Astronomy and Computing 15 (2016) 1–15

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

SPOKES: An end-to-end simulation facility for spectroscopic cosmological surveys

B. Nord[a,*], A. Amara[b], A. Réfrégier[b], La. Gamper[b], Lu. Gamper[b], B. Hambrecht[b], C. Chang[b], J.E. Forero-Romero[f], S. Serrano[g], C. Cunha[c,e], O. Coles[i], A. Nicola[b], M. Busha[e], A. Bauer[g], W. Saunders[h], S. Jouvel[g], D. Kirk[i], R. Wechsler[c,d,e]

[a] Fermilab Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500, United States
[b] ETH Zurich, Department of Physics, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
[c] Department of Physics, Stanford University, Stanford, CA 94305, United States
[d] SLAC National Accelerator Laboratory, 2575 Sand Hill Rd., MS 29, Menlo Park, CA 94025, United States
[e] Kavli Institute for Particle Astrophysics and Cosmology, 452 Lomita Mall, Stanford University, Stanford, CA 94305, United States
[f] Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia
[g] Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Facultat de Ciències, Torre C5 par-2, Barcelona 08193, Spain
[h] Australian Astronomical Observatory, PO Box 915 North Ryde NSW 1670, Australia
[i] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 4 April 2014
Accepted 2 February 2016
Available online 3 March 2016

Keywords:
Computation
Cosmology
Simulation
Spectroscopy
Extragalactic
Galaxies

ABSTRACT

The nature of dark matter, dark energy and large-scale gravity pose some of the most pressing questions in cosmology today. These fundamental questions require highly precise measurements, and a number of wide-field spectroscopic survey instruments are being designed to meet this requirement. A key component in these experiments is the development of a simulation tool to forecast science performance, define requirement flow-downs, optimize implementation, demonstrate feasibility, and prepare for exploitation. We present SPOKES (SPectrOscopic KEn Simulation), an end-to-end simulation facility for spectroscopic cosmological surveys designed to address this challenge. SPOKES is based on an integrated infrastructure, modular function organization, coherent data handling and fast data access. These key features allow reproducibility of pipeline runs, enable ease of use and provide flexibility to update functions within the pipeline. The cyclic nature of the pipeline offers the possibility to make the science output an efficient measure for design optimization and feasibility testing. We present the architecture, first science, and computational performance results of the simulation pipeline. The framework is general, but for the benchmark tests, we use the Dark Energy Spectrometer (DESpec), one of the early concepts for the upcoming project, the Dark Energy Spectroscopic Instrument (DESI). We discuss how the SPOKES framework enables a rigorous process to optimize and exploit spectroscopic survey experiments in order to derive high-precision cosmological measurements optimally.

1. Introduction

Progress in cosmology over recent decades has led to some of the most pressing questions in fundamental science today, such as those related to the nature of dark matter, dark energy, and gravity on cosmological scales. To address these questions, several wide-field spectroscopic surveys are in progress or being planned, including WiggleZ (Drinkwater et al., 2010), the Hobby–Eberly Telescope Dark Energy EXperiment (HETDEX; Adams et al., 2010), the Prime Focus Spectrograph (PFS; Takada et al., 2012), the Big Baryon Oscillation Spectroscopic Survey (BigBOSS; Schlegel, 2011), the Dark Energy Spectrometer (DESpec; Abdalla et al., 2012), the Dark Energy Spectroscopic Instrument (DESI[1]) and the 4 m Multi-Object Spectroscopic Telescope (4MOST; de Jong

---

* Corresponding author.
E-mail address: nord@fnal.gov (B. Nord).

[1] http://desi.lbl.gov.

http://dx.doi.org/10.1016/j.ascom.2016.02.001

et al., 2012). The goal of these experiments is to provide three-dimensional maps of the large-scale structure of the universe by measuring the angular positions and redshifts of galaxies in large cosmological volumes.

To reach the levels of precision that are needed to address the fundamental open questions in cosmology, these experiments must meet stringent requirements on both statistical power and control of systematic errors. These requirements, therefore, drive all aspects of these experiments from instrument design to survey optimization. Simulation tools play a key role in the design and optimization process: they are important for forecasting science performance of a given experimental configuration and, moreover, to demonstrate the feasibility of a mission design. Rigorous simulation tools can also allow the science team to prepare for the science interpretation and exploitation of the data.

Such simulation tools have been developed for a number of cosmological surveys. For example, the optical imaging project, Sloan Digital Sky Survey (SDSS; York et al., 2000) employed the Monte-Carlo technique to test deblending in the image processing pipeline prior to the survey taking place.[2] SDSS also used simulations of galaxies, stars and QSOs to prepare and calibrate analysis pipelines for object classification (Fan, 1999; Strateva et al., 2001), and for measurements of the galaxy luminosity function (Blanton et al., 2001).

The Dark Energy Survey (DES; Flaugher et al., 2012) will rely on large-scale simulated catalogs to forecast cosmological constraints (e.g., Bernstein et al., 2012), develop science analysis pipelines (e.g., Chang et al., 2014), and improve the survey strategy (Neilsen, 2012). Galaxy catalogs, along with pixel-level image simulations, also permit the development of image reduction pipelines. Next-generation experiments, like the Large Synoptic Survey Telescope (LSST; LSST Dark Energy Science Collaboration, 2012), will employ photon-level simulations to account for sources of noise, such as atmospheric turbulence (Connolly et al., 2010; Claver et al., 2012; Chang et al., 2012). In addition, operational procedures, like survey strategy, have benefited from extensive simulations (Delgado et al., 2006; LSST Science Collaboration, 2009; Gibson et al., 2011; Honscheid et al., 2012).

As simulations and forward modeling methods play an increasingly important role in survey design and analysis, new frameworks for simulations have been developed. For example, the Monte-Carlo-Control-Loop (MCCL) method, proposed by Réfrégier and Amara (2013), aims to build a robust set of control loops, based on simulations, for verifying that complex measurement methods meet systematic requirement levels. Such system-level optimizations have underscored the need for fast simulations leading to efforts to develop simulations that are fast enough to support such integrated development. An example of this is Ultra Fast Image Generator (UFig; Bergé et al., 2013), which has been developed to quickly and efficiently produce simulated wide-field survey images.

Spectroscopic surveys can take advantage of the same kinds of mock galaxy catalogs as imaging surveys for forecasting. However, there are more operations and additional levels of complexity in spectroscopic surveys: for example, targets must be pre-selected before the surveying can begin, and for each tile on the sky, fibers are allocated to sources; moreover, these operations are intertwined, such that decisions regarding one will affect one or more of the others.

In response to these challenges, spectroscopic experiments have undertaken several design approaches. For example, some recent surveys, such as SDSS-III's Baryon Oscillation Spectroscopic Survey (BOSS) and the 6dF Galaxy Survey (6dFGS), focused simulation efforts toward optimizing the fiber allocation and tiling algorithms (Campbell et al., 2004; Blanton et al., 2003). Studies for BigBOSS have performed target selection on mock catalogs and simulated two-dimensional images of galaxy spectra in an effort to develop the tools to extract spectra from images (Schlegel, 2011). 4MOST has developed the Facility Simulator (Boller and Dwelly, 2012), which links together the survey strategy and fiber allocation to convert an input catalog (from an imaging survey) into one that would result from a 4MOST survey.

In this paper, we describe the SPectrOscopic KEn Simulation (SPOKES), an end-to-end simulation facility for spectroscopic cosmological surveys. SPOKES is built on an integrated infrastructure, modular function organization, coherent data handling, and fast data access. These key features allow reproducibility of pipeline runs, enable ease of use, and provide flexibility to update functions within the pipeline. The pipeline's framework is also cyclic: it can be easily executed in a loop, offering the possibility to make the science output an efficient measure for design optimization and feasibility testing. While the framework is general, we use the design of the DESpec experiment concept (Abdalla et al., 2012) as a baseline for development and for benchmarking results. DESpec was one of the early concepts for the upcoming DESI experiment.

We present here the architecture, and the science and computational performance results of the SPOKES simulation pipeline. SPOKES and all the modules are written in the Python programming language.[3]

The paper is organized as follows. In Section 2, we describe the challenges that spectroscopic surveys need to meet in order to reach the required precision, as well as the principal ingredients in a framework that simulates surveys. In Section 3, we present SPOKES and show how its design addresses these challenges. In Section 4, we present science and performance results of the simulation pipeline. Our conclusions are summarized in Section 5. Details regarding the data format choices and the input cosmological simulation are described in the Appendix.

2. Challenges for spectroscopic survey simulations

2.1. Challenges

Future wide-field spectroscopic surveys offer great promise to address the fundamental questions described above. Their exploitation, however, will pose the following challenges that need to be addressed in order to achieve the required accuracy.

- *High precision:* The next generation of spectroscopic surveys, along with other Stage IV dark energy experiments (Albrecht et al., 2006), aim to measure the dark energy equation of state parameter, w, to percent-level precision. This sets ambitious requirements — e.g., that these experiments cover large cosmological volumes and maintain tight control over errors.
- *Systematics:* As the statistical power of surveys increases, numerous sources of systematic errors become significant. These include errors in the calibration of the survey selection function, inhomogeneous photometric target selection, masking, etc. These systematics need to be carefully calibrated and controlled so that they become subdominant compared to statistical errors.

---

[2] http://www.astro.princeton.edu/~rhl/photo-lite.pdf.

[3] http://www.python.org.