DEVELOPING ATMOSPHERIC RETRIEVAL METHODS FOR DIRECT IMAGING SPECTROSCOPY OF GAS GIANTS IN REFLECTED LIGHT I: METHANE ABUNDANCES AND BASIC CLOUD PROPERTIES

R. E. LUPU

BAER Institute / NASA Ames Research Center, Moffet Field, CA 94035, USA

M. S. MARLEY

NASA Ames Research Center, Moffet Field, CA 94035, USA

N. LEWIS

Space Telescope Science Institute, Baltimore, MD, USA

M. LINE

Univ. California at Santa Cruz, CA, USA

W. TRAUB

Jet Propulsion Laboratory, Pasadena, CA

K. ZAHNLE

NASA Ames Research Center, Moffet Field, CA 94035, USA

ABSTRACT

Reflected light spectroscopy and photometry of cool, directly imaged extrasolar giant planets are expected to be performed in the next decade by space-based telescopes equipped with optical wavelength coronagraphs and integral field spectrographs, such as the Wide-Field Infrared Survey Telescope (*WFIRST*). We are developing a new atmospheric retrieval methodology to help assess the science return and inform the instrument design for such future missions, and ultimately interpret the resulting observations. Our retrieval technique employs an albedo model coupled with both a Markov chain Monte Carlo Ensemble Sampler (*emcee*) and a multimodal nested sampling algorithm (*MultiNest*) to map the posterior distribution. This combination makes the global evidence calculation more robust for any given model, and highlights possible discrepancies in the likelihood maps. Here we apply this methodology to simulated spectra of cool giant planets. As a proof-of-concept, our current atmospheric model contains 1 or 2 cloud layers, methane as a major absorber, and a H_2-He background gas. This 6-to-9 parameter model is appropriate for Jupiter-like planets and can be easily expanded in the future. In addition to deriving the marginal likelihood distribution and confidence intervals for the model parameters, we perform model selection to determine the significance of methane and cloud detection as a function of expected signal-to-noise, in the presence of spectral noise correlations. After internal validation, the method is applied to realistic reflected-light spectra of Jupiter, Saturn, and HD 99492 c, a likely observing target. We find that the presence or absence of clouds and methane can be determined with high accuracy, while parameters uncertainties are model-dependent.

Keywords: methods:statistical — planets and satellites:atmospheres — planets and satellites: composition — techniques:spectroscopic

1. INTRODUCTION

Space-based telescopes equipped with coronagraphic imagers can separate light scattered by orbiting planets from that of their primary stars. The detection of light that penetrates deeply into an atmosphere rather than merely skims its upper layers, as with transit methods, potentially permits more extensive and informative characterization of atmospheric gaseous absorbers as well as cloud and haze layers. However the interpretation of the scattered light signal will in practice be limited by a multitude of uncertainties beyond the basic limitations

Roxana.E.Lupu@nasa.gov

of data quality. Among these are the uncertain or unknown planetary radii, masses, and cloud layers. Here, in the first of what we plan to be a series of papers, we present the initial development of a atmospheric retrieval methodology that quantifies the resultant uncertainties and clarifies the precision with which planet gravity, temperature, composition, and cloud structure and other parameters can be discerned.

Direct imaging offers the possibility of characterizing planets around nearby stars and at larger orbital distances than is possible for transit observations. Directly imaged planets are under less stellar irradiation than traditional transit observation targets and can either be young, warm, and self-luminous, or older and much colder than the ones studied by transit methods. While a multitude of space coronagraph missions have been studied or proposed over the last two decades, the only mission currently in development by NASA with the capability of imaging cool giant planets in reflected light is *WFIRST*.

Current estimates are that a coronagraph-equipped *WFIRST* mission will be able to obtain photometry and spectra for at least a dozen known radial velocity (RV) planets as well as search for lower mass planets. An example of the diversity of the known RV planets favorable for direct imaging is shown in Figure 1. This sample was drawn from the Exoplanet Encyclopedia and will likely increase with future discoveries from RV or *WFIRST* surveys. In this figure the known $M \sin i$, measured by RV methods, is plotted against estimated effective temperature in order to understand the phase space of atmospheric conditions that might be expected among these most favorable planets. The planets' inclination (i) will be determined form the direct imaging observations, therefore constraining their approximate masses and, with the aid of the mass-radius relationship, their surface gravities. Vertical color bands show the approximate ranges over which various atmospheric compounds form clouds. While many Jupiter and Saturn-like worlds, with ammonia clouds, are expected, some planets with water, alkali, and even methane clouds may also be observed.

The Coronagraph Instrument onboard *WIFRST*, in combination with an Integral Field Spectrometer, will provide us with images (430–970 nm) and low-resolution (spectral resolution $R \sim 70$) reflected light spectra of gaseous planets around nearby Sun-like stars (600-970 nm). Unlike transit spectroscopy that only probes the top of the atmosphere to ~ 1 mbar (e.g., Kreidberg et al. 2014), reflected light can probe deep into the atmosphere of these gas giants (e.g., Marley et al. 2014), and therefore offers a more comprehensive view of compositions, temperatures, and cloud layers.

Most planets in Figure 1 have equilibrium effective
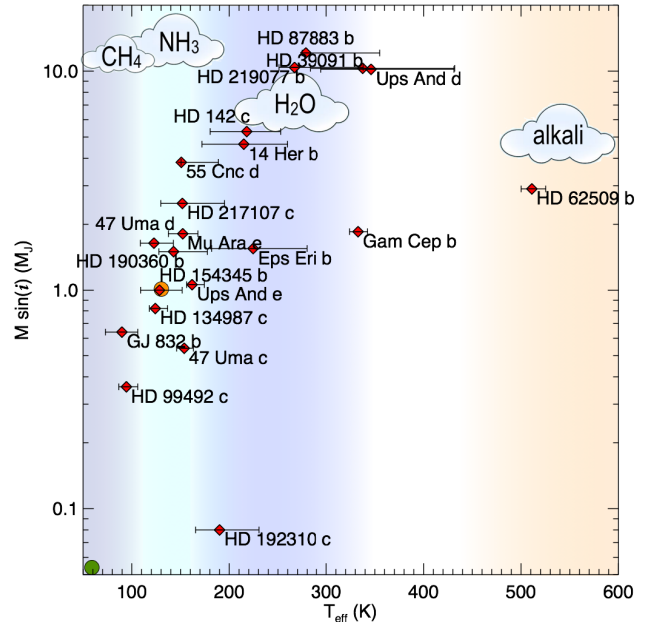
Figure 1. $M\sin(i)$ and ranges of estimated T_{eff} of a selection of announced RV planets that are favorable for direct imaging. The orange circle represents Jupiter while the green one hints at Uranus which actually falls below the lower axis. Estimated T_{eff} computed from planet orbits, Jupiter's albedo, and estimated internal heat flows given available estimates of the primary stars' ages. Bands show major cloud species expected in various ranges of T_{eff}. The existence of two of the planets shown, Ups And e and Eps Eri b, is controversial.

temperatures of $\sim 150 - 350$ K. Assuming these worlds are comparable to solar system gas giants, their $600 - -970$ nm spectra will be dominated by cloud decks of water or ammonia and gaseous absorption by methane and possibly water. Photochemical hazes will doubtless be important as well. There is a long and comprehensive history of interpretation of such spectra of solar system planets dating back to Sato & Hansen (1979) and before. For Jupiter-like atmospheres the continuum scattered flux level at these wavelengths is set by scattering from the bright clouds while Rayleigh scattering is more important at the bluest wavelengths. The bright continuum is punctuated by gaseous methane absorption features of varying strengths. The relative strengths of the various methane absorption bands, combined with the continuum flux level set by the clouds, together constrain the cloud properties and methane column abundance. Shortward of 600 nm, the photometric measurements will give us information about the shape of the continuum, dominated by Rayleigh, haze, and cloud scattering. If both CH_4 and H_2O features are present in the spectra, we can constrain the C/O ratio, value related to

the place of planet's formation in the circumstellar disk (Bond et al. 2010; Helling et al. 2014; Öberg et al. 2011).

Extracting such information from low to moderate spectral resolution data at modest signal-to-noise ratios will be a challenge. Cloud properties and location, absorber abundances, planetary radius (and thus gravity), and the atmospheric thermal profile will all be unknown. While forward modeling techniques, such as Cahoy et al. (2010) can give insight into the range of possible spectra, extraction of cloud properties and absorber abundances will require the application of retrieval methods to the available data.

We aim to develop the necessary theoretical and computational framework to enable such retrievals. As this will be a complex endeavor we approach the problem in steps. Here we present a first step in the development of this framework, focusing on the retrieval of gross cloud properties, atmospheric gravity, and methane mixing ratio. In future papers we will add retrievals for orbital phase, planet size, additional absorbers and atmospheric thermal profile.

In the remainder of this paper we provide more detailed background on reflected light spectra of giant planets, present the conceptual model and Markov Chain Monte Carlo retrieval method, and the results of this study. The paper is organized as follows: Section 2 provides more context and background to the problem. Section 3 describes our albedo code and the forward models used in the retrievals; Section 4 describes our noise model used to generate the input datasets; Section 5 contains the Bayesian retrieval scheme, followed by its validation in Section 6. Other retrieval results for more realistic spectra of known gas giants are shown in Section 7, and the conclusions are summarized in Section 8.

2. BACKGROUND

In this section we provide a brief overview to a few of the key concepts used throughout the remainder of the paper.

2.1. *Geometric Albedo*

The analysis of extrasolar planet reflection spectra owes much to the solar system literature. However this literature also brings its own set of conventions, not all of which translate smoothly to the exoplanet context. For expediency we nevertheless choose here to follow these conventions, although we recognize that as exoplanet direct imaging evolves into its own sub-field that this terminology will likely evolve to shed some vestigial structures.

A foremost concept is the geometric albedo, the ratio of light received from a planet when observed at full phase to that which would be measured from a perfectly reflective Lambert disk of the same size as the planet. Because
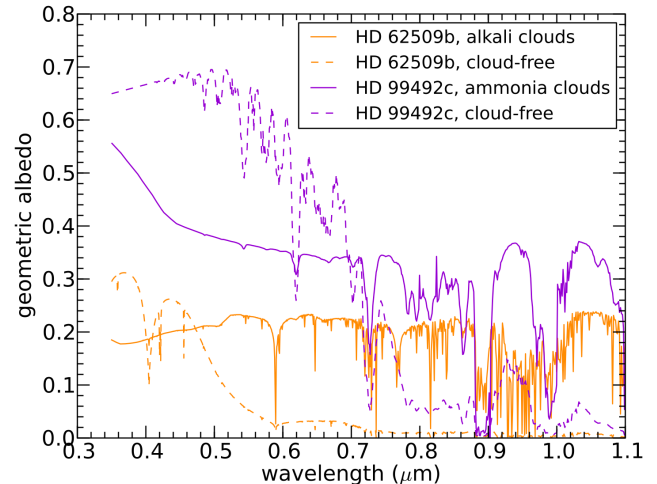

Figure 2. Model albedo spectra for two of the planets shown in Figure 1, HD 62509b and HD 99492c. For each planet two model cases are shown, one with and one without the clouds expected from equilibrium condensation chemistry considerations. Absorption bands seen in the HD 99492c model are methane while Na and K features are also prominent in HD 62509b.

the angular distribution of light scattered by a real atmosphere differs from that scattered by a Lambert disk, the geometric albedo of even perfectly scattering atmospheres is not unity. For a conservative, infinitely deep Rayleigh scattering atmosphere the geometric albedo is 0.750. The fractional reflectivity measured at a star-planet-observer angle differing from 180° is given by the product of the geometric albedo and the planetary phase function.

There are two important reasons why geometric albedo spectra will not be directly measured for directly imaged planets. First, while transiting planets can be observed at full phase just before they are eclipsed on the "far" side of their orbits, directly imaged planets will never be observed even close to full phase because they would lie too close to the primary star to be resolved from the star. Second, the radius of a planet will not be directly measured, rather only the product of planet area and reflectivity will be constrained as a function of wavelength. Thus it is an oversimplification to discuss geometric albedo spectra for directly imaged extrasolar planets. Nevertheless to simplify the model development for this work, we consider here only the planetary spectrum at full phase, cast as geometric albedo spectra. In the second paper of this series (Nayak et al., in prep.) we will explore issues arising from the phase dependence of planetary reflectivity (see Cahoy et al. (2010)) and the unknown planetary radius.

Figure 2 shows model geometric albedo spectra we cal-
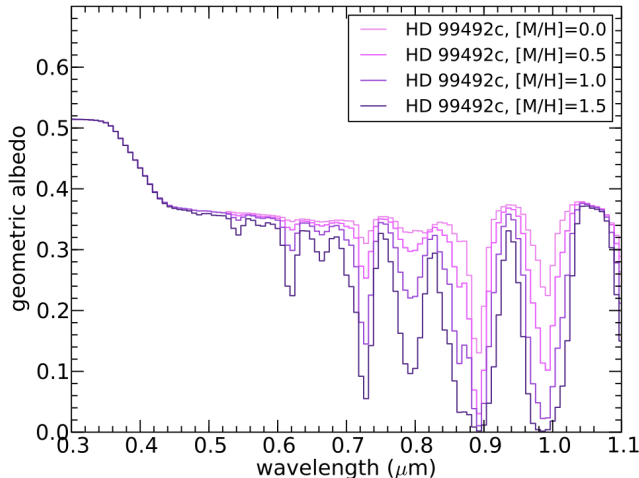
Figure 3. Model spectra at R=70 for HD 99492c at four metallicities, ranging from solar to 30x solar. With increasing metallicity the atmospheric methane bands deepen and weaker bands become more apparent. Thus it is crucial to detect multiple methane bands of varying strengths so that sufficient dynamic range is available to constrain a variety of atmospheric methane abundances.

culated for two of the planets shown in Figure 1 following the method of Cahoy et al. (2010). We find that HD 62509 b (Pollux b) is warm enough that the only likely atmospheric cloud condensate would be clouds of alkali species like Na_2S. Conversely HD 99492 c is cold enough that thick ammonia, and possibly even methane, clouds are expected. The figure compares computed albedo spectra with and without the expected clouds. Figure 3 presents additional model spectra for increasing atmospheric metallicities, that can be expected given the diversity of extrasolar planets. The atomic abundance ratios (C/H or O/H) can reveal links between the planet's and the host star's compositions. More spectral variation as a function of mass, orbit, metallicity, and phase are described in detail in Cahoy et al. (2010) and Sudarsky et al. (2000). Distinctive differences, diagnostic of important atmospheric processes, between the spectra of known planets can clearly be expected. This study explores how well an instrument like the coronagraph on *WFIRST* would be able to constrain planet atmospheric composition.

2.2. *Retrieval Approaches*

Our atmospheric retrieval procedure involves combining a well-tested planetary albedo code (McKay et al. 1989; Marley et al. 1999; Cahoy et al. 2010) that can take into account multiple absorbers, cloud and Rayleigh scattering, as well as arbitrary incident and observed angles, with state-of-the-art Bayesian inference tools, namely the

Markov chain Monte Carlo (MCMC) ensemble sample *emcee* (Goodman & Weare 2010; Foreman-Mackey et al. 2013) and the multimodal nested sampling algorithm *MultiNest* (Feroz & Hobson 2008; Feroz et al. 2009, 2013) that can be used interchangeably.

We believe that this is the first time ever such powerful retrieval techniques have been designed to *simultaneously* measure molecular abundances and cloud properties and their correlations from scattered light spectra. *NEMESIS* (Rodgers 2000; Irwin et al. 2008) is the only other existing retrieval method for planetary atmospheres in reflected light that has recently been applied to exoplanet characterization (Barstow et al. 2014). By contrast, *NEMESIS* uses non-linear optimal estimation to derive the best-fit model parameters and their uncertainties, and does not incorporate cloud properties in the retrieval process.

As shown by Line et al. (2013, 2014), the Bayesian inference tools are somewhat more appropriate for highly non-gaussian posterior distributions that are expected for future exoplanet observations, given the limited data and complex atmospheric models. Moreover, clouds play a significant role in the atmospheres of both gas giants in our Solar System and the exoplanets considered as future observing targets, given their expected effective temperatures. By including simple cloud properties (optical depth, albedo, depth in the atmosphere, etc.) as model parameters alongside molecular abundances, we can fully explore the degeneracies in the atmospheric structure, given the spectrum.

For our initial retrieval tests we constructed two highly idealized cloud models, one with a single cloud deck of arbitrary opacity, and the other with a scattering haze overlying a completely opaque cloud layer. Such atmospheric models are adequate for the types of planets addressed in this paper, and unquestionably can be improved in future work. Our goal is to determine if consistent results for scientifically interesting quantities (abundances, cloud properties) could even be obtained using reflected light spectra from a space based coronagraph, given the likely modest signal-to-noise and spectral resolution.

3. FORWARD MODEL

Our spectral albedo code for giant planets was originally developed by Marley et al. (1999) and is based on the methods of McKay et al. (1989). This code was subsequently modified and improved by Cahoy et al. (2010), who investigated the albedo variations as a function of star-planet distance, metallicity, mass, and phase angle. This original albedo code uses as input parameters the exoplanet's gravity and depth-dependent temperature, pressure, composition, and cloud properties which are in turn computed by a 1-D radiative-convective equilibrium model (Marley et al. 1999; Cahoy et al. 2010).

The atmosphere is divided in 60 layers, with the bottom pressure marking the point beyond which photon scattering is negligible. The clouds are parametrized by wavelength-dependent optical depth τ_{cld}, single scattering albedo ($\bar{\omega}_{cld}$), and scattering asymmetry factor ($\bar{g}_{cld}$), resulting from a full Mie scattering treatment of particle sizes predicted by a cloud model (Ackerman & Marley 2001). The single scattering albedo represents the ratio between the amounts of scattering and total particle extinction, and the asymmetry factor, $\bar{g}_{cld}$, is a measure of the degree of forward scattering.

To simulate a spherical planet, we cover the illuminated surface of a sphere with 100 plane–parallel facets (Cahoy et al. 2010), where each facet may have different incident and observed angles, $\mu_0 = \cos\theta_0$ and $\mu_1 = \cos\theta_1$, where θ_0 and θ_1 are the angles between the local normal vector and the star and observer, respectively. Increasing the number of facets proportionally increases the computing time, and only leads to a modest increase in accuracy. In this case, the albedo code takes about 3 s to run, which is reasonable to use in combination with an MCMC sampler. Although the general case permits $\theta_0 \neq \theta_1$, for the work reported here we set $\theta_0 = \theta_1$ in order to compute geometric albedo, which by definition is the reflectivity at zero phase angle. In a future work (Nayak et al, in preparation) we will consider observations at arbitrary phase angle.

Following the approach of Horak (1950) and Horak & Little (1965), we use two-dimensional planetary coordinates and Chebyshev-Gauss integration to integrate over the emergent intensities and calculate the albedo spectra. The radiative transfer is performed point by point for each of the points sampling the planetary disk. The scattering source function (Toon et al. 1989; Meador & Weaver 1980) includes the contributions of both diffuse and direct scattering:

$$S(\tau, \mu_1) = \frac{\bar{\omega}}{4\pi} F_0 p(\mu_1, -\mu_0) e^{-\tau/\mu_0}$$
$$+ \int_{-1}^{1} \frac{\bar{\omega}}{2} I(\tau, \mu') p(\mu_1, \mu') d\mu', \quad (1)$$

where F_0 is the Solar flux at to top of the atmosphere, normalized to 1, and $p(\mu_1, \mu_2)$ is the scattering phase function. The two terms on the right-hand side represent the single and multiple scattering components, respectively.

We use a two-stream quadrature (Toon et al. 1989) to solve for the diffuse, angle-independent radiation field. This solution is then used as an approximation to the source function, which is then back-propagated to the top of the atmosphere, while adding the angular dependence given by the scattering phase function. This is a completely scalar approach and does not include any polarization effects.

Based on our experience and the results of Cahoy et al. (2010), we expect that the most relevant model parameters for Jupiter-like exoplanets in reflected light will be the methane abundance, surface gravity, and cloud properties. In a future paper we will consider other gaseous opacity sources. The code uses the opacity for methane in the visible following Karkoschka (1994), and the collision-induced absorption for H_2-H_2, H_2-He and H_2-CH_4 as summarized in Freedman et al. (2008).

The total gaseous absorption optical depth is then $\tau_{abs} = \tau_{CH4} + \tau_{CIA}$. In spite of newer methane line lists, difficulties remain in calculating the high-energy transitions of methane and Karkoschka (1994) is still the best reference for the methane opacity in the visible, and is used to reproduce Solar System measurements. We define $\tau_{total} = \tau_{scat} + \tau_{abs}$, where the total optical depth to scattering is $\tau_{scat} = \tau_{Ray} + \tau_{cloud}$.

Following Cahoy et al. (2010), for the direct scattering (or single scattering term in Equation 1) we use a two-term Heyney-Greenstein scattering phase function with high forward scattering and moderate backscattering:

$$p_{TTHG} = \left(1 - \frac{\bar{g}^2}{4}\right) p_{HG}(\bar{g}, \Theta) + \frac{\bar{g}^2}{4} p_{HG}(-\bar{g}/2, \Theta), \quad (2)$$

where

$$p_{HG}(\bar{g}, \Theta) = \frac{1 - \bar{g}^2}{(1 + \bar{g}^2 - 2\bar{g}\cos\Theta)^{3/2}} \quad (3)$$

and Θ is the scattering angle, related to the planet's phase angle α by $\alpha = \pi - \Theta$, and $\bar{g}$ is the scattering asymmetry factor associated with the scattering by cloud particles, $\bar{g} = \bar{g}_{cld} \times \tau_{cld}/\tau_{scat}$. Rayleigh scattering is calculated following Hansen & Travis (1974), with $\bar{g}_{Ray} = 0.5$, and $\bar{\omega}_{Ray} = 1$. The total layer single scattering albedo then becomes $(\bar{\omega}_{Ray}\tau_{Ray} + \bar{\omega}_{cld}\tau_{cld})/\tau_{total}$, for every layer in the atmosphere. Further details of the radiative-transfer modeling are described in Marley et al. (1999); Cahoy et al. (2010).

For the multiple scattering term in Equation 1, the diffuse scattering phase function is written as a Legendre polynomial expansion, assuming azimuthal independence:

$$p(\mu, \mu') = 1 + 3\bar{g}\mu\mu' + \bar{g}_2(3(\mu\mu')^2 - 1)/2, \quad (4)$$

where μ and μ' denote the scattered and incident angle, respectively, and $\bar{g}_2$ contains the Rayleigh scattering contribution. In this case, for the diffuse radiation μ and μ' are chosen such that the right solution is obtained in the Rayleigh limit.

For retrieval purposes, we have preserved the radiative transfer and scattering prescription of the original albedo code, but made large simplifications to the input parameters. The simplified model used in the present study
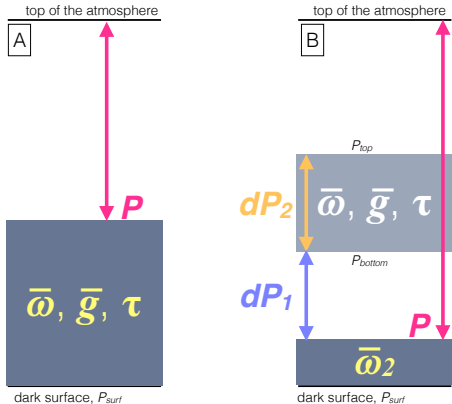
Figure 4. Visual representation of our 1-cloud (panel A) and 2-cloud (panel B) models. The definitions of model parameters and their use in the albedo code are given in Sections 3.1.1 and 3.1.2, respectively.

has constant molecular abundances throughout the atmosphere, with H_2 and He. The pressure-temperature profile $T(P)$ of the atmosphere is kept fixed in this study since we do not expect that our spectral range of interest ($0.4 - 1$ μm) will contain any information for constraining it (see also Barstow et al. (2014)). The wavelength dependence of the cloud parameters is also ignored (gray assumption for τ_{cld}, $\bar{g}_{cld}$, and $\bar{\omega}_{cld}$). The depth dependence is limited to parametrizing the cloud height and cloud top pressure, as described below.

In actuality of course the temperature-pressure profile will vary and this will primarily affect the atmospheric scale height. Here our variation of atmospheric gravity, g, stands in for variations in both $T(P)$ and g. As we add complexity to the model we will explore the sensitivity of retrievals to a varying $T(P)$.

3.1. Cloud Models

For the purposes of atmospheric retrieval we consider two different cloud treatments as illustrated in Figure 4. The simpler of the two models a single cloud layer while the more complex allows for two distinct clouds/hazes. We describe each model in turn below.

3.1.1. 1-Cloud Model

The one cloud model is parameterized as a semi-infinite layer with a cloud top at pressure P in the atmosphere and characterized by the single scattering albedo $\bar{\omega}$, scattering asymmetry factor $\bar{g}$, and the gray optical depth τ in the top model layer in which the cloud is found. For

simplicity of notation, we have dropped the subscript 'cld' from the quantities $\bar{\omega}_{cld}$, $\bar{g}_{cld}$, τ_{cld}, as defined in the previous section. This structure is shown in panel A of Figure 4.

The pressure of the cloud top is allowed to vary freely, not exceeding the atmospheric boundaries. Our typical input pressure-temperature profile has 60 layers dividing the depth of the atmosphere. We find the model layer in which this pressure level is located, j_c ($1 \leq j_c \leq N$), and scale the cloud optical depth in this layer by the position of the cloud top pressure relative to the pressure at the bottom of the layer. The next deeper layer ($j = j_c + 1$) will have cloud optical depth $\tau_j = \tau_{j_c} \times (P_{j+1}/P_j)$, where the layer number j increases with depth in the atmosphere from 0 to N and P_j denotes the pressure at the top of layer j. The cloud optical depths in the following layers all the way to the bottom are calculated iteratively as $\tau_{j+1} = \tau_j \times (P_{j+2}/P_{j+1})$. Thus in this model τ is essentially a measure of how opaque the cloud top is, and the optical depth per unit mass is constant over the entire vertical extent of the cloud. Large values of τ imply a rapid transition from cloudless atmosphere to cloud, whereas small values imply a more gradual increase of cloud opacity.

The cloud single scattering albedo $\bar{\omega}$ and scattering asymmetry factor $\bar{g}$ are kept constant as a function of wavelength and depth in the atmosphere, below the layer containing the top of the cloud, e.g. $\bar{\omega}_j = ... = \bar{\omega}_N = \bar{\omega}$ for $j \geq j_c$. This model will be referred in what follows as the "1-cloud model", and is characterized by 6 parameters: f_{CH4}, g, P, $\bar{\omega}$, $\bar{g}$, and τ, where g is the planet's surface gravity, to be distinguished from $\bar{g}$, and f_{CH4} is the methane abundance.

3.1.2. 2-Cloud Model

Increasing complexity, we created a model appropriate for a cloud deck overlain by a haze layer with a very simple 2 layer structure shown in panel B of Figure 4. Such a model is roughly capable of reproducing the structure observed in Solar System planets, and is a slight modification of the model used in the classic analysis of Jupiter's atmosphere by Sato & Hansen (1979).

The parameters describing the lower cloud are its top pressure P and single scattering albedo ($\bar{\omega}_2$). Following the same approach as in Section 3.1.1, the pressure of the top of the bottom cloud is found in layer j_c, the optical depth below this level is scaled in the same way, except now $\tau = 1$, and is not variable, thus modeling the case of a distinct cloud top. This ensures that the bottom cloud is always optically thick, and makes it effectively act as a reflective surface, with a reflectivity controlled by $\bar{\omega}_2$, and situated at a variable depth given by P.

The position of the upper cloud (or haze layer) relative to the bottom cloud is parametrized by the pres-

sure difference between the top of the lower cloud and the bottom of the upper cloud (dP_1) and the pressure difference between the top and the bottom of the upper cloud (dP_2). For computational convenience, these quantities are defined in log space, and are related to the size and location of the top cloud by $P_{\text{bottom}} = 10^{(P-dP_1)}$ and $P_{\text{top}} = 10^{(P-dP_1-dP_2)}$, where P_{top} and P_{bottom} are the pressures at the top and at the bottom of the upper cloud, respectively (see Panel B, Figure 4).

Similar to the 1-cloud approach, we find the layers in which the top and bottom pressure of the upper cloud are located and the corresponding fractions, or locate the cloud in a single layer, if necessary. For all the layers between the top and the bottom, the optical depth of the upper cloud is scaled as $\tau_j = \tau \times (P_{j+1} - P_j)/(P_{\text{bottom}} - P_{\text{top}})$, where τ is the input variable and is wavelength-independent. The single-scattering albedo $\bar{\omega}$ and asymmetry factor $\bar{g}$ are again kept constant as a function of wavelength and for all layers between P_{top} and P_{bottom}. This model will be referred in what follows as the "2-cloud model", and is characterized by 9 parameters: f_{CH4}, g, P, dP_1, dP_2, $\bar{\omega}$, $\bar{g}$, τ, and $\bar{\omega}_2$.

Note that the haze single scattering albedo is treated as a constant with wavelength. Thus hazes that absorb preferentially in the blue, lowering the albedo in the short-wavelength part of the spectrum, such as are commonly found in solar system giant planet atmospheres, are not taken into account here. These effects become more important below 0.5 μm, and are unlikely to affect the region of interest for this study (0.6 − 1 μm). We will address the wavelength dependence of the single scattering albedo in future work, especially when adding photometric points in the visible.

4. SIMULATED DATA

To simulate the direct imaging observations, we use a generic prescription for the total signal and associated noise expected in the planet's point spread function (PSF). This model is sufficient for investigating the effect of data quality (as quantified by the signal-to-noise ratio, SNR) on the size of uncertainties associated with the atmospheric parameters and on the significance of methane and cloud detection. We consider this to be a sufficiently general synthetic data model, that can be easily swapped once a specific instrument simulator becomes available (e.g. Robinson et al. 2016). The plots in Figure 5 exemplify our simulated data for a Jupiter-like planet around a Sun-like star, at a distance of 25 pc from our Solar System, using the method detailed below.

Let the total number of counts on the detector, within the planet's PSF, be the sum of planet counts n_{pl}, speckle noise counts n_{spec}, the zodiacal light n_{zodi}, and the total detector background counts from all other sources. The spectral bins are chosen such that the resolving power

$R = 70$ is constant across the $0.4 − 1.0$ μm bandpass. For each spectral bin, we define

$$signal(\text{e}^-) = n_{\text{pl}} \times t,$$
$$noise(\text{e}^-) = [n_{\text{total}} \times t + (f_{pp} \times n_{\text{raw speckle}} \times t)^2]^{1/2},$$
$$(5)$$

where

$$n_{\text{total}}(\text{e}^-/\text{s}) = [n_{\text{pl}} + n_{\text{zodi}} + n_{\text{raw speckle}}$$
$$+ D_c \times m_{\text{pix}} + CIC \times m_{\text{pix}}/t_{\text{frame}}] \times ENF^2$$
$$+ (N_R/G)^2 \times m_{\text{pix}}/t_{\text{frame}},$$
$$(6)$$

n_{total} is the total number of counts within the planet's PSF, t is the total integration time, and the other quantities characterize the detector background noise, with "typical values" for an electron multiplying (EM) CCD detector: $m_{\text{pix}} = 5$ pixels, $D_C = 0.001$ e$^-$(pixel s)$^{-1}$, $N_R = 3$ RMS e$^-$(pixel frame)$^{-1}$, $t_{\text{frame}} = 300$ s, $CIC = 0.001$ e$^-$(pixel frame)$^{-1}$, $ENF = 1.414$, $G = 1000$, and $t = 14000$ s. The remaining f_{pp} quantifies the speckle reduction efficiency and can take values roughly between 1/10 and 1/30. We use a value $f_{pp} = 1/20$ in this paper.

Assuming the stellar spectrum to be a blackbody at 6000 K, and using the model geometric albedo of the planet, we have calculated the expected number of photons in each spectral bin. This number was converted to a count rate, using estimated count rates of $n_{\text{pl}} = 0.012$ e$^-$/s, $n_{\text{zodi}} = 0.012$ e$^-$/s, $n_{\text{spec}} = 0.010$ e$^-$/s, which contain information about the expected quantum efficiency. The number counts coming from all contributions to the total signal are shown in Figure 5, top left panel. The observed spectrum is simulated assuming that the planet and zodi counts have a Poisson distribution (per channel), while the speckle and detector noise counts have a Gaussian distribution (Figure 5 center left). In other words, the simulated data points are drawn from their respective distributions.

In addition, we consider the possibility of noise correlations among different spectral regions. Since the speckle positions relative to the central star change with wavelength, we expect that at the position of the planet in the observed image certain wavelengths will be more affected by speckle noise than others. In our model, we assume that this will affect only the Gaussian-distributed counts, which are dominated by speckle counts, and not Poisson-distributed ones, which consist of planet and zodi counts. Therefore, the total noise contribution of the Gaussian-distributed counts (their distribution around the mean) was split into 2 components, one spectrally correlated, and one spectrally uncorrelated. The correlated noise component was generated as a Gaussian random process with a squared-exponential kernel and correlation length scale of either 25 or 100 nm. These length scales are

appropriate for our chosen spectral range and expected spatial resolution, and the choice of a random process reflects the existing uncertainty in the exact behavior of the speckle noise correlation. Furthermore, we assumed that both correlated and uncorrelated components have equal contributions to the total scatter in the data points, and therefore their distributions will have mean zero and equal variance. This combination of spectrally correlated and uncorrelated noise is shown in the top right panel of Figure 5.

We define the signal-to-noise reference value (SNR_0=signal/noise, from Equation 5) as corresponding to the integrated number of counts in a 10%-wide bandpass centered at 550 nm. Therefore, the integration time needed to achieve a given SNR_0 can be calculated as

$$t(\text{s}) = \frac{SNR_0^2 \times n_{0\text{total}}}{n_{0\text{pl}}^2 - (SNR_0 \times f_{pp} \times n_{0\text{raw speckle}})^2}, \quad (7)$$

where the index 0 denotes the fact that these values are calculated for the 550 nm reference bandpass. We calculate the integration time t_0 necessary to obtain a SNR_0 of 5, 10, or 20, respectively, which is then used to calculate the expected number of counts and scale the signal and noise across the entire bandpass. The final error bars are computed individually for each simulated data point using Equation 5. As shown in Figure 5, the resulting spectrum will have a $SNR < SNR_0$ on average, but we will take the SNR_0 as the reference value in what follows. The values for SNR_0 and speckle noise correlation length as defined above serve as a parametrization of the data space over which we perform our retrievals. The combination of the three SNR values and two possible speckle noise correlation lengths result in 6 simulated datasets for each planet model.

Lacking more detailed information about the instrument, in the above we have assumed that the entire bandpass is observed simultaneously and the quantum efficiency (detector response) is constant across the bandpass. Although these conditions will not be satisfied in a real observation, they amount to assuming that we can achieve the final SNR distribution with wavelength shown in the bottom right panel of Figure 5. This is just one of the many possible realizations of SNR variation over the bandpass, and this is likely to be unique to each dataset, which will likely be a combination of different observing modes. It is to be expected that the best fit parameter values from our retrievals will depend on the noise distribution with wavelength, as well as on the individual random point generation for each simulated dataset. A more adequate instrument simulator will be needed to estimate the actual science return from a future mission.

5. ATMOSPHERIC RETRIEVAL SCHEME

The allowed ranges and best fit values for the forward model parameters, given the data, are determined using two Bayesian posterior sampling algorithms, namely the affine invariant ensemble Markov chain Monte Carlo sampler, *emcee* (Goodman & Weare 2010; Foreman-Mackey et al. 2013), and the multimodal nested sampling algorithm *MultiNest* (Feroz & Hobson 2008; Feroz et al. 2009, 2013). These approaches permit efficient sampling of highly correlated, non gaussian, and high-dimensional parameter spaces, and are very readily scaleable to multiprocessor computing.

The different approaches taken by the two algorithms in sampling the posterior parameter space can help us avoid the pitfalls of either one. While *emcee* starts with a first guess and can become trapped in a local minimum, MultiNest starts with a grid of points covering the entire prior parameter space and proceeds by narrowing down the maximum likelihood regions. On the other hand, *MultiNest* could favor highly-peaked, multimodal, Gaussian-like distributions, while *emcee* is more agnostic to the shape of the posterior and can reveal additional tails and correlations. The total evidence for any given model (the integral over the posterior distribution) is automatically calculated by *MultiNest* as a part of the algorithm, but requires extra steps and can be tricky to compute for *emcee*. Ideally, the two methods will converge to the same solution.

Overall, we consider the two approaches complimentary, and offer greater confidence in avoiding potential biases. Recently, Allison & Dunkley (2014) have compared in detail these sampling techniques and found that nested sampling is the most efficient while still providing good accuracy, and the affine-invariant MCMC sampler can be competitive when massively parallelized. They both outperform by far traditional Metropolis-Hastings algorithms. For completeness, we provide a brief description of the two posterior sampling algorithms in the Appendix.

A second component of the retrieval process consists of model comparison, with the purpose of quantifying not only the uncertainties in the model parameters, but also the evidence in support of a chosen model. In this step we can assess whether the 1-cloud or 2-cloud model presented in Section 3.1.1 and 3.1.2 offer a better representation of the data and calculate the significance associated with the cloud or methane detection. The choice between two competing models $\mathcal{M}_X$ and $\mathcal{M}_Y$ then comes down to comparing their probabilities by constructing the *Bayes factor*

$$B_{XY} = \frac{\mathcal{P}(\mathcal{M}_X \mid \mathcal{D})}{\mathcal{P}(\mathcal{M}_Y \mid \mathcal{D})} = \frac{\mathcal{Z}_X}{\mathcal{Z}_Y} \frac{\mathcal{P}(\mathcal{M}_X)}{\mathcal{P}(\mathcal{M}_Y)}, \quad (8)$$

where the relevant quantities are defined in the Ap-

pendix. Usually the last term in Equation 8 is 1 (both models have the same probability). We use the guidelines provided by Jeffreys (1961); Raftery (1996) for assessing the evidence in support of model $\mathcal{M}_X$ vs $\mathcal{M}_Y$ in terms of Bayes factor:

$$
\begin{aligned}
2\log B_{XY} < 0&: \text{ Negative (supports } \mathcal{M}_Y),\\
0 < 2\log B_{XY} < 2&: \text{ Inconclusive},\\
2 < 2\log B_{XY} < 5&: \text{ Positive},\\
5 < 2\log B_{XY} < 10&: \text{ Moderate},\\
2\log B_{XY} > 10&: \text{ Very Strong}.
\end{aligned} \tag{9}
$$

Since the posterior distribution in general does not have an analytic form, the difficulty arises when attempting to compute $\mathcal{Z}$ for each model under consideration. In general, the evidence evaluation from an existing MCMC posterior is limited by the poor sampling of regions of low likelihood. This problem can be overcome using thermodynamic integration, at computational costs $10 - 100\times$ higher than a regular MCMC (e.g. Trotta 2008; Calderhead & Girolami 2009). However, as long as the Bayes factor is found within the ranges in Equation 9, the precise value of B_XY is not important. In general, some rough assumptions are made on the functional shape of the prior and posterior distributions to be able to approximate the value of this integral. While these approximations are not very accurate, Cornish & Littenberg (2007) show that for high signal-to noise data (SNR $\gtrsim 9$) all methods converge toward the same values. In this paper we estimate $\mathcal{Z}$ using three different methods: the Schwarz-Bayes information criterion (BIC, Schwarz 1978), the Laplace approximation (Lopes & West 2004; Cornish & Littenberg 2007), and the Numerical Lebesgue Algorithm (NLA) described by Weinberg (2012). We refer the reader to the Appendix for a summary of these methods and relevant definitions. The scatter among the results given by these three methods are indicative of the reliability of these approximations for various models and SNR regimes. In general, we observe that the values converge when the evidence for a given model is very strong. Further, these results obtained from the MCMC samples are validated by comparison with the evidence values calculated by default with the nested sampling algorithm.

5.1. *Priors*

The parameters retrieved for each of the cloud models are described in Sections 3.1.1 and 3.1.2. In addition to the cloud properties, we are retrieving the methane abundance and surface gravity. Water and alkali abundances will be included as model parameters in future work; however, for the applications considered in this paper (e.g. Jupiter, Saturn), methane is the main absorber. We define the atmospheric methane mixing ratio, *fCH4*,

as the volume mixing ratio of methane. Since in a giant planet atmosphere 98% of the atmospheric constituents are H_2 and He, this uniquely defines the atmospheric methane content. Such an approach would not be possible for a terrestrial planet of course.

We allowed gravity to vary because in the realistic case neither the size of the planet nor the planetary mass will be known precisely. We allowed an exceptionally large range of gravities to be tested by the retrievals. In a realistic case the planet mass will be known to substantially better than a factor two by the orbital astrometry solution. From the mass radius relationship for gas giant planets and albedo scaling arguments the radius will likely be known to within 50%, which dominates the gravity uncertainty. Thus for a Jupiter twin the gravity ($g = 25$ m s^{-2}) would plausibly be known to be < 100 m s^{-2}, not < 1000 m s^{-2} as is the constraint placed in most of the results shown here. This turned out to be very important as, all else being equal, a large methane mixing ratio is required at high gravity to produce equivalent absorption band depths as a lower abundance at lower gravity.

The only restriction on the vertical cloud structure (P, dP_1, and dP_2) is that it does not exceed the total vertical extent of the atmosphere. The cloud albedos and asymmetry factor are allowed to take any value between 0 and 1, while the optical depth of the upper cloud varies between 10^{-3} and 10^3. This optical depth is also varied in the 1-cloud model, but the lower cloud in the 2-cloud model is assumed optically thick (see Section 3.1.2).

The pressure-temperature profile of the atmosphere is kept constant, since there is not information in the spectra at these wavelengths ($0.4 - 1.0$ μm) to constrain it. We are considering replacing this fixed profile by a parametrized one, to better account for the effect of surface gravity (Line et al. 2013).

5.2. *Implementation*

The forward models described in Sections 3.1.1 and 3.1.2 have been coded in Fortran and converted into a Python-callable library using f2py (now part of the NumPy package). The retrieval scheme integrates this library with either emcee or PyMultiNest, alternatively. Both MCMC and nested sampling version are set up for efficiency and can be scaled to run from a laptop to a computer cluster. The Fortran code is also parallelizable, but this does not provide a significant increase in speed as long as the MCMC is parallelized. Our retrievals were run on the *NASA Pleiades* Supercomputer, where we highly optimized the code for the forward models, and took advantage of the parallel nature of the algorithms to run on up to 216 processors at the same time (one 24-core node per model parameter). The MultiNest algorithm is found to converge rapidly even when run on

just 1-2 nodes.

We have quantified the methane and cloud detections by calculating the ratios of their respective Bayes factors, as described in Section 5. For each case (SNR and spectral correlation length combination), a set of four different forward models was used: the 2-cloud model with 9 parameters (Section 3.1.2), the 1-could model with 6 parameters (Section 3.1.1), a model without clouds (the cloud subroutines are turned off in the previous models), and a model without methane (the methane abundance is set to 10^{-20} in the previous models). Therefore, for each planet example, we ran a set of 24 retrievals using *emcee*. In addition, we performed the same retrievals using *MultiNest* for the models with a spectral correlation length of 25 nm mainly to cross-check the Bayesian evidence values calculated from the MCMC chains. In cases of good convergence, *MultiNest* also provided parameter constraints in agreement with *emcee* at a lower computational costs.

6. RETRIEVAL VALIDATION

In order to validate our retrieval procedure, we generate albedo spectra using the 1-cloud and 2-cloud models presented in Section 3.1.1 and 3.1.2, respectively. We use the 1-cloud forward model to generate 2 types of spectra: one for an optically thin cloud very deep in the atmosphere, equivalent to a cloud-free atmosphere; and one for an optically thick cloud at moderate height. The third case is generated with the 2-cloud model. The model spectra are then converted to simulated observations using the noise prescription described in Section 4. For each of these three cases we investigate the ability to retrieve the input model parameters, as a function of SNR and noise correlation length. For each of the three cases we ran retrievals using the full 1-cloud and 2-cloud models, a forward model with the clouds turned off (referred to as "no clouds"; defaults to 0 for all $\bar{g}$, $\bar{\omega}$, and τ's), and a forward model with negligible methane abundance (referred to as "no methane", $fCH4 = 10^{-20}$). For convenience of notation, we will refer to these four model retrievals as **1c**, **2c**, **-c**, and **-m**, where a **2c-m** notation for example would stand for "2-cloud forward model without methane". Each SNR and spectral noise correlation length combination was run through the retrieval procedure four times to enable model comparison and assess the significance of methane and cloud detection. Tables 1 and 2 summarize the input parameter values for each of the simulated spectra, and the confidence intervals for each parameter obtained after running the retrieval procedure.

6.1. *Cloud-free case*

We construct the albedo of a cloud-free planet using the 1-cloud model in Section 3.1.1, where the optical depth τ is set to 10^{-8} and the top pressure of the cloud to 10 bar. The other parameters used to generate the model spectrum are listed in Table 1. Using the noise prescription in Section 4, we generate simulated datasets for SNR values of 5, 10, and 20, and spectral noise correlation lengths of 25 and 100 nm. The data realizations can be seen in the left panel of Figure 8. The retrieval is performed over the wavelength range 0.4-1.0 μm, indicated by the green line in Figure 8. Figures 6 and 7 show the retrieval results. The marginal probability distributions for the model parameters are shown in the top panel in Figure 6. The associated confidence intervals are bounded by the 16% and 84% quantiles of the cumulative probability distributions and are shown in the bottom panel of the same figure. These confidence intervals are also listed in Table 1.

We find that for a cloud-free planet both the methane abundance $fCH4$ and surface gravity g are well constrained. The methane abundance is constrained to within a factor of ~ 2.6 at a SNR of 5 and within a factor of ~ 1.15 at a SNR of 20. The surface gravity is constrained to within a factor of ~ 4 at a SNR of 5 and within a factor of ~ 1.2 at a SNR of 20. As expected, the cloud albedo $\bar{\omega}$ and scattering asymmetry factor $\bar{g}$ are not constrained, since they do not contribute to the observed spectrum.

The 2-dimensional posterior probability distributions shown in Figure 7 trace the changes in the parameter constraints as the SNR increases from 5 to 20. This is also reflected by the decrease in the size of confidence intervals shown in the bottom panel of Figure 6. The distributions clearly become narrower and more peaked as the SNR increases. This projection also shows that the pressure of the top of the cloud deck in the model is partly correlated with the optical depth τ. A larger top cloud pressure (deeper cloud) allows for a larger range of optical depths. This can be intuitively understood since a deep cloud will have little effect on the observed spectrum even when its optical depth is larger. The range of spectra obtained using parameters drawn from the posterior probability distributions are shown by the red contours in Figure 8. We also note the excellent agreement between the MCMC and nested sampling methods, where the nested sampling results are shown by the black-grey contours in Figure 7, and by the yellow lines in Figure 6.

The posterior constraints on the cloud parameters P, τ, $\bar{\omega}$, and $\bar{g}$ already indicate that the spectrum does not support the presence of an observable cloud. This is further confirmed by the Bayesian evidence analysis. We sample the posterior probability distributions for a set of 4 models: **1c**, **1c-m**, **1c-c**, and **2c**, as defined above. The pairwise Bayes factors for these models are shown in Figure 9. Clearly, methane is detected with a high significance even when the spectral SNR is 5 (yellow trian-

gles). However, the presence of a cloud is not supported. The models containing 2-clouds, 1-cloud, or no clouds are equally able of describing the data, since even in a multiple cloud model the optical depth of the clouds can be very low, effectively acting as a no-cloud model. No preference for a given cloud model in this case means that the presence of a cloud is not necessary to explain the observed spectrum. In this sense, the Bayesian evidence for all these models should be approximately equal, and the scatter in the Bayes factors in Figure 9 shows the poor performance of the evidence approximations when the significance is low. The black symbols in Figure 9 show the Bayes factors obtained using the evidence calculated by the nested sampling algorithm. The agreement is excellent for the high-significance methane detection, but lays within the large scatter for the cloud-model comparison.

6.2. *Single-cloud case*

By raising the optical depth τ to 1, and the cloud top pressure to 0.2 bar, we can use the 1-cloud model to generate the albedo spectrum of a planet with an observable cloud deck. The simulated observations of such a planet are shown in the middle panel of Figure 8. The results of this retrieval are shown in Figures 10 and 11, and in the bottom half of Table 1. In this case the methane abundance is still well constrained, although within a wider range than for the no-cloud case, namely within a factor of ~ 5 for a SNR of 5 up to within a factor of ~ 3 for a SNR of 20. The original abundance value is well within the predicted ranges, where the SNR=10 case with a correlation length of 100 nm seems to be an outlier.

The surface gravity of the planet is no longer constrained in this case, but is found to be correlated with the top cloud pressure (Figure 11). The power of the posterior sampling lays in discovering such correlations between model parameters. Figure 11 also shows the correlation between the cloud albedo $\bar{\omega}$ and scattering asymmetry factor $\bar{g}$, and between the top cloud pressure and its optical depth. Essentially, and optically thick cloud also constrains the cloud top pressure between ~ 0.01 and 1 bar, while an optically thin cloud would require the cloud top pressure to be very close to the top of the atmosphere. Independent constraints on the surface gravity, such as provided by RV measurements would narrow the allowed range for the cloud top pressure, which in turn would constrain the cloud optical depth. Lacking this information, we essentially obtain a lower limit for the optical depth and an upper limit for the top cloud pressure.

The other very well constrained parameter is the cloud albedo $\bar{\omega}$. The confidence intervals on this parameter are only of the order $\pm 5\%$ to 2% depending on the SNR and particular noise realization. The correlation with the

scattering asymmetry factor leads to a slight asymmetry in these confidence intervals, but the range of allowed values is still remarkably narrow. Just like the surface gravity, the scattering asymmetry factor $\bar{g}$ is virtually unconstrained. Similarly to the no-cloud case, there is excellent agreement between the MCMC and nested sampling results.

The high-significance cloud detection is revealed in the Bayes factor plot in Figure 12. The Bayesian evidence is calculated for the posterior distributions corresponding to the models **1c**, **1c-m**, **1c-c**, and **2c**. The Bayes factors favor the models with 1 cloud relative to the one without clouds (blue circles), and the model with methane relative to the one without (yellow triangles). The cloud detection significance is $> 10\sigma$ even when the data has a SNR of 5, showing that the presence of a cloud deck is required by the observations. The methane detection significance is similar to that in Section 6.1. Similarly, the retrieval cannot distinguish between a 1-cloud or a 2-cloud model (green stars), since a 2-cloud model can be reduced to a 1-cloud model as the gap between the 2 cloud decks becomes small and the optical depth of the top cloud becomes large.

6.3. *Two-cloud case*

The final validation case consists of a spectrum generated using the 2-cloud model in Section 3.1.2. The input parameters for the original spectrum are listed in Table 2, and the simulated datasets are shown in the right panel of Figure 8. The retrieved marginal probability distributions and confidence intervals are shown in Figure 13. In this case, the uncertainty in the methane abundance does not shrink considerably before the SNR reaches a value of 20. The confidence interval for fCH_4 extends over a factor of ~ 30 ($\sim 60 - 70$ for nested sampling) when the SNR is 5-10, but drops to a factor of 2 when the SNR reaches 20. Similarly to the 1-cloud case, the surface gravity is not constrained by the data. The multi-dimensional correlation between fCH_4, P, and g seen in Figure 14 (at SNR=10) shows the benefit in reducing the allowed range in g, via RV and astrometry measurements, which will then propagate into narrowing the allowed range in P, and further in fCH_4. For a SNR=20 dataset, the uncertainties in fCH_4 and P are simultaneously reduced (Figure 14). In this case, the top pressure of the bottom cloud (P) is also constrained to within a factor of ~ 3.

In terms of cloud properties, the scattering asymmetry factor $\bar{g}$ of the upper cloud and its albedo $\bar{\omega}$ are completely unconstrained, while the uncertainty for the albedo of the lower cloud ($\bar{\omega}_2$) is only of the order of 1% even when the data has a SNR of 5. The MCMC algorithm places an upper limit on the optical depth of the upper cloud, which is consistent with the lack of

constraints for the other upper cloud parameters, but very tight constraint on the bottom cloud albedo. Intuitively, as seen in the previous two examples, the parameters of the upper cloud can be constrained as long as this cloud is optically thick, while the properties of the lower cloud (its albedo) can be determined as long as the upper cloud is optically thin. However, especially at lower SNR (see Figure 14), the nested sampling algorithm identifies a second set of solutions, with an optically thick upper cloud, associated with a lower methane abundance and a deeper lower cloud. This result suggests that this degeneracy will not be broken unless the scatter in the data points is greatly reduced. Aside from this new mode identified by the nested sampling algorithm, the two Bayesian approaches are again in excellent agreement. The presence of the second mode can be further investigated by starting the MCMC chains in this part of the parameter space.

We have calculated the Bayes factors and compared the models **2c**, **1c**, **2c-c**, and **2c-m**. Similar to the 1-cloud case, the presence of methane and cloud deck are detected at very high significance ($\sigma > 4$) even for a dataset with a SNR of 5, as shown in Figure 15. The results are again unable to distinguish between a 1-cloud and a 2-cloud model, since the first is a special-case limit of the second (green stars). However, both the 1-cloud and the 2-cloud models are equally favored with respect to any cloud free model (blue circles, pink triangles).

6.4. *Importance of SNR and Spectral Noise Correlation Length*

We stress that the quoted significance of the detection itself has no other information on the confidence intervals associated with the model parameters. These confidence intervals, as well as possible correlation and multi-modality, are clearly affected by the SNR of the dataset. The change in the confidence intervals with SNR is shown in Figures 6, 10, and 13. Overall, while the presence of methane is clearly *detected* even at a SNR of 5, its *abundance* is well constrained (to within factors of 2-3) only at a SNR of 20. At lower SNR, the uncertainty in the methane abundance is mainly related to correlations with other models parameters, such as the surface gravity and the position of the cloud deck (P). This situation is improved in the case of a clear atmosphere, where the methane abundance and surface gravity are simultaneously constrained. However, the *presence* of a cloud deck is easy to confirm even at a SNR of 5 (as shown by the Bayes factor plots). This suggests that when the presence of clouds is indicated by early observations, an attempt to further increase in SNR is justified in order to constrain the methane abundance.

Our results do not indicate any influence of the spectral noise correlation length on the retrieval results. The uncertainties on the model parameters are similar (see Figures 6, 10, and 13, and Tables 1 and 2). There is a slight bias towards higher values for the retrieved methane abundance in the no-cloud and 1-cloud cases, for a spectral noise correlation length of 100 nm, but it is not clear whether this is an effect of the noise correlation length scale or of the particular noise realization in the simulated dataset. Multiple noise realizations for a given correlation length scale would be required to validate this effect.

7. REALISTIC TEST CASES

For the retrieval tests we used two types of input data, solar system giants and model planets. We used the solar system albedo spectra for Jupiter and Saturn from Karkoschka (1994), and a theoretical radiative-convective equilibrium model for HD 99492 c. All of these objects have methane dominated optical reflection spectra. We have applied our albedo retrieval method to a set of 24 cases, comprising 6 combinations of SNR (5, 10, 20) and correlation lengths (25 and 100 nm), the same as for the validation cases. The Solar System-like planets are assumed to be at 25 pc from the Earth, while the distance to the HD 99492 c system is 18 pc. The retrievals use data between 0.6 and 1 μm to more closely match the projected bandpass of *WFIRST* (unlike the validation cases where we used the 0.4-1.0 μm bandpass). For each case we run the MCMC ensemble sampler with 24 walkers per parameters, for a total of 3800 steps, and we select the last 400 steps for determining the posterior probability distributions. We also use the nested sampling algorithm for the spectra with noise correlation length of 25 nm. The pressure-temperature profiles are kept fixed during the retrievals, as shown in Figures 19, 25, and 29.

7.1. *HD 99492 c*

We start by looking at the model planet HD 99492 c, as the real-world example most closely resembling our 1-cloud model. HD 99492 c is thought to be a gas giant with a mass of 0.36 ± 0.02 M$_{Jup}$, and a semimajor axis of 5.4 ± 0.1 AU, orbiting a K2V star. However, its detection has been challenged recently due to high stellar activity (Kane et al. 2016).

We first determined the pressure-temperature profile for HD 99492 c by computing a 1D radiative-convective equilibrium model with a comprehensive cloud treatment built-in (Ackerman & Marley 2001). This code computes a self-consistent cloud with vertically varying abundances and particle sizes of each condensible species. This theoretical structure is shown in the right-hand panel in Figure 19. We then input the resulting pressure-temperature profile into a fine-grid albedo code to produce an albedo spectrum comparable to the solar system data. This high

resolution spectrum is then converted to simulated data following the prescription in Section 4, for each chosen combination of SNR and noise correlation length.

Figure 16 shows the summary of the retrieval results for the gas giant HD 99492 c, with the quantiles listed in Table 3. An example for the posterior probability distributions for the retrieval using the 2-cloud model is shown in Figure 17. In the 2-cloud scenario, the posterior in bi-modal, similar to that found in Section 6.3, and we show the most important parameters for the two modes separately in the two panels. The notable difference is that for the mode with a *low* optical depth for the top cloud (τ), the albedo of the bottom cloud ($\bar{\omega}_2$) is very well constrained, while for the mode with a *high* optical depth for the top cloud, the albedo of the top cloud ($\bar{\omega}$) is very well-constrained, to within $\sim 6\%$. This is easily understood, since in the case of low optical depth we can "see through" the top cloud, and the albedo of the bottom cloud surface is what determines the spectrum, while the opposite is true when the top cloud is optically thick.

We also note that an optically thin top cloud favors a lower methane abundance, since now we integrate over a larger CH_4 column. The position of the *best fit* parameter values for each mode was marked in green to emphasize that the best fit parameter *combination* is different from the set of median values of the marginal distributions, which are listed in Table 3. The range of spectra generated using random parameter sets from the posterior are shown in Figure 20.

In Figure 18 we show both the covariance plot for the retrieval using the 1-cloud model, as the more representative for the planet's vertical structure, and the best-fit spectra for the different models and modes. In the covariance plot the blue lines show the parameter values that are closest to the theoretical planet structure. We note that this 1-cloud retrieval solution resembles the high-τ mode of the 2-cloud posterior, only with a tighter correlation between P and g. In this case we find a lower bound for the pressure of the cloud surface, but a lack of constraints for g. Similar to the validation case, we can see that a tighter prior in g would translate into better limits on P (via correlation), and a narrower allowed range for fCH_4. The best-fit spectra reveal the complete degeneracy of these solutions (red, blue and yellow lines overlapping). The differences between the retrieved and original spectra (black line) are due to a more comprehensive treatment of gas and cloud opacities in the original model. Additional constraints placed by available photometric points shortward of 0.6 μm will be investigated in future work.

The degeneracy between the best-fit solution given by the 2-cloud and 1-cloud models is also apparent in Figure 19, where the two cloud decks in the left panel over-

lap, within the error bars, and basically occupy the same vertical regions as the 1-cloud deck in the middle panel. This plot suggests that for a planet like HD 99492 c our simple cloud model can only provide a lower bound on the pressure at the top of the cloud deck (i.e. upper bound to the height in the atmosphere above the planet's surface) and an upper bound on the methane abundance (i.e. the methane abundance is inversely correlated to the cloud top pressure, such that the total CH_4 column is constant). Independent priors on the top cloud pressure (from equilibrium structure) and surface gravity (from radius and mass measurements) would help mitigate these uncertainties. We note a similar bias toward high fCH_4 values in the case of Saturn below, which could be due to some deficiencies in our simplified cloud model and will be investigated in future work.

As before, we show the Bayes factors between different model choices in Figure 21. The presence of methane and a cloud deck is confirmed at very high significance. The 2-cloud model is more disfavored relative to the 1-cloud model, likely due to the presence of additional unnecessary parameters.

7.2. *Jupiter*

Arguably, a Jupiter-like planet is the closest real-world case to our 2-cloud forward model. We have simulated data for a Jupiter-like planet at 25 pc from the Sun using the spectra from Karkoschka (1994). The results of our retrievals are shown in Figure 22. This plot shows that the parameters that are best constrained by the data are fCH_4, P, and $\bar{\omega}_2$. We note the narrowing of the distributions and therefore the tightening of the constraints for SNR=20 (red lines), also shown by the size of the confidence intervals in the bottom plot. The derived CH_4 abundance is consistent with the generally adopted value of $(2.37 \pm 0.57) \times 10^{-3}$ (or -2.625 in log) in Jupiter (Wong et al. 2004). Also, the derived single scattering albedo of the lower cloud, $\bar{\omega}_2$, matches the observed value of 0.997 (e.g., Sato & Hansen 1979). The mean vales of these parameters are sensitive to the particular noise realization of each simulated dataset. Unconstrained parameters are g and $\bar{g}$, and an upper limit is derived for τ, showing that the upper cloud is likely optically thin. The confidence intervals are summarized in Table 4, and the range in spectra allowed by the posterior samples are shown in Figure 20.

Although the MCMC algorithm strongly favors a single-mode posterior with an optically thin upper cloud, the nested sampling algorithm identifies two posterior modes, the second one having an optically think upper cloud. This is reflected by the large confidence intervals shown in Figure 22 (yellow). The second, high optical depth mode, becomes favored by the nested sampling algorithm at SNR=20. Figure 23 shows posterior co-

variance plots for some selected parameters for SNR=20, and noise correlation length 25 nm Jupiter data, using both the 2-cloud and 1-cloud models. The blue lines indicate the parameter values that correspond to currently adopted values for Jupiter ($fCH_4 = 2.37 \times 10^{-3}$ and $g = 24.79$ m s^{-2}). It is apparent that the nested sampling (gray contours) favors a solution that best resembles the 1-cloud model, with a deep, optically thick cloud and unphysically low gravity. The correlations at the top of left panel in Figure 23 show that a narrower allowed range in g for known RV planets both constrain the methane abundance to match the real value and strongly disfavor the second, optically thick mode. The spectra corresponding to these best-fit solutions are shown in Figure 24. This plot shows that the spectra are degenerate relative to these solutions at wavelengths between 0.6 and 1 μm, but physical arguments can be made to eliminate certain solutions. We note the need for wavelength-dependent continuum opacity, especially for using photometry data shortward of 0.6 μm.

The Jupiter cloud structure, as retrieved by our 2-cloud and 1-cloud models is compared to the theoretical vertical structure for Jupiter in Figure 25. The cloud and haze layers shown in the right panel of Figure 25 approximately match the positions described elsewhere in the literature (e.g., Simon-Miller et al. 2001; Sato et al. 2013). The hazes are likely to have a wavelength-dependent continuum opacity, unlike our simple cloud model, and our notation was chosen to emphasize that the upper haze layer is likely absorbing and the lower haze/cloud layer is likely bright (reflective) at the wavelengths relevant in our study. We note that the upper cloud roughly matches the position of a hydrocarbon haze in the upper layers of the atmosphere, and the lower cloud deck overlaps with the bright haze and ammonia/water ice clouds in the deeper atmosphere. This deep cloud is also identified by the 1-cloud model retrieval, but without the opacity contribution of the upper haze/cloud, the retrieved suface gravity of the planet would be unphysical.

The significance of the cloud and methane detection is shown in the middle panel of Figure 21. The methane is detected at high significance for all SNR, while the cloud detection becomes *very strong* only when SNR>10. Due to the degeneracy of the solutions (see Figure 24), the Bayes factor does not favor the 2-cloud vs. the 1-cloud model except at very high signal-to-noise. However, based on the previous arguments related to the surface gravity, it is reasonable to select the 2-cloud model in this case, and we expect a more clear distinction to appear once independent constraints on the surface gravity are provided.

We conclude that the two-layer cloud model is necessary for Jupiter, constraining the methane abundance to within factors of ~ 20 at SNR=5 and factors of ~ 3 at SNR=20, possibly much better when tighter limits on the surface gravity are available. The single scattering albedo of the lower cloud is constrained within 0.5% even at the lowest SNR. This gives us an indication for the composition of the lower cloud, since particles with high reflectivity are necessary to explain the large value of $\bar{\omega}_2$.

7.3. Saturn

Our third and final case study is Saturn, which falls between HD 99492 c and Jupiter in terms of retrieval results. We use again data from Karkoschka (1994) to generate simulated observations using the method in Section 4. The summary plots for the retrieval results are shown in Figure 26, with the confidence intervals listed in Table 5. The posterior distribution for the 2-cloud retrieval is now clearly bi-modal, with one mode corresponding to a low optical depth for the upper cloud, and the other to an optically thick upper cloud. The large confidence intervals plotted in the bottom panel of Figure 26 are due to this bi-modality. The range of the possible spectra with parameters drawn from the posterior are shown in the right panel of Figure 20.

For clarity, the two modes have been separated and the covariances of the most relevant parameters shown in Figure 27 (middle and right panels). In the left panel of Figure 27 we show the retrieved posterior distribution for the 1-cloud forward model, with the blue lines indicated the parameter values that correspond to the currently known properties of Saturn ($fCH_4 = 4.5 \times 10^{-3}$ and $g = 10.44$ m s^{-2}). The green lines in the middle and right panels show the *best fit* solutions for each of the two modes. As seen in the case of HD 99492 c, the mode with low optical depth constrains the albedo of the lower cloud ($\bar{\omega}_2$), while the optically thick mode constrains the albedo of the upper cloud ($\bar{\omega}$). However, in contrast to HD 99492 c, the 1-cloud retrieval mostly resembles the *low optical depth* mode of the 2-cloud retrieval. In this case, the reflecting surface (P) is found relatively high ($10^{-3} - 1$ bar), with a position correlated with the methane abundance and g. The 1-cloud model also constrains the optical depth within a relatively narrow range of $\sim 0.1 - 1$. The surface gravity g is unconstrained by both the 1-cloud and 2-cloud retrievals, but independent constraints would translate into narrower confidence intervals for both P and fCH_4, as in the cases described above, especially considering the low optical depth mode. A more peaked distribution for fCH_4 is only obtained for the 2-cloud mode of low optical depth (right panel), while in the other two cases only lower limits can be inferred. The methane abundance is overall consistent with measured values, but biased towards higher values in the high optical depth mode, because the entire cloud structure is then obscuring most of the atmosphere.

Figure 28 shows the complete degeneracy between the 1-cloud retrieved solution and the two modes of the 2-cloud retrieval. Photometry shortward of 0.6 μm could be helpful for constraining haze properties. Based on these data, we cannot distinguish between the two possible modes, and the presence of the second cloud is not required. The retrieved cloud structure using the 1-cloud and 2-cloud models is presented in Figure 29 and compared with the structure derived from the literature in the right panel (e.g., Roman et al. 2013). The lack of evidence for a second cloud is also suggested by the overlap of the 2-cloud structure in the left panel, similar to the situation for HD 99492 c. By contrast, the cloud optical depth is low in this case, and therefore the transition from a clear atmosphere to cloud-obscured is very gradual. Overall, the retrieved cloud structure strongly overlaps with the theoretical structure, and all solutions are consistent with highly reflective layers present in the atmosphere. This is supported by the Bayes factors in the right panel of Figure 21, where both methane and a cloud layer are detected with high significance for all SNR. The evidence for the second cloud is inconclusive, since these solutions are degenerate. We suggest that some evidence is provided by the tighter distribution in Figure 27, right panel vs. left panel, and a more relevant Bayes factor calculation would be between the 1-cloud model and each of the two modes of the 2-cloud model separately.

8. SUMMARY AND CONCLUSIONS

We have used a Bayesian retrieval method to quantify the confidence intervals on the atmospheric methane abundance and cloud structure of extrasolar giant planets, using a simple atmospheric model with either 1 or 2 cloud decks. The results show that the modeling framework can be further refined and optimized to address realistic planets. Notably, we are trying to reproduce complex atmospheric structures by using simple 1-dimensional model approximations and low signal-to-noise, integrated light data. Nevertheless we find that reflected light spectra of the quality expected from a space based direct imaging exoplanet mission is sufficient to place interesting constraints on important planetary atmosphere characteristics, particularly methane mixing ratio and, in some cases, cloud albedo. In particular, the presence of clouds and/or methane absorption is detected at high significance even for a SNR of 5. However, a low SNR value will only give us such simple assessments, while a higher SNR is needed to determine accurate abundances and extracting useful information about mass-metallicity relationships. The retrieval methods presented are powerful for determining correlations among parameters and identifying which ones are unconstrained by the data. In this case demonstrating

the value in the synthetic datasets, even at low signal to noise ratios.

We found that our retrieval methods could reliably infer methane abundances to within factors of ten of the true value in most cases and could often accurately constrain cloud scattering properties, thus providing a clue to the cloud composition. Gravity, however, is not well constrained by optical spectra in the presence of clouds. Observing planets with known masses therefore removes an important source of uncertainty and allows much greater precision in the inference of atmospheric abundances. Furthermore, cases in which the cloud model was inadequate are readily apparent in the retrieval output. This prompted us to calculate the Bayesian evidence for a set of models for each simulated spectrum. This is a method to quantify the significance associated with the methane and cloud detection, and the assumed cloud model (1-cloud vs. 2-cloud) in each case. Although time-consuming, this is a very powerful test that will become a necessity for interpreting future observations, as the complexity of our model atmospheres and understanding of planetary diversity is increasing. Our preliminary applications on realistic planets show that it is worthwhile to investigate different vertical cloud structures, such as the 1-cloud vs. the 2-cloud models. This can help us address degeneracies and identify unnecessary parameters. In summary, our first study on the characterization of extrasolar giant planets in reflected light found that retrieval methods using simple, gray cloud models can be applied to optical spectra of exoplanets to retrieve molecular abundances and cloud properties. We found that generally the retrieval results are equally sensitive to the particular noise realization as to the chosen spectral correlation length.

8.1. Ongoing and Future Work

For this initial study we made a number of simplifications to the analysis to make our task tractable and obtain a first look at parameter correlations. However future work should address these simplifications and their roles in the fidelity of the retrievals. Foremost among those that should be explored include: planetary radius uncertainty, thermal profile uncertainty, and orbital phase uncertainty. The second paper in this series (Nayak et al, in preparation), will address the radius and phase uncertainties. In addition the retrieval of more atmospheric abundances should be explored, particularly water and alkali gasses. We will also investigate the possibility of adopting a somewhat more general cloud model.
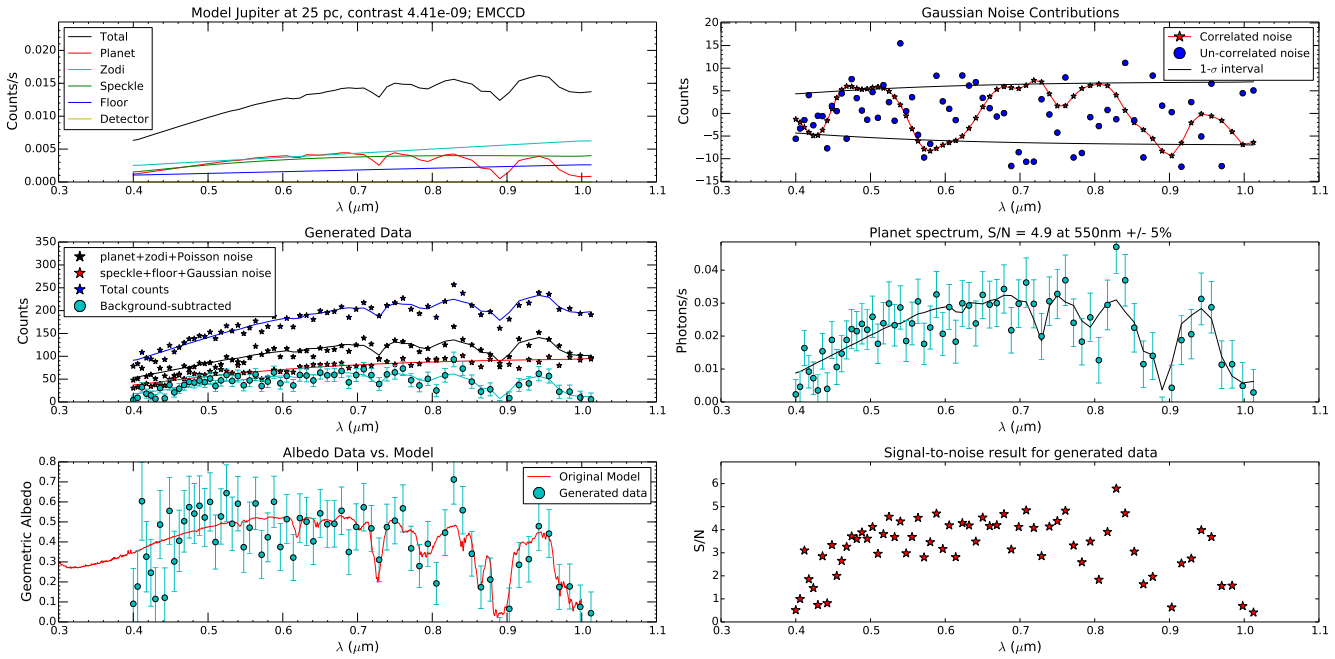
Figure 5. In each set, the panels are as follows: Top left expected count rates from all different sources: planet red, zodi- cyan, speckle-green, detector noise blue and yellow (to small to see). The total count rate is shown in black. Middle left total number of counts after calculating the integration time needed to get a SNR of 5. The model counts are solid lines, and the simulated data are stars. Top right correlated and un-correlated noise contributions. These are added-in when generating the red star data points in the middle-left panel. Middle right simulated data converted to photon rate, after background subtraction (cyan), compared to the input model (black). Bottom left simulated data converted back to geometric albedo, after division by the stellar spectrum (cyan), vs actual albedo of Jupiter (red). Bottom right SNR of the simulated data in each wavelength bin. The nominal SNR (5) corresponds to a 10% band around 550 nm.
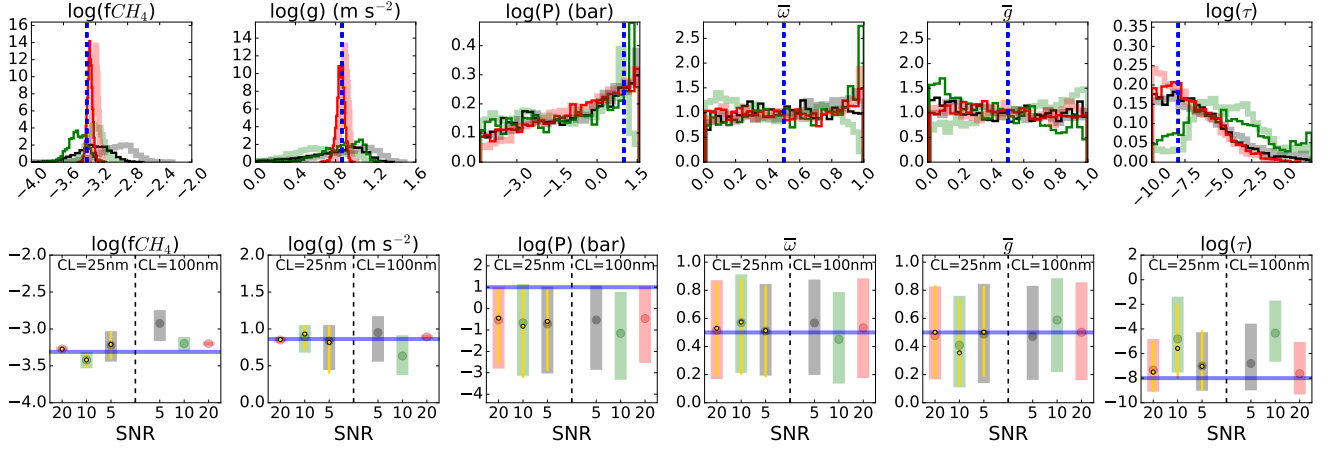
Figure 6. Upper: 1-D marginal likelihood distributions for the six parameters in the 1-cloud model (**1c**) for the cloud-free case in Section 6.1. The SNR values are color-coded, with black, green, and red for SNR 5, 10, and 20, respectively. The thin solid histograms show the distributions corresponding to a noise correlation length of 25 nm, and the thick semi-transparent ones for a noise correlation length of 100 nm. Lower: Confidence intervals for the model parameters retrieved using MCMC. The color coding matches the upper panel, the yellow lines show the 1σ intervals from the nested sampling retrievals, and the blue horizontal line show the input parameter value in the original albedo model. The 2 spectral correlation lengths are in labeled in the left/right parts of the plots. These values are also summarized in Table 1. Note that the confidence intervals are calculated from the distribution quantiles, and do not reflect possible upper/lower limits or unconstrained parameters that can be seen in the histograms.
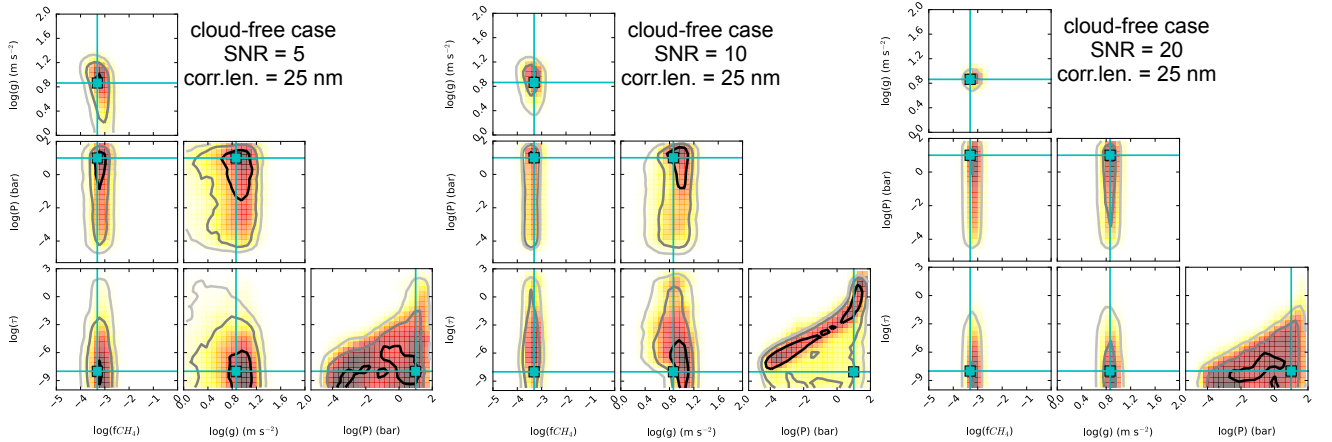
Figure 7. 2-D marginal posterior probability distributions for SNR=5, 10 and 20, and spectral noise correlation length of 25 nm, for the cloud free case in Section 6.1, using the **1c** forward model. Since the $\bar{g}$ and $\bar{\omega}$ parameters are unconstrained in this case, we only plot the remaining ones. The red color map corresponds to distributions obtained using the MCMC algorithm, and the black/gray contours to nested sampling. The teal lines show the real solution.
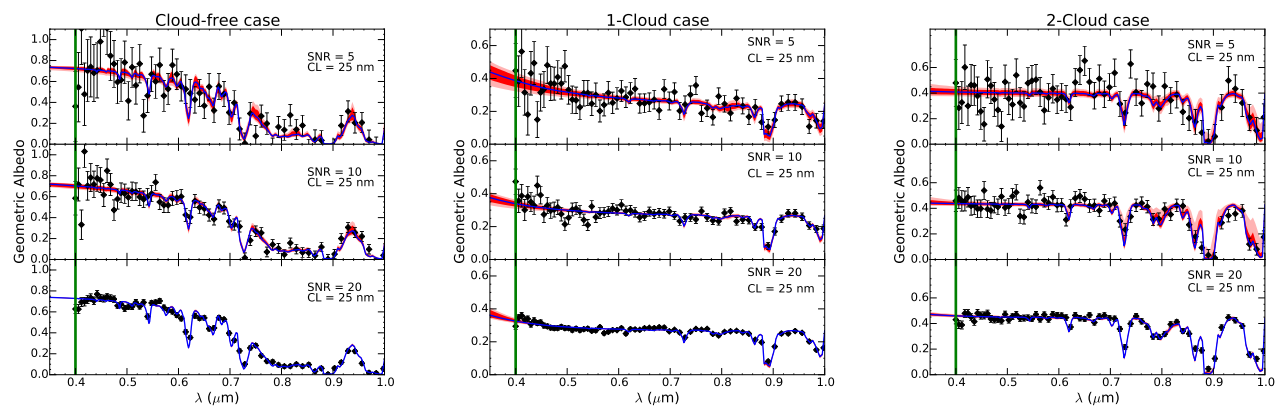
Figure 8. Simulated data and best fit spectra for the cloud free case in Section 6.1 (left) and the single cloud case in Section 6.2 (middle), using the **1c** forward model, and for the for the 2-cloud case in Section 6.3 (right), using the **2c** forward model. The data correspond to SNR=5, 10, 20, from top to bottom and a spectral correlation noise of 25 nm. The results for a correlation length of 100 nm are similar. The solid and semi-transparent red regions represent $1 - \sigma$ and $2 - \sigma$ intervals, respectively. These intervals represent the standard deviation a set of 500 spectra generated using random samples from the converged MCMC distribution. The blue line represents the median of this set. The retrieval was performed over the $0.4 - 1.0$ μm region, as indicated by the green vertical line.
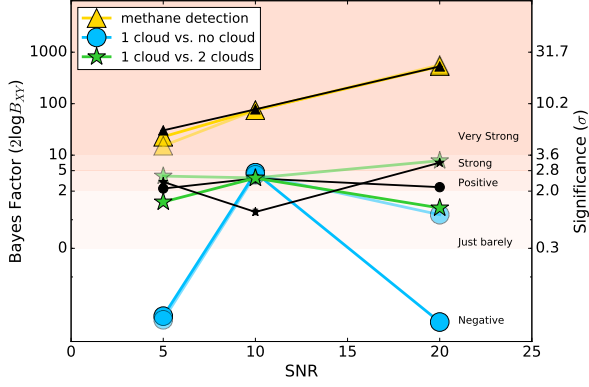
Figure 9. Bayes factors and associated significance levels, as defined in Section A.1, for the cloud free case in Section 6.1. The vertical shading grades follow the intervals defined in Equation 9. The yellow triangles correspond to the ratios $\mathcal{Z}_{1c}/\mathcal{Z}_{1c-m}$, the blue circles to $\mathcal{Z}_{1c}/\mathcal{Z}_{1c-c}$, and the green stars to $\mathcal{Z}_{1c}/\mathcal{Z}_{2c}$. The colored symbols represent the results derived from the MCMC samples, with the solid color corresponding to a noise correlation length of 25 nm, and the semi-transparent to a noise correlation length of 100 nm. For comparison, the black symbols use the evidence values provided by the nested sampling algorithm for the cases with a noise correlation length of 25 nm. The symbols correspond to the same Bayes factors shown in color. The values calculated using nested sampling have associated error bars, but too small in general to see on this plot.
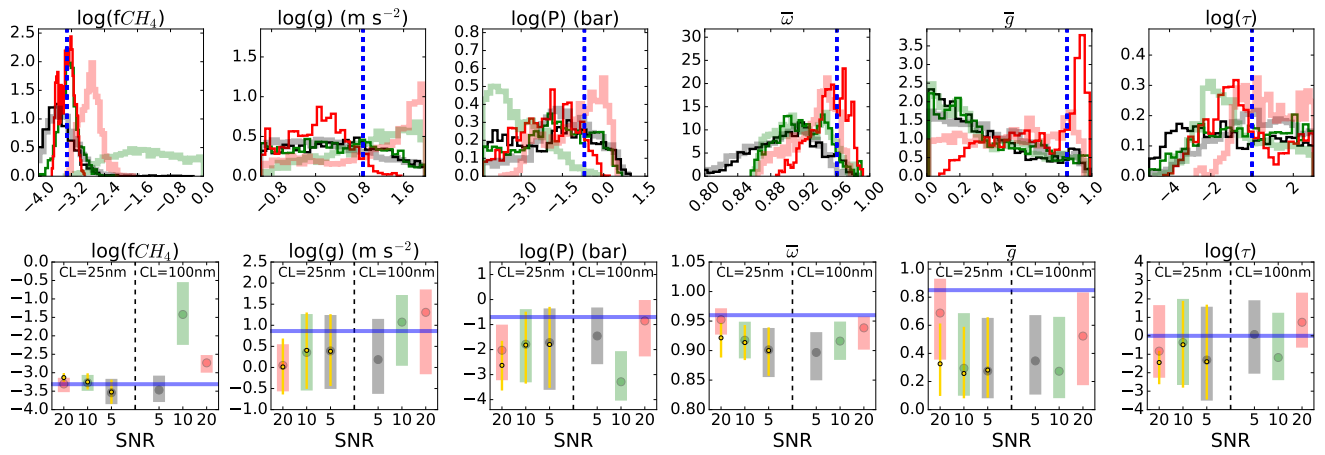
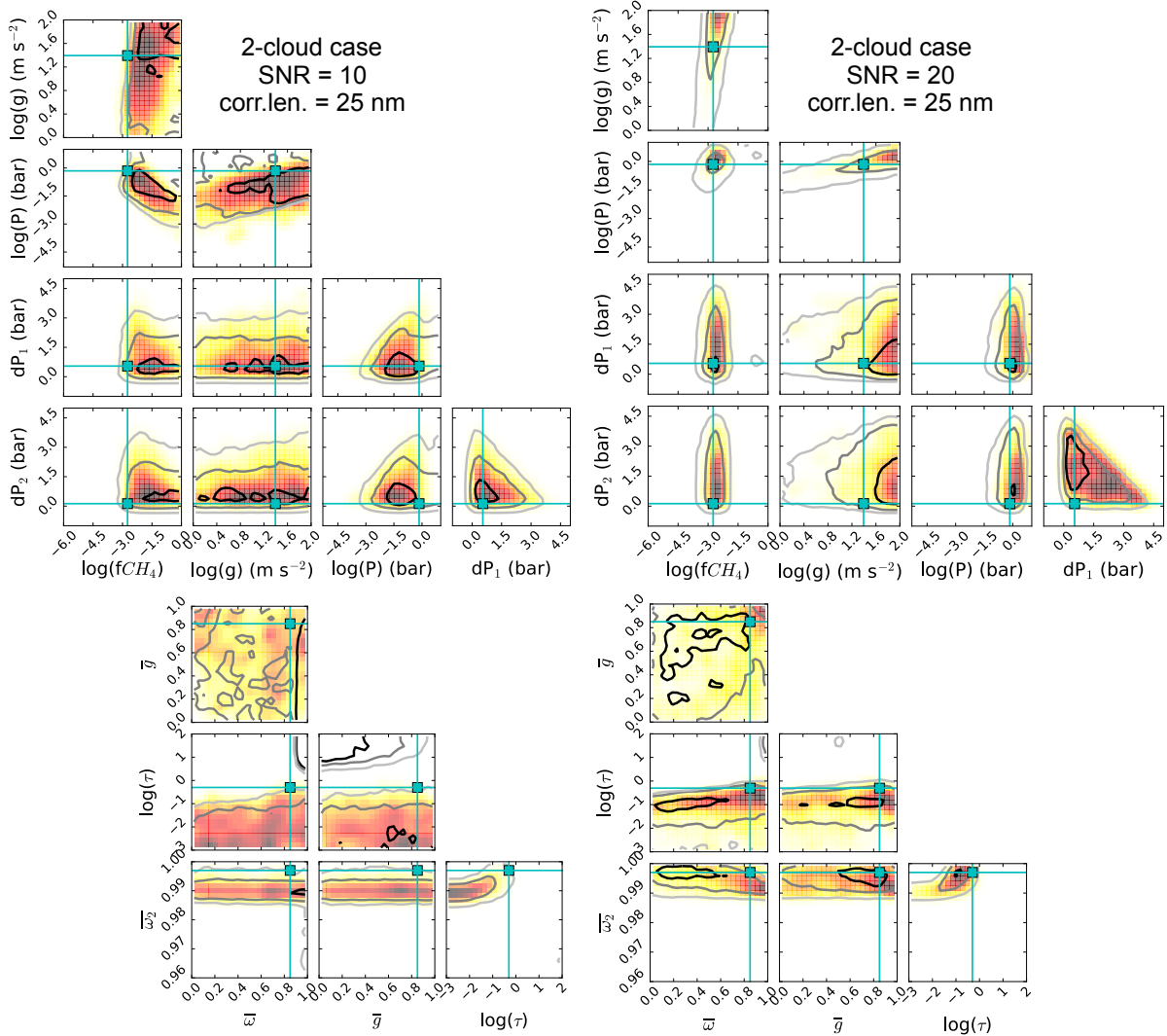**Figure 10**. Same as Figure 6, for the 1-cloud case in Section 6.2.
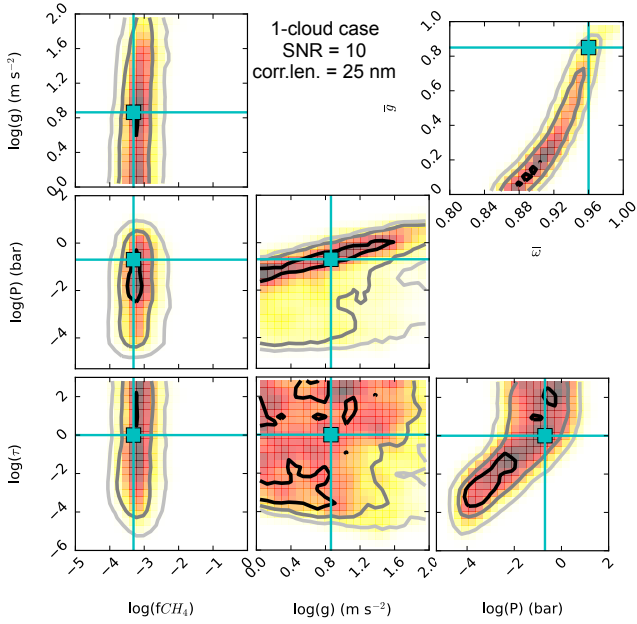
Figure 11. Sample 2-D marginal posterior probability distributions for SNR=10 and spectral noise correlation length of 25 nm, for the single cloud case in Section 6.2, using the **1c** forward model. The red color map corresponds to distributions obtained using the MCMC algorithm, and the black/gray contours to nested sampling. The teal lines show the real solution.
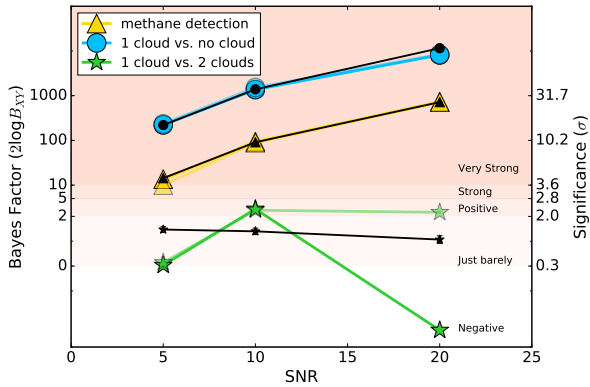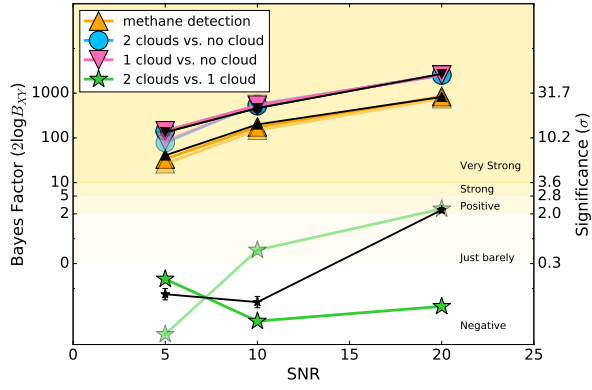

Figure 12. Same as Figure 9, for the 1-cloud case in Section 6.2. In this case, there is no ambiguity in model selection with a cloud clearly detected at $\sim 20\sigma$ significance even when the SNR of the input data is only 5.
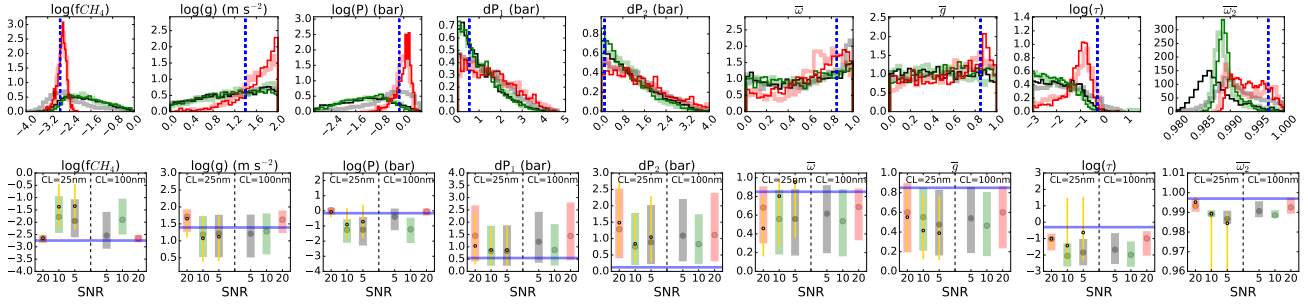
Figure 13. Similar to Figure 6, for the 2-cloud case in Section 6.3. The parameters correspond the the 2-cloud model (**2c**) in Section 3.1.2. The 1σ intervals obtained using nested sampling can be affected by possible bi-modal distributions (see also Figure 14).


Figure 14. Sample 2-D marginal posterior probability distributions for SNR=10 and 20, and spectral noise correlation length of 25 nm, for the 2-cloud case in Section 6.3, using the **2c** forward model. The red color map corresponds to distributions obtained using the MCMC algorithm, and the black/gray contours to nested sampling. The teal lines show the real solution.

Figure 15. Similar plot to Figure 9, for the 2-cloud case in Section 6.3. The color scheme has been modified to emphasize the case where a 2-cloud structure is assumed as default. The orange triangles correspond to the ratios $\mathcal{Z}_{2c}/\mathcal{Z}_{2c-m}$, the blue circles to $\mathcal{Z}_{2c}/\mathcal{Z}_{2c-c}$, the pink triangles to $\mathcal{Z}_{1c}/\mathcal{Z}_{1c-c}$, and the green stars to $\mathcal{Z}_{2c}/\mathcal{Z}_{1c}$. As in the previous examples, the methane and cloud are clearly detected even with a SNR=5 dataset.
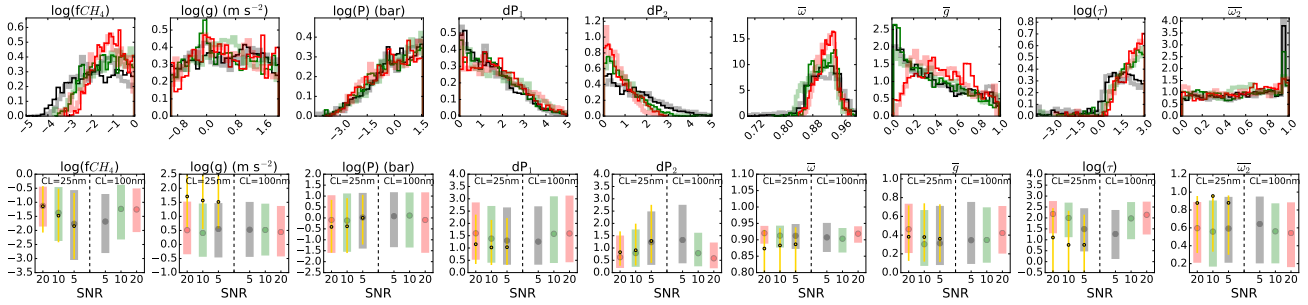
Figure 16. Same as Figure 13, for the HD 99492 c model in Section 7.1. In a realistic scenario, the "true" parameters values would not be known, and therefore are not shown.
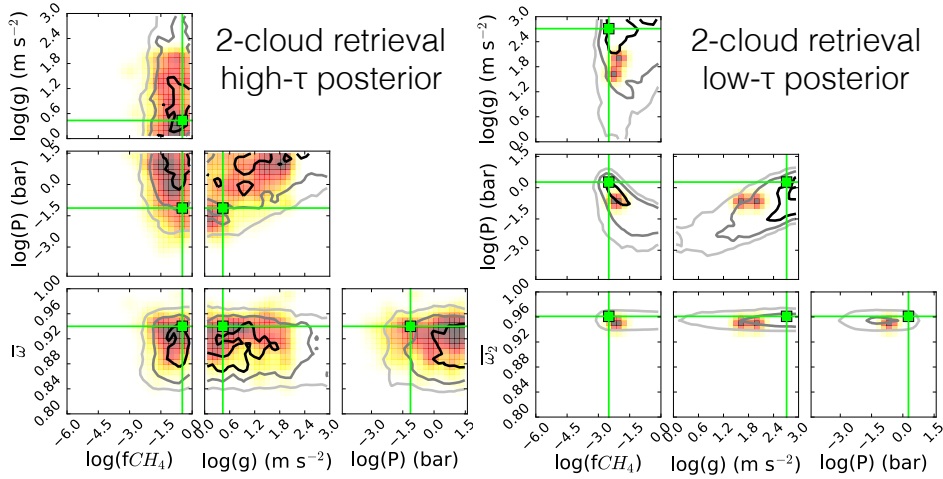

Figure 17. 2-D marginal posterior distributions for HD 99492 c (SNR=20, CL=25 nm), using a 2-cloud model. The full posterior is bi-modal, with a second, low optical depth mode better identified by the nested sampling algorithm (gray contours). For clarity, we plot the two modes separately, the high optical depth on the left, and the low optical depth on the right. The bright green lines mark the position of the *best fit solution* for each mode.
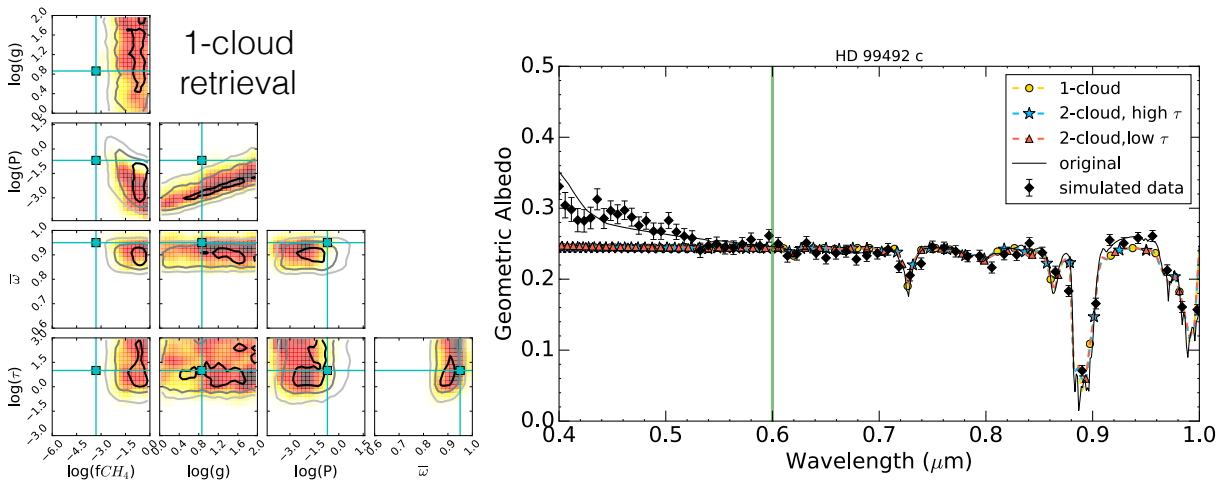

Figure 18. Best-fit spectra and 2-D marginal posterior distributions for HD 99492 c (SNR=20, CL=25 nm), using a 1-cloud model. The 2-cloud best fit parameters for the two modes are indicated in green in Figure 17. The blue lines on the left plot show the 1-cloud parameter values that best match the "theoretical model" on the right panel in Figure 19.
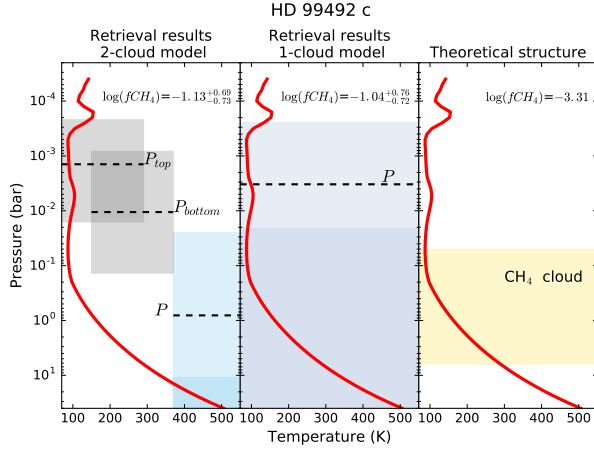
Figure 19. Cloud structure for gas giant HD 99492 c, as retrieved using the 2-cloud model (left), and the 1-cloud model (right). The semi-transparent regions are associated with the error bars for the cloud top (bottom) pressures, and the labeling follows the convention in Figure 4. In the left panel, the positions of the cloud layers have been offset for clarity, with the gray regions overlapping to emphasize the fact the both P_{top} and P_{bottom} refer to the same cloud deck, while the blue regions correspond to the second cloud deck defined in Figure 4. The theoretical structure is shown in the right panel, with the region occupied by the cloud calculated using the radiative-convective equilibrium code. The pressure-temperature profile calculated by this code and kept fixed in the retrievals is shown in red in all three panels. The theoretical and retrieved CH_4 abundance is shown at the top
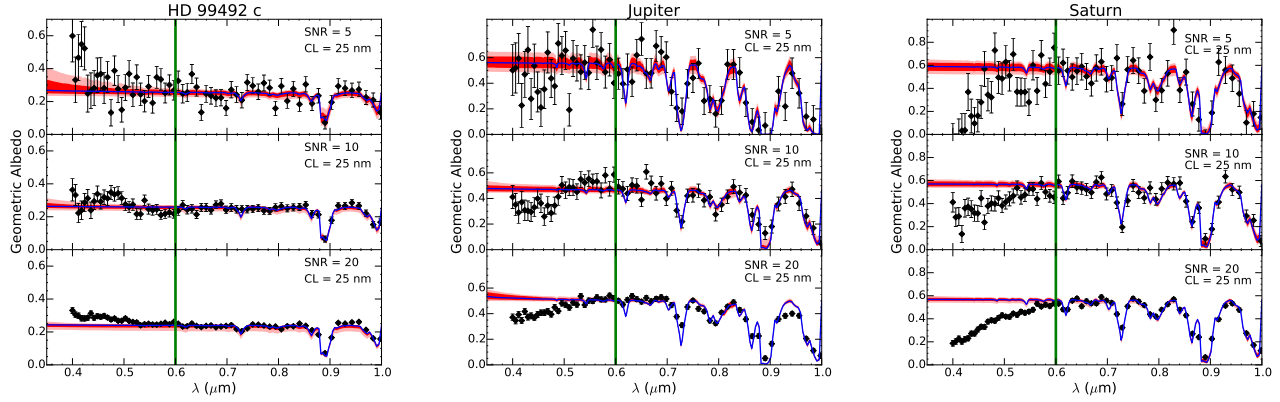
Figure 20. Simulated data and best fit spectra for HD 99492 c (left), Jupiter (middle), Saturn (right), using the **2c** forward model. The data correspond to SNR=5, 10, 20, from top to bottom and a spectral correlation noise of 25 nm. Same conventions as in Figure 8. The retrieval was performed over the $0.6 - 1.0$ μm region, as indicated by the green vertical line.
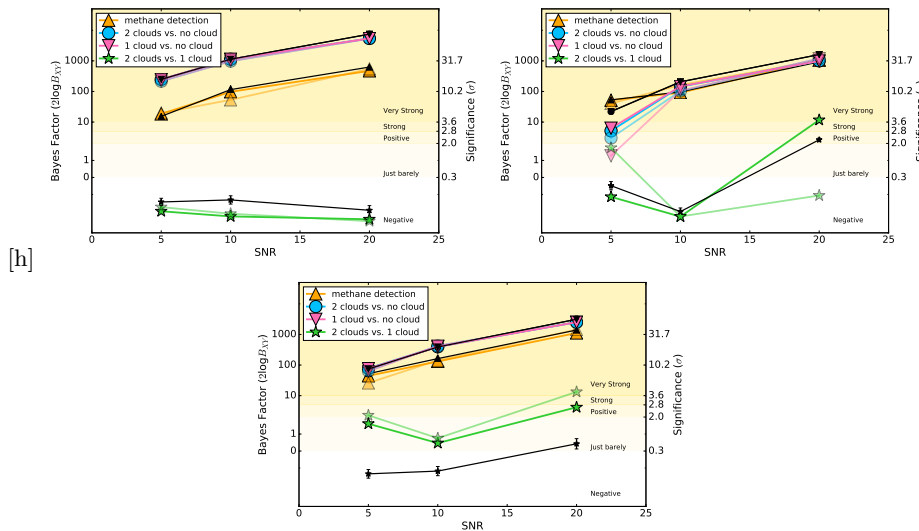

Figure 21. Same as Figure 15, for the applications in Section 7. The plots correspond to HD 99492 c, Jupiter, and Saturn , from left to right. As in the previous examples, the methane and cloud are clearly detected even with a SNR=5 dataset.
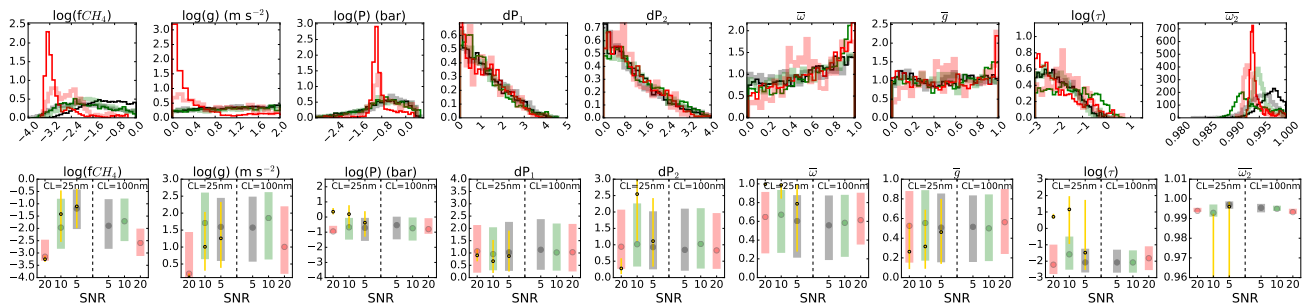
Figure 22. Same as Figure 16, for the Jupiter albedo in Section 7.2.
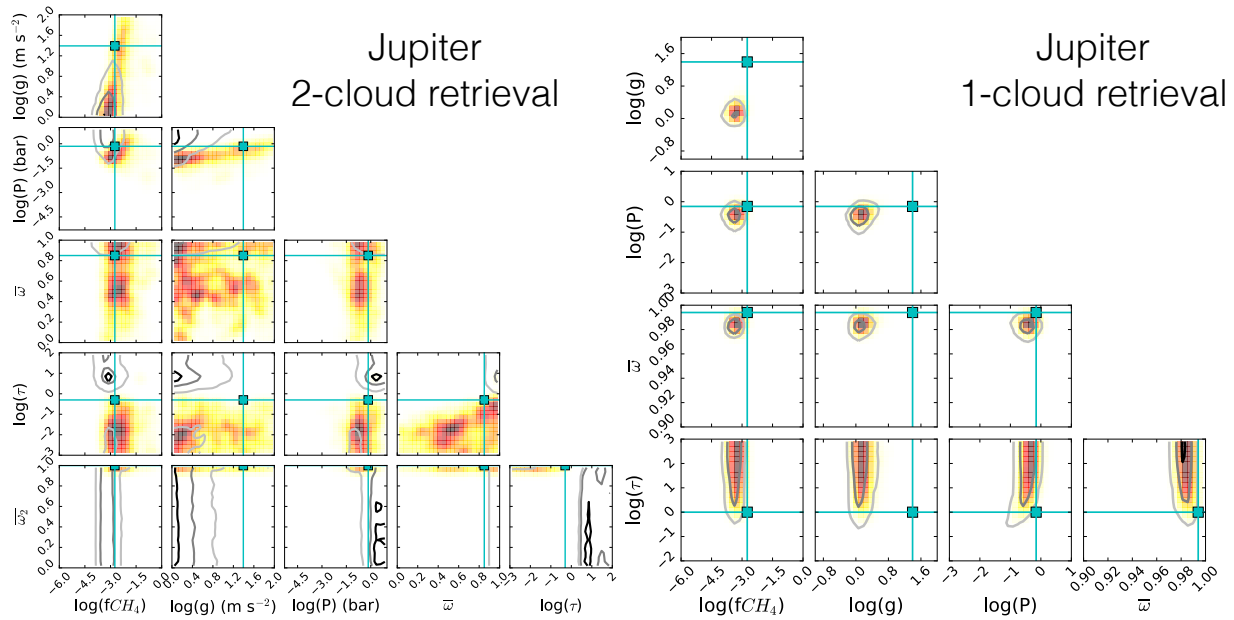
Figure 23. 2-D marginal posterior distributions for Jupiter (SNR=20, CL=25 nm), using a 2-cloud model (left) and a 1-cloud model (right). The two posterior sampling methods lock onto different modes, one with low optical depth (MCMC, red colormap), and the other with high optical depth (nested sampling, gray contours). The best-fir solution for each method is shown in Figure 24. The blue lines on the left plot show the 1-cloud parameter values that best match the "theoretical model" on the right panel in Figure 25 (e.g. known values for *g* and *fCH₄*).
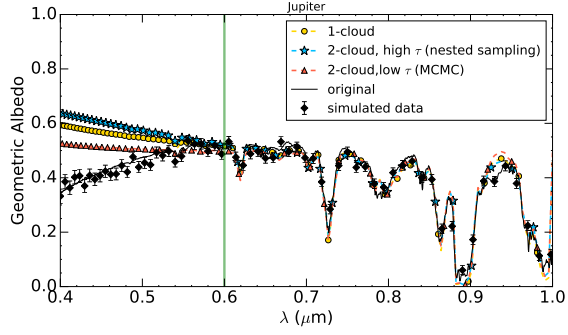
Figure 24. Best-fit spectra for Jupiter (SNR=20, CL=25 nm), retrieved using the 2-cloud and 1-cloud models. The legend indicates that the low optical depth fit is favored by the MCMC method, while the high optical depth fit is favored by nested sampling (see also Figure 23). The vertical green line indicates that the retrieval is performed only on data between 0.6 and 1 μm.

Figure 25. Cloud structure for Jupiter, as retrieved using the 2-cloud model (left), and the 1-cloud model (right). The conventions are described in the Figure 19 caption. The theoretical structure is shown in the right panel, with the cloud structure closely resembling available literature (e.g., Simon-Miller et al. 2001; Sato et al. 2013). The pressure-temperature profile is approximated as purely radiative in the top layes of the atmosphere (dashed line).
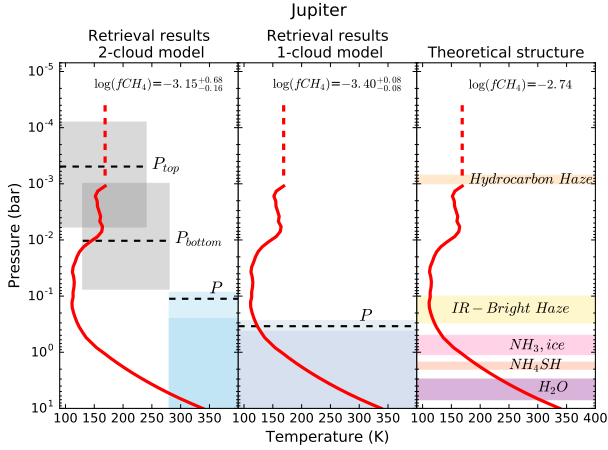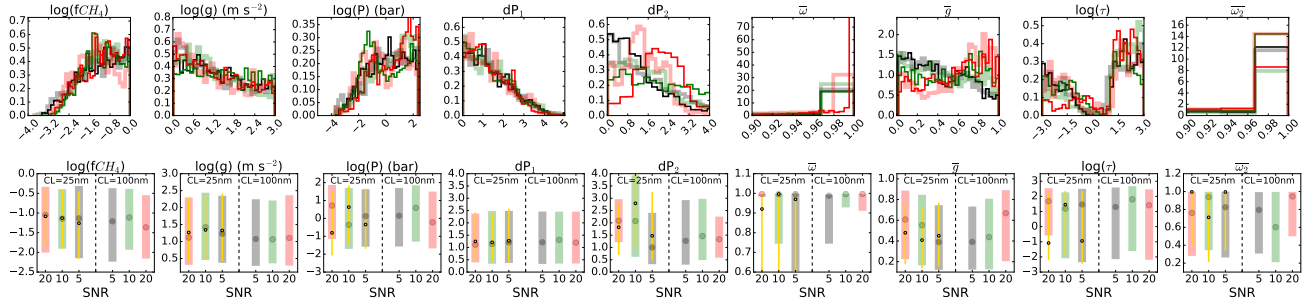
Figure 26. Same as Figure 16, for the Saturn albedo in Section 7.3.
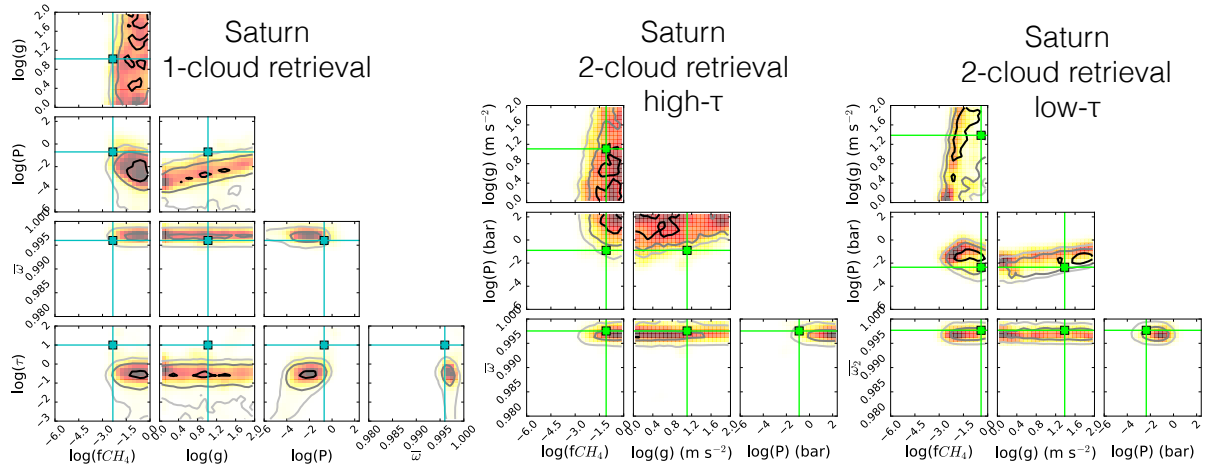
Figure 27. 2-D marginal posterior distributions for Saturn (SNR=20, CL=25 nm), using a 1-cloud model (left) and a 2-cloud model (middle and right). The posterior for the 2-cloud model is bi-modal, and the two modes are shown separately, for clarity. The bright green lines mark the position of the *best fit solution* for each mode (corresponding to the spectra in Figure 28), while the blue lines on the left plot show the 1-cloud parameter values that best match the "theoretical model" on the right panel in Figure 29 (e.g. known values for *g* and *fCH₄*).
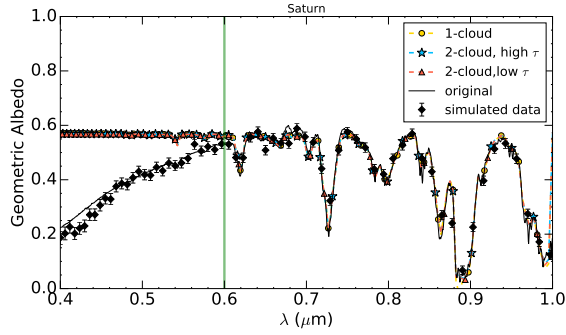
Figure 28. Best-fit spectra for Saturn (SNR=20, CL=25 nm), retrieved using the 2-cloud and 1-cloud models. The 2-cloud posterior is bimodal, with the low optical depth and high optical depth best fit solutions shown separately (see also Figure 27).
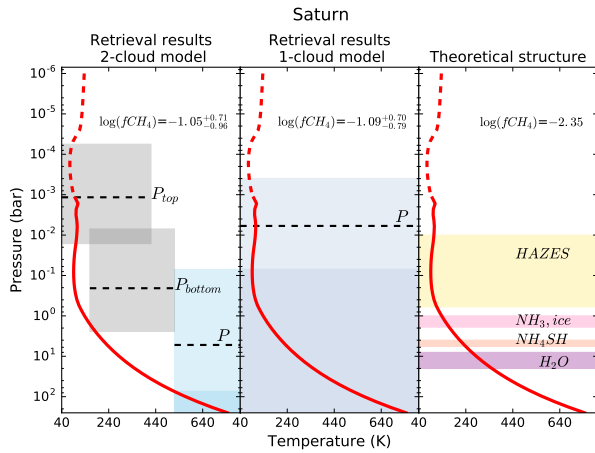
Figure 29. Cloud structure for Saturn, as retrieved using the 2-cloud model (left), and the 1-cloud model (right). The conventions are described in the Figure 19 and 25 captions. The theoretical structure is shown in the right panel, with the cloud structure closely resembling available literature (e.g., Roman et al. 2013).

Table 1. Retrieval verification results for the 1-cloud model.

Parameter	Original Value	SNR=5		SNR=10		SNR = 20	
		CL[a]=25nm	CL=100nm	CL=25nm	CL=100nm	CL=25nm	CL=100nm
Cloud-free case							
$\log(fCH_4)$	-3.31	$-3.22^{+0.19}_{-0.22}$ $(-3.21^{+0.18}_{-0.20})$[b]	$-2.92^{+0.18}_{-0.24}$	$-3.42^{+0.11}_{-0.11}$ $(-3.42^{+0.10}_{-0.10})$	$-3.20^{+0.09}_{-0.10}$	$-3.27^{+0.03}_{-0.03}$ $(-3.27^{+0.03}_{-0.03})$	$-3.20^{+0.03}_{-0.03}$
$\log(g)$ (m s^{-2})	0.86	$0.84^{+0.21}_{-0.39}$ $(0.82^{+0.22}_{-0.42})$	$0.95^{+0.22}_{-0.39}$	$0.90^{+0.16}_{-0.22}$ $(0.93^{+0.11}_{-0.21})$	$0.63^{+0.28}_{-0.26}$	$0.85^{+0.03}_{-0.04}$ $(0.86^{+0.03}_{-0.04})$	$0.89^{+0.03}_{-0.03}$
$\log(P)$ (bar)	1.00	$-0.71^{+1.74}_{-2.32}$ $(-0.60^{+1.53}_{-2.32})$	$-0.53^{+1.61}_{-2.34}$	$-0.67^{+1.82}_{-2.47}$ $(-0.83^{+1.89}_{-2.41})$	$-1.15^{+1.94}_{-2.18}$	$-0.53^{+1.55}_{-2.27}$ $(-0.45^{+1.46}_{-2.14})$	$-0.46^{+1.51}_{-2.08}$
$\bar{\omega}$	0.50	$0.51^{+0.33}_{-0.32}$ $(0.51^{+0.32}_{-0.33})$	$0.57^{+0.31}_{-0.37}$	$0.57^{+0.34}_{-0.36}$ $(0.57^{+0.31}_{-0.37})$	$0.45^{+0.34}_{-0.31}$	$0.52^{+0.36}_{-0.35}$ $(0.53^{+0.32}_{-0.34})$	$0.53^{+0.35}_{-0.36}$
$\bar{g}$	0.50	$0.49^{+0.36}_{-0.35}$ $(0.50^{+0.33}_{-0.32})$	$0.47^{+0.36}_{-0.31}$	$0.41^{+0.35}_{-0.30}$ $(0.35^{+0.38}_{-0.24})$	$0.59^{+0.30}_{-0.37}$	$0.48^{+0.35}_{-0.31}$ $(0.50^{+0.33}_{-0.32})$	$0.50^{+0.36}_{-0.34}$
$\log(\tau)$	-8.00	$-7.01^{+2.74}_{-2.01}$ $(-7.04^{+2.92}_{-1.93})$	$-6.81^{+3.24}_{-2.19}$	$-4.81^{+3.44}_{-2.73}$ $(-5.58^{+3.56}_{-2.49})$	$-4.34^{+2.65}_{-2.32}$	$-7.35^{+2.53}_{-1.76}$ $(-7.51^{+2.49}_{-1.64})$	$-7.64^{+2.58}_{-1.69}$
1-Cloud case							
$\log(fCH_4)$	-3.31	$-3.54^{+0.38}_{-0.31}$ $(-3.52^{+0.34}_{-0.32})$	$-3.47^{+0.39}_{-0.32}$	$-3.27^{+0.21}_{-0.22}$ $(-3.25^{+0.23}_{-0.20})$	$-1.42^{+0.88}_{-0.82}$	$-3.31^{+0.17}_{-0.22}$ $(-3.13^{+0.12}_{-0.11})$	$-2.73^{+0.21}_{-0.27}$
$\log(g)$ (m s^{-2})	0.86	$0.39^{+0.85}_{-0.90}$ $(0.38^{+0.88}_{-0.82})$	$0.19^{+0.97}_{-0.81}$	$0.36^{+0.91}_{-0.90}$ $(0.41^{+0.90}_{-0.91})$	$1.08^{+0.64}_{-1.04}$	$0.05^{+0.50}_{-0.62}$ $(0.01^{+0.67}_{-0.65})$	$1.31^{+0.54}_{-1.47}$
$\log(P)$ (bar)	-0.70	$-1.72^{+1.36}_{-1.89}$ $(-1.80^{+1.51}_{-1.73})$	$-1.46^{+1.14}_{-1.13}$	$-1.79^{+1.40}_{-1.52}$ $(-1.82^{+1.33}_{-1.54})$	$-3.29^{+1.22}_{-0.75}$	$-2.03^{+1.02}_{-1.20}$ $(-2.63^{+0.98}_{-1.01})$	$-0.85^{+0.84}_{-1.42}$
$\bar{\omega}$	0.96	$0.90^{+0.04}_{-0.05}$ $(0.90^{+0.04}_{-0.04})$	$0.90^{+0.03}_{-0.05}$	$0.92^{+0.03}_{-0.03}$ $(0.91^{+0.03}_{-0.03})$	$0.92^{+0.03}_{-0.03}$	$0.95^{+0.02}_{-0.03}$ $(0.92^{+0.03}_{-0.03})$	$0.94^{+0.02}_{-0.04}$
$\bar{g}$	0.85	$0.27^{+0.38}_{-0.19}$ $(0.28^{+0.38}_{-0.20})$	$0.35^{+0.33}_{-0.24}$	$0.29^{+0.39}_{-0.20}$ $(0.26^{+0.33}_{-0.18})$	$0.27^{+0.39}_{-0.19}$	$0.69^{+0.24}_{-0.33}$ $(0.33^{+0.29}_{-0.23})$	$0.52^{+0.31}_{-0.35}$
$\log(\tau)$	0.00	$-1.31^{+2.89}_{-2.20}$ $(-1.40^{+3.09}_{-2.05})$	$0.07^{+1.85}_{-2.13}$	$-0.36^{+2.36}_{-2.31}$ $(-0.48^{+2.38}_{-2.33})$	$-1.18^{+2.43}_{-1.23}$	$-0.83^{+2.49}_{-1.44}$ $(-1.45^{+0.63}_{-1.18})$	$0.73^{+1.61}_{-1.38}$

[a]CL here is a shorthand notation for the spectral noise correlation length.

[b]Numbers in parentheses show the nested sampling results.

Table 2. Retrieval verification results for the 2-cloud model.

Parameter	Original Value	SNR=5		SNR=10		SNR = 20	
		CL=25nm	CL=100nm	CL=25nm	CL=100nm	CL=25nm	CL=100nm
$\log(fCH_4)$	-2.74	$-1.95^{+0.88}_{-0.67}$ $(-1.35^{+0.89}_{-0.95})$	$-2.54^{+0.96}_{-0.53}$	$-1.79^{+0.85}_{-0.65}$ $(-1.37^{+0.92}_{-0.86})$	$-1.90^{+0.86}_{-0.60}$	$-2.66^{+0.14}_{-0.19}$ $(-2.65^{+0.15}_{-0.21})$	$-2.65^{+0.14}_{-0.17}$
$\log(g)$ (m s^{-2})	1.39	$1.22^{+0.55}_{-0.70}$ $(1.12^{+0.60}_{-0.72})$	$1.21^{+0.56}_{-0.70}$	$1.19^{+0.54}_{-0.66}$ $(1.07^{+0.63}_{-0.69})$	$1.28^{+0.53}_{-0.68}$	$1.71^{+0.22}_{-0.44}$ $(1.65^{+0.24}_{-0.56})$	$1.62^{+0.27}_{-0.39}$
$\log(P)$ (bar)	-0.15	$-1.25^{+0.84}_{-1.04}$ $(-0.72^{+1.12}_{-1.07})$	$-0.39^{+0.54}_{-0.89}$	$-1.25^{+0.70}_{-0.86}$ $(-0.90^{+1.08}_{-0.85})$	$-1.23^{+0.78}_{-0.90}$	$0.06^{+0.15}_{-0.32}$ $(-0.07^{+0.20}_{-0.37})$	$-0.04^{+0.20}_{-0.27}$
dP_1 (bar)	0.54	$0.82^{+1.05}_{-0.60}$ $(0.87^{+0.97}_{-0.62})$	$1.21^{+1.22}_{-0.85}$	$0.83^{+1.03}_{-0.60}$ $(0.87^{+0.95}_{-0.60})$	$0.87^{+1.05}_{-0.63}$	$1.45^{+1.23}_{-1.05}$ $(1.04^{+1.38}_{-0.74})$	$1.44^{+1.36}_{-0.99}$
dP_2 (bar)	0.12	$0.89^{+1.13}_{-0.66}$ $(1.04^{+1.25}_{-0.75})$	$1.09^{+1.12}_{-0.78}$	$0.76^{+1.02}_{-0.57}$ $(0.84^{+0.94}_{-0.58})$	$0.83^{+0.91}_{-0.60}$	$1.28^{+1.24}_{-0.88}$ $(1.49^{+1.27}_{-1.00})$	$1.11^{+1.28}_{-0.81}$
$\bar{\omega}$	0.85	$0.56^{+0.32}_{-0.39}$ $(0.95^{+0.03}_{-0.59})$	$0.62^{+0.30}_{-0.42}$	$0.56^{+0.31}_{-0.38}$ $(0.80^{+0.18}_{-0.54})$	$0.54^{+0.34}_{-0.37}$	$0.68^{+0.23}_{-0.38}$ $(0.46^{+0.33}_{-0.30})$	$0.69^{+0.20}_{-0.35}$
$\bar{g}$	0.85	$0.48^{+0.35}_{-0.31}$ $(0.39^{+0.38}_{-0.27})$	$0.54^{+0.32}_{-0.35}$	$0.55^{+0.30}_{-0.34}$ $(0.42^{+0.37}_{-0.29})$	$0.47^{+0.34}_{-0.31}$	$0.60^{+0.30}_{-0.40}$ $(0.55^{+0.29}_{-0.35})$	$0.60^{+0.27}_{-0.37}$
$\log(\tau)$	-0.30	$-1.85^{+0.82}_{-0.79}$ $(-0.63^{+2.18}_{-1.70})$	$-1.67^{+1.00}_{-0.88}$	$-2.06^{+0.78}_{-0.65}$ $(-1.43^{+2.92}_{-1.08})$	$-1.99^{+0.81}_{-0.73}$	$-1.02^{+0.33}_{-0.70}$ $(-1.01^{+0.28}_{-0.60})$	$-1.00^{+0.45}_{-1.04}$
$\bar{\omega}_2$	0.997	$0.987^{+0.004}_{-0.003}$ $(0.984^{+0.005}_{-0.638})$	$0.991^{+0.005}_{-0.003}$	$0.989^{+0.002}_{-0.001}$ $(0.989^{+0.002}_{-0.564})$	$0.988^{+0.003}_{-0.001}$	$0.993^{+0.003}_{-0.003}$ $(0.995^{+0.003}_{-0.004})$	$0.993^{+0.005}_{-0.003}$

Table 3. Retrieval results for HD 99492 c.

Parameter	SNR=5		SNR=10		SNR = 20	
	CL=25nm	CL=100nm	CL=25nm	CL=100nm	CL=25nm	CL=100nm
$\log(fCH_4)$	$-1.76^{+1.20}_{-1.29}$ $(-1.85^{+1.18}_{-1.18})$	$-1.68^{+0.98}_{-1.12}$	$-1.37^{+0.92}_{-1.00}$ $(-1.48^{+0.96}_{-0.95})$	$-1.24^{+0.86}_{-1.09}$	$-1.13^{+0.69}_{-0.73}$ $(-1.14^{+0.72}_{-0.94})$	$-1.25^{+0.75}_{-0.80}$
$\log(g)$ (m s^{-2})	$0.55^{+0.92}_{-1.01}$ $(1.51^{+0.97}_{-0.95})$	$0.52^{+0.99}_{-0.97}$	$0.41^{+1.05}_{-0.85}$ $(1.56^{+0.95}_{-1.02})$	$0.52^{+0.93}_{-0.91}$	$0.51^{+1.02}_{-0.86}$ $(1.71^{+0.88}_{-1.10})$	$0.44^{+0.93}_{-0.87}$
$\log(P)$ (bar)	$0.02^{+1.13}_{-1.42}$ $(0.00^{+1.05}_{-1.22})$	$0.08^{+1.10}_{-1.42}$	$-0.12^{+1.24}_{-1.43}$ $(-0.38^{+1.28}_{-1.22})$	$0.11^{+1.06}_{-1.46}$	$-0.09^{+1.13}_{-1.51}$ $(-0.41^{+1.24}_{-1.18})$	$-0.09^{+1.26}_{-1.50}$
dP_1 (bar)	$1.30^{+1.35}_{-0.98}$ $(1.03^{+1.19}_{-0.72})$	$1.26^{+1.44}_{-0.94}$	$1.38^{+1.33}_{-0.98}$ $(1.02^{+1.12}_{-0.71})$	$1.58^{+1.52}_{-1.19}$	$1.60^{+1.26}_{-1.07}$ $(1.15^{+1.20}_{-0.80})$	$1.59^{+1.54}_{-1.17}$
dP_2 (bar)	$1.24^{+1.25}_{-0.93}$ $(1.28^{+1.47}_{-0.89})$	$1.33^{+1.43}_{-0.95}$	$0.79^{+0.96}_{-0.56}$ $(0.91^{+1.08}_{-0.62})$	$0.79^{+0.83}_{-0.53}$	$0.63^{+0.88}_{-0.44}$ $(0.83^{+0.85}_{-0.55})$	$0.58^{+0.64}_{-0.41}$
$\bar{\omega}$	$0.91^{+0.04}_{-0.04}$ $(0.89^{+0.05}_{-0.49})$	$0.91^{+0.04}_{-0.04}$	$0.91^{+0.03}_{-0.04}$ $(0.88^{+0.05}_{-0.50})$	$0.90^{+0.03}_{-0.04}$	$0.92^{+0.02}_{-0.03}$ $(0.87^{+0.06}_{-0.46})$	$0.92^{+0.02}_{-0.03}$
$\bar{g}$	$0.31^{+0.40}_{-0.23}$ $(0.36^{+0.37}_{-0.25})$	$0.35^{+0.41}_{-0.25}$	$0.30^{+0.36}_{-0.23}$ $(0.38^{+0.36}_{-0.26})$	$0.35^{+0.35}_{-0.24}$	$0.46^{+0.27}_{-0.25}$ $(0.38^{+0.33}_{-0.26})$	$0.42^{+0.29}_{-0.22}$
$\log(\tau)$	$1.49^{+0.95}_{-1.02}$ $(0.77^{+1.38}_{-3.59})$	$1.27^{+1.10}_{-1.14}$	$2.00^{+0.70}_{-0.88}$ $(0.77^{+1.56}_{-3.75})$	$1.98^{+0.75}_{-0.95}$	$2.18^{+0.59}_{-0.89}$ $(1.10^{+1.35}_{-4.19})$	$2.14^{+0.60}_{-0.90}$
$\bar{\omega}_2$	$0.592^{+0.354}_{-0.382}$ $(0.880^{+0.083}_{-0.580})$	$0.644^{+0.307}_{-0.441}$	$0.558^{+0.348}_{-0.391}$ $(0.956^{+0.006}_{-0.623})$	$0.562^{+0.313}_{-0.358}$	$0.596^{+0.293}_{-0.386}$ $(0.878^{+0.078}_{-0.559})$	$0.542^{+0.343}_{-0.382}$

Table 4. Retrieval results for Jupiter.

Parameter	SNR=5		SNR=10		SNR = 20	
	CL=25nm	CL=100nm	CL=25nm	CL=100nm	CL=25nm	CL=100nm
$\log(fCH_4)$	$-1.15^{+0.74}_{-0.87}$ $(-1.10^{+0.72}_{-0.89})$	$-1.91^{+1.13}_{-0.91}$	$-1.95^{+1.14}_{-0.83}$ $(-1.42^{+0.96}_{-1.13})$	$-1.70^{+0.88}_{-0.81}$	$-2.80^{+0.48}_{-0.35}$ $(-3.25^{+0.14}_{-0.11})$	$-2.60^{+0.61}_{-0.52}$
$\log(g)$ (m s^{-2})	$1.63^{+0.86}_{-1.04}$ $(1.26^{+1.07}_{-0.87})$	$1.62^{+0.88}_{-1.01}$	$1.74^{+0.87}_{-1.08}$ $(1.01^{+1.03}_{-0.70})$	$1.83^{+0.79}_{-1.19}$	$0.76^{+0.90}_{-0.62}$ $(0.09^{+0.13}_{-0.06})$	$1.00^{+1.19}_{-0.78}$
$\log(P)$ (bar)	$-0.71^{+0.67}_{-0.86}$ $(-0.37^{+0.76}_{-0.93})$	$-0.52^{+0.56}_{-0.92}$	$-0.67^{+0.62}_{-0.86}$ $(0.19^{+0.59}_{-0.94})$	$-0.74^{+0.71}_{-0.83}$	$-0.79^{+0.53}_{-0.30}$ $(0.35^{+0.24}_{-0.18})$	$-0.78^{+0.70}_{-0.31}$
dP_1 (bar)	$1.06^{+1.23}_{-0.77}$ $(0.87^{+1.04}_{-0.63})$	$1.13^{+1.32}_{-0.82}$	$0.95^{+1.15}_{-0.65}$ $(0.67^{+0.86}_{-0.48})$	$1.01^{+1.16}_{-0.72}$	$0.62^{+0.99}_{-0.46}$ $(0.90^{+0.25}_{-0.24})$	$1.00^{+1.20}_{-0.75}$
dP_2 (bar)	$0.93^{+1.09}_{-0.67}$ $(1.11^{+1.31}_{-0.78})$	$0.88^{+1.30}_{-0.66}$	$1.00^{+1.23}_{-0.70}$ $(2.55^{+1.14}_{-1.43})$	$1.05^{+1.07}_{-0.77}$	$0.43^{+0.71}_{-0.27}$ $(0.28^{+0.31}_{-0.17})$	$0.83^{+1.13}_{-0.57}$
$\bar{\omega}$	$0.60^{+0.29}_{-0.39}$ $(0.79^{+0.21}_{-0.51})$	$0.55^{+0.32}_{-0.36}$	$0.67^{+0.26}_{-0.43}$ $(0.99^{+0.00}_{-0.15})$	$0.58^{+0.30}_{-0.35}$	$0.84^{+0.11}_{-0.33}$ $(1.00^{+0.00}_{-0.00})$	$0.61^{+0.29}_{-0.27}$
$\bar{g}$	$0.52^{+0.34}_{-0.36}$ $(0.46^{+0.34}_{-0.32})$	$0.50^{+0.33}_{-0.36}$	$0.53^{+0.35}_{-0.37}$ $(0.32^{+0.37}_{-0.23})$	$0.49^{+0.34}_{-0.33}$	$0.88^{+0.11}_{-0.48}$ $(0.26^{+0.29}_{-0.18})$	$0.56^{+0.33}_{-0.33}$
$\log(\tau)$	$-2.04^{+0.81}_{-0.64}$ $(-1.48^{+3.22}_{-1.05})$	$-2.12^{+0.78}_{-0.60}$	$-1.59^{+1.08}_{-0.93}$ $(1.16^{+0.81}_{-2.11})$	$-2.08^{+0.75}_{-0.63}$	$-1.12^{+0.62}_{-0.95}$ $(0.71^{+0.18}_{-0.11})$	$-1.83^{+0.78}_{-0.72}$
$\bar{\omega}_2$	$0.997^{+0.002}_{-0.002}$ $(0.996^{+0.002}_{-0.494})$	$0.995^{+0.002}_{-0.002}$	$0.993^{+0.004}_{-0.002}$ $(0.645^{+0.348}_{-0.435})$	$0.995^{+0.002}_{-0.001}$	$0.995^{+0.001}_{-0.001}$ $(0.379^{+0.313}_{-0.254})$	$0.993^{+0.002}_{-0.001}$

Table 5. Retrieval results for Saturn.

Parameter	SNR=5		SNR=10		SNR = 20	
	CL=25nm	CL=100nm	CL=25nm	CL=100nm	CL=25nm	CL=100nm
$\log(fCH_4)$	$-1.15^{+0.83}_{-0.99}$ $(-1.26^{+0.81}_{-0.90})$	$-1.20^{+0.85}_{-1.00}$	$-1.14^{+0.77}_{-0.76}$ $(-1.13^{+0.72}_{-0.77})$	$-1.10^{+0.69}_{-0.86}$	$-1.29^{+0.73}_{-0.63}$ $(-1.08^{+0.70}_{-0.77})$	$-1.37^{+0.83}_{-0.83}$
$\log(g)$ (m s^{-2})	$1.24^{+1.18}_{-0.86}$ $(1.33^{+1.09}_{-0.90})$	$1.13^{+1.11}_{-0.82}$	$1.43^{+1.00}_{-0.98}$ $(1.34^{+1.01}_{-0.88})$	$1.07^{+1.14}_{-0.72}$	$1.18^{+1.14}_{-0.80}$ $(1.27^{+1.04}_{-0.83})$	$1.14^{+1.23}_{-0.84}$
$\log(P)$ (bar)	$0.12^{+1.51}_{-1.70}$ $(-0.34^{+2.00}_{-1.35})$	$0.19^{+1.53}_{-1.86}$	$-0.32^{+2.03}_{-1.37}$ $(0.63^{+1.20}_{-2.07})$	$0.60^{+1.23}_{-1.90}$	$-0.37^{+2.11}_{-1.21}$ $(-0.81^{+2.31}_{-1.26})$	$-0.19^{+1.75}_{-1.49}$
dP_1 (bar)	$1.20^{+1.32}_{-0.85}$ $(1.27^{+1.31}_{-0.88})$	$1.22^{+1.23}_{-0.88}$	$1.18^{+1.28}_{-0.81}$ $(1.21^{+1.19}_{-0.80})$	$1.31^{+1.16}_{-0.97}$	$1.21^{+1.24}_{-0.90}$ $(1.25^{+1.17}_{-0.82})$	$1.22^{+1.23}_{-0.85}$
dP_2 (bar)	$0.99^{+1.43}_{-0.68}$ $(1.47^{+1.78}_{-1.00})$	$1.22^{+1.68}_{-0.90}$	$2.08^{+1.85}_{-1.46}$ $(2.79^{+1.96}_{-1.79})$	$1.47^{+1.60}_{-0.98}$	$1.75^{+0.96}_{-0.91}$ $(1.82^{+0.96}_{-1.13})$	$1.32^{+0.94}_{-0.70}$
$\bar{\omega}$	$1.00^{+0.00}_{-0.36}$ $(0.97^{+0.03}_{-0.59})$	$0.99^{+0.01}_{-0.25}$	$1.00^{+0.00}_{-0.25}$ $(1.00^{+0.00}_{-0.41})$	$1.00^{+0.00}_{-0.06}$	$1.00^{+0.00}_{-0.05}$ $(0.92^{+0.08}_{-0.54})$	$1.00^{+0.00}_{-0.08}$
$\bar{g}$	$0.39^{+0.36}_{-0.27}$ $(0.45^{+0.32}_{-0.30})$	$0.39^{+0.34}_{-0.27}$	$0.55^{+0.30}_{-0.38}$ $(0.41^{+0.32}_{-0.27})$	$0.46^{+0.35}_{-0.31}$	$0.67^{+0.25}_{-0.41}$ $(0.48^{+0.29}_{-0.30})$	$0.67^{+0.22}_{-0.46}$
$\log(\tau)$	$1.49^{+1.01}_{-3.77}$ $(-0.94^{+3.28}_{-1.50})$	$1.28^{+1.30}_{-3.41}$	$1.20^{+1.09}_{-3.33}$ $(1.46^{+0.89}_{-3.35})$	$1.80^{+0.88}_{-3.38}$	$1.28^{+1.05}_{-2.83}$ $(-1.08^{+3.31}_{-1.14})$	$1.42^{+1.04}_{-3.56}$
$\bar{\omega}_2$	$0.812^{+0.186}_{-0.552}$ $(0.996^{+0.003}_{-0.658})$	$0.806^{+0.188}_{-0.507}$	$0.946^{+0.052}_{-0.593}$ $(0.712^{+0.285}_{-0.500})$	$0.610^{+0.386}_{-0.405}$	$0.968^{+0.029}_{-0.598}$ $(0.996^{+0.001}_{-0.608})$	$0.949^{+0.048}_{-0.449}$

APPENDIX

A. APPENDIX

In Bayesian inference, the allowed ranges of model parameters are given by the posterior probability distribution of the parameter vector $\boldsymbol{\theta}$,

$$\mathcal{P}(\boldsymbol{\theta}) = \frac{\mathcal{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}}, \tag{A1}$$

where $\mathcal{P}(\boldsymbol{\theta}) \equiv \mathrm{Pr}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M})$, $\mathcal{L}(\boldsymbol{\theta}) \equiv \mathrm{Pr}(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M})$ is the likelihood, $\pi(\boldsymbol{\theta}) \equiv \mathrm{Pr}(\boldsymbol{\theta} \mid \mathcal{M})$ is the prior on model parameters, and $\mathcal{Z} \equiv \mathrm{Pr}(\mathcal{D} \mid \mathcal{M})$ is the Bayesian evidence. Here $\mathcal{D}$ and $\mathcal{M}$ denote the data and the model, respectively. Normalization of the posterior distribution requires that

$$\mathcal{Z} = \int \mathcal{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{A2}$$

The calculation of $\mathcal{Z}$ is not necessary for parameter estimation, and best-fit parameter values with associated confidence intervals are obtained from the un-normalized $\mathcal{P}(\boldsymbol{\theta})$. In general, the posterior $\mathcal{P}(\boldsymbol{\theta})$ is difficult or impossible to calculate analytically, and in practice the shape of this distribution is approximated by taking a large number of samples. The methods described below are optimized to sample more efficiently the regions of parameter space where $\mathcal{L}(\boldsymbol{\theta})$ is large, such that a good approximation of $\mathcal{P}(\boldsymbol{\theta})$ is obtained with a minimum number of samples. The Bayesian evidence $\mathcal{Z}$ is by definition model-dependent, and provides the information necessary for model selection. The evaluation of this multi-dimensional integral is also often difficult, and addressed by various approximations (Section A.1).

A.1. *Model selection*

In Bayesian inference, the probability associated with a given model $\mathcal{M}$, given the data, is defined as $\mathcal{P}(\mathcal{M} \mid \mathcal{D}) = \mathcal{P}(\mathcal{D} \mid \mathcal{M})\mathcal{P}(\mathcal{M}) = \mathcal{Z}\mathcal{P}(\mathcal{M})$. In our calculations of Bayesian evidence we have employed the approximations described below.

In the Laplace-Metropolis approximation (Lopes & West 2004), $\mathcal{Z}$ is computed using the covariance matrix $\boldsymbol{C}$ of the posterior, or the minimum volume ellipsoid enclosing the posterior distribution

$$\mathcal{Z} \simeq \mathcal{L}_{max}(\boldsymbol{\theta})(2\pi)^{n/2}\sqrt{\det \boldsymbol{C}}, \tag{A3}$$

where n is the dimension of the parameter space, and $\mathcal{L}_{max}(\boldsymbol{\theta})$ is the maximum likelihood value. This approximation clearly breaks down when the posterior is multi-modal.

The BIC estimate is a result obtained in the asymptotic limit for distributions in the exponential family, and gives the largest penalty to models with a large number of parameters. In this approximation

$$\ln \mathcal{Z} \simeq \ln \mathcal{L}_{max}(\boldsymbol{\theta}) - \frac{n}{2} \ln N_D, \tag{A4}$$

where N_D is the number of data points. In most cases, this offers a simple, order-of magnitude estimate for $\mathcal{Z}$. Finally, the NLA computes the evidence using the equality

$$\begin{aligned}
\frac{1}{\mathcal{Z}} &= \int \frac{\mathcal{P}(\boldsymbol{\theta})}{\mathcal{L}(\boldsymbol{\theta})}d\boldsymbol{\theta} \\
&= \int_{Y_0}^{Y_N} M(Y)dY + M(Y_0)Y_0,
\end{aligned} \tag{A5}$$

where the last term contains a Lebesgue integral with $Y = \mathcal{L}(\boldsymbol{\theta})^{-1}$ and measure $M(y)$

$$M(y) = \int_{Y(\boldsymbol{\theta})>y} \mathcal{P}(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{A6}$$

This conversion to a Lebesgue integral has the clear advantage of replacing the n-dimensional integral by a 1-dimensional one. This approach is also used by the nested sampling algorithm (Section A.3) where $\mathcal{Z}$ is computed as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)dX; \; X(\lambda) = \int_{\mathcal{L}(\boldsymbol{\theta})>\lambda} \pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{A7}$$

Since the final MCMC sample is distributed as the posterior probability $\mathcal{P}(\boldsymbol{\theta})$, in Equation A5, M can be approximated as $M(Y_i) \approx \frac{1}{N}\sum_{j=1}^N \mathbf{1}_{Y_j>Y_i}$ for each $\mathcal{L}_i$, where $\mathbf{1}$ is the indicator function. With this approximation we have

$$\mathcal{Z} \approx \left(\frac{1}{N}\sum_j \frac{1}{L_j} \right)^{-1}, \tag{A8}$$

which is also known as the harmonic mean estimator (HME). This disadvantages of this estimator are well known in the literature (e.g., Raftery et al. 2007; Calderhead & Girolami 2009). Due to the presence of $1/\mathcal{L}_j$ terms this method is unstable for very small likelihood values that dominate the sum. The proposed solution is to restrict the integration space only to well-sampled regions of high likelihood. Therefore this method suffers from problems intrinsic to MCMC sampling. In addition, Calderhead & Girolami (2009) show that even in well-behaved scenarios, the HME can produce biased (lower) results. To avoid these issues, the nested sampled approach (Equation A7) is the preferred alternative to thermodynamic integration.

The B_XY factor can also be estimated directly using the reverse jump MCMC (e.g., Lopes & West 2004), or the Savage-Dickie density ratio (e.g., Trotta 2007). The reverse jump MCMC is essentially a chain moving between different models, and can be either slow to converge or inaccurate for a small number of samples. The last method can provide high accuracy for nested models, as long as the parameter priors are separable, which is not generally true for our atmospheric models.

To draw the analogy with the frequentist approach, the Bayes factor for nested models can be shown to satisfy the relation (Trotta 2008; Sellke et al. 2001)

$$B_XY \le -\frac{1}{e\mathsf{p}\ln\mathsf{p}}, \tag{A9}$$

where $e = \exp(1)$, and p is the p-value. Equivalently, this probability can be expressed as the number of standard deviations from the mean $x\sigma$, assuming a Gaussian distribution, $\mathsf{p} = \mathrm{erf}(x/\sqrt{2})$. This upper bound is the significance σ value we refer to in our model comparison examples.

A.2. *Markov chain Monte Carlo*

MCMC methods are widely used in investigating multi-dimensional, non-gaussian and highly correlated posteriors, since they don't require any *a priori* assumption about the shape of the posterior probability distribution. The most common form is the Metropolis-Hastigs algorithm, where the chain is created as a random walk towards the region of maximum likelihood. Each sample is generated from a proposal distribution centered on the current point, and accepted with a probability $\mathrm{pr} = \min(1, \mathcal{L}(\boldsymbol{\theta}')/\mathcal{L}(\boldsymbol{\theta}))$. If the new sample is rejected, the position of the chain remains unchanged. The chain is initialized by a first guess $\boldsymbol{\theta_0}$, and after a burn-in period reaches a stationary state where the sample distribution reflects the shape of the posterior (more samples are drawn from high-likelihood regions). The un-normalized posterior distribution is simply the histograms of all the MCMC samples after the burn-in stage, and the marginal probability distributions for all parameters can be derived from it. Although much more efficient than just a simple Monte Carlo technique, MCMC still has a series of drawbacks: the convergence is not easily testable and can require a very large number of samples; due to its Markov chain nature, it is not easily parallelizable in this form; can be sensitive to the initial guess and get stuck in local minima; sample correlation can affect the final distribution.

The affine-invariant MCMC ensemble sampler proposed by Goodman & Weare (2010) solves some of these problems. In this paper we use the version of this algorithm *emcee* implemented in Python by Foreman-Mackey et al. (2013)[1]. This algorithm uses multiple chains, or "walkers" run in parallel for a faster exploration of the parameter space. The K chains are initialized in a n-dimensional Gaussian distribution around the initial guess. At each step, the position

[1] http://dan.iel.fm/emcee/

of a given walker W_i is determined by randomly choosing a different walker from the set W_j and generating the new position $W_j + Z(W_i - W_j)$, with Z is distributed as

$$Z \sim \frac{1}{\sqrt{z}}, \text{ for } z \in \left[\frac{1}{a}, a\right] \text{ and } 0 \text{ otherwise,} \tag{A10}$$

where $a = 2$ is the scale parameter. This new position is accepted with the probability pr = $\min(1, Z^{K-1}\mathcal{L}(\boldsymbol{\theta_i'})/\mathcal{L}(\boldsymbol{\theta_i}))$. Alternate sets of walkers can be updated in parallel, greatly enhancing computing time. This method produces more independent (uncorrelated) samples than the traditional MCMC. Essentially, with a few hundred walkers each iteration can be considered a snapshot of the full posterior, after the burn-in time. The multiple walkers can also more easily spread out to explore the parameter space, such that a large number of iterations is not necessary. We adopted this method for speed, reliability, and ease of implementation for retrieving model parameters. However, it does not provide a direct estimate of $\mathcal{Z}$ and we have to resort to the approximations presented in Section A.1.

A.3. *Multimodal nested sampling*

The multimodal nested sampling method was devised by Skilling (2004), further refined by Shaw et al. (2007); Feroz & Hobson (2008), and implemented into the *MultiNest* package by Feroz et al. (2009)[2], with an easy-to-use Python wrapper (Buchner et al. 2014)[3]. It was initially designed as a tool for more reliable Bayesian evidence calculation, but was also found to provide low-noise estimates of the posterior distribution, and thus constraints on the model parameters.

Nested sampling starts with N "live points" uniformly spaced across the entice initial prior volume, mapped into a unit hypercube. At every iteration i, the "live points" with the lowest likelihood value $\mathcal{L}_i$ are iteratively replaced by requiring that new ones have $\mathcal{L} > \mathcal{L}_i$. In order to ensure that last condition is satisfied, the iso-likelihood contour is approximated by a set of (possibly overlapping) ellipsoids containing the active points, and new samples are drawn from within this new volume until one is found that satisfies $\mathcal{L}' > \mathcal{L}_i$. This new point then replaces the one with $\mathcal{L}_i$ in the set of active points. The volume occupied by the points with $\mathcal{L}_i > \mathcal{L}_{i-1}$ at iteration i is a random variable that can be approximated by its expectation value as $\ln X_i \approx -(i \pm \sqrt{i})/N$ (Feroz & Hobson 2008) and used in the evaluation of the Bayesian evidence $\mathcal{Z}$ as a 1-dimensional integral (Equation A7):

$$\mathcal{Z} = \sum_{i=1}^{M} \mathcal{L}_i w_i + \bar{\mathcal{L}} X_M, \tag{A11}$$

where the last term represents the contribution of the current set of active points, and w_i are the weights for the trapezoidal rule $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$.

The error in $\mathcal{Z}$ is estimated (Skilling 2004) as $\sqrt{H/N}$, where

$$H \approx \sum_{i=1}^{M} \frac{\mathcal{L}_i w_i}{\mathcal{Z}} \ln \frac{\mathcal{L}_i}{\mathcal{Z}}, \tag{A12}$$

and M is the number of iterations. The posterior distribution is approximated by the total set of active and discarded points and their weights $p_i = \mathcal{L}_i w_i/\mathcal{Z}$, where $w_i = X_M/N$ for the active points. The mean and covariance of the parameters are then

$$\bar{\boldsymbol{\theta}} = \sum_{i=1}^{M+N} p_i \boldsymbol{\theta_i}, \tag{A13}$$

$$\boldsymbol{C} = \sum_{i=1}^{M+N} p_i (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T, \tag{A14}$$

In addition to providing the Bayesian evidence as a by-product, *MultiNest* also employs a well-defined convergence criterion that can significantly reduce the number of required posterior samples, and therefore the running time. Convergence is achieved when the estimated change in likelihood $\Delta\mathcal{Z}_i = \max(\mathcal{L}_i)X_i$ is less than a user-specified

---

[2] https://ccpforge.cse.rl.ac.uk/gf/project/multinest/

[3] https://github.com/JohannesBuchner/PyMultiNest

tolerance. Generally, the number of likelihood evaluations until convergence grows exponentially with the number of dimensions of the parameter space. This makes the algorithm unfeasible for a large number of dimensions ($\gtrsim 10$). However, at avery step new samples can be drawn in parallel, significantly increasing computational speed. In practice, we find that *MultiNest* can be run for a much shorter time than *emcee* to converge, mainly because *emcee* does not have a self-stopping criterion and is left to run long enough to cover the entire parameter space and obtain sufficient independent samples. Similar to MCMC, in some cases the acceptance rate is low for *MultiNest*, and therefore convergence is also slow.

ACKNOWLEDGMENTS

REFERENCES

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872

Allison, R., & Dunkley, J. 2014, MNRAS, 437, 3918

Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al. 2014, ApJ, 786, 154

Bond, J. C., Lauretta, D. S., & O'Brien, D. P. 2010, Icarus, 205, 321

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Cahoy, K. L., Marley, M. S., & Fortney, J. J. 2010, ApJ, 724, 189

Calderhead, B., & Girolami, M. 2009, Computational Statistics & Data Analysis, 53, 4028

Cornish, N. J., & Littenberg, T. B. 2007, Phys. Rev. D, 76, 083006

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-prints, arXiv:1306.2144

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Freedman, R. S., Marley, M. S., & Lodders, K. 2008, ApJS, 174, 504

Goodman, J., & Weare, J. 2010, Comm. App. Math. Comp. Sci, 5, 65

Hansen, J. E., & Travis, L. D. 1974, Space Sci. Rev., 16, 527

Helling, C., Woitke, P., Rimmer, P. B., et al. 2014, Life, 4, arXiv:1403.4420

Horak, H. G. 1950, ApJ, 112, 445

Horak, H. G., & Little, S. J. 1965, ApJS, 11, 373

Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, J. Quant. Spec. Radiat. Transf., 109, 1136

Jeffreys, H. 1961, Theory of Probability, Third Edition. (Oxford University Press)

Kane, S. R., Thirumalachari, B., Henry, G. W., et al. 2016, ApJL, 820, L5

Karkoschka, E. 1994, Icarus, 111, 174

Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, Nature, 505, 69

Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014, ApJ, 783, 70

Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137

Lopes, H. F., & West, M. 2004, Statistica Sinica, 14, 41

Marley, M., Lupu, R., Lewis, N., et al. 2014, ArXiv e-prints, arXiv:1412.8440

Marley, M. S., Gelino, C., Stephens, D., Lunine, J. I., & Freedman, R. 1999, ApJ, 513, 879

McKay, C. P., Pollack, J. B., & Courtin, R. 1989, Icarus, 80, 23

Meador, W. E., & Weaver, W. R. 1980, Journal of Atmospheric Sciences, 37, 630

Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, ApJL, 743, L16

Raftery, A. E. 1996, Hypothesis testing and model selection., ed. W. R. Gilks, D. J. Spiegelhalter, & S. Richardson (London, UK: Chapman and Hall), 163–188

Raftery, A. E., Newton, M. A., Satagopan, J. M., & Krivitsky, P. 2007, Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity (with Discussion)., ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Oxford University Press), 1–45

Robinson, T. D., Stapelfeldt, K. R., & Marley, M. S. 2016, PASP, 128, 025003

Rodgers, C. D. 2000, Series on Atmospheric Oceanic and Planetary Physics, Vol. 2, Inverse Methods for Atmospheric Sounding - Theory and Practice (World Scientific Publishing Co. Pte. Ltd.), doi:10.1142/9789812813718

Roman, M. T., Banfield, D., & Gierasch, P. J. 2013, Icarus, 225, 93

Sato, M., & Hansen, J. E. 1979, Journal of Atmospheric Sciences, 36, 1133

Sato, T. M., Satoh, T., & Kasaba, Y. 2013, Icarus, 222, 100

Schwarz, G. 1978, Ann. Stat., 5, 461

Sellke, T., Bayarri, M., & Berger, J. O. 2001, American Statistician, 55, 62

Shaw, J. R., Bridges, M., & Hobson, M. P. 2007, MNRAS, 378, 1365

Simon-Miller, A. A., Banfield, D., & Gierasch, P. J. 2001, Icarus, 154, 459

Skilling, J. 2004, AIP Conference Proceedings, 735, 395

Sudarsky, D., Burrows, A., & Pinto, P. 2000, ApJ, 538, 885

Toon, O. B., McKay, C. P., Ackerman, T. P., & Santhanam, K. 1989, J. Geophys. Res., 94, 16287

Trotta, R. 2007, MNRAS, 378, 72

—. 2008, Contemporary Physics, 49, 71

Weinberg, M. D. 2012, Bayesian Anal., 7, 737

Wong, M. H., Mahaffy, P. R., Atreya, S. K., Niemann, H. B., & Owen, T. C. 2004, Icarus, 171, 153