafumi TAKATA¹

¹*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka City, Tokyo, 181-8588, Japan*

²*The Institute for Cosmic Ray Research, University of Tokyo, 5-1-5 Kashiwa-no-Ha, Kashiwa City, Chiba, 277-8582, Japan*
yuji.shirasaki@nao.ac.jp

(Received 2009 0; accepted 2009 0)

Abstract

We present the result of projected cross correlation analysis of AGNs and galaxies at redshifts from 0.3 to 3.0. We used the Japanese Virtual Observatory (JVO) to obtain the Subaru Suprime-Cam images and UKIDSS catalog data around known AGNs. We investigated $\sim 1,000$ AGNs, which is about ten times larger samples than previous studies covering the redshifts larger than 0.6. We found significant excess of galaxies around the AGNs at redshifts of 0.3 to 1.8. For the lowest redshift sample ($z < 0.8$), we obtained correlation length of $r_0 \sim 5$ Mpc, which indicates that the AGN at these redshifts resides in a similar environment to typical local galaxies. We also obtained redshift dependence of r_0 in each luminosity group, which show the AGN produced at higher redshift resides in a denser environment than the lower redshift AGN does. For AGNs at the highest redshift sample ($z > 1.8$), we could not obtain a significant clustering signature and put an upper limit of $r_0 < 17$ Mpc (90% C.L.). These results are compatible with the major merger scenario which incorporates the downsizing of mass assembly.

Key words: astronomical data bases: miscellaneous — galaxies: clusters: general — galaxies: nuclei — cosmology: large-scale structure of universe

超巨大ブラックホールと銀河の進化モデルを検証

- ✓ 銀河中心；超巨大ブラックホール(SMBH, $>10^6 M_{\text{SUN}}$)
- ✓ 銀河の成長：銀河と銀河の衝突・合体
- ✓ BH の成長：銀河の衝突時にBHも合体 → SMBHへ成長
- ✓ 中心の BH に質量降着が発生 → 明るく輝く AGN (QSO)
- ✓ AGN 周辺 → 銀河が密集し衝突が起きやすい環境のはず
- ✓ $z < 0.6$ (～50億年前): 特に密集した環境とはいえない
- ✓ より遠方ではどうか?: AGN の生成率が最大となる $z \sim 2.0$ (～100億年前) では?

AGN と銀河のクラスタリング解析

Step 0 大量データ＋並列計算

- すばる Suprime-Cam の全データ解析

Step 1 分散データベース連携

- AGN 領域画像の検索

Step 2 ワークフローによる自動処理、並列処理

- 各 AGN 周辺における銀河数密度分布の計算

Step 3 天文学上の解析

- クラスタリング強度の計算

解析

✓ 10TBの生データの整約

- データ転送 10日 @100Mbps
- 計算 1年 (1CPU)
- 10 TB以上ものストレージ装置が必要
- 30 並列で計算 → 10 日
- 並列計算システムを構築するスキル(購入するお金)
- 研究者個人で用意するのは容易ではない。
- データセンターで標準パイプラインで処理
- 最大48並列で計算可能
- 解析システムとデータアーカイブシステムとは 48Gbps の容量をもつ専用スイッチで直結
- 整約したデータは VO サービスとして公開
- 大型望遠鏡の特定装置の全データをすべて整約したのは世界初？

Step 1: AGN 周辺画像の検索・取得

- ✓ 複数のデータサイトにまたがるテーブルジョイン
- ✓ サイト A にある AGN の座標 (約20万件) に対しサイト B にある画像 (約6千件) の検索

```
SELECT qso.*, img.*  
FROM ivo://jvo/vizier/VII/235:qso_veron_2006 qso,  
      ivo://jvo/subaru/spcam:spcam_mos_view AS img  
WHERE qso.z >= 1.0 AND qso.z < 1.2  
      AND  
      img.region = Circle((qso.raj2000, qso.dej2000), 0.01)
```

- ✓ すばる望遠鏡による観測データがある AGN のリスト
→ ~6,000件
- ✓ JVOを使わない場合 → 20万回のクエリが必要！
- ✓ 20回程度に分割して検索 → 一時間程度で検索終了

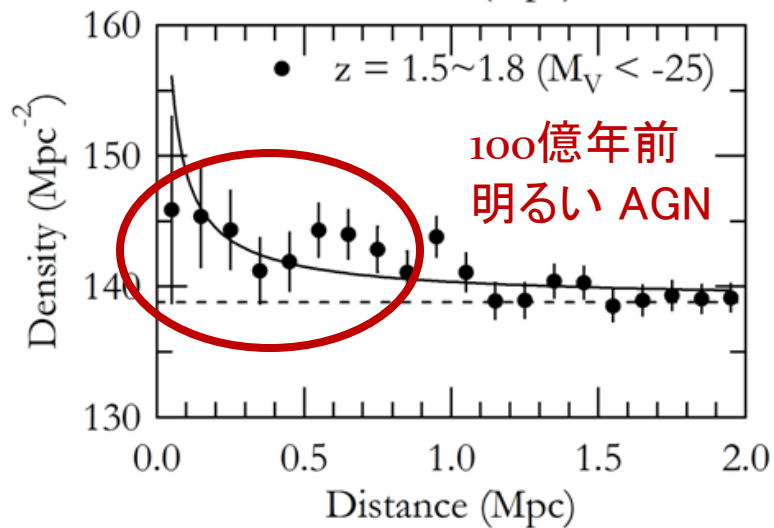
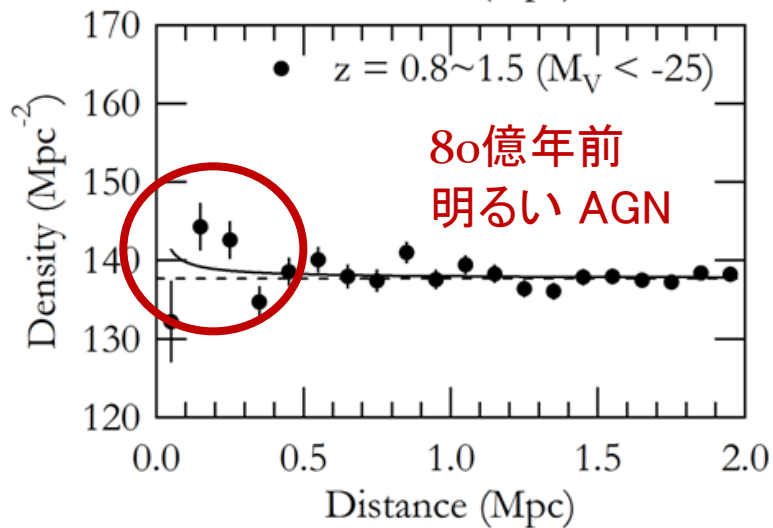
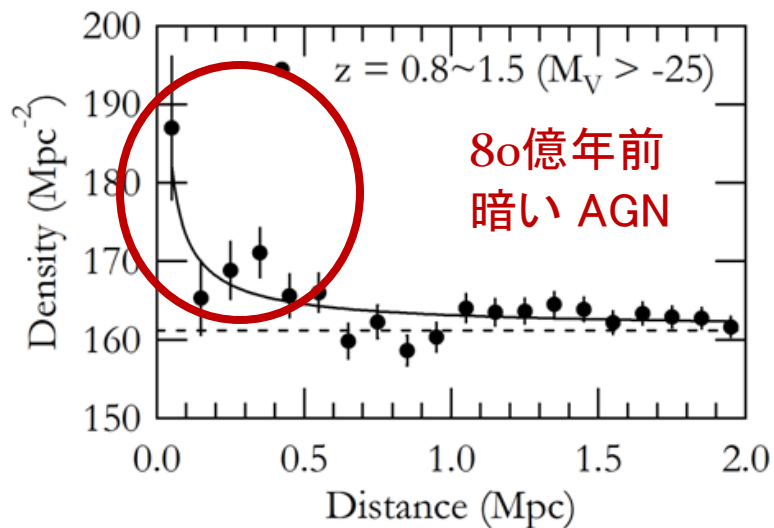
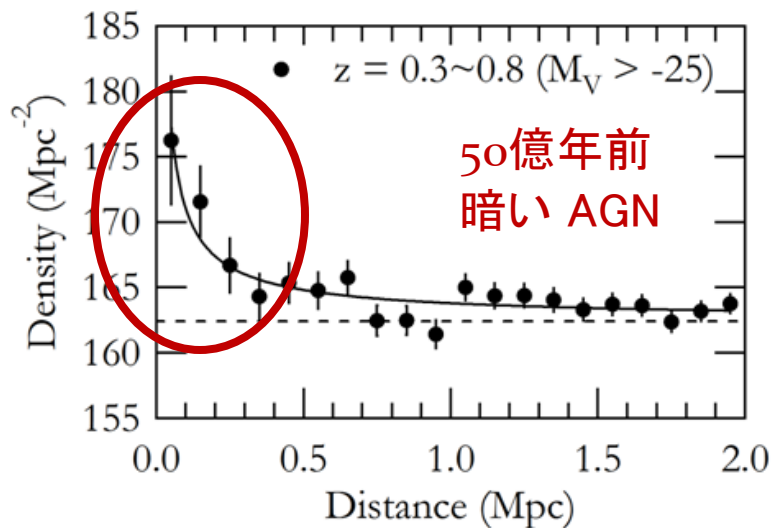
Step 2: 画像解析 (AGN毎)

- AGN の座標値を入力とするワークフローを作成、実行。
- ワークフローはシェルスクリプトで記述。
- ワークフローの実行はユーザ側の計算機で。
- 各フローは 30 並列で実行。数日で終了。
 1. 指定した領域の画像データを取得 (VOサービス)
 2. 画像データのクオリティーチェック
 3. 画像データから天体検出、多波長カタログ作成
 4. AGN を中心とした銀河数分布と有効検出エリアの計算
 5. 数密度分布作成

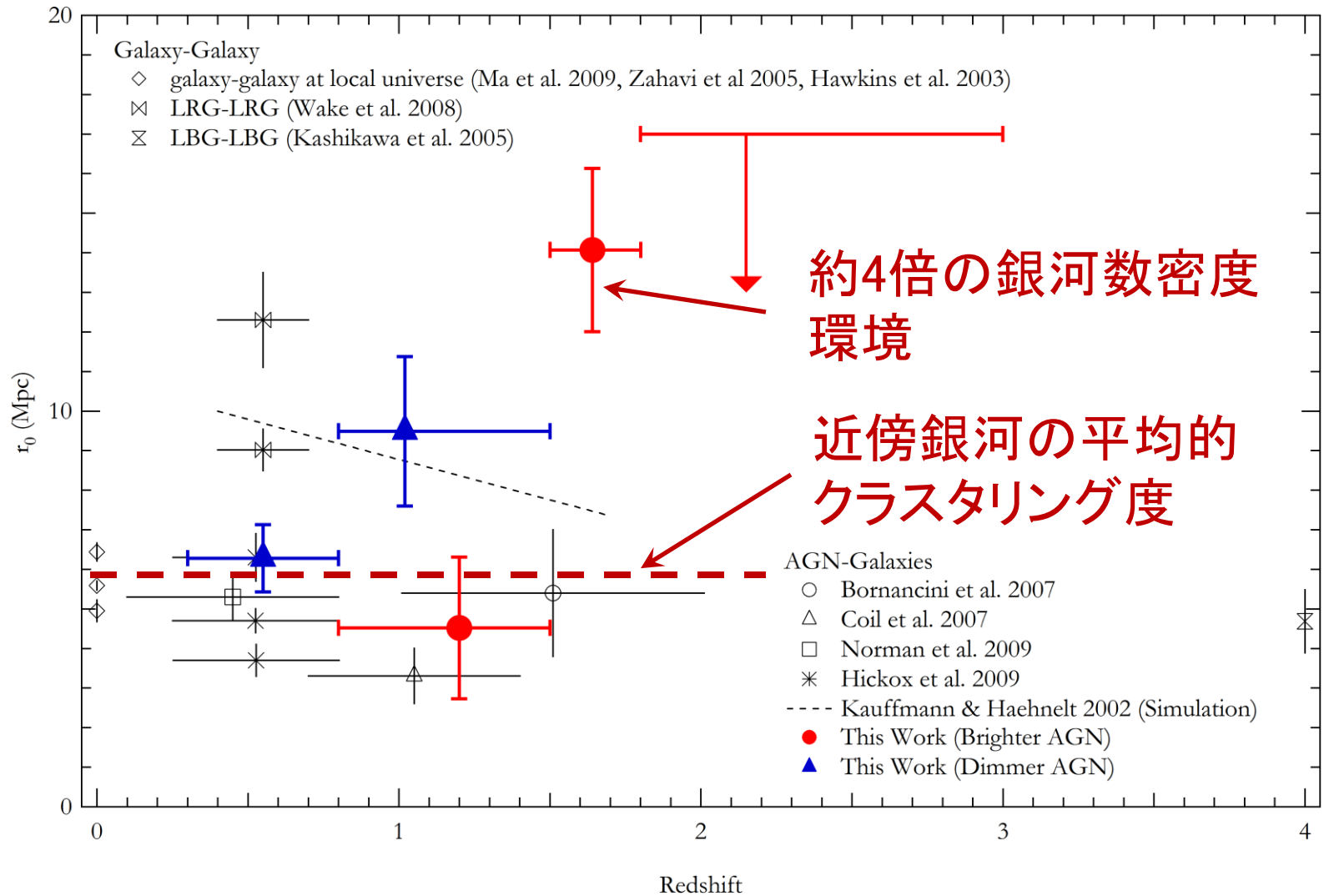
Step 3: 平均数密度分布の計算

- ✓ Step 2 で求めた数密度分布を足し合わせ平均化する。
- ✓ シングルスレッド計算 → 数分で終了する処理
- ✓ AGN の赤方偏移、絶対光度で分類し、それぞれで平均数密度を求める。
- ✓ AGN-銀河の相関距離(クラスタリング強度)を計算。
- ✓ 近傍銀河等の相関距離と比較、AGN は高密度領域にあるか？

AGN周辺の銀河数密度分布



銀河のクラスタリング度 vs 赤方偏移



データベース科学実践の結論

- ✓ 世界第一線級の研究が行えるデータの配信が最重要である。
 - データアーカイブシステムが広く使われるためには、使いやすさ以前にコンテンツの充実が不可欠
- ✓ データ取得が自動でできることがその次に重要
 - スクリプト言語などで書かれたワークフローから自動でデータ取得が行えることで、大量のデータを利用した研究が可能になる。
- ✓ 必要な情報処理技術
 - 大量データを並列処理する技術が必須
 - 大量データを解析ノードに転送する部分がネック
 - データと並列解析システムは同一ネット上に置く
 - ワークフロー → GUI ???, 個人的にはスクリプトで書くほうが効率的…