Astronomy & Astrophysics

Spectral and atmospheric characterization of 51 Eridani b using VLT/SPHERE[⋆,⋆⋆]

M. Samland[1,2], P. Mollière[1,2], M. Bonnefoy[3], A.-L. Maire[1], F. Cantalloube[1,3,4], A. C. Cheetham[5], D. Mesa[6],
R. Gratton[6], B. A. Biller[7,1], Z. Wahhaj[8], J. Bouwman[1], W. Brandner[1], D. Melnick[9], J. Carson[9,1], M. Janson[10,1],
T. Henning[1], D. Homeier[11], C. Mordasini[12], M. Langlois[13,14], S. P. Quanz[15], R. van Boekel[1], A. Zurlo[16,17,14],
J. E. Schlieder[18,1], H. Avenhaus[16,17,15], J.-L. Beuzit[3], A. Boccaletti[19], M. Bonavita[7,6], G. Chauvin[3], R. Claudi[6],
M. Cudel[3], S. Desidera[6], M. Feldt[1], T. Fusco[4], R. Galicher[19], T. G. Kopytova[1,2,20,21], A.-M. Lagrange[3],
H. Le Coroller[14], P. Martinez[22], O. Moeller-Nilsson[1], D. Mouillet[3], L. M. Mugnier[4], C. Perrot[19], A. Sevin[19], E. Sissa[6],
A. Vigan[14], and L. Weber[5]

(Affiliations can be found after the references)

ABSTRACT

Context. 51 Eridani b is an exoplanet around a young (20 Myr) nearby (29.4 pc) F0-type star, which was recently discovered by direct imaging. It is one of the closest direct imaging planets in angular and physical separation (∼0.5″, ∼13 au) and is well suited for spectroscopic analysis using integral field spectrographs.
Aims. We aim to refine the atmospheric properties of the known giant planet and to constrain the architecture of the system further by searching for additional companions.
Methods. We used the extreme adaptive optics instrument SPHERE at the Very Large Telescope (VLT) to obtain simultaneous dual-band imaging with IRDIS and integral field spectra with IFS, extending the spectral coverage of the planet to the complete Y- to H-band range and providing additional photometry in the K12-bands (2.11, 2.25 μm). The object is compared to other known cool and peculiar dwarfs. The posterior probability distributions for parameters of cloudy and clear atmospheric models are explored using MCMC. We verified our methods by determining atmospheric parameters for the two benchmark brown dwarfs Gl 570D and HD 3651B. We used archival VLT-NACO (L') Sparse Aperture Masking data to probe the innermost region for additional companions.
Results. We present the first spectrophotometric measurements in the Y and K bands for the planet and revise its J-band flux to values 40% fainter than previous measurements. Cloudy models with uniform cloud coverage provide a good match to the data. We derive the temperature, radius, surface gravity, metallicity, and cloud sedimentation parameter f_{sed}. We find that the atmosphere is highly super-solar ([Fe/H] = 1.0 ± 0.1 dex), and the low $f_{sed} = 1.26^{+0.36}_{-0.29}$ value is indicative of a vertically extended, optically thick cloud cover with small sized particles. The model radius and surface gravity estimates suggest higher planetary masses of $M_{gravity} = 9.1^{+4.9}_{-3.3} M_J$. The evolutionary model only provides a lower mass limit of >2 M_J (for pure hot-start). The cold-start model cannot explain the luminosity of the planet. The SPHERE and NACO/SAM detection limits probe the 51 Eri system at solar system scales and exclude brown-dwarf companions more massive than 20 M_J beyond separations of ∼2.5 au and giant planets more massive than 2 M_J beyond 9 au.

Key words. stars: individual: 51 Eridani – planets and satellites: atmospheres – methods: data analysis – techniques: high angular resolution – techniques: image processing

1. Introduction

The number of extrasolar giant planets found with ground-based high-contrast imaging techniques is growing steadily (e.g., Marois et al. 2008, 2010b; Lagrange et al. 2010; Rameau et al. 2013; Bailey et al. 2014; Wagner et al. 2016) and the advent of dedicated high-contrast imaging instruments such as Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE; Beuzit et al. 2008) and Gemini Planet Imager (GPI; Macintosh et al. 2014) has made it possible to study and characterize these planets and substellar companions in detail with

low- to mid-resolution spectrometry and/or narrowband photometry (e.g., Ingraham et al. 2014; Chilcote et al. 2015; Apai et al. 2016; Maire et al. 2016a; Vigan et al. 2016; Zurlo et al. 2016). At the same time, modeling of giant planet and brown dwarf atmospheres has made important progress with the development of cloudy models for colder objects (e.g., Allard et al. 2012; Morley et al. 2012; Baudino et al. 2015).

51 Eridani b is the first discovered planet with the GPI instrument (Macintosh et al. 2015) and was characterized using J- and H-band spectra taken with GPI and Keck/NIRC2 photometry in the L' band. This object occupies a unique place in parameter space as a young, low-mass ($M < 10 M_J$), methane-rich, cold (∼700 K), but seemingly cloudy planet. This peculiar object is located at an angular separation from its host star ($\rho \sim 0.5″$), which is suited for spectroscopic characterization within the small field of view (FoV) of integral field spectrographs (IFS), but far enough away to achieve good signal-to-noise ratio (S/N) despite its contrast. Given these characteristics,

it will become a benchmark object for current and future instruments as well as for the calibration of atmospheric models.

Its host star is part of a multiple system together with an M-dwarf binary (Montet et al. 2015) and is located in the well-studied β Pictoris moving group (Zuckerman et al. 2001). The age estimates range from 12 to 23 million years (Myr; e.g., Simon & Schaefer 2011; Binks & Jeffries 2014; Mamajek & Bell 2014; Bell et al. 2015), and we follow the adopted age of the discovery paper as 20 ± 6 Myr for all components of the system. A recent dynamical mass estimate of the distant binary M-dwarf companion GJ 3305 predicts an older age of the GJ 3305 AB system of 37 ± 9 Myr. An astrometric follow-up paper by De Rosa et al. (2015) has confirmed that the planet is co-moving with 51 Eri. The tentative orbital solutions (semi-major axis $a = 14^{+7}_{-3}$ au, orbital period $T = 41^{+35}_{-13}$ yr, inclination $i = 138^{+15}_{-13}$) suggest that the planet does not share the inclination of the distant M-dwarf companion GJ 3305 (Montet et al. 2015).

The host star also has an infrared excess that can be modeled by two components corresponding to a warm belt of debris at 5.5 AU and another colder one at 82 AU (Patel et al. 2014; Riviere-Marichalar et al. 2014). As such, the architecture of 51 Eri is reminiscent of our solar system and of other benchmark systems such as HR 8799 and HD 95086.

In this work we present new near-infrared (NIR) spectra and photometric data obtained with the SPHERE instrument at the Very Large Telescope (VLT) in Chile, as part of the consortium guaranteed-time exoplanet imaging survey SHINE (SpHere INfrared survey for Exoplanets; Chauvin et al., in prep.). The SPHERE observations are described in Sect. 2 and the data reduction in Sect. 3. The spectrophotometric analysis of 51 Eri b is discussed in detail in Sect. 4.

Finally, we present sensitivity limits to additional closer-in companions in Sect. 5, extended to the innermost region by archival Sparse Aperture Masking (SAM) data taken with the VLT-NACO instrument in the L' band, and end with our summary and conclusions in Sect. 6. The astrometric analysis of the planet is deferred to a future paper.

2. Observations

The SPHERE (Beuzit et al. 2008) instrument is an extreme adaptive optics system (SAXO; Fusco et al. 2014) that feeds three science instruments: the infrared dual-band imager and spectrograph (IRDIS; Dohlen et al. 2008), an IFS (Claudi et al. 2008), and the visible light imaging polarimeter (ZIMPOL; Thalmann et al. 2008). We observed 51 Eri four times between September 2015 and January 2016 as part of the SPHERE GTO program using IRDIS in the dual-imaging mode (DBI; Vigan et al. 2010) and the IFS operating simultaneously (IRDIFS and IRDIFS_EXT modes, see Table 1). The observations were obtained with an apodized pupil Lyot coronagraph (Soummer 2005; Boccaletti et al. 2008), consisting of a focal mask with a diameter of 185 milli-arcsec. The pupil stabilized mode was used close to meridian passage to exploit angular differential imaging (ADI) post-processing (Marois et al. 2006) with the goal of attenuating residual speckle noise. The usual SPHERE survey observation strategy was employed: 1) photometric calibration: imaging of star offset from coronagraph mask to obtain PSF for relative photometric calibration at the beginning and end of the observation sequence; 2) centering: imaging with the star behind the coronagraphic mask with four artificially induced satellite spots using the deformable mirror (Langlois et al. 2013) for deriving the star center location directly before and after the science sequence; 3) coronagraphic

sequence; 4) sky background observation using same configuration as coronagraphic sequence. Finally, north angle offset and pixel scale were determined using astrometric calibrators as part of the SPHERE GTO survey for each run (Maire et al. 2016b). All the other calibration files (e.g., dark, flat, and spectral calibration) are obtained during the day following the observation via the instrument internal calibration hardware. Four IRDIS observations were conducted in three different filter setups: once in broadband H (BB_H), twice in dual-band H23 (H2 $\lambda_c = 1589$ nm, $FWHM = 53$ nm; H3 $\lambda_c = 1667$ nm, $FWHM = 56$ nm), and once in dual-band K12 (K1 $\lambda_c = 2103$ nm, $FWHM = 102$ nm; K2 $\lambda_c = 2255$ nm, $FWHM = 109$ nm, see also Table 2). The YJ setup (YJ: 0.95–1.35 μm, spectral resolution $R \sim 54$) was used three times and the YH mode (YH: 0.95–1.65 μm, spectral resolution $R \sim 33$) once. Observing conditions were variable for the two September data sets, but yielded the best data quality. Both December and January observations were conducted in bad seeing conditions with a strong jet stream that caused saturation near the edge of the coronagraphic mask when using the standard exposure times.

3. Data reduction and spectrophotometric extraction

Basic reduction of both the IRDIS and IFS data (background subtraction, flat fielding, bad pixel removal, centering, and spectral calibration for IFS) was performed using the pipeline of the SPHERE data center hosted at the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) using the SPHERE Data Reduction Handling (DRH) pipeline (version 15.0; Pavlov et al. 2008). The calibrated output consists of data cubes for each waveband, recentered onto a common origin using the satellite spot reference. The unsaturated stellar PSF frames taken before and after the coronagraphic sequence were reduced via the same routines and also corrected for the neutral density filter transmission[1]. The variation of the stellar flux measurement of the host star ($\lesssim 5\%$ for all used data) is propagated in the uncertainties of the companion flux measurement.

3.1. IFS data reduction and spectra extraction

In addition to the DRH pipeline, custom IDL routines (Mesa et al. 2015) have been used for the basic reduction following Maire et al. (2016a) with an additional step added to further refine the wavelength calibration using the shift in satellite spot position. The analysis of DBI and/or IFS data with aggressive spectral differential imaging (SDI; Racine et al. 1999) algorithms, such as algorithms that include all other spectral channels as reference to model the speckle pattern, may lead to wavelength-dependent biases in the signal of a planet that cannot be modeled in a straightforward way (Maire et al. 2014). In order to avoid biasing the extracted spectrum while still retaining a good S/N, we opted for a non-aggressive method for the removal of the speckle noise in two steps. We first reduced the data via ADI post-processing and noted the channels in which, because of the peak in methane and water absorption, flux is neither expected nor observed. We then went back to the cosmetically reduced data cubes and used these selected channels as reference for a first classical SDI (cSDI) step, i.e., scaled to the same λ/D and mean flux for each channel, and we subtract these from all other channels. For the YJ spectrum, we used the

---

[1] https://www.eso.org/sci/facilities/paranal/instruments/sphere/doc.html

Table 1. Observing log.

UT date	Instr. mode	IRDIS filter	IFS band	IRDIS DIT[a] (s, #)	IFS DIT[a] (s, #)	T_{exp}[b] (min)	Field Rot.[c] (deg)	Sr[d]
2015-09-25	IRDIFS_EXT	K12	YH	16×256	16×256	68.3	41.66	0.80–0.90
2015-09-26	IRDIFS	*H*	YJ	4×918	64×64	68.3	43.64	0.80–0.90
2015-12-26	IRDIFS	H23	YJ	16×256	32×128	68.3 (34.7[b])	37.19	0.75–0.85
2016-01-16	IRDIFS	H23	YJ	16×256	64×64	68.3	41.76	0.75–0.90

Notes. [a] Detector integration time. [b] Exposure time after bad frame removal; about half the frames are unusable. [c] All observation were centered on the meridian passage of the target with an airmass between 1.08 and 1.10. [d] Strehl ratio measured at 1.6 μm as measured by AO system.
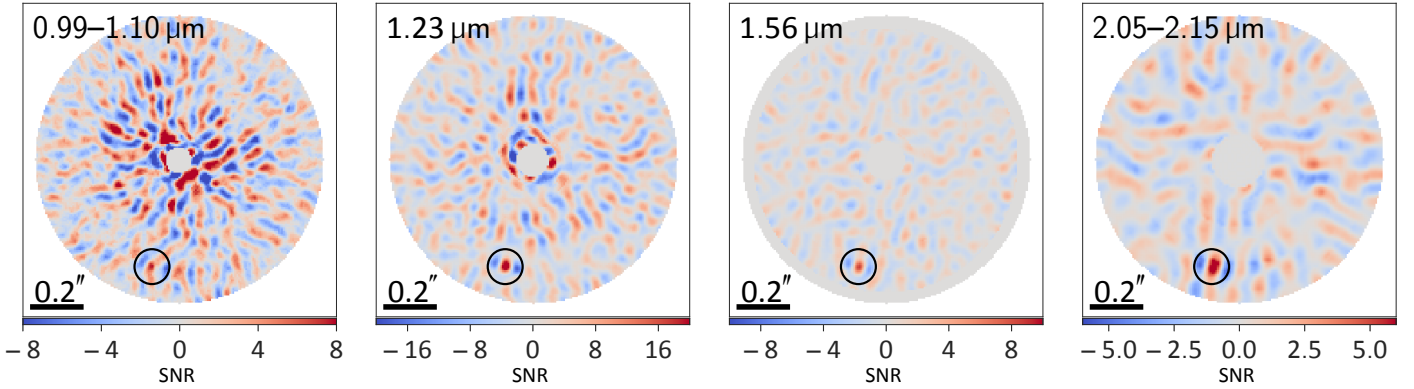

Fig. 1. S/N maps created by ANDROMEDA for IFS and IRDIS K1. The maps are in order of ascending wavelength. The first two are extracted from the YJ-IFS data and the third from the YH-IFS data. The *Y*-band image (*left*) shows the median combined map between 0.99–1.10 μm as S/N is low, whereas the second and third image, which correspond to the peak in *J* and *H*, show single channels. The *right panel* shows the IRDIS K1 filter. Standard astronomical orientation is used, where up is north and left is east. The black circle indicates the position of the planet. The azimuthal negative wings around the planet signal is the characteristic planet signature that ANDROMEDA is fitting for in ADI data and not undesirable self-subtraction as in the PCA/LOCI approach (Cantalloube et al. 2015).

channels between the *Y* and *J* band (1.11–1.17 μm) as reference. For the YH spectrum, we used the 1.14 μm channel to reduce all shorter wavelengths (*Y* band) and the 1.41 μm channel for the rest of the spectrum (*J* and *H* band). Because the YH spectrum spans a big wavelength range, we used two different reference channels, depending on the wavelength, to ensure that the effect of chromatic aberration on the speckle subtraction was minimized. These SDI preprocessed data cubes with attenuated speckle noise were then used as input for the following ADI reduction via various algorithms. We used the Specal pipeline (R. Galicher, priv. comm.) developed for the SHINE survey as a first-quick look reduction. For the spectral extraction we tested three different reduction approaches: PCA (Soummer et al. 2012; Amara & Quanz 2012, Specal implementation used), TLOCI (Marois et al. 2014, Specal implementation used), and ANDROMEDA (Cantalloube et al. 2015). We chose to focus our analysis on the spectra extracted with the ANDROMEDA algorithm, which was used for the first time on SPHERE/IFS data. This algorithm provides robust YJ and YH spectra and has a number of advantages compared to other reduction methods. In ANDROMEDA the signal is explicitly modeled, therefore no post-processing is necessary to extract an unbiased planetary signal, S/N, and detection limits, i.e., no self-subtraction correction by injection of artificial signals (see, e.g., Lagrange et al. 2010; Marois et al. 2010a) is needed. Furthermore, in contrast to other methods, ANDROMEDA has only one tunable parameter N_{smooth} (set to 8 pixels) and it only marginally affects the determined noise level at close separations (thus the S/N of a detection) and could affect the astrometry, but not the signal itself. We confirmed that 51 Eri b is located far enough from the center for none of this to be the case. Cantalloube et al. (2015) lists additional parameters, but these are either set directly

by the wavelength of the observation or can be set to default owing to the much higher stability of SPHERE compared to NACO. As such, the ANDROMEDA reduction is very reproducible in the sense that it is less prone to subjective choices of parameters that influence the data reduction. Figure 1 shows the planet in the ANDROMEDA reduction at four different wavelengths, corresponding to the *Y*, *J*, *H*, and K1 feature.

In addition to the reductions for every spectral channel, we also produced a collapsed "detection image" (see Fig. 2) to measure precisely the position of the planet at high S/N and to look for additional companions. For these images, instead of median combining all spectral channels, we follow the method introduced by Thiébaut et al. (2016). We first produce S/N maps for all spectral channels, which are a by-product of the ANDROMEDA algorithm, threshold them above zero, and sum up the squares of the thresholded S/N maps to make clean, collapsed images. In addition to getting the precise position for a planet, the advantage of this approach is that it is not necessary to make any assumptions about the exact spectral shape of a potential point source and it is suitable for visual inspection for further potential candidates.

The spectra and photometry extracted using ANDROMEDA and used for the atmospheric characterization are shown in Fig. 4 and discussed in Sect. 3.4. Reductions with alternative algorithms are shown and further discussed in Appendix A.

3.2. Broad and dual-band imaging

In addition to the basic reduction the cubes are corrected for distortion and for the north angle offset determined from the astrometric calibrations (Maire et al. 2016b). For the unsaturated
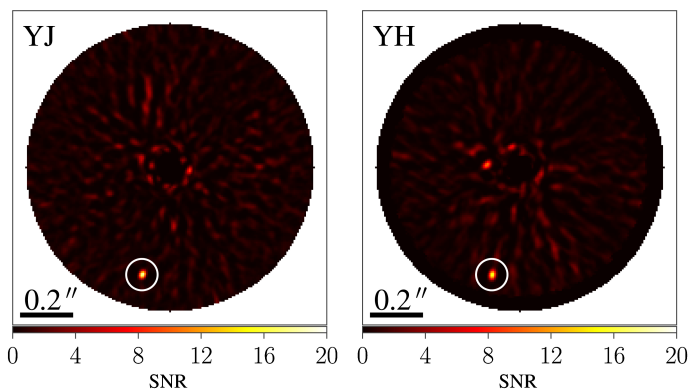
Fig. 2. Detection images made using the method described in Thiébaut et al. (2016) with the S/N maps created by ANDROMEDA for IFS-YJ and YH data sets. The *left image* shows the collapsed image for the YJ band, the *right image* YH band. The circle white indicates the position of 51 Eri b.
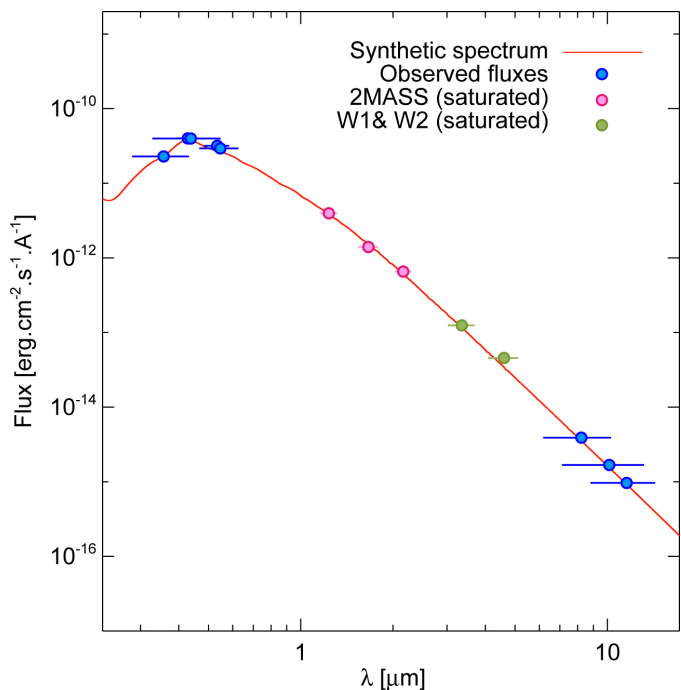

Fig. 3. BT-NextGen synthetic spectrum of host star 51 Eri, scaled to match SED of optical and mid-infrared photometry. Owing to saturation of the star in these bands, the 2MASS J, H, and K_s, as well as W1-W2 were excluded from the fit.

calibration frames of IRDIS, we used a custom routine that does not interpolate bad pixels in the PSF from the surrounding pixels, but replaces the respective bad pixels with the value obtained by fitting a Moffat function to the PSF in all frames.

The cosmetically cleaned and centered data cubes were then used as input for further ADI/SDI post-processing pipelines. We again used three different approaches to reduce the data: PCA, MLOCI (Wahhaj et al. 2015), and ANDROMEDA. All three data reductions were consistent within their respective uncertainties. However, we chose to use ANDROMEDA for the final reduction of the IRDIS photometry presented here to be consistent with the IFS reduction.

Additionally we include the broadband L' photometric data point observed with the W. M. Keck Observatory Near Infrared Camera 2 (NIRC2; L' band, $\lambda_c = 3780$ nm, $FWHM = 700$ nm) reported in Macintosh et al. (2015). The absolute magnitude $L' = 13.82 \pm 0.27$ mag was converted to flux $f_{L'} = (1.82 \pm 0.45) \times 10^{-17}$ W m^{-2} μm^{-1} with the same distance as for the rest of the analysis (29.4 $\pm$ 0.3 pc).

3.3. Conversion of the planet contrasts to physical fluxes

In order to convert the measured star to planet contrast in IFS and IRDIS data to physical fluxes we used a synthetic photometry approach. This can be summarized in three steps:

1. We built the spectral energy distribution (SED) of the star (see Fig. 3) from Tycho B_T, V_T (Hoeg et al. 1997), Johnson filter U, V, B (Mermilliod 2006), WISE W3 photometry (Cutri et al. 2013), and IRAS 12 μm photometry (Helou & Walker 1988). The 2MASS J, H, K_s (Cutri et al. 2003) as well as W1-W2 photometry could not be used because of saturation of the central region of the star. The 2MASS K_s band is not flagged as saturated in the catalog, but can clearly be seen to be saturated in the individual 2MASS K_s-band images. On the other hand, W4 had to be excluded owing to noticeable infrared excess.
2. We scaled a BT-NextGen model (Allard et al. 2012) with $T_{\rm eff} = 7200$ K, $\log g = 4.0$ dex, and [Fe/H] = 0.0 dex, to fit the above mentioned flux values via χ^2 minimization. The chosen model parameters are close to those determined from high-resolution spectra of the star ($T_{\rm eff} = 7256$ K, $\log g = 4.13$ dex, and [Fe/H] = 0.0 dex; Prugniel et al. 2007).

3. We determined the mean stellar flux in the used SPHERE/IRDIS bandpasses and IFS bins (YJ: 0.95–1.35 μm, spectral resolution $R \sim 54$; YH: 0.95–1.65 μm, spectral resolution $R \sim 33$), by applying the spectral response curve of the instrument, i.e., the normalized wavelength dependent end-to-end transmission including optical elements (e.g., beam-splitters and coronagraph) and filters to the flux-calibrated synthetic spectra. We used the whole spectral response curve for the IRDIS bands. For IFS, because the spectral response is almost flat inside each respective spectral channel, we approximated the spectral response as a Gaussian of a width corresponding to the resolution of the spectrograph in the respective mode.

Our approach differs from that taken in Macintosh et al. (2015), in that we use a stellar atmosphere model for the flux calibration SED and not a blackbody spectrum. Comparing the two approaches over the NIR wavelength range of interest, we observe deviations due to spectral features on the order of $\sim$3%.

3.4. Spectrum of 51 Eridani b

The SED of 51 Eri b showing all of our observations is presented in Fig. 4. Our IRDIS photometry is summarized in Table 2, where values are given for ADI and SDI plus ADI data reduction. For completeness we also plot the GPI spectra published in the discovery paper (Macintosh et al. 2015). With the SPHERE data, we extended the spectral coverage of the atmosphere to the Y band, provided the first photometry in the K band, and substantially improved the S/N in the J band. All of these are of paramount importance for deriving atmospheric parameters and cloud characteristics, as is discussed in Sects. 4.1 and 4.2. In further analysis we use both IFS spectra, the four ADI-only narrowband photometric data points in the H and K band. Additionally we also use the GPI-H spectrum, as it extends the
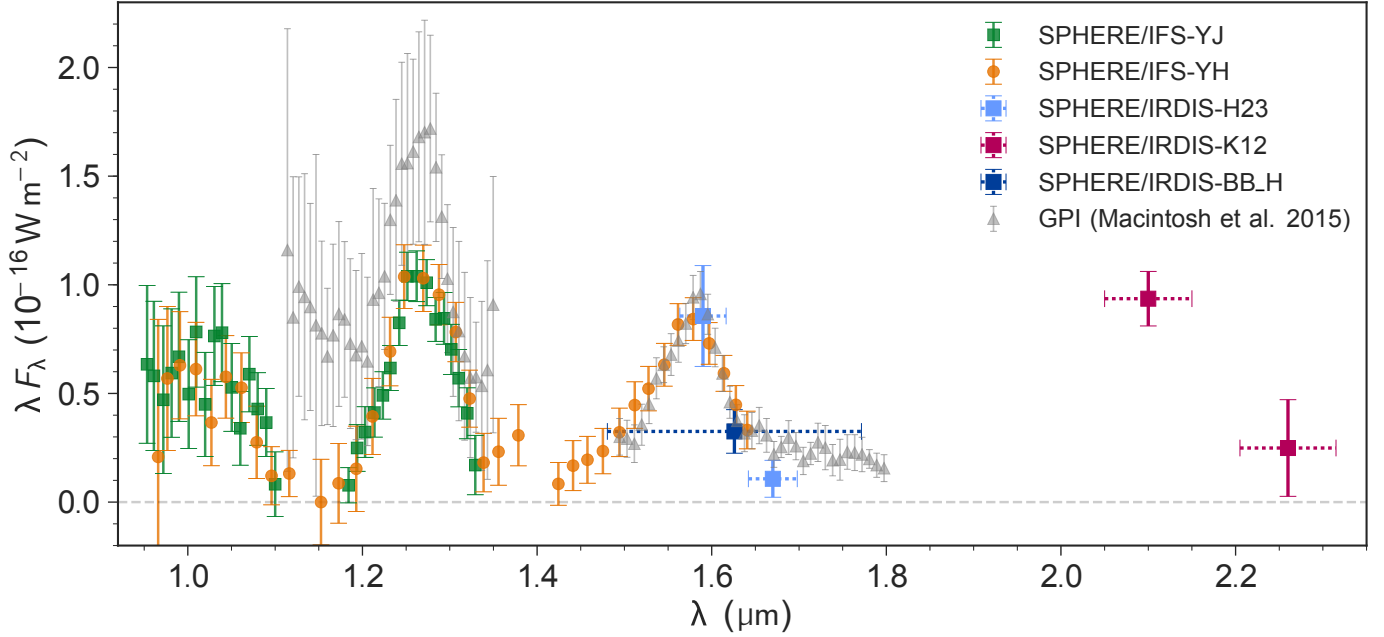
Fig. 4. Spectral energy distribution for 51 Eri b constructed from our YJ-, YH-IFS spectra and IRDIS photometry in H23, BB_H, and K12. Channels used as reference for classical SDI were removed as they are biased. In addition to the SPHERE data we plot the two GPI spectra in the *J* and *H* band, respectively. The flux in the *J* band is consistent between the two SPHERE spectra, but significantly different compared to GPI. Uncertainties are given as 1σ and are assumed to be Gaussian.

Table 2. IRDIS photometry.

Filter	λ (μm)	Width (μm)	Contrast	err. contrast	App. flux (Wm$^{-2}\,\mu$m^{-1})	Err. flux (Wm$^{-2}\,\mu$m^{-1})	Abs. magnitude[a]
ADI							
BB_H	1.626	0.291	1.20×10^{-6}	3.70×10^{-7}	2.00×10^{-17}	6.17×10^{-18}	17.11 ± 0.29
H2	1.589	0.048	3.10×10^{-6}	8.39×10^{-7}	5.39×10^{-17}	1.46×10^{-17}	16.07 ± 0.26
H3[b]	1.667	0.056	4.25×10^{-7}	3.36×10^{-7}	6.42×10^{-18}	5.09×10^{-18}	>17.59
K1	2.103	0.102	6.73×10^{-6}	9.02×10^{-7}	4.46×10^{-17}	5.97×10^{-18}	15.21 ± 0.14
K2[b]	2.255	0.109	2.08×10^{-6}	1.64×10^{-6}	1.10×10^{-17}	9.84×10^{-18}	>15.79
SDI+ADI							
H2–H3	1.589	0.048	2.32×10^{-6}	3.99×10^{-7}	4.04×10^{-17}	6.94×10^{-18}	16.39 ± 0.17
K1–K2	2.103	0.102	5.90×10^{-6}	5.04×10^{-7}	3.91×10^{-17}	2.67×10^{-18}	15.35 ± 0.08

Notes. Values under header ADI are obtained using ADI processing only. Values under SDI+ADI are obtained using SDI followed by ADI. Uncertainties are given as 1σ. The contrast uncertainties include speckle noise as dominant noise term and the variation in measured host star flux based the two unsaturated stellar images as a minor contribution. [a] With distance modulus $\mu = m - M = 2.34$ using Vega magnitude system. Distance uncertainty is negligible in magnitude measurement. [b] Forced photometry: magnitude 1σ upper limits obtained by adding the respective flux measurement and uncertainty.

wavelength coverage of the *H* band toward longer wavelengths, as well as the *L'* band photometry of Macintosh et al. (2015). We are not using the broadband *H*-band observation in our later analysis because it does not further constrain the spectral shape. The SPHERE YH spectrum is in excellent agreement in the overlapping part with the IRDIS H2 and BB_H photometry taken on different dates, as well as the previous high S/N *H*-band spectrum obtained by GPI. The only discrepancy we see is between the *J*-band flux reported in Macintosh et al. (2015) and our data. The question presenting itself is therefore the origin of the difference observed in the *J* band for which there are two possibilities: first, strong (~40%) atmospheric variability in the atmosphere of the planet; or second, systematic offsets in the absolute calibration between the data sets. The *J* band is known to be more sensitive to temporal amplitude changes in the atmosphere

of L/T-type objects than the *H* and *K* bands (e.g., Radigan et al. 2012; Biller et al. 2015), but even so, given that we see no significant difference in the *H* band, we think it is unlikely that 40% variability in the *J* band is in agreement with our consistent values for the *H*-band flux. We therefore believe that the difference in the *J* band between our observations and previous observations is a result of systematics. The reduction of the YH spectrum via different algorithms shows consistent results (Fig. A.1) giving us confidence in the overall reliability of the data, data reduction, and calibration.

While we use the ANDROMEDA method in this paper, (Macintosh et al. 2015) uses TLOCI. We also compared different reduction methods in Appendix A and notice a difference in *J*-band flux between our two data sets depending on the reduction method; for example, using ANDROMEDA, the flux

measured in the two data sets is within 10%, which is consistent in their respective uncertainties, while for TLOCI this differs by ~40%. It is possible that absolute calibration is more difficult for the other algorithms compared here because additional steps to account for algorithm throughput are necessary.

4. Spectrophotometric analysis

4.1. Empirical comparison to known objects

Our SPHERE YJ and YH spectra of 51 Eri b confirm the presence of several deep water and methane absorption bands typical of T-dwarfs from 1.1 to 1.2, 1.3 to 1.5, and longward of 1.6 μm. We compared the YH spectrum and K1 photometry of 51 Eri b to that of L- and T-type objects from the SpeXPrism library (Burgasser 2014) completed with spectra from Mace et al. (2013) and Best et al. (2015). The comparison spectra were smoothed to the resolution of the IFS YH spectrum and their flux was integrated within the wavelength intervals covered by each channel of the IFS and the IRDIS K1 filter. We used the G goodness of fit indicator for the comparison (Cushing et al. 2008), with the implementation following Vigan et al. (2016), accounting for the filter widths and the uncertainty on the 51 Eri b spectrophotometry.

The results are shown in Fig. 5. The best fits are obtained for late-L and early-T objects, in agreement with the placement of the planet in color-color and color–magnitude diagrams (see below). The best fitting object is PSO J207.7496+29.4240, a peculiar T0 object, and possibly an unresolved binary, from the Best et al. (2015) library. A visual inspection of the fit reveals that while the object is able to reproduce the overall JHK spectral slope, 51 Eri b has deeper methane plus water absorptions. The fit of the YH+K1 spectrophotometry is influenced by the strong overluminosity of the $K1$ band caused by the reduced collision-induced absorption (CIA) of H_2. The Y-band flux is also known to be modulated by the surface gravity and metallicity (Burgasser et al. 2006a; Liu et al. 2007). To mitigate this in the comparison to higher log g objects, we decided to rerun the fit on the YJ spectrum and on the part of the YH spectrum excluding the Y band (hereafter JH spectrum). The results are shown in Fig. 6.

The T7–T8 objects represent the best match to the planet YJ spectrum only. Among the sample of T5.5–T7 objects, the brown dwarfs SDSSpJ111010.01+011613.1, 2MASSIJ0243137-245329, 2MASSJ12373919+6526148, and 2MASSIJ1553022+153236 are minimizing G and therefore represent the best fits to the YJ spectrum. SDSSpJ111010.01+011613.1 and 2MASSIJ0243137-245329 belong to the growing class of red T dwarfs (Gagné et al. 2014; Stephens et al. 2009). SDSSpJ111010.01+011613.1 has been proposed as a member of the AB Doradus moving group (Gagné et al. 2015). The two other objects are a binary (unresolved in the SpeX slit; Burgasser et al. 2006c) and a magnetically active object with strong H_α emission, respectively; these display some variability in the J band (Burgasser et al. 2000; Artigau et al. 2003; Kao et al. 2016).

The JH spectrum is best represented by SDSSJ141530.05+ 572428.7, a T3 dwarf from the SpeXPrism libraries, which is again a candidate unresolved binary (Burgasser et al. 2010). Therefore, there appears to be a correlation between the spectral type of the best fit template found and the maximum wavelength of the photometric points included in the fit. We interpret this as, first, the consequence of the unusual red slope of the near-infrared SED of the planet compared to the templates and, second, the fit limited to the shortest wavelengths becomes more
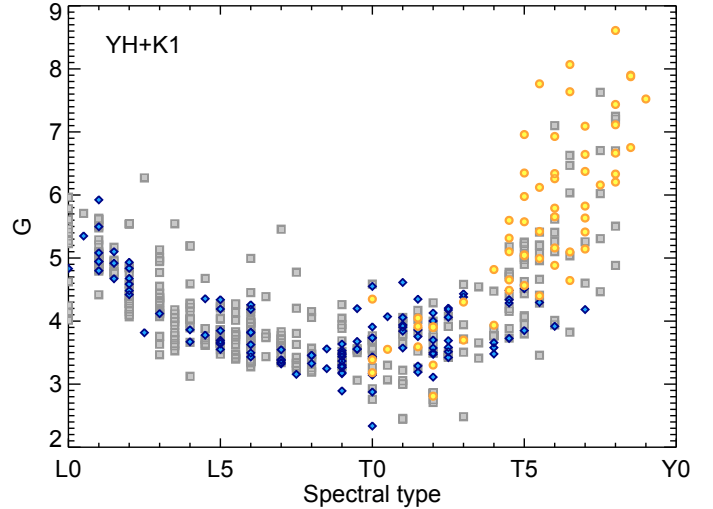
Fig. 5. Goodness of fit G for the comparison of 51 Eri b YH+K1 spectrophotometry with that of template spectra of L and T dwarfs from the SpeXPrism (gray squares), Mace et al. (2013) yellow circles, and Best et al. (2015; blue diamonds) libraries.

Table 3. IFS photometry.

Filter	λ (μm)	Width (μm)	Contrast (10^{-6})	Abs. magnitude[a]
J	1.245	0.240	1.03 ± 0.67	17.40 ± 0.71
J3	1.273	0.051	2.22 ± 0.53	16.52 ± 0.26
H2	1.593	0.052	2.70 ± 0.70	16.22 ± 0.28

Notes. Photometric magnitudes for IRDIS filters derived from IFS spectra. Uncertainties are given as 1σ. [a] With distance modulus $\mu = m - M = 2.34$ using Vega magnitude system. Distance uncertainty is negligible in magnitude measurement.

sensitive to the $CH_4 + H_2O$ absorption from 1.1 to 1.2 μm, which is characteristics of late-T dwarfs. To conclude, the planet SED is often reproduced by candidate unresolved binaries, which is a class of objects that was also found to provide a good fit to the HR 8799b and c planets (Bonnefoy et al. 2016).

We took advantage of the SPHERE spectra to generate synthetic photometry for the narrowband filters of SPHERE overlapping with the wavelength range of the IFS spectra (assuming simple top-hat profile): J ($\lambda_c = 1245$ nm, $FWHM = 240$ nm), J3 ($\lambda_c = 1273$, nm, $FWHM = 51$ nm), and H2 ($\lambda_c = 1593$ nm, $FWHM = 52$ nm). Photometric magnitudes in these bands enable a homogeneous comparison of the planet properties with those of known reference objects.

The photometry was obtained considering a flux calibrated spectrum of Vega (Bohlin 2007) and ESO Skycalc web application[2] (Noll et al. 2012; Jones et al. 2013). We find $J = 19.74 \pm 0.71$ mag, J3 $= 18.86 \pm 0.26$ mag, and H2 $= 18.56 \pm 0.28$ mag (Table 3). We combined this synthetic photometry with that obtained in K1 (17.55 ± 0.14 mag) to show the position of the planet in color–color and color–magnitude diagrams (Figs. 7 and 8). The CMD are built with low-resolution spectra taken from the literature and published parallaxes. For the field dwarfs, we used the spectra from Leggett et al. (2000) and from the SpeXPrism library (Burgasser 2014). The SpeXPrism spectra were
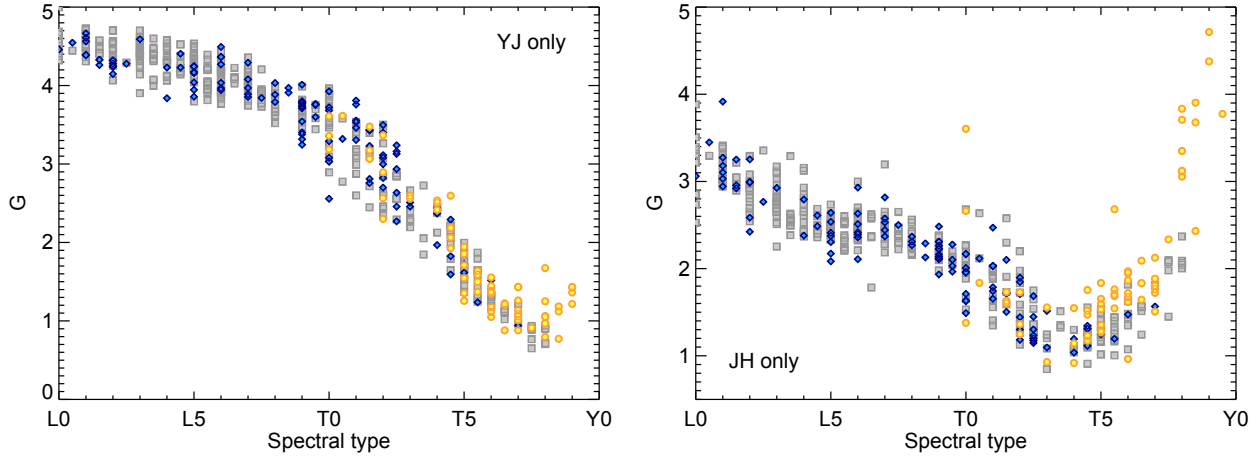
---
[2] http://www.eso.org/observing/etc/bin/gen/form?INS.MODE=swspectr+INS.NAME=SKYCALC

Fig. 6. Same as Fig. 5 but considering the YJ spectrum (*left*) and JH spectrum (*right*) of 51 Eri b only.
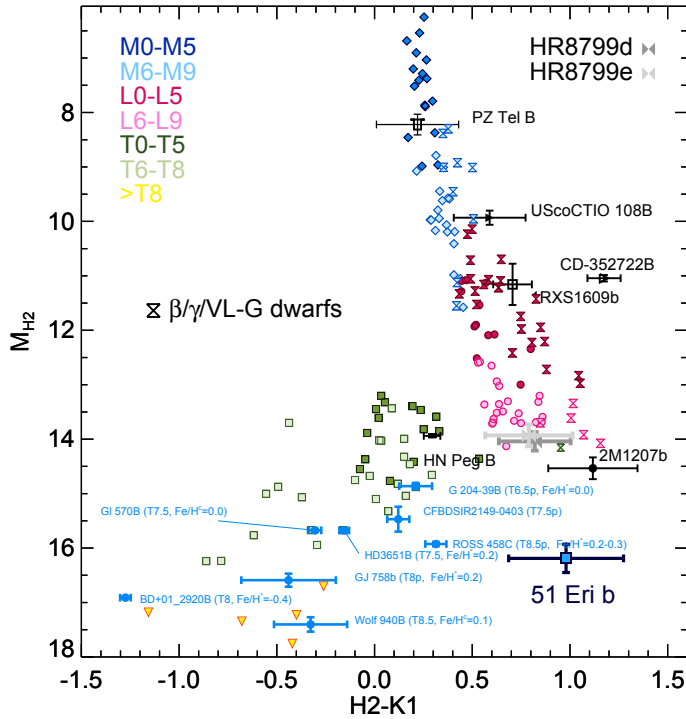

Fig. 7. Placement of 51 Eri b in color–magnitude diagram. All magnitudes except for K1 are derived from the IFS spectra.
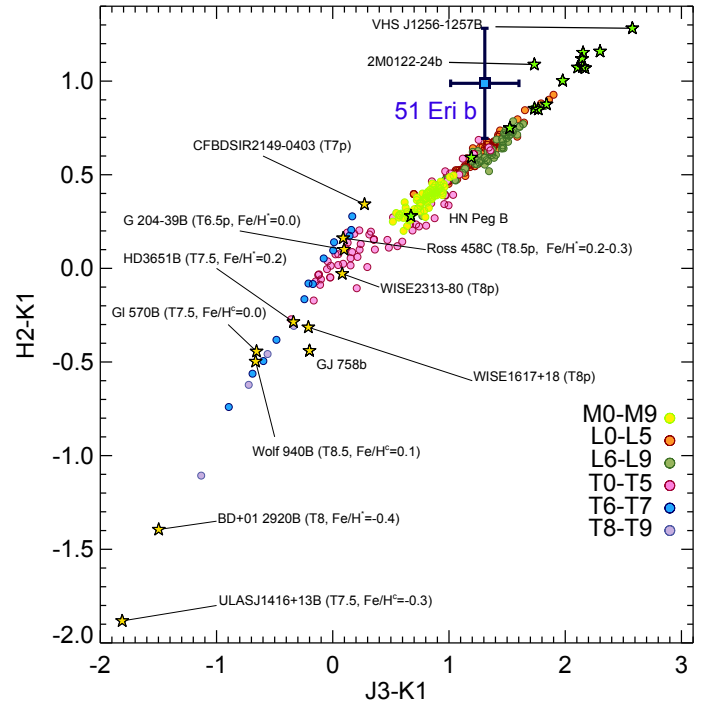

Fig. 8. Placement of 51 Eri b in color–color diagram. All magnitudes except for K1 are derived from the IFS spectra.

calibrated in flux using the *H*-band 2MASS photometry of the targets. We used parallaxes from the literature (mostly from Monet et al. 1992; Faherty et al. 2012) and newly revised values from Liu et al. (2016) where applicable. We repeated this procedure for young, low-gravity and/or dusty M, L, and T dwarfs (spectra taken for the most part from Allers & Liu 2013 and parallaxes from Faherty et al. 2012 and Zapatero Osorio et al. 2014). We added the known T-type companions (and the isolated object CFHTBD2149; Delorme et al. 2017) with known distances and with some knowledge of their metallicity either from the primary star [Fe/H]* or from the companion spectrum [Fe/H]^c. Bonnefoy et al. (in prep., and references therein) provide a full description.

The planet has the luminosity of T6–T8 dwarfs but much redder colors that are consistent with those of late-L dwarfs in J3/J3-K1 and H2/H2-K1 color–magnitude diagrams (CMD). In these diagrams, the benchmark T6.5−T8.5 objects

suspected to be metal rich and/or younger than the field (CFBDSIRJ214947.2-040308.9, GJ 758b, ROSS 458C, SDSSJ175805.46+463311.9/G 204-39B; Delorme et al. 2012; Vigan et al. 2016; Burningham et al. 2011; Faherty et al. 2010) also have redder colors than the sequence of field dwarfs. Although they are not as red as those of 51 Eri b. In color–color diagrams (Fig. 8), 51 Eri b falls at the location of the L/T transition objects in color–color diagrams, although the planet luminosity and the presence of a methane bands in its spectrum is inconsistent with this object being at the L/T transition. Instead, it suggests that the object has a color deviation that is similar to the color deviation seen for young and/or dusty late-L dwarfs (green stars in Fig. 8) with respect to regular late-L dwarfs. The peculiar T7 dwarf CFBDSIRJ214947.2-040308.9 is also deviating from the sequence of T dwarfs but to a lower extent. We interpret the deviation for 51 Eri b as a consequence
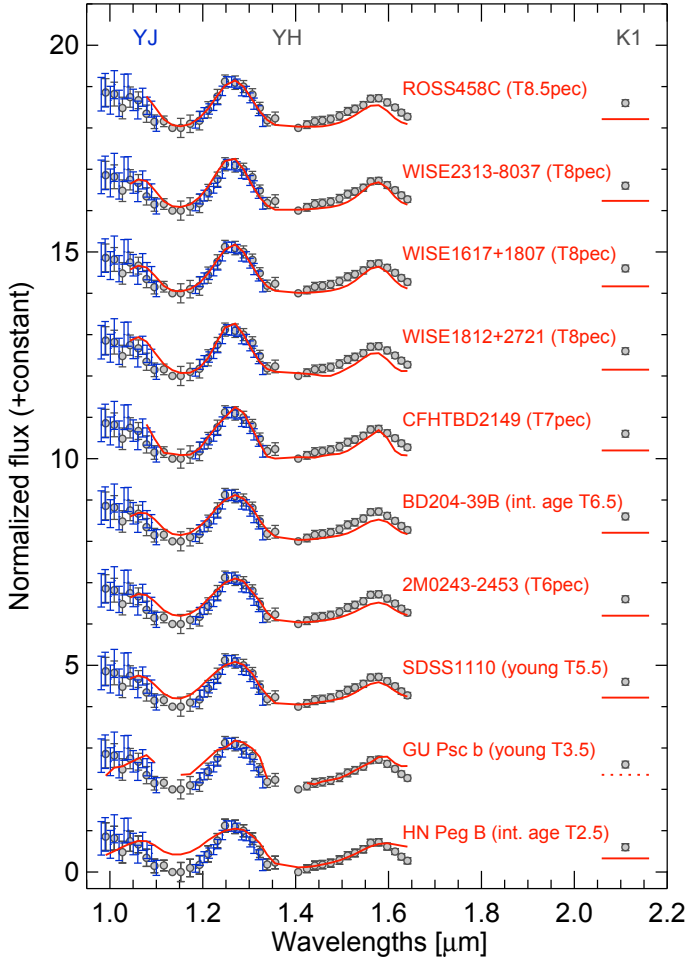
Fig. 9. Comparison of the flux density F_λ of 51 Eri b to those of selected peculiar T-type objects whose flux density has been normalized to match that of the planet between 1.2 and 1.3 μm.

of the reduced opacities caused by CIA of H_2 (Borysow 1991), which occurs in low gravity and metal-enriched atmospheres and that primarily affects the K band (Allard et al. 2001). We cannot exclude that it could also be caused by a haze of submicron-sized particles as proposed for low-gravity L/T transition objects (see Marocco et al. 2014; Bonnefoy et al. 2016; Hiranaka et al. 2016) and consistent with our atmospheric modeling analysis in Sect. 4.2.

We compare in Fig. 9 the spectrophotometry of 51 Eri b to those of extra T-type objects known to be younger than the field or dusty objects (Burgasser et al. 2011; Naud et al. 2014). No object could simultaneously reproduce the YH-band features and the K1 flux of 51 Eri b in agreement with the previous analysis. We also note a strong departure of the 0.95–1.05 μm flux of the planet whose origin is unclear, but further discussed in detail in Sect. 4.2.5. The depth of the 1.1–1.2 and 1.3–1.5 μm bands of 51 Eri b is only reproduced by those of objects later than T7.

In summary, the empirical approach: 1) confirms the peculiarity of 51 Eri b; 2) further suggests that the planet shares the properties of late-T dwarfs; 3) suggests that some of the properties of the planet are related to low-surface gravity and young age or super-solar metallicity; and 4) is limited by the lack of objects from clusters and young moving groups with spectral types later than T5. These findings are in good agreement with our atmospheric modeling as described in the next section.

4.2. Atmospheric modeling with petitCODE

In order to characterize 51 Eri b we carried out dedicated calculations with *petitCODE*, which is a self-consistent 1D radiative-convective equilibrium code, solving for the atmospheric temperature structure and abundances, assuming equilibrium chemistry. For every converged structure the *petitCODE* calculates an emission and transmission spectrum, where the latter is of no importance for studying 51 Eri b, given that the planet is not transiting. The first version of the code has been reported on in Mollière et al. (2015) and updates have been shortly described in Mancini et al. (2016a,b). The current version of the code is described in detail in Mollière et al. (2017).

In its current form the code includes molecular and atomic line and continuum (CIA) opacities and an implementation of the cloud model by Ackerman & Marley (2001). The *petitCODE* treats the non-ideal line shapes of Na and K atoms using the wing profiles by Nicole Allard; see Mollière et al. (2015) for a more detailed description. As possible cloud species $MgAl_2O_4$, $MgSiO_3$, Mg_2SiO_4, Fe, KCl, and Na_2S can be included with the optical constants taken from Palik (2012) for $MgAl_2O_4$, Scott & Duley (1996), Jaeger et al. (1998) for $MgSiO_3$, Servoin & Piriou (1973) for Mg_2SiO_4, Henning & Stognienko (1996) for Fe, Palik (2012) for KCl, and Morley et al. (2012) for Na_2S. Finally, the implementation of the Ackerman & Marley (2001) cloud model deviates from the description in the original paper in the sense that the mixing length is set equal to the atmospheric pressure scale height in all cases. This is different from the Ackerman & Marley (2001) description, where the mixing length can be up to 10 times smaller than the pressure scale height in the radiative regions. In the regions above the cloud deck the cloud mass fraction is proportional to $P^{f_{sed}/\alpha}$, where f_{sed} is the ratio of the mass averaged settling velocity of the cloud particles and the atmospheric mixing speed and α is the ratio between the mixing length of the eddy diffusion process and the atmospheric scale height. In our implementation of the Ackerman & Marley (2001) model it holds that $\alpha = 1$. The given power law can be derived from solving the homogeneous part of the differential equation for the condensate density (Eq. (4); Ackerman & Marley 2001). Therefore, for a given f_{sed} value, clouds in the *petitCODE* implementation are more extended than in the Ackerman & Marley (2001) description. Further, the atmospheric mixing speed is equal to K_{zz}/L, where K_{zz} is the atmospheric eddy diffusion coefficient and L is the associated mixing length, or mean free path, of the mixing process. Because the *petitCODE* implementation sets $L = H_P$, where H_P is the pressure scale height, the mixing velocity is smaller than in the Ackerman & Marley (2001) description, which favors smaller cloud particles at a given f_{sed} value. Therefore, adopting $L = H_P$ results in effectively smaller f_{sed} values when comparing cloud properties of the original Ackerman & Marley (2001) description to the *petitCODE* at the same f_{sed} value.

Two dedicated grids were calculated for 51 Eri b (see Table 4). The first grid is a clear grid (subsequently PTC-Clear, i.e., cloud free), assuming scaled solar compositions for the planetary abundances. We varied the effective temperature T_{eff} between 500 and 1700 K, the surface gravity by assuming $\log g$ values between 3 and 6 (with g in cgs units), and the metallicities [Fe/H] between -1.0 and 1.4.

The second grid is a cloudy grid (PTC-C, "cloudy") for which we assumed a mixing coefficient $K_{zz} = 10^{7.5}$, which is similar to the value used in Macintosh et al. (2015). Here the varied grid parameters are $T_{eff} = 500$–850 K, $\log g = 3$–5, [Fe/H] = 0.0–1.4, and $f_{sed} = 0.5$–2.0 (1–5 for initial

Table 4. Model grids used as input for MCMC exploration.

Model	T_{eff} (K)	ΔT (K)	$\log g$ $\log_{10}$ (cgs)	$\Delta \log g$ $\log_{10}$ (cgs)	[Fe/H] (dex)	Δ[Fe/H] (dex)	f_{sed}	Δf_{sed}
petitCODE (clear)	500–1700	50	3.0–6.0	0.5	−1.0–1.4	0.2	n/a	n/a
petitCODE (cloudy)	500–850	50	3.0–5.0	0.5	0.0–1.4	0.2	0.5–2.0[a]	0.5[a]
Morley+12	600–800	100	4.0–5.5	0.5	0.0	n/a	2–5	1

Notes. The radius of the planet was included as an additional analytic fit-parameter regardless of the model ranging from $0.1\,R_J$ to $2\,R_J$. [a] For the initial exploration a wider grid between 1 and 3 was used with a step size of 1. Smaller values were consistently preferred, leading to the final grid values.

exploration). Following Morley et al. (2012), the opacities of MgAl$_2$O$_4$, MgSiO$_3$, Mg$_2$SiO$_4$, and Fe were neglected for this cool grid, such that for the clouds only KCl and Na$_2$S opacities were considered.

Finally, our calculations were carried out assuming equilibrium chemistry for the gas composition and for identifying the cloud deck locations within the atmospheres. It is well known that for planets, compared to the higher mass brown dwarfs, non-equilibrium effects, and the associated quenching of CH$_4$ and NH$_3$ abundances may be more important (see, e.g., Zahnle & Marley 2014). Because we clearly detect methane in the atmosphere of 51 Eri b and we find best fit $\log g > 4$ (see Sect. 4.2.3), we conclude that CH$_4$ quenching is not very strong in this object; this is in agreement with the results presented in Zahnle & Marley (2014) for higher $\log g$ objects.

In addition to the two grids outlined above we compare our results with the cloudy model atmospheres described in Morley et al. (2012). However, as the grid does not include super-solar metallicity the resulting parameters are skewed (Figs. C.3 and D.3). We focus our discussion on the *petitCODE* models.

A summary of the used grids can be found in Table 4 and our *petitCODE* model grids are available on CDS.

4.2.1. Determination of the spectral covariance matrices

When comparing the spectrum obtained with an IFS instrument with a model, taking into account the spectral covariance of the residual speckle noise has been shown to be of great importance for assessing the uncertainty of the fitted atmospheric model parameters (Greco & Brandt 2016). Following the methods presented by these authors, we determine the mean spectral correlation between all spectral channels within an annulus of width $1.5\lambda/D$ at the separation of the planet (with the planet masked out by a $2\lambda/D$ radius mask),

$$\psi_{ij} = \frac{\langle I_i I_j \rangle}{\sqrt{\langle I_i^2 \rangle \langle I_j^2 \rangle}}, \qquad (1)$$

where $\langle I_i \rangle$ is the average intensity inside the annulus at wavelength λ_i. The correlation matrix can then be used to obtain the covariance matrix C, which is used in computing the Gaussian log-likelihood $\ln \mathcal{L}$ (or χ^2) for the MCMC model fit according to

$$-2 \ln \mathcal{L} \equiv \chi^2 = (S - F)^T C^{-1} (S - F), \qquad (2)$$

where S is the observed spectrum and F the model spectrum. In the case of uncorrelated noise C is equal to the unity matrix and χ^2 reduces to the more familiar sum over the residuals squared, which is not correct for correlated IFS data. The correlation matrix for the YH spectrum is shown in Fig. 10. We can
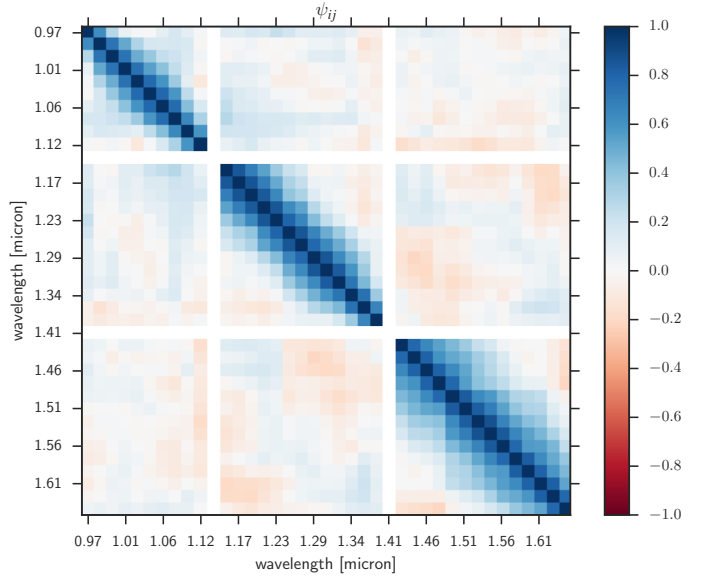

Fig. 10. Correlation matrix ψ_{ij} showing the correlation between each pair of spectral channels (1: completely correlated; −1: completely anti-correlated; 0: uncorrelated). The 1.14 μm channel was used as reference for SDI for all wavelength channels shorter than this, whereas the 1.41 μm channel was used as reference for all other channels.

see that each channel at the separation of 51 Eri b is strongly correlated with three to four of its adjacent channels in both directions. Contrary to Greco & Brandt (2016), we note that there are also anti-correlations present, which are due to the use of classical SDI and the larger spectral coverage available with the SPHERE IFS. The SHERE spectral coverage, unlike GPI spectra, spans multiple bands and band gaps.

As we do not have access to the reduced GPI data of 51 Eri b, we assume the fiducial model for a GPI-H spectrum reduced using simultaneous SDI and ADI as given in Greco & Brandt (2016) to calculate the correlation matrix at the angular separation of 51 Eri b.

4.2.2. Markov chain Monte Carlo exploration of atmospheric models and parameters

We use the python implementation of the affine-invariant Markov chain Monte Carlo (MCMC) algorithm *emcee* (Goodman & Weare 2010; Foreman-Mackey et al. 2013) to explore the posterior probability distribution of model parameters for various atmospheric model grids (see Table 4). Our custom procedure can handle model grids of an arbitrary number of parameters, number of photometric data points and/or spectra, as well as their respective covariance matrices. The only restriction is that we require the model grid to be regularly spaced in each

individual parameter to allow for efficient N-dimensional linear interpolation, where N is the number of free atmospheric model grid parameters. As the atmosphere of 51 Eri b is not well characterized yet, we use flat priors over parameter ranges listed in Table 4. Planetary radii are fitted as a separate analytic parameter. The log-likelihood for each spectrum with their complete covariance matrix and each photometric data point are evaluated separately. These values are then summed to obtain the overall log-likelihood of the model given the data, and there is no statistical weighting between the data sets. Rather than defining a wavelength dependent weighting scheme, this is more properly taken into account by using the real covariances between the data. This effectively down-weights the relative importance of the many spectral data points with respect to the fewer, yet independent, photometric data points. Uncertainties are assumed to be uncorrelated between the separate data sets. The likelihood evaluation is carried out in luminosity space, taking into account the additional uncertainty of the systems distance (29.4 ± 0.3 pc), which can be important for the radius uncertainty of the planet, which otherwise would be slightly underestimated. In this case, owing to the proximity and brightness of the host star, the distance uncertainty is only on the order of 1%, and thus does not impact the radius uncertainty much, but for objects at larger distance this can become a significant factor.

We follow a different approach when treating upper limits compared to many previous studies. We treat data points "below the detection limit" not as non-detection or upper limits in the fit because one does not look for a previously undiscovered source. We know where to measure the flux. Knowing the position of the source contains strong prior information, and even data points that are below the formal detection limit contain useful information; this can be seen by the fact that even data points that are technically not 3σ detections follow the model predictions very well. These non-detection points can still contain significant flux and in the "worst case" are consistent with negligible flux within their 1σ uncertainties. We use this approach of "forced photometry" (Lang et al. 2016), which is a method successfully used in other fields of astronomy, such as the study of faint galaxies and quasars (e.g., Venemans et al. 2015), to replace the more common practice of simply excluding data points below the classical detection threshold for point sources of unknown position because this would mean mixing two unrelated statistical quantities in an unjustified way and would effectively lead to throwing away informative data. Also replacing these measurements with an upper limit as is common – while seemingly the conservative choice – is not necessarily the optimal choice. In direct imaging, reporting only the upper limit is equivalent to just reporting the uncertainty for the measurement without reporting the measurement itself. Applying forced photometry for all measurements means consistently reporting both measurement and uncertainty. This has the advantage that all the data are treated uniformly and no arbitrary choice about a cutoff value for "detection" has to be chosen. The problem is illustrated by Fig. 1, where one would not claim the discovery of an unknown planet at the position of 51 Eri b given only the Y-band image, but the fact that the occurrence of a clear excess in flux is located at the exact position of the planet visible in the other bands is much more informative and illustrates the importance of prior knowledge of the position of a planet for characterization.

However, to put the importance of the Y-band measurement in this particular case into perspective, it should be pointed out that while it is true that all points included for forced photometry contain some information on the spectrum of the planet owing to their low S/N and because other parts of the spectrum already put

very strong constraints on the Y-band model fluxes, they do not impact the derived atmospheric parameters significantly. This is especially true for cases in which the rest of the spectrum poorly constrains the spectral shape at wavelengths where the flux measurements are not above the detection limit; for example, models comparisons that seek to distinguish between the presence or absence of a physical model component, such as thermal inversion or significant non-equilibrium chemistry.

For all of these reasons, we also include the measured flux in the methane "non-detection" bands H3 and K2 (see Table 2). The only data that we do not include in the fit are the spectral channels that were used as a reference in the SDI step of data reduction, as these are biased, and the first three IFS channels as they are most affected by degrading overall system performance and telluric lines.

4.2.3. Discussion of physical parameters

The best fitting models for the cloudy model grid (PTC-C) are shown in Fig. 11, where the black line represents the best fit and the gray lines showing the spectrum for 16 randomly drawn parameter combinations from the posterior probability distribution. As the most extensive model of the three, the posterior probability distribution of the PTC-C model is shown for each of the model parameters along with their marginalized values in Fig. 12. Cloud-free models are incapable of explaining all of the observed spectral features simultaneously: models that explain the Y, J, and H peaks are not able to explain the K1- and L'-band data (see Fig. D.2). They also result in model predictions that are unphysical for young giant planets, for example, high $\log g = 5.35^{+0.15}_{-0.12}$ and very low radius $R = 0.40 \pm 0.02\,R_J$ (see Fig. D.2). Cloudy models vastly improve the consistency with the data over the whole spectral range for which data are available. Our discussion below centers on the results obtained on the *petitCODE* cloudy models. These results cover the complete parameter space relevant for 51 Eri b, including metallicity and cloud sedimentation values (f_{sed}). The results of all tested models are summarized in Table 5.

Temperature, radius, and surface gravity. We obtain a value of $T_{eff} = 760 \pm 20$ K for the effective temperature, $R = 1.11^{+0.16}_{-0.14}\,R_J$ for the radius, and $\log g = 4.26 \pm 0.25$ (cgs-units) for the surface gravity of 51 Eri b. The effective temperature and radius of the planet are expectedly correlated ($T_{eff} \propto R^{-0.5}$ for black bodies) as they both relate to the luminosity of the planet. With a temperature that is likely above 700 K, it appears to be above the temperature for which sulfur chemistry becomes an important factor for 51 Eri b as discussed in Zahnle et al. (2016). The radius is consistent with the radius of Jupiter and may be slightly larger as expected for young objects (Chabrier et al. 2009; Mordasini et al. 2012a).

Mass. To check the consistency of the best fitting solution with our physical understanding we can use multiple approaches to derive the mass of the planet. Using the posterior sampling of surface gravities and planet radii, according to $M = g/g_J \cdot (R/R_J)^2$, where $g_J = 24.79$ m s^{-2} and $R_J = 6.9911 \times 10^7$ m are the surface gravity and (volumetric mean) radius of Jupiter, respectively, we get a mass estimate of $M_{gravity} = 9.1^{+4.9}_{-3.3}\,M_J$. Another approach is to use the luminosity of the planet derived similarly from the posterior sampling of the radius and effective temperature according to $L/L_\odot \sim (R/R_\odot)^2 \cdot (T/T_\odot)^4$, where
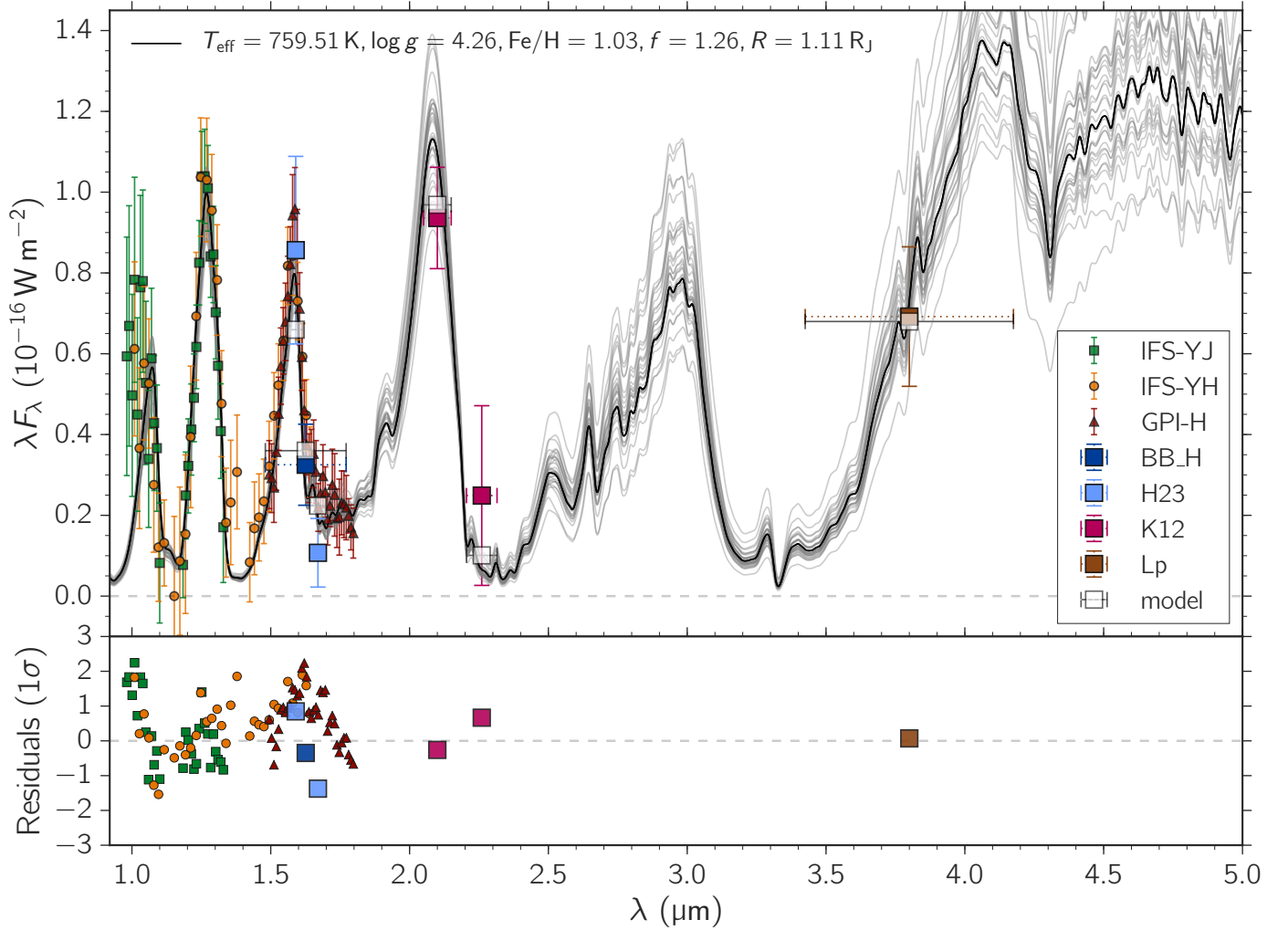
Fig. 11. *petitCODE* cloudy model interpolated to the parameters best describing the data according to the posterior probability distribution (black line), as well as the SPHERE spectrophotometric data, GPI *H*-band spectrum, and *L′* data point from Macintosh et al. (2015). For photometric data points, the *x*-error bar reflects the filter width rather than uncertainties. The gray lines represent 32 randomly drawn samples from the posterior probability distribution to reflect the spread of plausible model parameter combinations that fit the data. Photometric points describe the average flux in the respective filter, whereas the orange points describe the average flux in the respective filter for the best fitting model. The residuals are shown in multiples of 1σ uncertainties of the data.

$R_\odot = 9.728 \, R_J$ and $T_\odot = 5772$ K are the radius and effective temperature of the sun, we get a luminosity of $L/L_\odot \sim (3.94^{+0.66}_{-0.55}) \times 10^{-6}$ or $\log(L/L_\odot)$ between −5.470 and −5.338 (compared to −5.4 and −5.8, Macintosh et al. 2015), which can be converted to a mass assuming a formation (initial entropy) model. Figure 13 shows the luminosity–mass relationship derived from a complete core accretion population synthesis model (Mordasini et al. 2012b,a). The shaded region corresponds to the above-mentioned luminosity and (surface gravity derived) mass range. Three populations at 20 Myr using different assumptions are shown, where two populations correspond to the traditional hot-start and cold-start populations, and one corresponds to an intermediate warm-start population. In the hot-start case, energy from the accretion shock is not radiated away efficiently and is deposited in the planet. In the cold-start case, all energy from gas accretion is radiated away, furthermore core masses are restricted to $<17 \, M_{\rm Earth}$ to mimic the traditional cold-start model by Marley et al. (2007). The warm-start population is similar to the cold-start population in terms of accretion physics, but allows higher core masses, which in turn leads to more energy deposition during the growth of the planet (Mordasini 2013). A

similar result would be achieved by allowing for a spread in the radiation efficiency of the accretion shock (Marleau et al. 2017). We can see that the observed luminosity range excludes the traditional cold-start model (consistent with theoretical predictions Marleau & Cumming 2014), but includes both objects of the hot- and warm-start case with a large spread in masses. Small masses between 2.4 and 5 M_J are preferred in the hot-start case, whereas a big spread of masses between 2.4 and 12 M_J are possible in the warm-start case. While in this synthesis model objects with smaller mass are more common in this luminosity range, big masses are not excluded from the point of view of the formation model.

The above discussion and Fig. 13 exemplify the problem of determining the mass of directly imaged exoplanets in the absence of low-mass and cool benchmark objects with independent mass measurement. In the two approaches, using the measured luminosity together with evolutionary models gives a statistical picture of the distribution of planets resulting from the planet formation synthesis modeling approach, and allows for a wide range of masses for the given age and luminosity, depending strongly on the accretion physics assumed. In principle, the mass
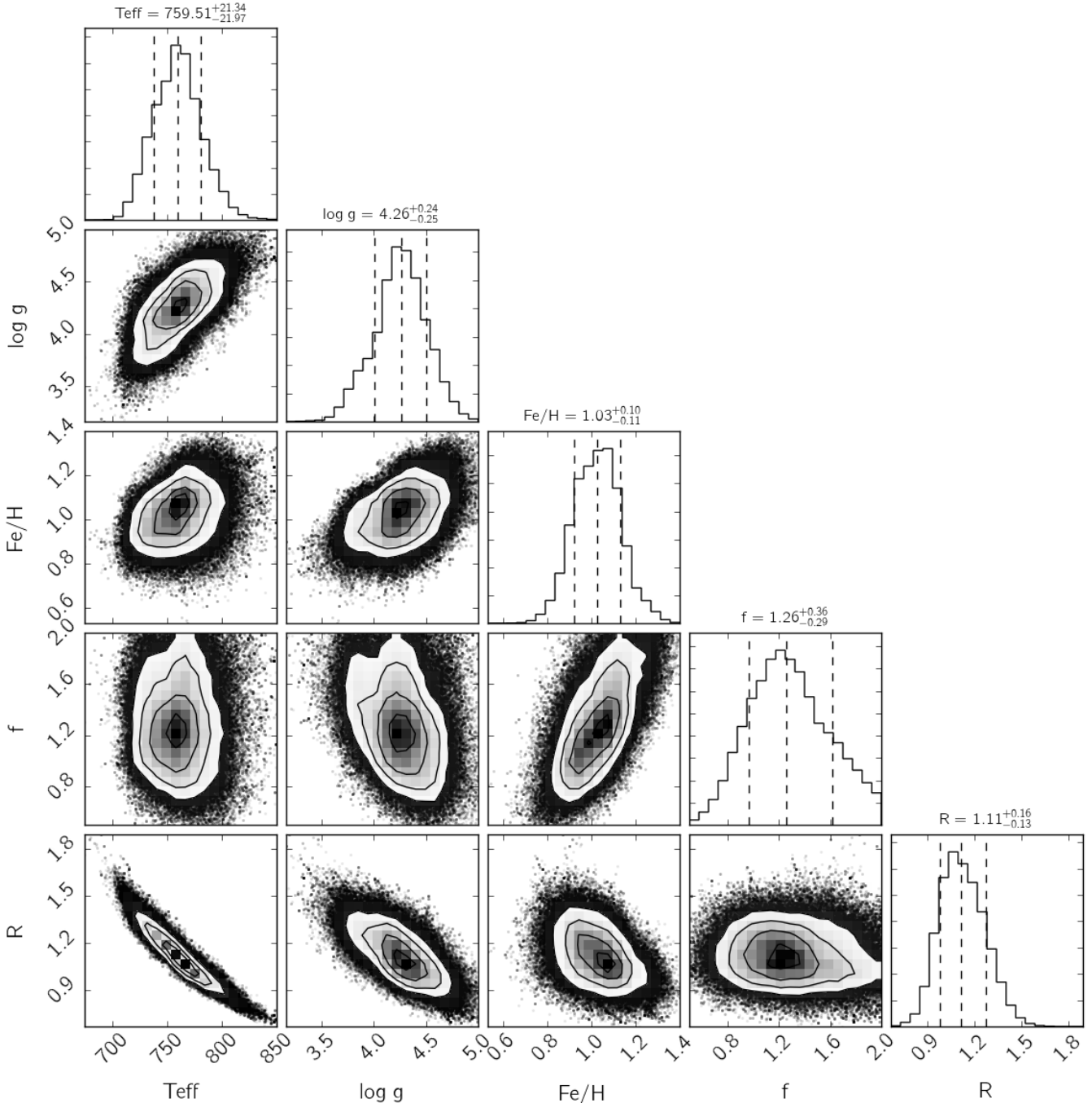
Fig. 12. Posterior probability distribution of the cloudy *petitCODE* grid with respect to each of its parameter pairs as well as the marginalized distribution for each parameter. The uncertainties are given as 16% to 84% quantiles as is common for multivariate MCMC results.

derived from the surface gravity and radius are more constraining, but depend strongly on the atmospheric model assumptions, which in the case of cold and cloudy objects still have many uncertainties. Assuming the physics model represents the real nature of the planet, the determined log *g* and radius is more consistent with a more massive planet that would be expected based on hot-start evolution models. It should be mentioned, however, that we can rule out the brown dwarf regime, as brown dwarfs at this age we would expect to see a significantly larger radius because they have deuterium burning as an additional heat source.

In conclusion it can be said that both mass estimates are highly model dependent and there are multiple big sources of

uncertainty in both approaches. For atmospheric models it is possible that the cloud physics are not sufficiently well modeled, leading to big uncertainties in the surface gravity determination. The surface gravity determination is also most strongly impacted by the *J*-band flux for which some amount of variability cannot be completely ruled out at this point. On the other hand, evolutionary models have a big intrinsic spread as they reflect the statistics of populations rather than single objects (e.g., different core-mass fractions). Initial conditions for planet formation and evolution are not well constrained. Another aspect that deserves further research is the current lack of evolutionary models that consider super-solar metallicity objects. While the composition

Table 5. Summary of model results.

Model	$T_{\rm eff}$	$\log g$	[Fe/H]	$f_{\rm sed}$	R	L	$\log L$	$M_{\rm gravity}$
	(K)	$\log_{10}$ (cgs)	(dex)		($R_{\rm J}$)	($10^{-6}\,L_\odot$)	$\log_{10}\,(L_\odot)$	($M_{\rm J}$)
PTC-C	760^{+21}_{-22}	$4.26^{+0.24}_{-0.25}$	$1.03^{+0.10}_{-0.11}$	$1.26^{+0.36}_{-0.29}$	$1.11^{+0.16}_{-0.13}$	$3.94^{+0.66}_{-0.55}$	-5.470 to -5.338	$9.1^{+4.9}_{-3.3}$
PTC-C w.o. Y-band	754^{+23}_{-23}	$4.25^{+0.32}_{-0.37}$	$1.04^{+0.11}_{-0.12}$	$1.33^{+0.38}_{-0.33}$	$1.13^{+0.17}_{-0.15}$	$3.92^{+0.68}_{-0.61}$	-5.479 to -5.337	$9.1^{+7.3}_{-4.3}$
PTC-Patchy[a]	757^{+24}_{-24}	$4.47^{+0.24}_{-0.26}$	$1.25^{+0.10}_{-0.16}$	$1.07^{+0.36}_{-0.31}$	$1.11^{+0.16}_{-0.14}$	$3.84^{+0.63}_{-0.56}$	-5.484 to -5.350	$14.5^{+8.7}_{-5.6}$
PTC-Clear	982^{+18}_{-15}	$5.35^{+0.15}_{-0.12}$	$1.36^{+0.03}_{-0.06}$	–	$0.40^{+0.02}_{-0.02}$	$1.43^{+0.06}_{-0.06}$	-5.863 to -5.827	$14.5^{+4.7}_{-3.1}$
Morley+12	684^{+16}_{-20}	$5.19^{+0.10}_{-0.11}$	–	$4.16^{+0.52}_{-0.67}$	$1.01^{+0.07}_{-0.06}$	$2.12^{+0.14}_{-0.13}$	-5.700 to -5.645	$64.9^{+19.1}_{-15.6}$
PTC-C (Macintosh et al.[b])	785^{+11}_{-17}	$3.35^{+0.29}_{-0.21}$	$0.83^{+013}_{-0.12}$	$2.54^{+0.32}_{-0.47}$	$1.12^{+0.08}_{-0.05}$	$4.55^{+0.34}_{-0.30}$	-5.372 to -5.311	$1.2^{+1.0}_{-0.4}$

Notes. See Appendix D for corner plots of the posterior probability distributions of the model parameters for each model. [a] Effective temperature calculated by $T_{\rm eff} = \left[CF \cdot T^4_{\rm cloudy} + (1 - CF) \cdot T^4_{\rm clear} \right]^{1/4}$ with $T_{\rm cloudy} = 751^{+24}_{-25}$, $T_{\rm clear} = 813^{+67}_{-40}$ and cloud fraction $CF = 0.91 \pm 0.05$. [b] Using our cloudy model and the same data as Macintosh et al. (2015) without covariance treatment.
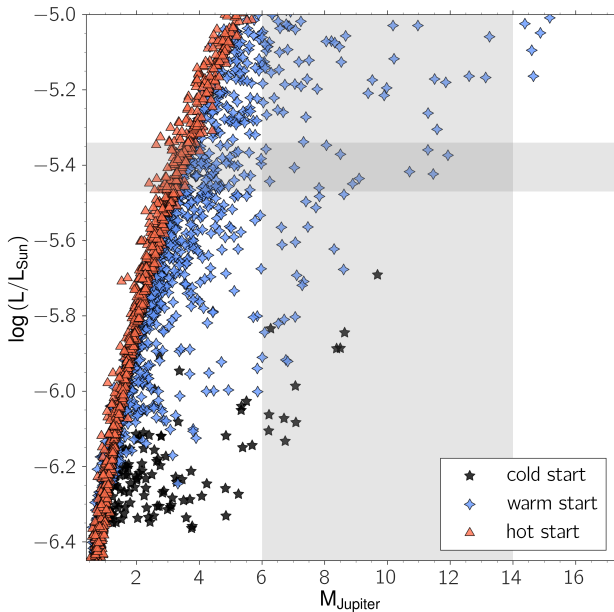
Fig. 13. Luminosity-mass relationship from core-accretion population synthesis model at 20 Myr. Populations using different entropy assumptions are plotted, corresponding to what is traditionally referred to as cold (black) and hot start (red). The warm start (blue) model corresponds to a cold gas accretion model, but allowing for higher core masses than 17 $M_{\rm Earth}$. Gray shaded regions correspond to the luminosity of 51 Eri b derived in this work and the mass range determined from surface gravity and radius.

is reflected in the core mass of the models used here, the thermal evolution does not include the increase in opacity caused by high metallicity. This is also an issue for all other evolution models currently available.

Metallicity. The metallicity [Fe/H] = 1.0±0.1 dex is super-solar, and significantly above that of the solar metallicity host star. This is similar, but even more pronounced than what has previously been observed in the cool object GJ 504 b Fe/H = 0.60 ± 0.12 around a slightly metal-rich host star, (Skemer et al. 2016). The metallicity determined here is in good agreement with

predictions of bulk composition for giant planets formed by core accretion (e.g., Mordasini et al. 2014). Other studies of massive directly imaged exoplanets also suggest super-solar metallicities, for example HR 8799 b (Lee et al. 2013).

Increasing the planetary metallicity strongly enhances the K-band brightness, redistributing a part of the flux from shorter to longer wavelengths. The reason for the metallicity-dependent K-band brightness is that a change in metallicity shifts the position of the planetary photosphere within the atmosphere because in hydrostatic equilibrium it holds that $d\tau = (\kappa/g)\,dP$, where τ is the optical depth, κ the opacity, g the planetary surface gravity, and P the pressure. If the pressure dependence of κ is neglected, an increase in κ, resulting from an increase in metallicity, shifts the photosphere ($\tau_{\rm Phot} \sim 2/3$) to lower pressures. Nonetheless, it is critical to note that the opacity varies as a function of pressure: the strength of pressure broadened molecular and atomic line wings is proportional to the pressure P, but for the many lines of water and methane the effect in the K band is of secondary importance. More importantly, the continuum opacity due to CIA of H_2–H_2 and H_2–He pairs is linear in pressure for all wavelengths. Further, CIA exhibits a peak in opacity at 2.3 μm, i.e., very close to the K band, whereas the CIA opacity in the neighboring H band is lower by a factor of 100. Consequently, as the photosphere is shifted to lower pressures, owing to an increased metallicity, the contribution of CIA to the total opacity in the K band diminishes, such that the opacity minimum resulting from the scissor-like crossover of the water and methane opacities becomes visible as an emission feature (Allard et al. 2001). Owing to the steep decrease of the CIA opacity toward smaller wavelengths the H band is unaffected by the increase in metallicity. As a final test we carried out runs neglecting the CIA opacities and were unable to reproduce the effect of the metallicity-dependent K band.

The strong influence of metallicity on key spectral features shows the importance of having a broad wavelength coverage of all features present in the near- to mid-infrared and using model grids that include non-solar metallicity as a free parameter. Finally, to make sure our methodology and models are not systematically biased toward providing high metallicity results, we analyzed two benchmark brown dwarfs (Gl 570D and HD 3651B; similar to Line et al. 2015) and confirmed that the metallicities derived are reasonable. The details of this analysis can be found in Appendix B.

Clouds. For the cloud sedimentation parameter, we derive a value of $f_{sed} = 1.26^{+0.36}_{-0.29}$. A lower f_{sed} results in more vertically extended optically thicker clouds with smaller particle sizes. While the slight differences in atmospheric model implementations make it difficult to compare this result exactly with previous research, f_{sed} as low as this (<2) are unusual for self-luminous substellar objects of low temperature, especially considering that our implementation of f_{sed} would result in a lower value in the Ackerman & Marley (2001) implementation (see model discussion in Sect. 4.2). Values of f_{sed} between 3 and 5 are usually reported, for example, for GJ 504 b (Skemer et al. 2016) and GJ 758 B ($f_{sed} = 5$ Vigan et al. 2016). Cushing et al. (2008) report values between 1 and 4 for their sample of L and T brown dwarfs, but all of these objects are significantly hotter than 51 Eri b and only models of solar metallicity are considered. The lack of similar objects and detailed analyses including metallicity as a free parameter make a real comparison difficult. Lack of metallicity as free parameter in the model can significantly alter the cloud parameter because it tends to compensate for the lack or overabundance of heavy elements in the spectrum. We encourage modelers to include low f_{sed}-values as well as metallicity in their consideration for future model grids.

The $f_{sed} = 1.26^{+0.36}_{-0.29}$ we obtain for 51 Eri b reflects a particle size distribution with mean values of 1 μm, and slightly below in the upper regions (below 10^{-2} bar). Owing to the width of $\sigma = 2$ of the lognormal size distribution, however, the opacities are dominated by the larger particles.

4.2.4. Patchy cloud models

As variability has been observed in a number of brown dwarfs, the idea that for cool substellar objects cloud coverage may be less than 100% should not be excluded a priori. Macintosh et al. (2015) used such a patchy cloud model, which can be expressed as a linear combination of a clear and a cloudy atmosphere. They also included non-equilibrium chemistry in the cloudless model. We tested this idea with the following simple composite model using *petitCODE*

$$F_{patchy} = CF \cdot F_{cloudy} + (1 - CF) \cdot F_{clear}, \qquad (3)$$

where CF is the cloud fraction and F_{cloudy} and F_{clear} are the flux of the cloudy and clear model, respectively. Under this model we have the following MCMC fitting parameters $\theta = (T_{cloudy}, T_{clear}, CF, \log g, [Fe/H], f_{sed}, R)$, i.e., we now fit for the cloud fraction and allow the two models to have different temperatures, as the cloudy and clear model fluxes probe different temperatures. Because both models must, however, describe the same physical planet, we keep the metallicity, as well as the surface gravity and radius, the same for both models. Furthermore, we impose $T_{cloudy} < T_{clear}$ as a prior, as the cloudy model flux is supposed to come from higher in the atmosphere than the clear flux, which in this model corresponds to holes in the clouds.

However, the result of this test shows no significant improvement of the fit for the resulting composite model spectrum as cloud coverage tends toward $>90\%$ (see Fig. D.4 for corner plot). As the resulting spectra (Fig. C.4) look almost the same, we conclude that a patchy cloud model may not be necessary to explain the observed spectrum, and at least at this point, the increase in model complexity is not justified. It should be pointed out that using the patchy cloud model improves the fit marginally when data from Macintosh et al. (2015) are used exclusively. This may be attributed to the higher J-band flux in GPI and resulting bluer spectrum.

To be clear, we do not wish to say that patchy cloud models in general do not work or should be avoided, but that for this particular planet, data set, and model, cloudy models alone seem to be capable of fitting the data very well. It may well be that inclusion of more physics (e.g., non-equilibrium chemistry) improves the results. It is also important to keep in mind that a simple linear combination of the clear and cloudy models, as carried out here, is not self-consistent and strictly speaking not physically correct (Marley et al. 2010). A detailed model comparison with a more rigorous patchy cloud model should be performed in the future to test whether further increasing the model complexity is justified by the gain in fit quality, for example, by using Bayesian evidence (e.g., nested sampling).

4.2.5. Unexplained spectral features

A number of features in the spectrum of 51 Eri b exist that cannot be explained with the current model, either pointing to unaccounted systematic effects in the data or the need for more sophisticated atmospheric models.

1. The Y-band peak in the data is stronger and seems to extend to smaller wavelengths than predicted by the model. The Y band is difficult to observe with good S/N, mostly because overall instrument performance degrades toward shorter wavelengths (e.g., worse AO correction and end-to-end instrument throughput). It is also subject to some unresolved telluric features at short wavelengths ($\lesssim 1$ μm) in the Earth's atmosphere. There are multiple plausible scenarios for the perceived discrepancy: a) residual speckle flux at planet position at these wavelengths; b) a genuine instrument systematic effect (e.g., unaccounted variations in system transmission); and c) a real physical phenomenon or improper treatment of potassium wings or abundances in model. If there is residual speckle flux (i.e., speckle noise) and the noise is spectrally correlated (as it is, the treatment of which is described in Sect. 4.2.1 and taken into account), we expect it to affect at least half of the Y-band channels consistently (as about six neighboring channels are correlated). Seeing visually that a number of points scatter "systematically" higher or lower than the model is actually what we should expect in this case at low S/N. It is important to remember that we can only plot 1D error bars, which looks like we have a systematic effect if we work under the assumption that the measurements are uncorrelated and should scatter randomly around the true values. The proper treatment of IFS covariances is a relatively new practice in this field and needs to be kept in mind. As such, the method of forced photometry should only be used in conjecture with proper covariance treatment. On the other hand, it makes it challenging to distinguish residual speckle flux from other instrument systematics that may only be present in certain wavelength regions. "Systematic deviations" that conform to the correlation length are not much of an issue for the parameter fitting because we already take this effect into account. To confirm this, we performed the same fit without including the Y-band data at all, which only marginally changes the results. This test shows that the relatively large and correlated uncertainties in the Y-band data are not very constraining at this point, which strengthens our confidence in the robustness of the analysis and our treatment of the noise; see Fig. D.1 for posterior distribution in the case in which Y-band data are excluded. Resolving the issue of which of the three scenarios (or mixture of them) is dominant will require

us to obtain more high S/N Y-band measurements. If the elevated flux level in the observation are shown to be persistent and significant in upcoming observations, this raises the possibility that the model treatment of potassium wings or abundances needs to be reconsidered and improved (e.g., better alkaline profiles and a mechanism for depletion of alkaline species).

2. We observe an emission feature at ∼1.35 μm that is not explained by the model. While it is possible that this is caused by instrument systematics or the fact that it is in a region of strong telluric absorption, it is striking that both the GPI and SPHERE observations show an increase in flux. A very similar feature in the deep water bands between the J and H peaks at 1.35–1.40 μm has been observed and discussed by King et al. (2010) in the ϵ Indi Ba and Bb brown dwarf binary members. These authors also list objects with descriptions of similar features, for example, the T1 spectral standard SDSS0151+1244 (Burgasser et al. 2006b), the T8.5 and T9 dwarfs ULAS1238 and ULAS1335 (Burningham et al. 2008), and some L dwarfs (e.g., 2MASS J1507–1627 (L5) Burgasser 2007). King et al. (2010) argue that this feature is due to the structure of the strongest part of the water absorption bands in the higher levels of the atmosphere, which may be the result of an underestimated local temperature in this region of the atmosphere. According to their toy model, raising the temperature, and therefore changing the temperature gradient can reconcile the modeled and observed flux levels, although they could not point to a reasonable physical mechanism, such as back-warming owing to an additional opacity source. If this feature is indeed a real feature, in the case of 51 Eri b, it may be related to its atmospheric cloud structure, but at this point this is very speculative. The fact that both the target planet and Earth's atmosphere contain complex telluric features at these wavelengths makes it difficult to draw strong conclusions.

3. The H-band feature has a broad tail toward shorter wavelengths and an extended wing toward longer wavelengths, which has a profound impact on the model fit. In general the H-band wings strongly favor models with higher log g and lower metallicity, whereas the rest of the spectrum favors lower log g and higher metallicity (especially the need for high metallicity to produce the K1-peak). Excluding the GPI H-band spectrum from the fit allows the PTC-C model as well as the Morley et al. (2012) models to fit the strength of all of the observed features well (except for the width and height of the Y-band peak and the width of the H-band peak). Including H-band wings in the fit puts strong weight on these features and the resulting best model is a compromise between fitting the H-band wings and the amplitude of the peak. While this spectrum fits the overall shape of the H band well, it does not match the absolute amplitude of the H peak. A zoom in on the wavelength range covered by IFS data is shown in Fig. C.1. This trade off shows how important extensive coverage of the spectral bands is for drawing physical conclusions.

5. Constraints on additional companions

5.1. VLT-NACO: Sparse aperture masking with L′ filter

To constrain the presence of any potential companions at smaller separations, we processed and analyzed archival sparse aperture masking (SAM) data taken with the VLT-NACO instrument. The observations were made on 2009-12-26 using the L' filter

and the 7-hole aperture mask. The calibrator stars HIP 22226, HIP 30034, HIP 24947, and HIP 32435 were used, and the conditions were between median and bad. Single exposures had detector integration times (DITs) of 0.2 s with a total of 3200 frames (NDIT = 200, 16 cubes). The calibrators had DITs of 0.25 with the same number of frames. Data were processed using the IDL aperture masking pipeline developed at the University of Sydney. The data processing steps are described in Tuthill et al. (2000), Kraus et al. (2008) and the references therein. Briefly, the images were sky subtracted, flat fielded, cleaned of bad pixels and cosmic rays, and then windowed with a super-Gaussian function. The closure phases were then measured from the Fourier transforms of the resulting cleaned cubes. The closure phases were calibrated by subtracting the average of those measured on several unresolved calibrator stars observed during the same night with the same instrument configuration. To estimate the detection limits of the SAM data, a Monte Carlo simulation was performed. Using a Gaussian distribution, we generated 10 000 simulated data sets consistent with the measured uncertainties. For each point on a grid of separation, position angle and contrast, we defined our detection limits to be the point at which 99.9% of the simulated data sets were fit better by a point source model than the binary model. These 3.3σ detection limits were then scaled to 5σ to simplify the comparison with the results from SPHERE. No additional point sources are detected.

5.2. SPHERE

We detect no additional point sources in the SPHERE data. Contrast curves for the more extended FoV achievable with IFS and IRDIS were compiled with different reduction methods. The methodology for deriving the contrast was kept as similar as possible between the algorithms. For the LOCI and PCA reductions the contrast curves correspond to the azimuthal 5σ self-subtraction corrected variance in the respective separation bin, whereas ANDROMEDA inherently models the detectable signal contrast and does not need self-subtraction correction. The effect of small-sample statistics at small separations (Mawet et al. 2014) and the coronagraphic throughput (A. Boccaletti, priv. comm.) have been accounted for. The IFS-YJ PCA reduction was performed with the more aggressive simultaneous ADI plus SDI algorithm for detection and the median combination of all channels is shown. The top panel of Fig. 14 shows the achieved contrast with both the innermost region explored by small aperture masking and the exploration region of SPHERE. It is not straightforward to convert the "detection images" shown in Fig. 2 into quantitative contrast curves and detection limits. As such they are not used for this purpose in this paper, but serve as a qualitative probe for additional candidates. However, no obvious candidates are seen.

For the conversion from contrast to mass limits, we used the same $JHKL'$ magnitudes, distance, and age for the host star as in Macintosh et al. (2015). We use evolutionary tracks of Baraffe et al. (2003) together with the atmosphere model of Baraffe et al. (2015) for the SPHERE data and BT-Settl models of Allard et al. (2012) for the NaCo data because Baraffe et al. (2015) does not include NaCo L' filters. The mass limit derived from the IFS data assumes a companion-to-star contrast, which is constant with wavelength as a conservative choice for which we already run into the lower mass limits of the used models. The mass limits (see bottom panel of Fig. 14) constrain the presence of additional components in the system very well. The SAM data reject >20 M_J companions at ∼2−4 au, while the
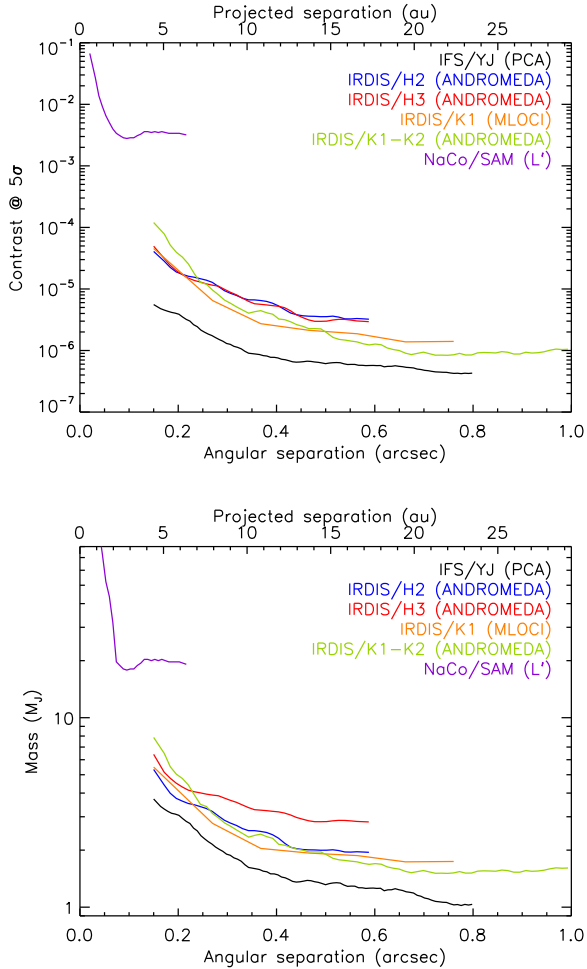
Fig. 14. 5σ-contrast (*top*) and mass (*bottom*) is plotted for NACO/SAM (L') as well as the observations with the best quality for IRDIS and IFS, respectively.

IFS data are sensitive to planets more massive than 4 M_J beyond 4.5 au and 2 M_J beyond 9 au.

6. Summary and conclusions

Our new study of 51 Eri b provides new and improved spectra and photometry and allow us to revise the previous flux measurements and to explore new wavelength bands, especially the Y and K band. The photometric measurements obtained with SPHERE are $J = 19.74 \pm 0.71$ mag, J3 = 18.86 ± 0.26 mag, H2 = 18.56 ± 0.28 mag, and K1 = 17.55 ± 0.14 mag. The broad wavelength coverage was made possible by combining data sets from the Y band up to the L' band, allowing us to take a comprehensive view of the planet for the first time; this showed how important thorough knowledge of all features is for understanding and modeling the system. Given separation of ~0.5″ of 51 Eri b, it is very suitable for high-contrast spectral observations and will become a benchmark object for current and future atmospheric models, especially once further spectra and photometry at longer wavelengths are obtained and all NIR features are mapped in detail. The models produced in this work provide strong predictions on the expected flux and shape of these features and validation or rejection of these predictions will further improve our understanding.

In this study, for the first time for SPHERE data, we combined the use of the recently developed ANDROMEDA algorithm to extract an unbiased planetary spectrum with a proper treatment of the spectral covariance and forced photometry with a state-of-the-art atmospheric model including clouds and varying metallicities combined with a detailed MCMC analysis.

We would like to advocate the use of forced photometry, together with the proper treatment of the noise covariance in the direct imaging community, to use all fluxes obtained at the known position of a planet; such data points can contain information even if the flux values obtained are below the detection limit, which is a quantity related to the probability of a previously unknown point source to be a real detection and cannot be directly applied as a cutoff threshold for an already known object. Furthermore, the usage of empirical covariances for IFS spectra can be used to take care of the relative weighting of spectral and photometric data, without having to rely on ad hoc weighting factors to artificially down-weight the spectral data with respect to independent photometric data. Our best fitting planet parameters for the cloudy models are $T_{eff} = 760 \pm 20$ K, $R = 1.11^{+0.16}_{-0.14} R_J$, $\log g = 4.26 \pm 0.25$ (cgs-units), highly super-solar metallicity [Fe/H] = 1.0 ± 0.1 dex, and $f_{sed} = 1.26^{+0.36}_{-0.29}$, indicating the presence of a vertically extended, optically thick cloud cover with small particle size. According to our models the planet seems to have an effective temperature above 700 K and thus sulfur chemistry, as discussed in Zahnle et al. (2016), probably does not play a major role. We note that the effective temperature is in general higher compared to Macintosh et al. (2015). The new parameters are suggestive of a higher mass for the planet than previously thought. The high surface gravity at a radius slightly bigger than that of Jupiter is consistent with a high-mass planet $M_{gravity} = 9.1^{+4.9}_{-3.3} M_J$, whereas the formation model that we consider is compatible with a wide range of masses depending on the initial conditions and does not strongly constrain the mass. Assuming the model atmosphere derived mass would mean that we can reject pure hot- and pure cold-start models. However, if 51 Eri b were in the brown dwarf mass regime we would expect to see a higher radius if 51 Eri b owing to deuterium burning, which makes this scenario unlikely. This discussion highlights the immense difficulty of precise mass determinations using direct imaging.

Tests performed for patchy cloud models showed that they do not improve the result significantly for the data used in comparison with a model of uniform cloud coverage and at this point do not seem to justify the increase in model complexity, which comes with the linear combination of clear and cloudy models. Further tests should be performed to explore patchiness in detail, for example using Bayesian evidence in the model comparison to account for overfitting and complexity. The consistency of the H-band flux over three independent measurements, speaks against strong variability in the J band, but to answer this question conclusively additional data is necessary. If there truly is variability, a more complex model, such as a patchy cloud model becomes necessary to explain the data. There is a strong need to consider super-solar metallicities in models of exoplanet atmospheres, beyond what is currently available. Beyond the characterization of the planet itself, neglecting super-solar metallicity will impact thermal evolution models of exoplanet and consequently limits placed on the occurrence rate of planets of certain mass through direct imaging. This impact will be especially noticeable if observations are performed in the K band.

The empirical comparison to other substellar objects confirmed the peculiarity of 51 Eri b. It is located in a unique place in the color–color and color-magnitude diagrams, which may be

related to low-surface gravity and/or young age effects, but also shares common properties with other late-T dwarfs. The empirical characterization approach is limited by a lack of comparable objects from clusters and young moving groups with spectral type later than T5. Finally, no additional point sources were detected in the data. However, the SPHERE/IFS observations together with the NACO/SAM data provide strong constraints on the existence of additional objects in the system, rejecting >20 M_J companions at ~2−4 au and planets more massive than 4 M_J beyond 4.5 au and 2 M_J beyond 9 au.

Future IFS observations in the Y, K, and L bands, with existing and upcoming instruments (e.g., SPHERE, GPI, ALES, and CHARIS), as well as photometric measurements with JWST/NIRCAM and MIRI at mid-infrared wavelengths, can significantly improve the constraints on the atmospheric parameters. This will make 51 Eri b one of the planets with the best spectral coverage available and can thereby serve as a benchmark for atmospheric model development.

References

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872
Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., & Schweitzer, A. 2001, ApJ, 556, 357
Allard, F., Homeier, D., & Freytag, B. 2012, Phil. Trans. R. Soc. Lond. Ser. A, 370, 2765
Allers, K. N., & Liu, M. C. 2013, ApJ, 772, 79
Amara, A., & Quanz, S. P. 2012, MNRAS, 427, 948
Apai, D., Kasper, M., Skemer, A., et al. 2016, ApJ, 820, 40
Artigau, E., Nadeau, D., & Doyon, R. 2003, in Brown Dwarfs, ed. E. Martín, IAU Symp., 211, 451
Bailey, V., Meshkat, T., Reiter, M., et al. 2014, ApJ, 780, L4
Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Baudino, J.-L., Bézard, B., Boccaletti, A., et al. 2015, A&A, 582, A83
Bell, C. P. M., Mamajek, E. E., & Naylor, T. 2015, MNRAS, 454, 593
Best, W. M. J., Liu, M. C., Magnier, E. A., et al. 2015, ApJ, 814, 118
Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Proc. SPIE, 7014, 701418
Biller, B. A., Vos, J., Bonavita, M., et al. 2015, ApJ, 813, L23
Binks, A. S. & Jeffries, R. D. 2014, MNRAS, 438, L11
Boccaletti, A., Abe, L., Baudrand, J., et al. 2008, in SPIE Conf. Ser., 7015, 1

Bohlin, R. C. 2007, in The Future of Photometric, Spectrophotometric and Polarimetric Standardization, ed. C. Sterken, ASP Conf. Ser., 364, 315
Bonnefoy, M., Zurlo, A., Baudino, J. L., et al. 2016, A&A, 587, A58
Borysow, A. 1991, Icarus, 92, 273
Burgasser, A. J. 2007, AJ, 134, 1330
Burgasser, A. J. 2014, in Astron. Soc. India Conf. Ser., 11, 7
Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., et al. 2000, AJ, 120, 473
Burgasser, A. J., Burrows, A., & Kirkpatrick, J. D. 2006a, ApJ, 639, 1095
Burgasser, A. J., Geballe, T. R., Leggett, S. K., Kirkpatrick, J. D., & Golimowski, D. A. 2006b, ApJ, 637, 1067
Burgasser, A. J., Kirkpatrick, J. D., Cruz, K. L., et al. 2006c, ApJS, 166, 585
Burgasser, A. J., Cruz, K. L., Cushing, M., et al. 2010, ApJ, 710, 1142
Burgasser, A. J., Cushing, M. C., Kirkpatrick, J. D., et al. 2011, ApJ, 735, 116
Burningham, B., Pinfield, D. J., Leggett, S. K., et al. 2008, MNRAS, 391, 320
Burningham, B., Leggett, S. K., Homeier, D., et al. 2011, MNRAS, 414, 3590
Cantalloube, F., Mouillet, D., Mugnier, L. M., et al. 2015, A&A, 582, A89
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138
Chabrier, G., Baraffe, I., Leconte, J., Gallardo, J., & Barman, T. 2009, in 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. E. Stempels, AIP Conf. Ser., 1094, 102
Chilcote, J., Barman, T., Fitzgerald, M. P., et al. 2015, ApJ, 798, L3
Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Proc. SPIE, 7014, 70143
Cushing, M. C., Marley, M. S., Saumon, D., et al. 2008, ApJ, 678, 1372
Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, 2MASS All Sky Catalog of point sources
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep.
De Rosa, R. J., Nielsen, E. L., Blunt, S. C., et al. 2015, ApJ, 814, L3
Delorme, P., Gagné, J., Malo, L., et al. 2012, A&A, 548, A26
Delorme, P., Dupuy, T., Gagné, J., et al. 2017, A&A, 602, A82
Dohlen, K., Langlois, M., Saisse, M., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Proc. SPIE, 7014, 70143
Faherty, J. K., Burgasser, A. J., West, A. A., et al. 2010, AJ, 139, 176
Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fusco, T., Sauvage, J.-F., Petit, C., et al. 2014, in Adaptive Optics Systems IV, Proc. SPIE, 9148, 91481
Gagné, J., Lafrenière, D., Doyon, R., Malo, L., & Artigau, É. 2014, ApJ, 783, 121
Gagné, J., Burgasser, A. J., Faherty, J. K., et al. 2015, ApJ, 808, L20
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Greco, J. P., & Brandt, T. D. 2016, ApJ, 833, 134
Helou, G., & Walker, D. W. 1988, Infrared astronomical satellite (IRAS) catalogs and atlases, Vol. 7, The small scale structure catalog, 1
Henning, T., & Stognienko, R. 1996, A&A, 311, 291
Hiranaka, K., Cruz, K. L., Douglas, S. T., Marley, M. S., & Baldassare, V. F. 2016, ApJ, 830, 96
Hoeg, E., Bässgen, G., Bastian, U., et al. 1997, A&A, 323, 57
Ingraham, P., Marley, M. S., Saumon, D., et al. 2014, ApJ, 794, L15
Jaeger, C., Molster, F. J., Dorschner, J., et al. 1998, A&A, 339, 904
Jones, A., Noll, S., Kausch, W., Szyszka, C., & Kimeswenger, S. 2013, A&A, 560, A91
Kao, M. M., Hallinan, G., Pineda, J. S., et al. 2016, ApJ, 818, 24
King, R. R., McCaughrean, M. J., Homeier, D., et al. 2010, A&A, 510, A99
Kraus, A. L., Ireland, M. J., Martinache, F., & Lloyd, J. P. 2008, ApJ, 679, 762
Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, Science, 329, 57
Lang, D., Hogg, D. W., & Schlegel, D. J. 2016, AJ, 151, 36
Langlois, M., Vigan, A., Moutou, C., et al. 2013, in Proc. Third AO4ELT Conf., eds. S. Esposito, & L. Fini, 63
Lee, J.-M., Heng, K., & Irwin, P. G. J. 2013, ApJ, 778, 97
Leggett, S. K., Allard, F., Dahn, C., et al. 2000, ApJ, 535, 965
Line, M. R., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, ApJ, 807, 183
Line, M. R., Marley, M. S., Liu, M. C., et al. 2016, ApJ, submitted [arXiv:1612.02809]
Liu, M. C., Leggett, S. K., & Chiu, K. 2007, ApJ, 660, 1507
Liu, M. C., Dupuy, T. J., & Allers, K. N. 2016, ApJ, 833, 96
Mace, G. N., Kirkpatrick, J. D., Cushing, M. C., et al. 2013, ApJS, 205, 6
Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proc. National Academy of Science, 111, 12661
Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64
Maire, A.-L., Boccaletti, A., Rameau, J., et al. 2014, A&A, 566, A126
Maire, A.-L., Bonnefoy, M., Ginski, C., et al. 2016a, A&A, 587, A56
Maire, A.-L., Langlois, M., Dohlen, K., et al. 2016b, in SPIE Conf. Ser., 9908, 990834

Mamajek, E. E., & Bell, C. P. M. 2014, MNRAS, 445, 2169
Mancini, L., Giordano, M., Mollière, P., et al. 2016a, MNRAS, 461, 1053
Mancini, L., Kemmer, J., Southworth, J., et al. 2016b, MNRAS, 459, 1393
Marleau, G.-D., & Cumming, A. 2014, MNRAS, 437, 1378
Marleau, G.-D., Klahr, H., Kuiper, R., & Mordasini, C. 2017, ApJ, 836, 221
Marley, M. S., Fortney, J. J., Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2007, ApJ, 655, 541
Marley, M. S., Saumon, D., & Goldblatt, C. 2010, ApJ, 723, L117
Marocco, F., Day-Jones, A. C., Lucas, P. W., et al. 2014, MNRAS, 439, 372
Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556
Marois, C., Macintosh, B., Barman, T., et al. 2008, Science, 322, 1348
Marois, C., Macintosh, B., & Véran, J.-P. 2010a, in SPIE Conf. Ser., 7736, 1
Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010b, Nature, 468, 1080
Marois, C., Correia, C., Galicher, R., et al. 2014, in SPIE Conf. Ser., 9148, 91480U
Mawet, D., Milli, J., Wahhaj, Z., et al. 2014, ApJ, 792, 97
Mermilliod, J. C. 2006, VizieR Online Data Catalog, 2168, 0
Mesa, D., Gratton, R., Zurlo, A., et al. 2015, A&A, 576, A121
Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47
Mollière, P., van Boekel, R., Bouwman, J., et al. 2017, A&A, 600, A10
Monet, D. G., Dahn, C. C., Vrba, F. J., et al. 1992, AJ, 103, 638
Montet, B. T., Bowler, B. P., Shkolnik, E. L., et al. 2015, ApJ, 813, L11
Mordasini, C. 2013, A&A, 558, A113
Mordasini, C., Alibert, Y., Georgy, C., et al. 2012a, A&A, 547, A112
Mordasini, C., Alibert, Y., Klahr, H., & Henning, T. 2012b, A&A, 547, A111
Mordasini, C., Klahr, H., Alibert, Y., Miller, N., & Henning, T. 2014, A&A, 566, A141
Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2012, ApJ, 756, 172
Naud, M.-E., Artigau, É., Malo, L., et al. 2014, ApJ, 787, 5
Noll, S., Kausch, W., Barden, M., et al. 2012, A&A, 543, A92
Palik, E. 2012, Handbook of Optical Constants of Solids No. Bd. 1 (Elsevier Science)
Patel, R. I., Metchev, S. A., & Heinze, A. 2014, ApJS, 212, 10
Pavlov, A., Möller-Nilsson, O., Feldt, M., et al. 2008, in Advanced Software and Control for Astronomy II, Proc. SPIE, 7019, 701939
Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, ArXiv Astrophysics e-prints [arXiv:astro-ph/0703658]
Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999, PASP, 111, 587
Radigan, J., Jayawardhana, R., Lafrenière, D., et al. 2012, ApJ, 750, 105
Rameau, J., Chauvin, G., Lagrange, A.-M., et al. 2013, ApJ, 772, L15
Ramírez, I., Allende Prieto, C., & Lambert, D. L. 2013, ApJ, 764, 78
Riviere-Marichalar, P., Barrado, D., Montesinos, B., et al. 2014, A&A, 565, A68
Saumon, D., Marley, M. S., Cushing, M. C., et al. 2006, ApJ, 647, 552
Scott, A., & Duley, W. W. 1996, ApJS, 105, 401
Servoin, J. L., & Piriou, B. 1973, Physica Status Solidi B Basic Research, 55, 677
Simon, M., & Schaefer, G. H. 2011, ApJ, 743, 158
Skemer, A. J., Morley, C. V., Zimmerman, N. T., et al. 2016, ApJ, 817, 166
Soummer, R. 2005, ApJ, 618, L161
Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJ, 755, L28
Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154
Thalmann, C., Schmid, H. M., Boccaletti, A., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Proc. SPIE, 7014, 70143
Thiébaut, E., Denis, L., Mugnier, L. M., et al. 2016, in Adaptive Optics Systems V, eds. E. Marchetti, L. M. Close, & J.-P. Véran, Proc. SPIE, 9909, 990957
Tuthill, P. G., Monnier, J. D., Danchi, W. C., Wishnow, E. H., & Haniff, C. A. 2000, PASP, 112, 555
Venemans, B. P., Bañados, E., Decarli, R., et al. 2015, ApJ, 801, L11

Vigan, A., Moutou, C., Langlois, M., et al. 2010, MNRAS, 407, 71
Vigan, A., Bonnefoy, M., Ginski, C., et al. 2016, A&A, 587, A55
Wagner, K., Apai, D., Kasper, M., et al. 2016, Science, 353, 673
Wahhaj, Z., Cieza, L. A., Mawet, D., et al. 2015, A&A, 581, A24
Zahnle, K. J., & Marley, M. S. 2014, ApJ, 797, 41
Zahnle, K., Marley, M. S., Morley, C. V., & Moses, J. I. 2016, ApJ, 824, 137
Zapatero Osorio, M. R., Béjar, V. J. S., Miles-Páez, P. A., et al. 2014, A&A, 568, A6
Zuckerman, B., Song, I., Bessell, M. S., & Webb, R. A. 2001, ApJ, 562, L87
Zurlo, A., Vigan, A., Galicher, R., et al. 2016, A&A, 587, A57

[1] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: samland@mpia.de
[2] International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD), 69117 Heidelberg, Germany
[3] Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
[4] ONERA, The French Aerospace Lab BP72, 29 avenue de la Division Leclerc, 92322 Châtillon Cedex, France
[5] Geneva Observatory, University of Geneva, Chemin des Maillettes 51, 1290 Versoix, Switzerland
[6] INAF–Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
[7] Institute for Astronomy, University of Edinburgh, Blackford Hill View, Edinburgh EH9 3HJ, UK
[8] European Southern Observatory, Alonso de Cordova 3107, Casilla 19001 Vitacura, Santiago 19, Chile
[9] Department of Physics & Astronomy, College of Charleston, 66 George Street, Charleston, SC 29424, USA
[10] Department of Astronomy, Stockholm University, AlbaNova University Center, 106 91 Stockholm, Sweden
[11] Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany
[12] Physikalisches Institut, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland
[13] CRAL, UMR 5574, CNRS, Université Lyon 1, 9 avenue Charles André, 69561 Saint Genis Laval Cedex, France
[14] Aix-Marseille Univ., CNRS, LAM, Laboratoire d'Astrophysique de Marseille, 13013 Marseille, France
[15] Institute for Astronomy, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
[16] Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile
[17] Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejercito 441, Santiago, Chile
[18] NASA Ames Research Center, Moffett Field, CA 94035, USA
[19] LESIA, Observatoire de Paris, PSL Research Univ., CNRS, Univ. Paris Diderot, Sorbonne Paris Cité, UPMC Paris 6, Sorbonne Univ., 5 place Jules Janssen, 92195 Meudon Cedex, France
[20] School of Earth & Space Exploration, Arizona State University, Tempe, AZ 85287, USA
[21] Ural Federal University, 620002 Yekaterinburg, Russia
[22] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Lagrange, 06304 Nice Cedex 4, France

Appendix A: Alternative reductions

Shown in Fig. A.1 are the extracted spectra with the different algorithms that were tested (see Sect. 3.1): ANDROMEDA (panel 1), PCA (panel 2), TLOCI (panel 3, Specal), and additionally PCA with simultaneous use of ADI and SDI reference images (panel 4). Additionally, panel 5 shows all reductions for the YJ data set and panel 6 all reductions for the YH data set. The ANDROMEDA, PCA, and TLOCI reductions all rely on the same cSDI prereduced data cube, whereas the simultaneous PCA ADI plus SDI reduction is completely independent based on the pipeline introduced in Mesa et al. (2015). The YH spectrum is fully consistent between all reduction methods and the only exception is the low quality of the H band extraction using PCA.

We notice more uncertainties in the absolute calibration of our YJ-spectral data, which changes depending on the exact algorithm used to reduce the data. The ANDROMEDA code yields compatible fluxes between the two observations, as does the independent reduction using spectral PCA with simultaneous ADI and SDI references. Whereas the PCA and TLOCI reductions using the same classical SDI prereduced frames, show higher peak fluxes in the J band. Although these reductions show comparable peak values to the GPI J-band spectrum, they do not follow the same spectral shape over the entirety of the GPI J-band spectrum.
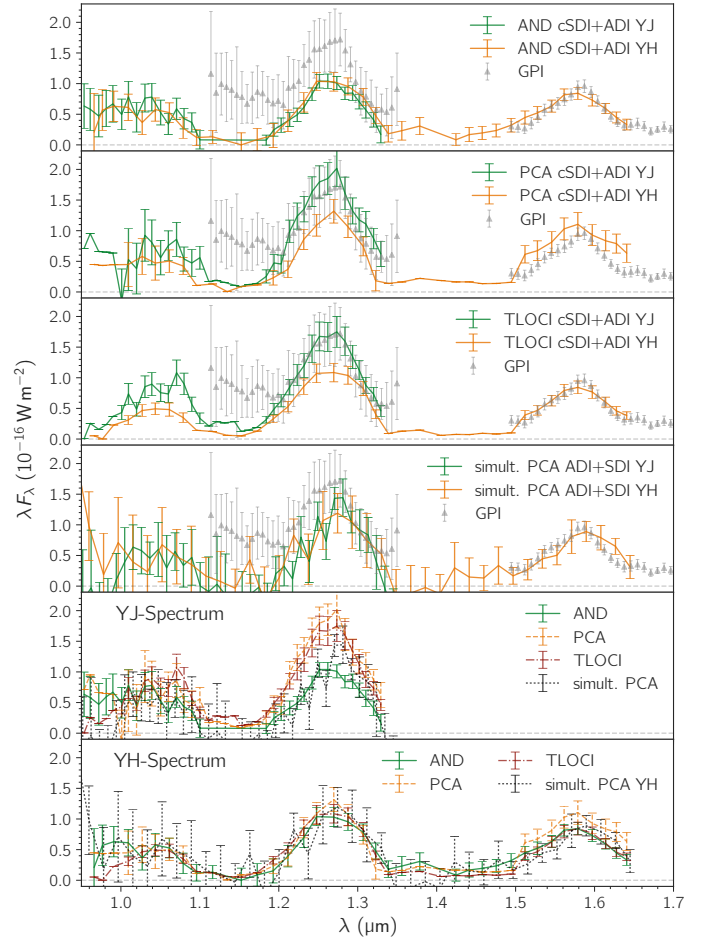
Fig. A.1. *From top to bottom*: the extracted spectra with ANDROMEDA (panel 1), PCA (panel 2), TLOCI (panel 3), and simultaneous ADI+SDI with PCA (panel 4), all reductions for YJ (panel 5), and all reductions for YH (panel 6). The GPI spectra are plotted for comparison in the first 4 panels. The PCA and TLOCI pipeline used automatically exchange non-significant detections with upper limits, in which case no uncertainty is displayed on the data point.

Appendix B: Testing metallicity determination with benchmark brown dwarfs

The two T7.5 brown dwarfs Gl 570D and HD 3651B are considered to be benchmark objects, because they are on wide orbits around extensively studied K stars with known properties (Line et al. 2015, 2016). Having formed from the same cloud as their host stars, these brown dwarfs provide the opportunity to compare the derived parameters for the brown dwarfs, especially their composition, with their respective host star. Given the high metallicity inferred by our model for 51 Eri b, we want to make sure that our methodology and models are not biased toward obtaining high metallicity results. Below we compare the host star metallicities with the brown dwarf metallicities obtained with our self-consistent equilibrium *petitCODE* models. We further compare the parameters with the parameters derived by Line et al. (2015).

To make a comparison with Line et al. (2015) easier, we follow the same methodology, using every third pixel to avoid correlations between neighboring data points and the same additional free fit-parameter b in the likelihood function, which accounts for the underestimated uncertainties in the data by adding a constant 10^b term to the flux uncertainties. A flat prior is assumed for this parameter. All systematics in the absolute photometric calibration and distance are absorbed into the brown dwarf "radius"-parameter R because, with a flat prior, it allows the spectrum to freely float up and down. As also pointed out in Line et al. (2015), the absolute calibration is not necessarily reliable, so the radius should *not* be seen as a physical quantity, but rather as a data scaling parameter. On the positive side, this means that the determination of $T_{\rm eff}$, $\log g$, and [Fe/H] is independent of the absolute photometry and distance of the objects and is purely determined by the model shape and relative strength of the features. A summary of the derived parameters for Gl 570D and HD 3651B for the Line et al. (2015) retrieval and the *petitCODE* clear models is shown in

Table B.1. The best fit *petitCODE* model spectra are shown in Fig. B.1 and the respective posterior probability distributions in Figs. B.2 and B.3. Line et al. (2016) gives a summary of literature metallicity values for the host star Gl 570A and Line et al. (2015) derived the metallicity for HD 3651A. For Gl 570D, Saumon et al. (2006) gave an averaged metallicity based on recent literature of [Fe/H] = 0.09 ± 0.04, and Casagrande et al. (2011) found a more recent value of 0.31 and −0.05 ± 0.17 from Line et al. (2015). The preponderance of evidence seems to suggest a slightly super-solar metallicity, whereas HD 3651A has a super-solar metallicity in the range of [Fe/H] = 0.18 ± 0.07 (Ramírez et al. 2013).

The metallicity we determined using the *petitCODE* models is within the given range of host star metallicities, showing that our model and fitting approach can be used to estimate metallicities reliably. Comparing the results to the free retrieval performed by Line et al. (2015), their metallicities fall on the lower end, whereas ours fall on the higher end of the metallicity range, which may reflect a difference in the free retrieval versus self-consistent model approach. For example, their retrieval requires an additional step to compute the chemical bulk metallicity from the retrieved abundances. On the other hand, our models assume a fixed solar C/O ratio. Both objects share very similar properties according to our fits, except that HD 3651B is more metal rich. This is consistent with our expectations as they are both classified as T7.5 dwarfs and the normalized spectra are virtually indistinguishable at the resolution of the SpeX instrument. Only the strength of the K-band flux, which is an indicator of metallicity mainly due to its sensitivity to collision-induced absorption (CIA) of H_2–H_2 and H_2–He pairs (see discussion in Sect. 4.2), is stronger in HD 3651B. Compared to Line et al. (2015) our models are about 50 K hotter in effective temperature. The derived surface gravity for Gl 570D is almost the same, whereas Line et al. (2015) arrive at a significantly higher surface gravity for HD 3651B.

Table B.1. Benchmark brown dwarfs.

Object	$T_{\rm eff}$	$\log g$	Chemically derived bulk [Fe/H]	$[{\rm Fe/H}]_{\rm BD} - [{\rm Fe/H}]_{\rm host}$ [a]
petitCODE clear model				
Gl 570D	769^{+14}_{-13}	4.67 ± 0.04	0.11 ± 0.04	$+0.16 \pm 0.18$
HD 3651B	783^{+13}_{-12}	4.64 ± 0.04	0.25 ± 0.04	$+0.07 \pm 0.08$
Line et al. (2015) retrieval				
Gl 570D	714^{+20}_{-23}	$4.76^{+0.27}_{-0.28}$	-0.15	-0.10 ± 0.17
HD 3651B	726^{+22}_{-21}	$5.12^{+0.09}_{-0.17}$	$+0.08$	-0.10 ± 0.07

Notes. Summary of modeling results for Gl 570D and HD 3651B using *petitCODE* clear models and comparison to atmospheric retrieval results by Line et al. (2015). [a] Difference of best fit model to middle of metallicity range of host star. Uncertainty includes the width of host star metallicity range in all cases, as well as uncertainty for brown dwarf metallicity in the *petitCODE* case. Host star metallicities used here: Gl 570D: −0.22–0.12 or −0.05 ± 0.17 (Line et al. 2015); HD 3651B: 0.11–0.25 or 0.18 ± 0.07 (Ramírez et al. 2013).
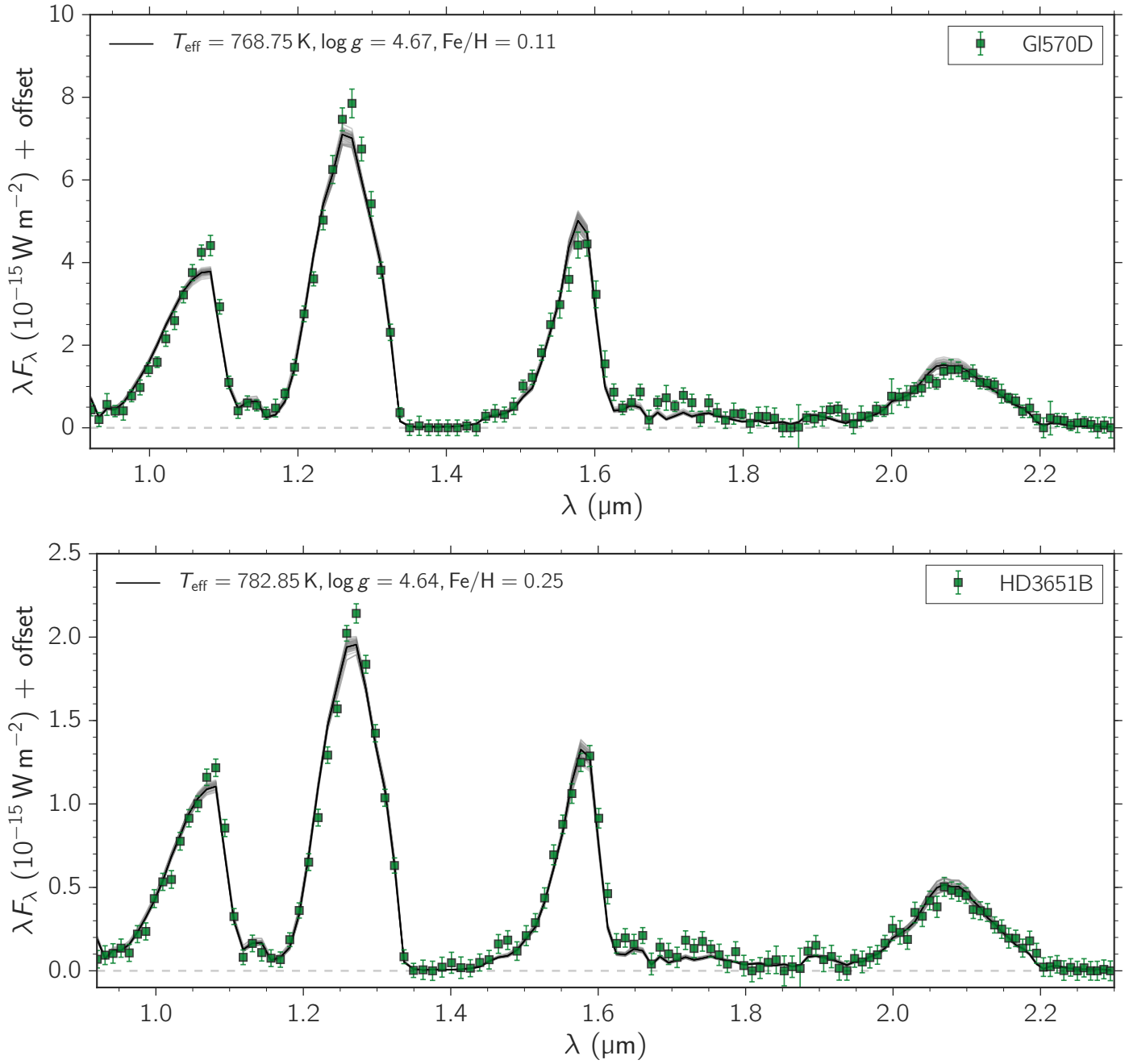
Fig. B.1. *Upper panel*: best fit *petitCODE* clear spectrum for Gl 570D; *lower panel*: same for HD 3651B. The overplotted gray lines represent the model scatter with spectra generated from randomly drawn samples of the posterior parameter distribution. Error bars plotted include the best fit value of the *b* parameter, correcting for the underestimated data uncertainty.
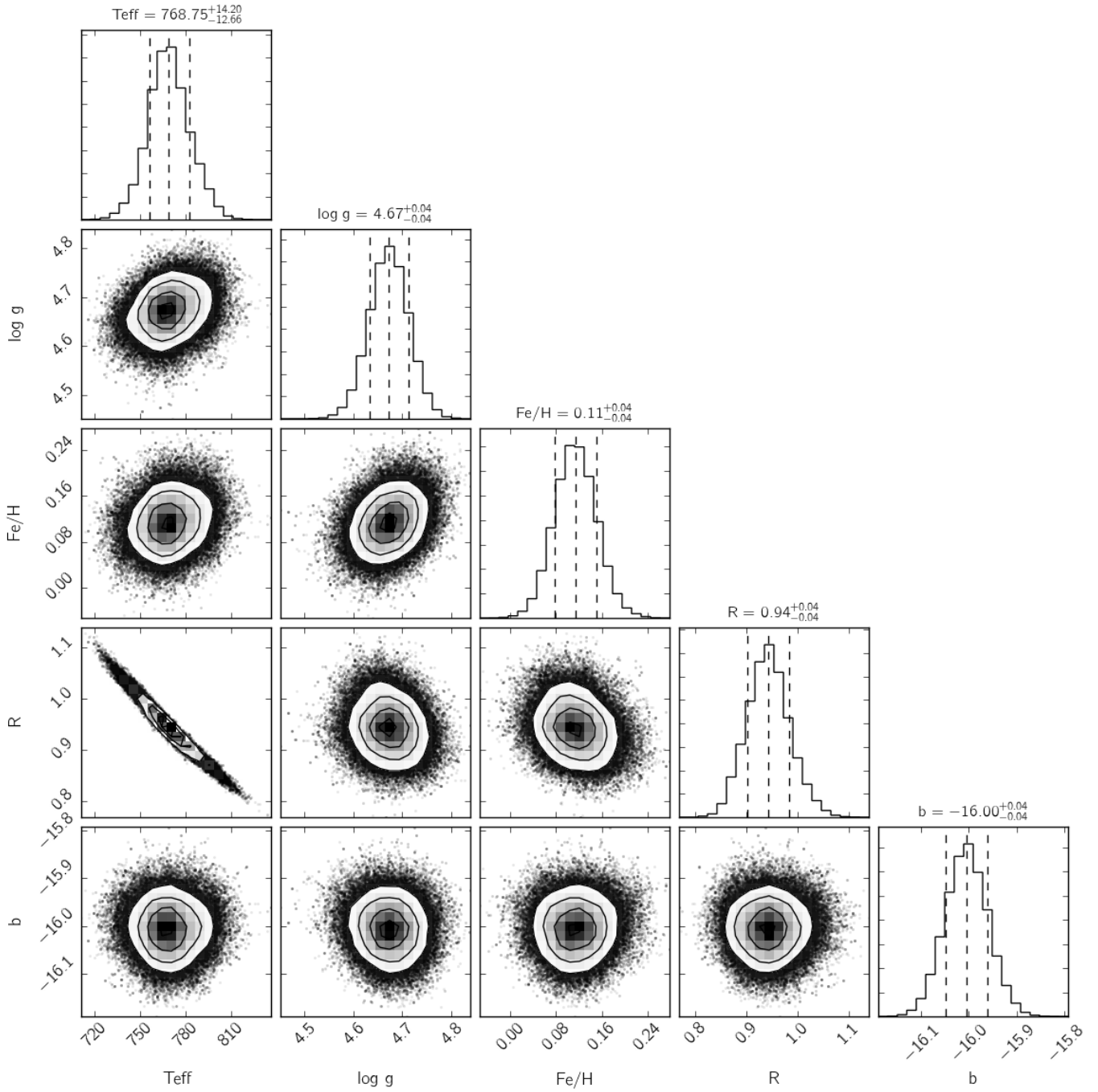
Fig. B.2. Posterior probability distribution of the clear *petitCODE* fitted to Gl 570D, including a further scale parameter *b* as an additive term to the flux uncertainty.
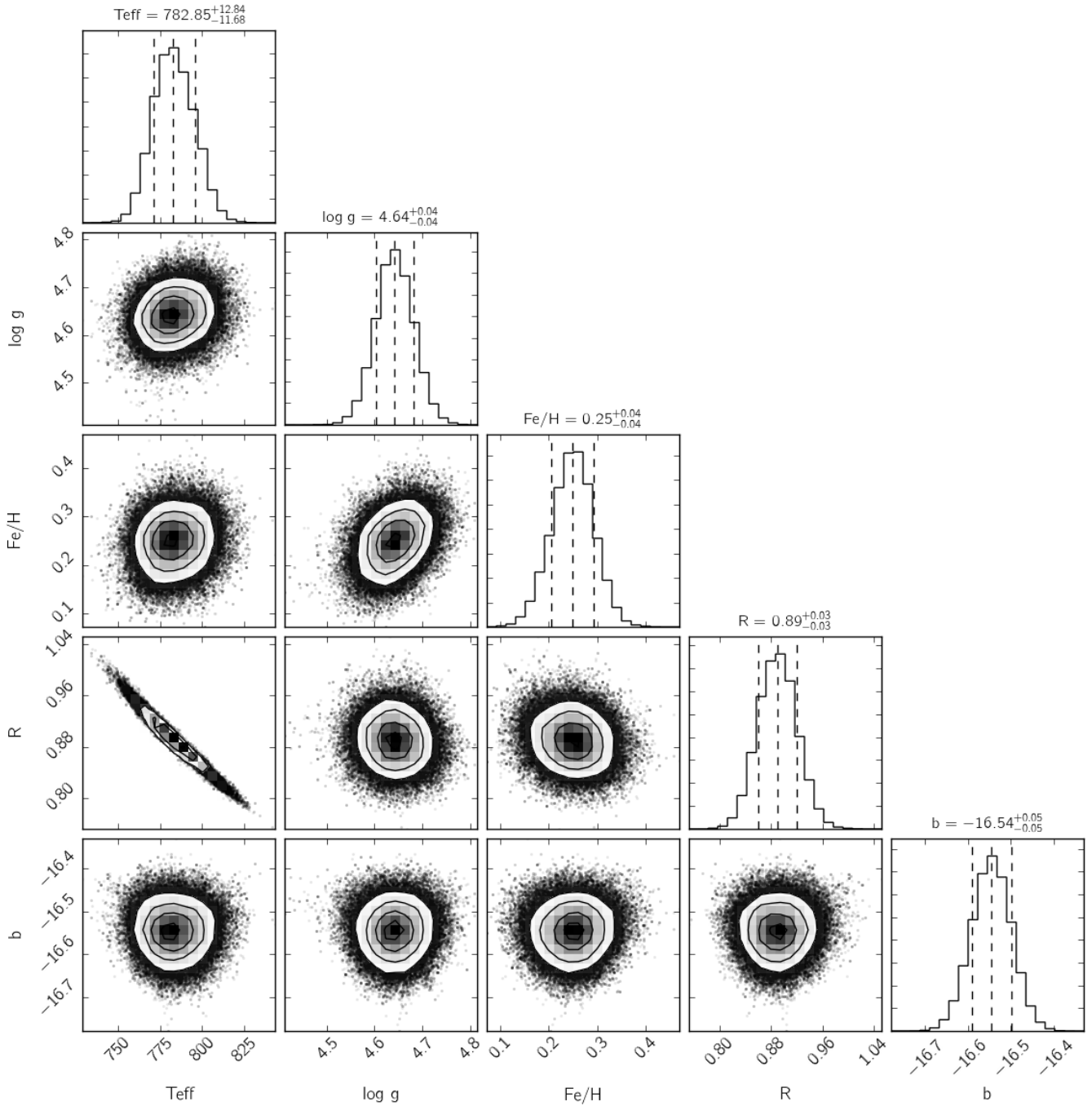
Fig. B.3. Posterior probability distribution of the clear *petitCODE* fitted to HD 3651B, including a further scale parameter *b* as an additive term to the flux uncertainty.
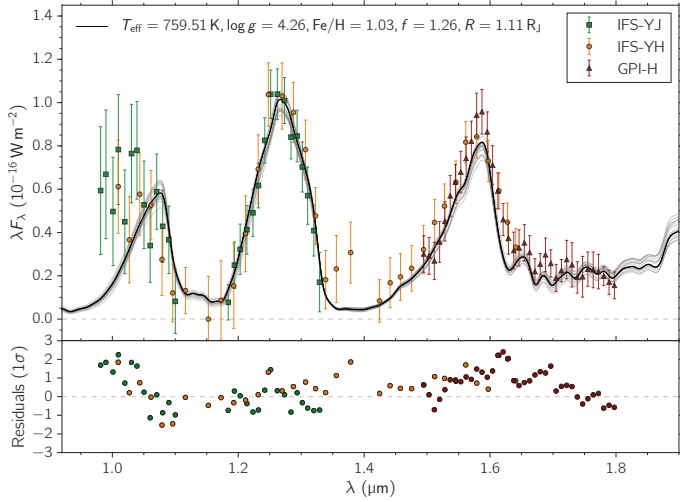
Fig. C.1. Same plot as Fig. 11, but zooming in on the wavelength range covered by spectral data.

Fig. C.2. Same plot as Fig. 11, but using the clear model.

Fig. C.3. Same plot as Fig. 11, but using the Morley et al. (2012) model.
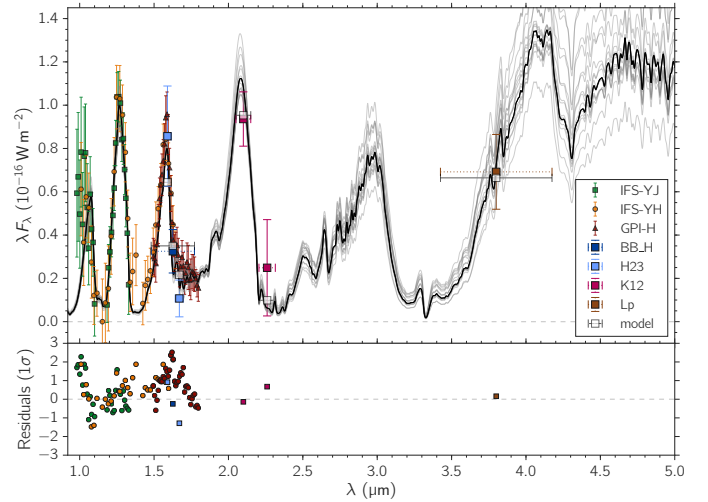
Fig. C.4. Same plot as Fig. 11, but using the patchy-cloud model described in Sect. 4.2.4. Best fitting parameters are: $T_{\mathrm{cloud}} = 750 \pm 25$ K, $T_{\mathrm{clear}} = 815^{+70}_{-40}$ K, $CF = 0.9 \pm 0.05$, $\log g = 4.5 \pm 0.3$, [Fe/H] $= 1.25$, $f_{\mathrm{sed}} = 1.10 \pm 0.15$, $R = 1.10 \pm 0.15\, R_{\mathrm{J}}$.

Appendix C: Model spectra

Figure C.1 shows a zoom in on the IFS spectra in the best fit cloudy *petitCODE* model. It can be seen that there is very good agreement between the model and the data in shape and amplitude, except for a systematic offset in the amplitude of the H band, which still exists and could not be modeled without negatively impacting the overall fit to the rest of the spectrum.

Figure C.2 shows the same plot as Fig. 11, but with the cloud-free *petitCODE* model. It is immediately apparent that the cloud-free model is incapable of explaining the long wavelengths of the spectrum (K1 and L' band), which results in unphysical parameters in temperature and radius and the lack of clouds in the model is compensated with extremely high metallicities.

Figure C.3 of the Morley et al. (2012) shows that the lack of metallicity as a free-parameter (only solar metallicity was

available) also skews the overall parameters, especially in order to fit the K1 peak. Since high metallicities are not allowed, the f_{sed} parameter increases in an attempt to compensate, again the resulting physical parameters are unreliable. All of this shows that a complex model that allows coverage of at least the basic physics (e.g., a cloud model with free parameters and non-solar metallicity), is a bare minimum to model these cold giant planets.

Figure C.4 shows the spectrum resulting from the patchy cloud model introduced in Sect. 4.2.4. The spectrum is almost indistinguishable from a pure cloudy model and does not improve the result significantly.

Appendix D: Corner plots

The corner plots for additional models and data sets are shown. Figure D.1 shows the parameter distribution for the cloudy *petitCODE* model in the case in which we exclude the *Y*-band data completely. We see that the *Y* band does not significantly constrain the models. Figure D.2 shows the corner plot for the clear *petitCODE* model and Fig. D.3 for the Morley et al. (2012) model. As discussed above, both lead to skewed results, because important physics is missing. Figure D.4 shows the corner plot for the patchy cloud model, a linear combination of cloudy and cloud-free models, which share the same parameters except for temperature and are linked by a cloud fraction parameter. Cloud fractions are very high and, as pointed out above, the resulting

As an additional experiment, Fig. D.5 shows the posterior distribution for the cloudy model grid when only data from Macintosh et al. (2015) are used. This fit does not include the co-variance matrices and should reduce roughly to a straightforward fit as the discovery paper described (except that the model can vary in metallicity). With the ~40% higher *J*-band flux and missing *K* band, we retrieve a very low surface gravity (same as the discovery paper), but significantly higher temperature outside of our model grid. In the a patchy cloud model of the original paper a higher *J*-band contribution can come from a clear model, but this is more difficult to explain in a pure cloudy model.
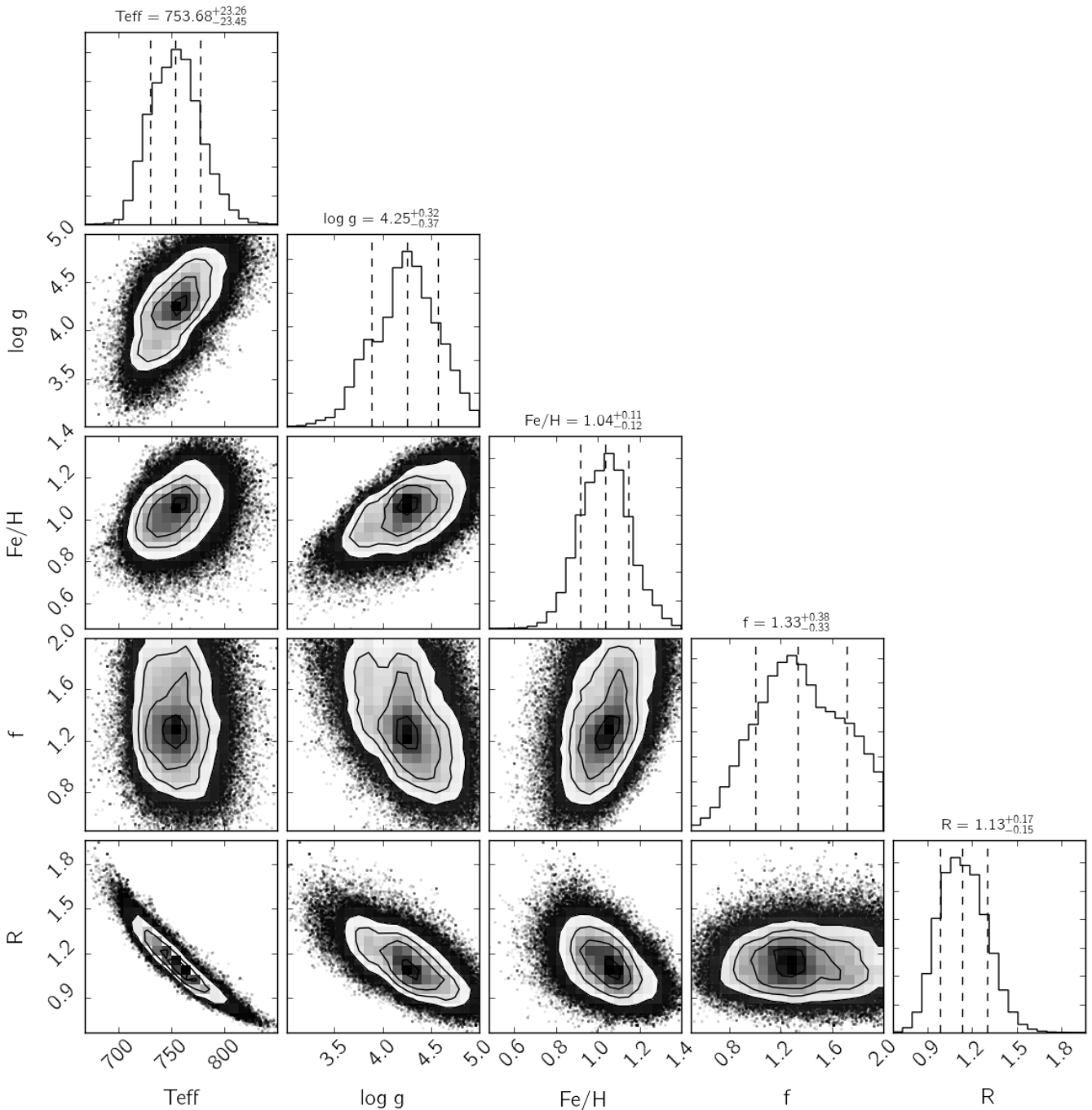

Fig. D.1. Same as Fig. 12, but excluding the *Y*-band data. Corner plot showing the posterior probability distribution of the cloudy *petitCODE* grid with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% quantiles as commonly done for multivariate MCMC results.
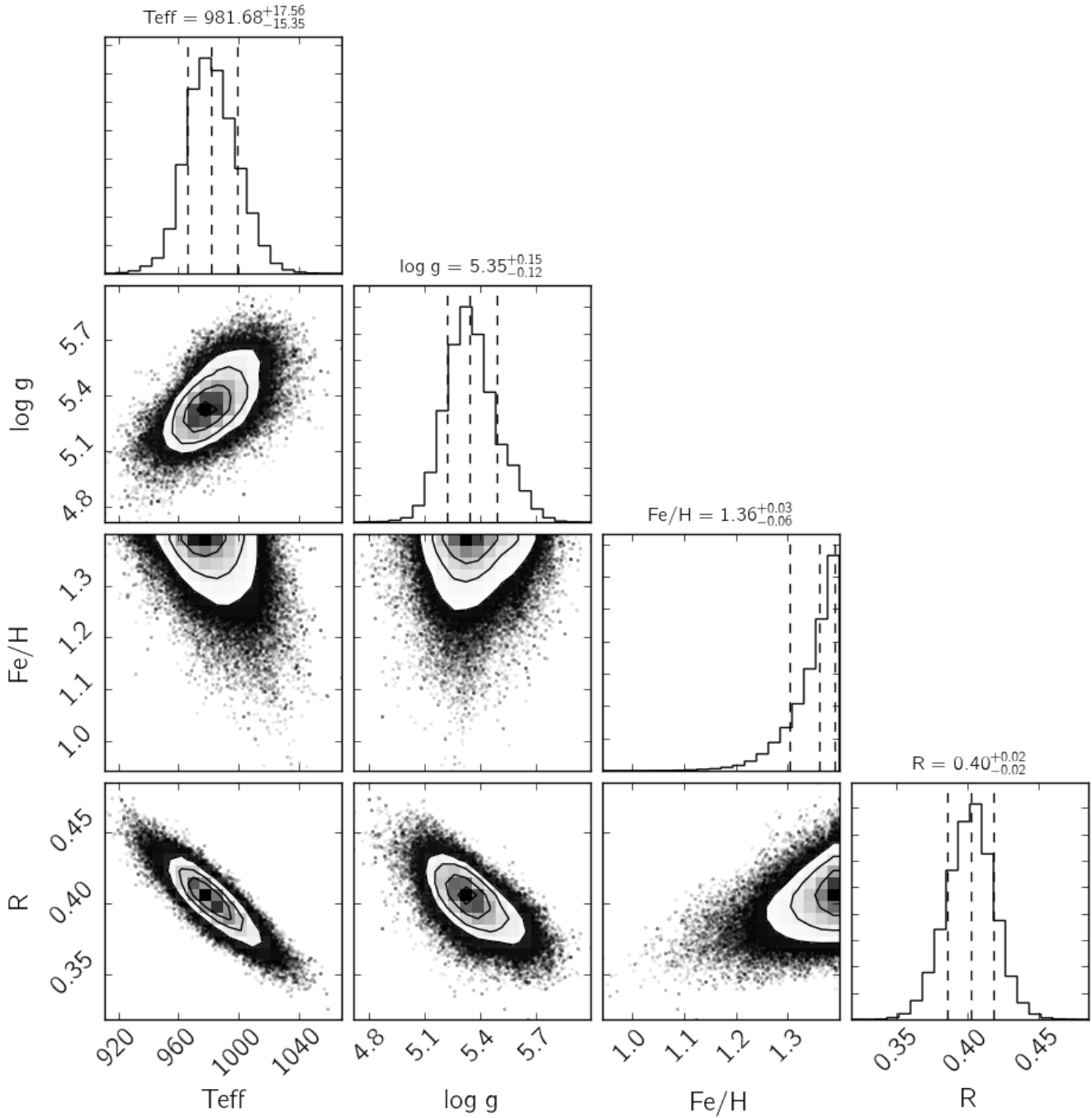
Fig. D.2. Posterior probability distribution of the clear *petitCODE* grid, with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% quantiles as commonly done for multivariate MCMC results. Note that a clear model atmosphere requires a small radius, which speaks against this model.
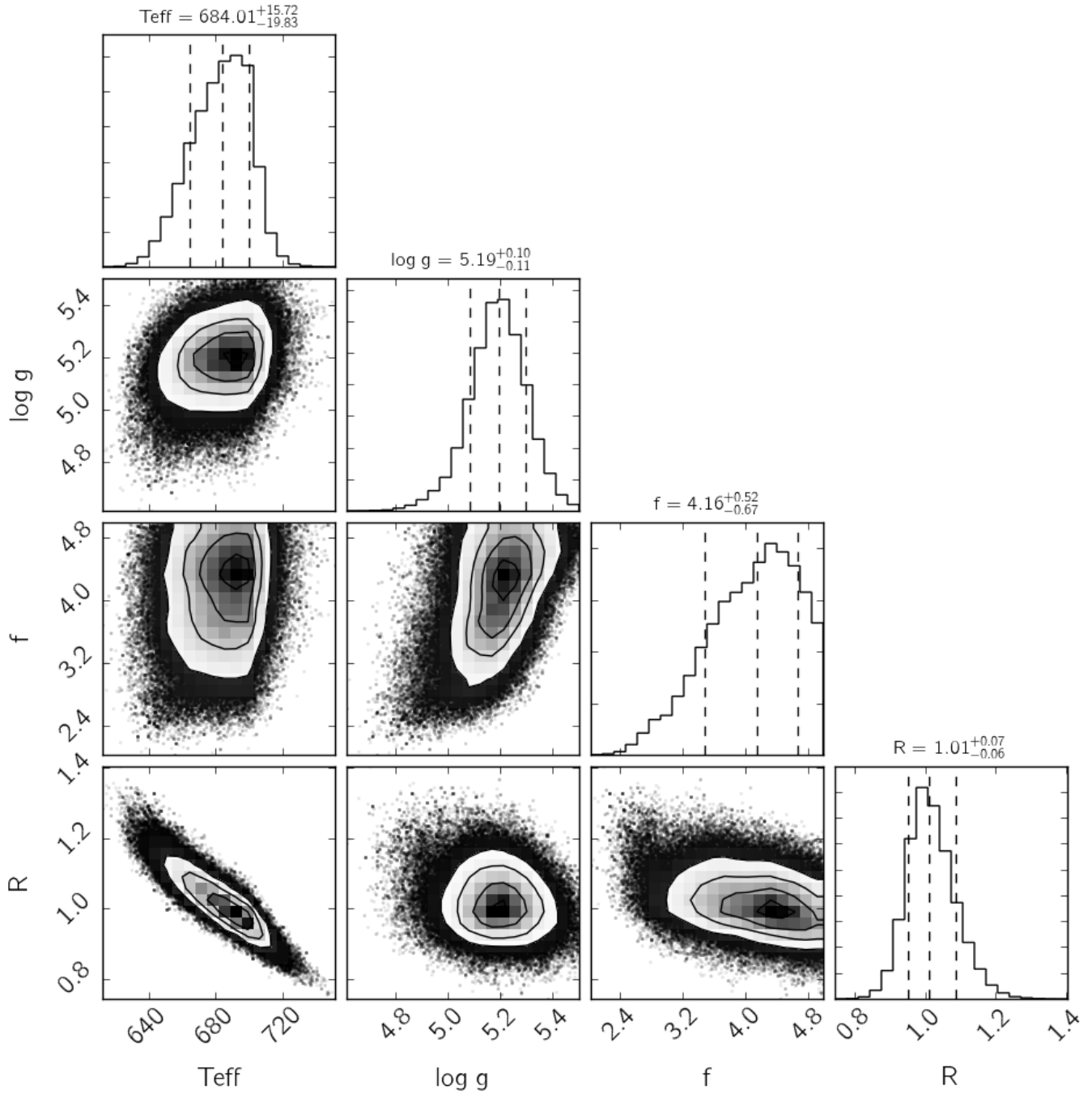
Fig. D.3. Posterior probability distribution of the Morley et al. (2012) grid, with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% quantiles as commonly done for multivariate MCMC results.
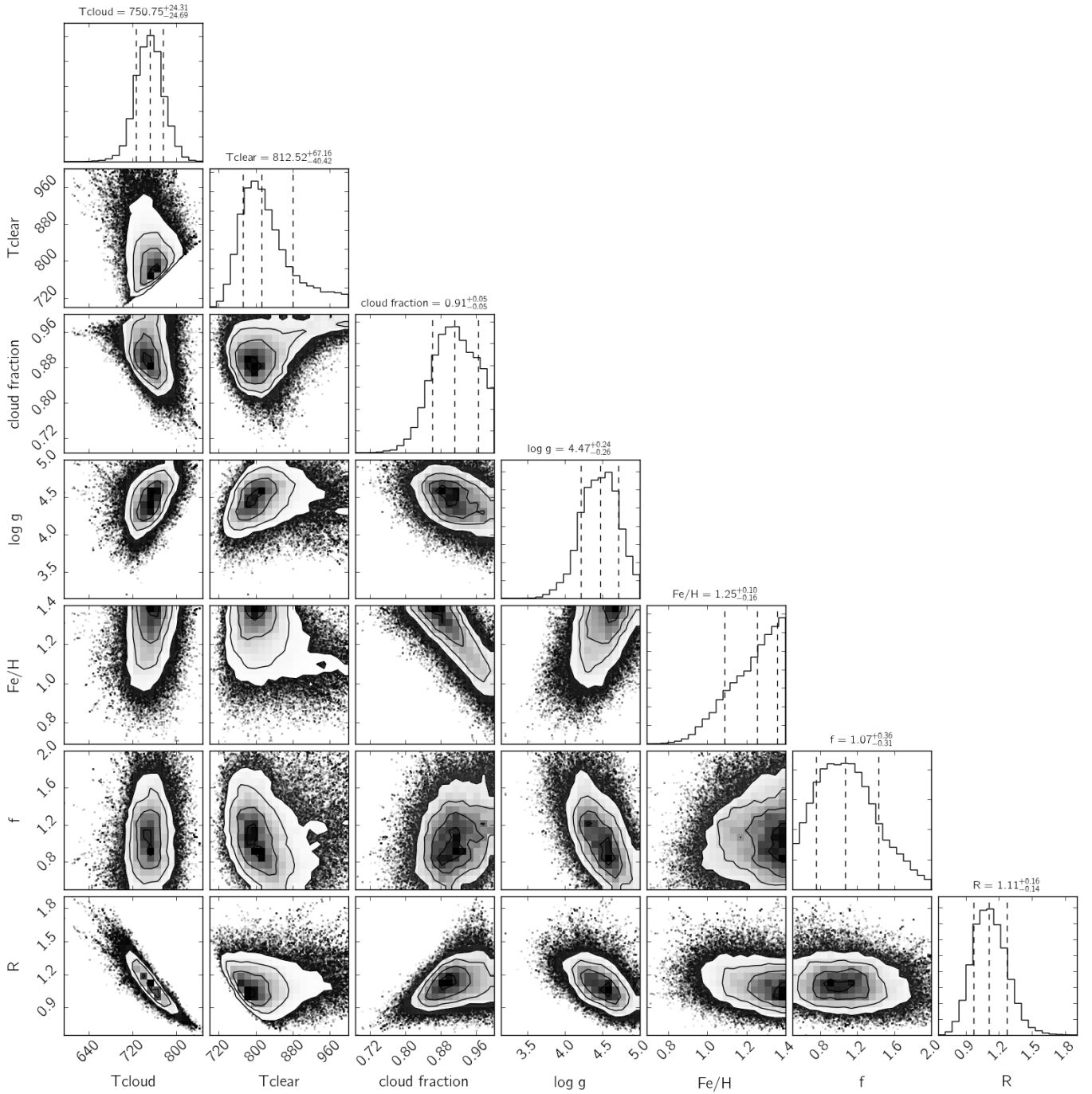
Fig. D.4. Posterior probability distribution of the patchy cloud model described in Sect. 4.2.4, with respect to each of its parameter pair and the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% quantiles as is common for multivariate MCMC results.
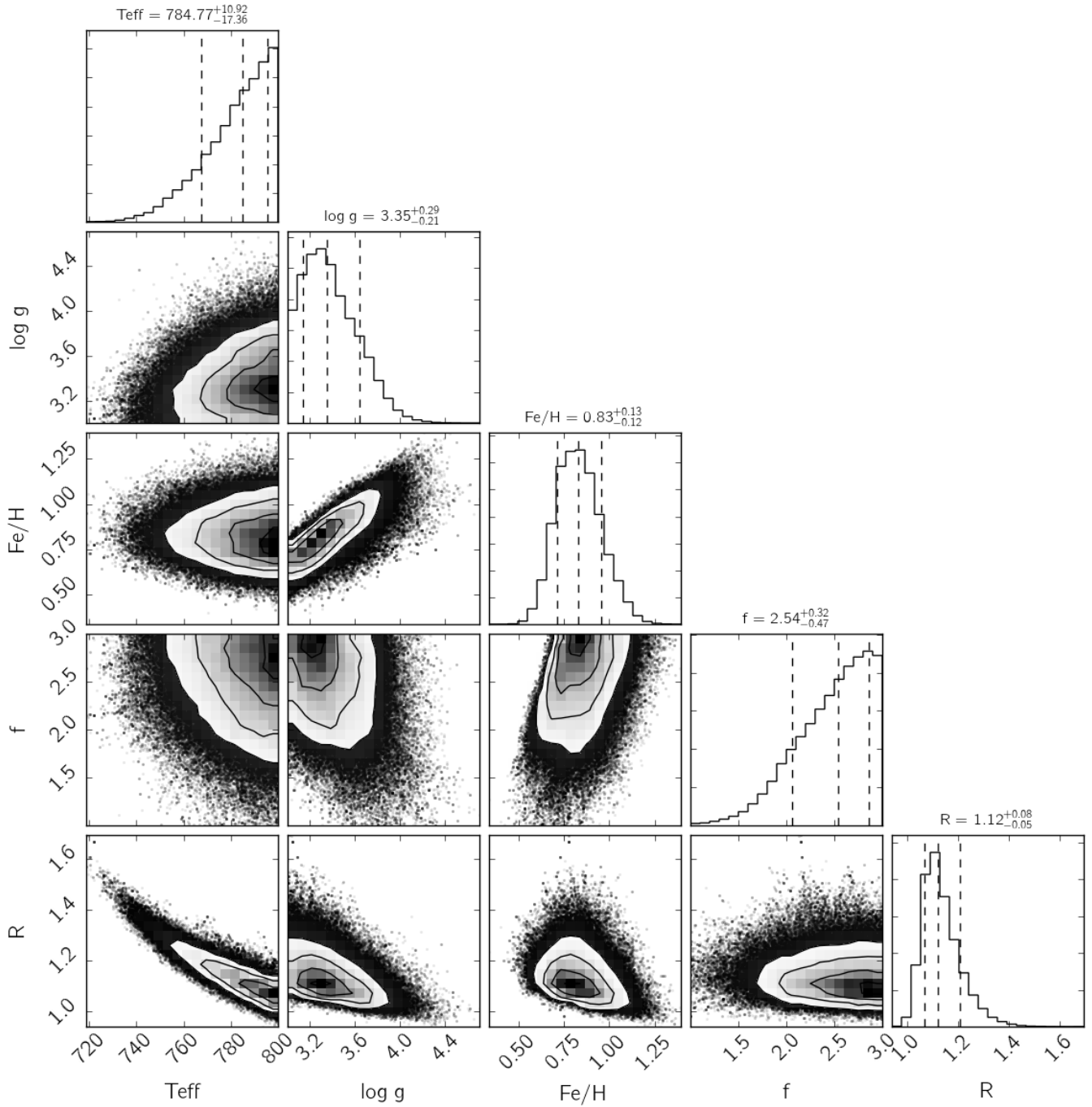
Fig. D.5. Posterior probability distribution using only data published in Macintosh et al. (2015; without taking the covariance into account) with respect to each of its parameter pairs as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% quantiles as is common for multivariate MCMC results.