Astronomy & Astrophysics

MIRACLES: atmospheric characterization of directly imaged planets and substellar companions at 4–5 μm

I. Photometric analysis of β Pic b, HIP 65426 b, PZ Tel B, and HD 206893 B[*]

T. Stolker[1], S. P. Quanz[1,**], K. O. Todorov[2], J. Kühn[3,1], P. Mollière[4,5], M. R. Meyer[6], T. Currie[7,8],
S. Daemgen[1], and B. Lavie[9]

[1] Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
  e-mail: tomas.stolker@phys.ethz.ch
[2] Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1090 GE Amsterdam, The Netherlands
[3] Center for Space and Habitability, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
[4] Leiden Observatory, Leiden University, Postbus 9513, 2300 RA Leiden, The Netherlands
[5] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[6] Department of Astronomy, University of Michigan, 1085 South University Avenue, Ann Arbor, MI 48109-1107, USA
[7] NASA-Ames Research Center, Moffett Field, CA, USA
[8] Subaru Telescope, National Astronomical Observatory of Japan, 650 North Aʻohōkū Place, Hilo, HI 96720, USA
[9] Observatoire astronomique de l'Université de Genève, 51 chemin des Maillettes, 1290 Versoix, Switzerland

ABSTRACT

Context. Directly imaged planets and substellar companions are key targets for the characterization of self-luminous atmospheres. Their photometric appearance at 4–5 μm is sensitive to the chemical composition and cloud content of their atmosphere.

Aims. We aim to systematically characterize the atmospheres of directly imaged low-mass companions at 4–5 μm. We want to homogeneously process the data, provide robust flux measurements, and compile a photometric library at thermal wavelengths of these mostly young, low-gravity objects. In this way, we want to find trends related to their spectral type and surface gravity by comparing with isolated brown dwarfs and predictions from atmospheric models.

Methods. We used the high-resolution, high-contrast capabilities of NACO at the Very Large Telescope (VLT) to directly image the companions of HIP 65426, PZ Tel, and HD 206893 in the NB4.05 and/or M' filters. For the same targets, and additionally β Pic, we also analyzed six archival VLT/NACO datasets which were taken with the NB3.74, L', NB4.05, and M' filters. The data processing and photometric extraction of the companions was done with `PynPoint` while the `species` toolkit was used to further analyze and interpret the fluxes and colors.

Results. We detect for the first time HIP 65426 b, PZ Tel B, and HD 206893 B in the NB4.05 filter, PZ Tel B and HD 206893 B in the M' filter, and β Pic b in the NB3.74 filter. We provide calibrated magnitudes and fluxes with a careful analysis of the error budget, both for the new and archival datasets. The L'–NB4.05 and L'–M' colors of the studied sample are all red while the NB4.05–M' color is blue for β Pic b, gray for PZ Tel B, and red for HIP 65426 b and HD 206893 B (although typically with low significance). The absolute NB4.05 and M' fluxes of our sample are all larger than those of field dwarfs with similar spectral types. Finally, the surface gravity of β Pic b has been constrained to $\log g = 4.17^{+0.10}_{-0.13}$ dex from its photometry and dynamical mass.

Conclusions. A red color at 3–4 μm and a blue color at 4–5 μm might be (partially) caused by H_2O and CO absorption, respectively, which are expected to be the most dominant gaseous opacities in hot ($T_{\rm eff} \gtrsim 1300$ K) atmospheres. The red characteristics of β Pic b, HIP 65426 b, and HD 206893 B at 3–5μm, as well as their higher fluxes in NB4.05 and M' compared to field dwarfs, indicate that cloud densities are enhanced close to the photosphere as a result of their low surface gravity.

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

1. Introduction

The population of directly imaged planetary and substellar companions provides an important window onto the formation and evolution of low-mass objects on long-period orbits (e.g., Bowler 2016). Large-scale surveys have been ongoing for more than a decade, yielding a dozen exoplanet detections (e.g., Marois et al. 2008; Lagrange et al. 2009; Macintosh et al. 2015; Chauvin et al. 2017) and constraints on their demographics beyond ∼10 au (e.g., Stone et al. 2018; Nielsen et al. 2019). Sensitivity limits continue to increase thanks to dedicated high-contrast imaging instruments (e.g., Macintosh et al. 2008; Jovanovic et al. 2015; Beuzit et al. 2019) and observing strategies (e.g., Käufl et al. 2018), as well as developments in image processing techniques and detection algorithms (e.g., Gomez Gonzalez et al. 2018; Flasseur et al. 2018). As a result, detailed spectrophotometric measurements have been carried out at near-infrared (NIR) wavelengths to

Table 1. Target information.

Target	Spectral type	Distance [a] (pc)	Age (Myr)	$L'/W1$ [b] (mag)	$M'/W2$ [b] (mag)	References
β Pic	A6V	19.75 ± 0.13	22 ± 6	3.454 ± 0.003	3.458 ± 0.009	(1), (2), (3), (4)
HIP 65426	A2V	109.21 ± 0.75	14 ± 4	6.761 ± 0.038	6.798 ± 0.019	(3), (5), (6), (7)
PZ Tel	G9IV	47.13 ± 0.13	24 ± 3	6.257 ± 0.049	6.285 ± 0.022	(3), (7), (8), (9)
HD 206893	F5V	40.81 ± 0.11	250^{+450}_{-200}	5.528 ± 0.066	5.437 ± 0.028	(1), (3), (7), (10)

Notes. [a]Distances are calculated from the *Gaia* DR2 parallaxes. [b]L' and M' magnitudes in the ESO system for β Pic b and WISE $W1$ and $W2$ magnitudes for all other targets.
References. (1) Gray et al. (2006), (2) Shkolnik et al. (2017), (3) Gaia Collaboration (2018), (4) Bouchet et al. (1991), (5) Houk (1978), (6) Chauvin et al. (2017), (7) Cutri et al. (2012), (8) Jenkins et al. (2012), (9) Torres et al. (2006), (10) Delorme et al. (2017).

constrain some of the physical and chemical atmospheric properties (e.g.., De Rosa et al. 2016; Samland et al. 2017). At the same time, multi-epoch observations are starting to place constraints on the orbital architecture of these objects (e.g., Wang et al. 2018), sometimes in tandem with stellar radial velocity and/or astrometry constraints (e.g., Bonnefoy et al. 2018; Dupuy et al. 2019).

Atmospheric studies of directly imaged planets have benefitted from observations of isolated brown dwarfs (e.g., Golimowski et al. 2004; Cushing et al. 2008; Stephens et al. 2009; Leggett et al. 2010; Yamamura et al. 2010) and extended modeling efforts of their evolution and atmospheric processes (e.g., Ackerman & Marley 2001; Saumon & Marley 2008; Allard et al. 2012; Baraffe et al. 2015). Without the bright glare of a central star, these objects have been detected in the solar neighborhood (Luhman 2013) down to temperatures of only a few hundred Kelvin (Skemer et al. 2016a; Morley et al. 2018). Gas giant planets and brown dwarfs are expected to share some of their main atmospheric characteristics, such as temperature, radius, and composition, as well as the physics and chemistry that govern their atmospheres (e.g., Marley & Robinson 2015). As a result, the directly imaged planets share a similar spectral sequence to that of their older counterparts. However, for a given temperature, their mass and related surface gravity are lower due to a difference in age, which leads to noticeable differences in their appearance. For example, the NIR colors of young low-mass companions appear redder while their absolute fluxes are lower (e.g., Metchev & Hillenbrand 2006; Marois et al. 2008; Barman et al. 2011). These characteristics can be attributed to thicker and/or vertically more extended clouds in a low-gravity environment (e.g., Currie et al. 2011), as well as the typical sizes of the condensed dust grains (e.g., Burrows et al. 2006; Marley et al. 2012). The surface gravity may also impact the chemical abundances because enhanced vertical mixing can lead to a nonequilibrium balance of CO and CH_4 (e.g., Yamamura et al. 2010; Zahnle & Marley 2014; Moses et al. 2016).

The advent of extreme adaptive optics (AO) assisted high-contrast imaging instruments has provided detailed insight into the NIR (1–2.4 μm) photometric and spectral characteristics of the family of directly imaged low-mass companions (e.g., Vigan et al. 2016; Chilcote et al. 2017; Greenbaum et al. 2018), but their mid-infrared (MIR; 3–5 μm) characteristics remain more sparsely sampled (e.g., Galicher et al. 2011; Bailey et al. 2013; Skemer et al. 2014, 2016b; Stone et al. 2016; Cheetham et al. 2019). While observations at these wavelengths are hampered by the bright thermal background emission, the flux contrast

with the central star is more favorable and an increasing number of photons is emitted at MIR wavelengths along the spectral sequence towards lower mass and cooler objects (see Fig. 1 from Skemer et al. 2014). Most importantly, complementary information about the chemical composition and cloud configuration are to be expected from photometry at 3–5 μm (e.g., Geißler et al. 2008; Galicher et al. 2011; Currie et al. 2014). The *James Webb* Space Telescope (JWST) will fully exploit the MIR regime (Boccaletti et al. 2015; Danielski et al. 2018) – in particular for companions with moderate contrast at wide separations – while ground-based telescopes provide direct access to the atmospheres of more close-in planets (Mawet et al. 2016; Kenworthy et al. 2018).

The MIRACLES (Mid-InfraRed Atmospheric Characterization of Long-period Exoplanets and Substellar companions) survey was designed for photometric characterization of directly imaged planetary and substellar companions at 4–5 μm. It was carried out with the NAOS-CONICA (NACO) AO system and infrared camera (Lenzen et al. 2003; Rousset et al. 2003) that was mounted on the Unit Telescope 1 (UT1) of the Very Large Telescope (VLT) at the Paranal Observatory in Chile. The sample consists of 15 targets that were observed with the Brackett-α and/or M' filters. We aim to homogeneously process and analyze the data in order to obtain robust fluxes and colors of the companions. For consistency, we also reprocess archival data in the 4–5 μm range where such data are available, as well as archival L' band data to cover a second spectral window. The current paper presents the first results as well as extended details on the data processing and calibration. Therefore, only a subset of the companions (β Pic b, HIP 65426 b, PZ Tel B, and HD 206893 B) is analyzed as a demonstration while a more in-depth analysis of the full sample will follow in a future work. Some of the main stellar characteristics that are relevant for this study are listed in Table 1.

In Sect. 2, we report the high-contrast imaging observations with VLT/NACO and the archival datasets that have been (re)processed. In Sect. 3, we describe the data processing with PynPoint, present the companion detections, and outline the procedure for the photometric extraction, calibration, and uncertainty estimation. In Sect. 4, we present the photometry and colors, and compare them with results for field and low-gravity brown dwarfs, other directly imaged companions, and synthetic model photometry. We also carry out a dedicated analysis of β Pic b and present the detection limits derived from our data. In Sect. 5, we discuss our measurements and provide an outlook on future opportunities.

Table 2. Observation details on the new and archival data.

Target	Filter	UT date	DIT [a] (s)	N_{images}	$\Delta\pi$ [b] (deg)	Airmass	Seeing [c] (arcsec)	τ_0 [d] (ms)	$FWHM$ [e] (mas)
New data									
HIP 65426	NB4.05	2018 Jun. 7	1.0	11 520	99.5	1.12–1.22	1.52(3.01)	3.5(0.6)	112.8 ± 0.9
PZ Tel	NB4.05	2018 Jun. 8	1.0	1160	35.3	1.11–1.19	0.82(0.06)	4.2(0.5)	112.5 ± 1.1
PZ Tel	M'	2018 Jun. 8	0.04	40 000	25.1	1.13–1.21	0.85(0.07)	4.9(1.0)	125.9 ± 2.2
HD 206893	NB4.05	2018 Oct. 24	0.9	3752	68.5	1.02–1.04	1.02(0.10)	3.1(0.6)	111.6 ± 0.9
HD 206893	M'	2018 Jun. 8	0.04	115 000	85.5	1.02–1.08	0.89(0.09)	4.5(0.6)	121.7 ± 1.1
Archival data									
β Pic	NB3.74 [f]	2012 Dec. 18	0.1	12 000	43.1	1.12–1.18	0.88(0.05)	4.3(0.6)	108.2 ± 1.1
β Pic	NB4.05	2012 Dec. 16	0.075	25 600	57	1.12–1.15	0.74(0.02)	5.5(1.0)	113.4 ± 1.1
HIP 65426	L'	2017 May 18	0.2	30 224	86.9	1.12–1.33	0.90(0.11)	3.8(0.5)	107.3 ± 1.0
HIP 65426	M'	2017 May 20	0.05	72 900	73.3	1.12–1.35	0.64(0.04)	5.2(1.2)	135.4 ± 0.9
PZ Tel	L'	2010 Sep. 26	0.3	14 400	42.4	1.11–1.21	3.60(2.54)	0.9(0.1)	99.2 ± 1.7
HD 206893	L'	2016 Aug. 9	0.3	18 000	96.7	1.04–1.09	0.86(0.09)	5.7(1.2)	117.0 ± 1.8

Notes. The upper part of the table lists the observed targets and the lower part the archival datasets that were analyzed. [a]The total integration time is given by the product of detector integration time (DIT) and the total number of images (N_{images}). [b]Total parallactic rotation. [c]Mean and standard deviation of the seeing as measured by the differential image motion monitor (DIMM) at 0.5 μm. [d]Mean and standard deviation of the coherence time. [e]Full width at half maximum of the unsaturated point spread function (see main text for details). [f]Archival data which has not yet been published.

2. Observations and archival data

2.1. High-contrast imaging with VLT/NACO at 4–5 μm

The data were acquired with VLT/NACO at the Paranal Observatory in Chile (ESO program IDs: 0101.C-0588(A) and 0102.C-0649(A)). We used the narrowband Brackett-α (NB4.05) filter ($\lambda_0 = 4.05$ μm, $\Delta\lambda = 0.02$ μm) and the broadband M' filter ($\lambda_0 = 4.78$ μm, $\Delta\lambda = 0.59$ μm) to probe two complementary parts of the 4–5 μm regime. The upper part of Table 2 provides an overview of the observed targets and their filters.

We used the pupil-stabilized mode of the instrument to detect the off-axis companions during post-processing with angular differential imaging (ADI). The targets were dithered between the top left and bottom right quadrant of the detector. The bottom left quadrant was excluded due to the persistent striping while the top right quadrant was avoided out of precaution since it was occasionally affected by an excess of striped detector noise. The number of integrations per dither position and readout corresponded to an effective exposure time of ∼1 min such that slowly evolving variations in the thermal background emission could be sampled. We used the *Uncorr* readout mode, which resets and reads the detector array once for each integration and is therefore most suitable for observations with a high background flux. The pixel scale for the L27 camera has not been recalibrated so we assume a value of 27.1 mas per pixel.

At thermal wavelengths (M' band in particular), the detector integration time (DIT) is typically limited by the brightness of the background emission except for bright targets such as β Pic. Therefore, the observations did not require the use of a coronagraph (which was also not available for the M' filter). The DIT that could be used in M' was 40 ms while integrations of 0.9 or 1.0 s were used with the NB4.05 filter (see Table 2). Given the short exposure time with the M' filter, we had windowed the detector to a field of view of 256 pixels in order to prevent frame loss and to limit the amount of overhead time. The data were obtained in cube mode which means that each individual exposure was stored such that a careful frame selection was possible.

The flux of the companions was measured relative to their central star (see Sect. 3.3.1) which requires unsaturated exposures of the stellar point spread function (PSF). The peak flux in M' remained at least 50% below the full well depth of the detector, which corresponds to 28 000 ADU with the HighBackground detector mode (the setting of the bias voltage of the array). For the NB4.05 data, the peak flux had values up to two-thirds of the full well depth (i.e., 15 000 ADU with the HighDynamic detector mode). Therefore, we obtained additional exposures with a smaller DIT for the flux calibration to ensure a robust sampling of the stellar flux: specifically, 480 exposures of 0.5–0.6 s for HIP 65426, 720 exposures of 0.5 s for PZ Tel, and 600 exposures of 0.4 s for HD 206893. After processing of the images (see Sect. 3.1), the stellar PSF was fitted with a 2D Moffat profile to determine the angular resolution of the data. The best-fit values of the full width at half maximum (FWHM) of each PSF are listed in Table 2, for which the averages of the major and minor axes were used.

The observing conditions were overall stable on the night of 2018 June 7 with a mean seeing of about 0.′′8–0.′′9. The seeing made a temporary sharp increase during the observations of HIP 65426 but remained below 1.′′0 for most of the sequence. Such low-quality images were removed during the subsequent frame selection (see Sect. 3.1). The conditions were poorer and more variable during the observations of HD 206893 in NB4.05 with the seeing fluctuating between 0.′′9 and 1.′′3. The mean coherence time was typically in the range of 3–5 ms (see Table 2). Generally, observations at 4–5 μm yield higher Strehl ratios than at shorter wavelengths because for a given phase aberration the relative impact on a PSF in the M' band is significantly smaller than in the optical. Consequently, speckle noise is typically less important for companion detections at 4–5 μm.

The transmission profiles of the considered NACO filters are shown in Fig. 1 in comparison with the WISE $W1$ and $W2$ filters, which cover the L' and M' bands, respectively. The WISE photometry is used in Sect. 3.3.3 to calibrate the extracted fluxes of the companions. The figure also shows the transmission at Cerro Paranal, which was computed with the SkyCalc interface
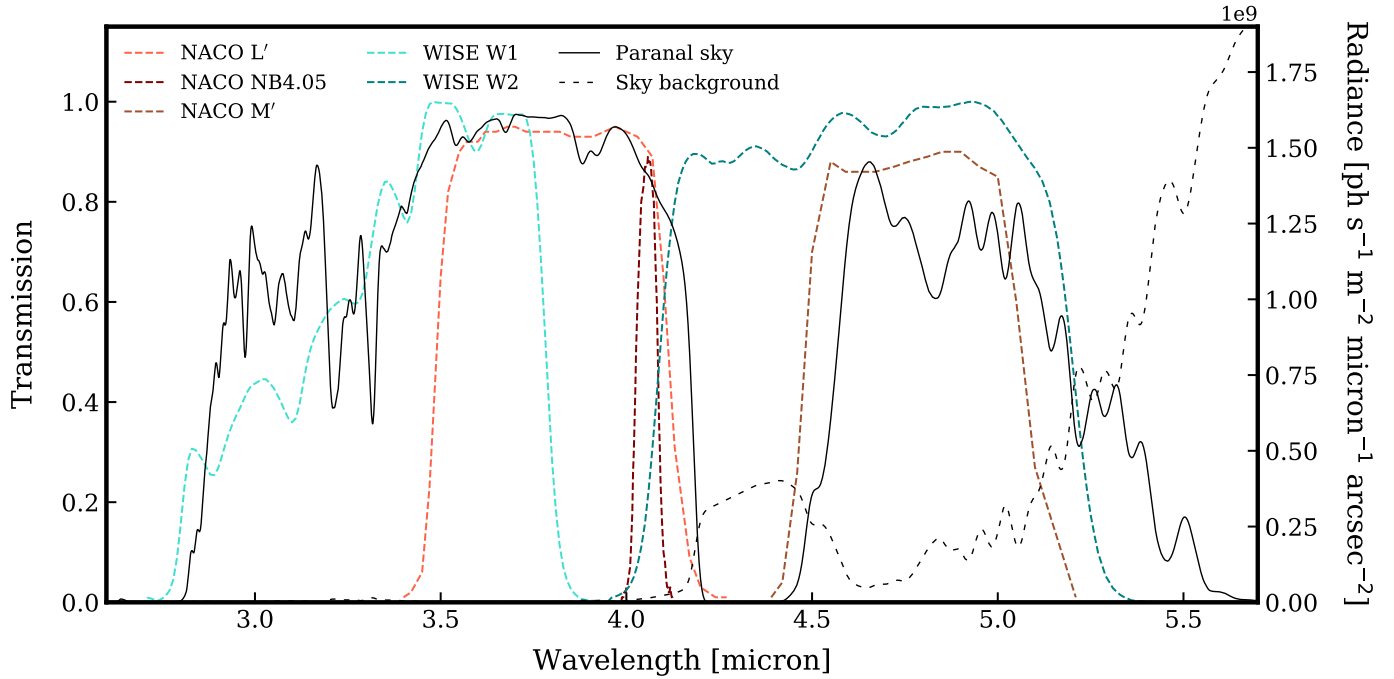
Fig. 1. Transmission profiles of the NACO L', NB4.05, and M' filters and the WISE $W1$ and $W2$ filters (colored dotted lines). The telluric transmission at the Paranal Observatory is also shown (black solid line). The right axis shows the sky background emission (black dashed line) which steeply increases beyond ~4.5 μm.

(Noll et al. 2012) for an airmass of 1.0 and a precipitable water vapor level of 2.5 mm (approximately the median value at the observatory). In addition to the limited telluric transmission in M' (~70%), the steep increase of the sky background emission across this band is also visible. The majority of the sky radiance at wavelengths longer than 4 μm is caused by molecular emission in the lower part of the Earth's atmosphere.

2.2. Complementary and reprocessed archival data

The five new NACO datasets that were obtained with the NB4.05 and M' filters were complemented with several archival datasets. Here we selected NB4.05 and/or M' datasets in case these were already available, L' band data for all targets, and a dataset of β Pic b taken with the narrowband NB3.74 filter (which overlaps the central part of the L' bandpass). We chose to reprocess these archival datasets in order to homogeneously analyze them with the newly obtained data such that the photometric extraction, calibration, and uncertainty estimation is done in a consistent manner. By including L' data, we obtain additional color information in the 3–4 μm spectral region, which is complementary to the 4–5 μm range.

The archival data that were used in this study are listed in the lower part of Table 2. Here we briefly summarize a few of the relevant data characteristics and observing conditions. For β Pic, we analyzed data that were obtained with the NB3.74 and NB4.05 filters (ESO program ID: 090.C-0396(B)) of which the NB4.05 data were published by Currie et al. (2013). The photometry in these filters is combined in the dedicated analysis of this object in Sect. 4.3. With integration times of 0.1 s and 0.075 s, the stellar flux had remained within the linear detector regime for both filters so we could use the full sequence of images for the photometric calibration.

For HIP 65426, we reprocessed the L' and M' band data (ESO program ID: 199.C-0065(A)) that were recently published

by Cheetham et al. (2019). Separate exposures with a DIT of 0.1 s (4 min in total) had been taken in L' for the flux calibration while the DIT of the M' data was sufficiently low such that the images themselves could be used as PSF template. The conditions were variable during the L' observations with a seeing in the range of 0″8–1″1, which may impact the precision of the flux calibration. The M' data on the other hand were obtained in good observing conditions.

For PZ Tel, we complemented our study with archival L' data (ESO program ID: 085.C-0277(B)) that were published by Beust et al. (2016). The data were taken in highly variable conditions with a mean seeing of 3″6 and values below 1″0 for only about 20% of the observing sequence. Nonetheless, the star had remained visible with sufficient AO correction but the flux of PZ Tel varied up to ~40%. Additional exposures were taken with a smaller DIT of 0.2 s (8 min in total) and a neutral density filter (ND_long) in the optical path.

Finally, we reprocessed the L' band data (ESO program ID: 095.C-0937(B)) of HD 206893 that were presented by Milli et al. (2017). The data were taken with the annular groove phase mask coronagraph with six additional off-axis exposures of the star with a DIT of 0.1 s for calibration purpose. In this case, the background emission was sampled by nodding the field of view away from the target every ~10 min.

3. Data reduction, detection, and calibration

3.1. Data processing with PynPoint

The data reduction was done with `PynPoint`[1], which is an end-to-end pipeline for processing and analysis of high-contrast imaging data (Amara & Quanz 2012; Stolker et al. 2019). In Sect. 3.3, we also use its functionalities for the extraction of the relative flux and uncertainty of the companions, and in Sect. 4.4

---

[1] https://github.com/PynPoint/PynPoint

for estimating the detection limits from our data. All results were obtained with version 0.8.1, which is currently the latest release[2].

After reading the raw FITS files into the database, we applied basic preprocessing and calibration procedures. The thermal background emission, dark current, and detector bias were subtracted with the mean of the adjacent data cubes in which the star was located at a different dither position. Remaining bad pixels were corrected by selecting 5σ outliers within a 9×9 pixel filter and replacing them with the mean of the neighboring values. The parallactic angle associated with each individual exposure was precisely calculated from the relevant header information.

Frames were registered by first cropping a subregion around the brightest pixel and subsequently, for a relative alignment, cross-correlating each image with ten randomly selected references images and shifting to the mean offset with the reference frames. Finally, for an absolute centering, the mean of the image stack was fitted with a 2D Moffat function and a constant shift was applied to each image. After the first registration step, we also applied a frame selection by measuring the integrated flux within an aperture (1 $FWHM$ in radius) centered at the approximate position of the star and removing poor-quality frames by sigma clipping frames of which the flux deviated by more than $1-2\sigma$ from the median. Typically, about 5–20% of the images were removed with the frame selection except for the NB3.74 dataset of β Pic and the L' dataset of PZ Tel. For the first, we removed 31% of the data due to the degrading observing conditions at the end of the sequence. For the latter, we only used 11% of the data, which were selected from the end of the sequence during which the photometry remained approximately stable and after which the flux exposures were taken.

Subsequently, we stacked subsets of images such that the final image stack contained approximately 500 images. This is required for the Markov chain Monte Carlo (MCMC) analysis of the contrast and position of the companions (see Sect. 3.3), which otherwise would be too computationally expensive. However, it also typically enhances the signal-to-noise ratio (S/N) for the companion detections in the L' and M' filters (see e.g., Meshkat et al. 2014; Quanz et al. 2015a). Images were also cropped before running the PSF subtraction and the MCMC analysis, but a larger field of view was used for the calculation of the detection limits (see Sect. 4.4).

The PSF subtraction was done with an implementation of full-frame principal component analysis (PCA; Amara & Quanz 2012; Soummer et al. 2012). First, pixel values were masked at radii larger than the image size and at separations within typically 1 $FWHM$ of the image center. Second, the stack of images were decomposed into a lower-dimensional basis set of orthogonal images by applying a singular value decomposition (SVD). Each image was then projected onto the basis of principal components (PCs) and the model was subtracted from the image itself to remove the quasi-static PSF and speckle noise from the star. We varied the number of PCs that were used for the PSF subtraction in the range of 1–50. Finally, all images were derotated towards a common field orientation and median-combined.

3.2. Companion detections

The residuals of the PSF subtraction are presented in the first and third columns of Fig. 2. We calculated the S/N for each number of PCs with the two-sample t-test, which includes a correction

---

[2] https://pypi.org/project/pynpoint/

term for the small sample statistics (Mawet et al. 2014):

$$S/N = \frac{\bar{x}_1 - \bar{x}_2}{s_2 \sqrt{1 + \frac{1}{n_2}}}, \tag{1}$$

where $\bar{x}_1$ is the flux at the position of the companion, $\bar{x}_2$ the average flux within the remaining nonoverlapping apertures at the same separation, s_2 the empirical standard deviation, and n_2 the number of reference apertures. We chose a conservative aperture diameter of 1 $FWHM$ and excluded the apertures directly adjacent to the companion aperture as they contained self-subtraction artifacts. The S/N was then optimized with the position of the companion aperture as a free parameter. The final flux calibration was done for a fixed number of PCs, which was chosen by maximizing the S/N and limiting the amount of variation in the retrieved contrast and position values (see Sect. 3.3.1).

While β Pic b and PZ Tel B are bright targets at 4–5 μm and detected with high S/N (maximum values in the range of 10–35), HIP 65426 b is significantly fainter and only reaches moderately above the background limit in the NB4.05 and M' filters (maximum S/N of 8.1 and 5.5, respectively). For these filters, we smoothed the images of HIP 65426 b with a Gaussian filter of similar FWHM to the angular resolution in order to lower pixel-to-pixel variations and enhance the planet detection (but the original images were used for the photometric extraction). HD 206893 B is brighter than HIP 65426 b but the residual speckle noise at its small angular separation ($\sim$2–3λ/D at the observed wavelengths) limited the S/N to values in the range of 5–8. Nonetheless, this companion can also be identified after subtracting a sufficient number of PCs from the data. Although PZ Tel B was detected in L' with an S/N of 51 when the full dataset was used, we only considered a small subset of the images to ensure a robust flux calibration (see Sect. 2) which resulted in a reduced S/N of 12.8.

In addition to the PSF subtraction residuals, we display in Fig. 3 the derotated and median-combined images of PZ Tel in all three filters. The L' band data were obtained $\sim$8 yr before the NB4.05 and M' data, and therefore the companion is detected in L' at a separation that is smaller by 190 mas. Although the peak flux of the PSF is slightly saturated at the shortest wavelengths, it can be seen from the images that PZ Tel B is approximately two orders of magnitude fainter than PZ Tel A in all filters. The image quality in L' is affected by the poor observing conditions while the NB4.05 and M' images appear of better quality with a well discernible Airy pattern.

3.3. Photometric extraction, calibration, and error estimation

3.3.1. Relative calibration by negative PSF injection

We require accurate and robust photometry of our companions to characterize their atmospheres. Photometric and astrometric measurements of directly imaged objects are challenging due to self-subtraction effects that are intrinsic to post-processing algorithms such as PCA. One way to overcome this is problem is by iteratively injecting negative copies of the unsaturated stellar PSF and minimizing the residuals at the position of the companion after the PSF subtraction. For this procedure, we masked the PSF beyond a radius of 2–4 $FWHM$, depending on its brightness relative to the background flux.

The contrast and position were first computed as a function of the number of PCs by minimizing the flux residuals with a downhill simplex method. Similarly to Stolker et al. (2019), we then used Bayesian inference with MCMC to sample the posterior
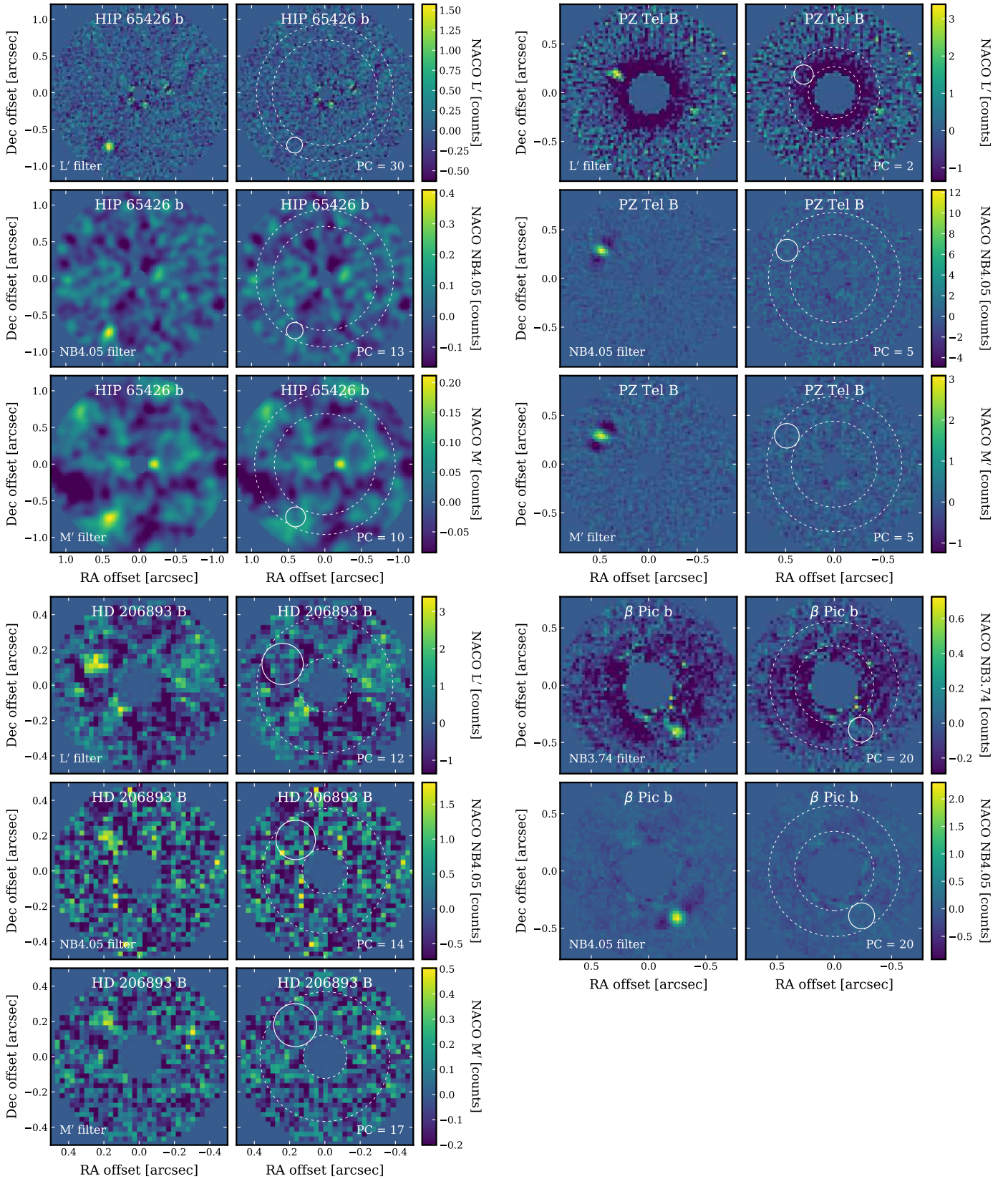
Fig. 2. Median-combined residuals of the PSF subtraction for HIP 65426 (*top left*), PZ Tel (*top right*), HD 206893 (*bottom left*), and β Pic (*bottom right*). Images are ordered for each target *from top to bottom* by the central wavelength of the filters. For each object, *left column*: regular residuals of the PSF subtraction, *right column*: residuals with the best-fit negative PSF template injected. The color scale is set by the maximum pixel value of each object and filter combination and adjusted by a factor −0.3 of the peak flux for the lower limit. The pixel values within the solid circles (at the positions of the companions) were minimized for the flux estimation while the noise was estimated from the area indicated by the dashed circles (see Sect. 3.3.1 for details). North and east are in upward and leftward direction, respectively, in all images.
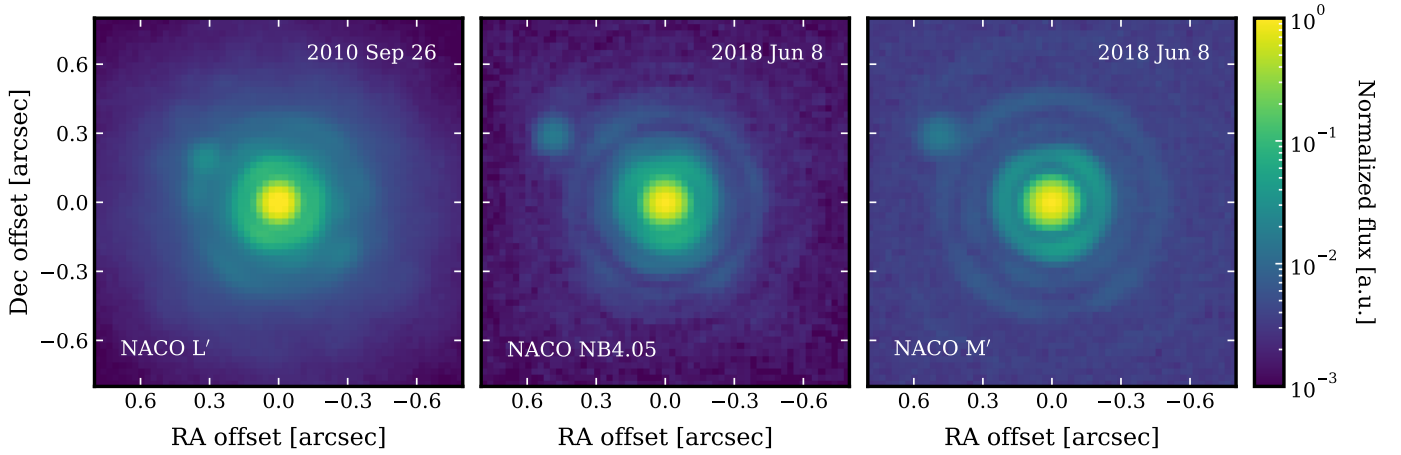
Fig. 3. Derotated and median-combined images of PZ Tel in the *L'* (*left*), NB4.05 (*center*), and *M'* (*right*) filters. The ~8 yr baseline with respect to the archival *L'* data reveals clear orbital motion. The colors are shown on a logarithmic scale which has been normalized to the peak flux. North and east are in upward and leftward direction, respectively.

distributions of the separation, position angle, and flux contrast of the companions in order to determine the most probable values and their uncertainties. For the MCMC, the number of PCs was fixed (i.e., the values indicated in Fig. 2) by selecting a value for which both the S/N was high and the retrieved contrast and position values showed minimal variation.

The log-likelihood function for the MCMC was defined as

$$\log \mathcal{L} \propto -\frac{1}{2} \sum_{i,j}^{N} \left(\frac{I_{ij}}{\sigma_{\text{pix}}} \right)^2, \quad (2)$$

where i and j are the pixel indices, N is the total number of pixels encircled by the aperture, I_{ij} is the pixel value at position ij, and σ_{pix} is the (constant) noise level associated with the pixels. The pixel values that are minimized have been selected within a circular aperture with a diameter of $2\,FWHM$ at the position of the companion (see Fig. 2). The noise on the other hand is calculated by derotating the residuals of the PSF subtraction in the opposite direction, median-combining the images, and computing the standard deviation of all pixels within an annulus covering the separation range of the circular aperture.

When defining the likelihood function, we make several assumptions. First, we assume that the expected value is zero when the injected negative PSF has fully removed the companion signal. Second, the pixel values are considered as independent measurements which may not be strictly true due to potential spatial correlations in the noise residuals after the PSF subtraction. Third, the reference pixels within the annulus are assumed to follow a Gaussian distribution. This is in contrast to the approach followed in Stolker et al. (2019), where the noise associated with each pixel was assumed to follow a Poisson distribution (see also Wertz et al. 2017).

We chose uniform priors for the three parameters and used the affine-invariance sampler implementation of `emcee` (Foreman-Mackey et al. 2013) to compute the posterior distributions of the contrast and position of the companions. For each dataset, we let 200 walkers explore the probability landscape with 500 steps per walker (i.e., requiring 10^5 PSF subtractions). The first 100 steps were removed after visual inspection of the walker's evolution. We then plotted the posterior distributions and adopted the median as the best-fit value and the 16th and 84th percentiles as the uncertainties. The mean acceptance fraction was typically ~0.6–0.65 and the integrated autocorrelation

time of each parameter in the range of ~30–40 steps. Figure 4 shows as an example the 1D and 2D marginalized distributions (created with `corner.py`; Foreman-Mackey 2016) of the NB4.05 dataset of HIP 65426 b and the *M'* dataset of HD 206893 B. For the other datasets, we provide an overview of all remaining posterior distributions in Fig. A.1.

3.3.2. Error budget and measurement bias

The MCMC analysis provides an estimate of the statistical uncertainty of the companion contrast and position with the noise sampled from the distribution of reference pixels. In the derivation of the final contrast, we also include a correction for the intrinsic bias of the measurement and several systematic error components that are not captured by the likelihood function of the MCMC.

Azimuthal variations in the noise residuals from speckles and background flux may cause a bias in the retrieved flux contrast. The intrinsic offset and its uncertainty was estimated with the injection and retrieval of artificial planets. This was done by first removing the companion flux with the best-fit results from the MCMC analysis. We then injected a PSF template with the same contrast and separation as the real companion while the position angle was stepwise changed by 1 deg over the full 360 deg. For each position, we retrieved the separation, position angle, and contrast of the artificial source by minimizing the χ^2 function from Eq. (2). The offset between injected and retrieved values were then calculated and we adopted the median as the bias of our measurement, and the 16th and 84th percentiles as the systematic uncertainty related to the remaining noise residuals after the PSF subtraction.

Figure 5 shows an example of the offset distributions for the NB4.05 dataset of HIP 65426. In this case, there is a bias in the contrast of 0.07 mag and the distribution of the offset is somewhat asymmetric with respect to the median. Specifically, the distribution is broader towards retrieved contrast values that are smaller than the injected value. Based on this analysis, we corrected the contrast by the bias offset and conservatively included the largest of the two error bars in the error budget (see Table 3).

The bias of the contrast is typically small or even negligible except for the NB3.74 dataset of β Pic and the *L'* dataset of PZ Tel. For these cases, we tested the use of different numbers of PCs, but this had only a small impact on the estimated bias and
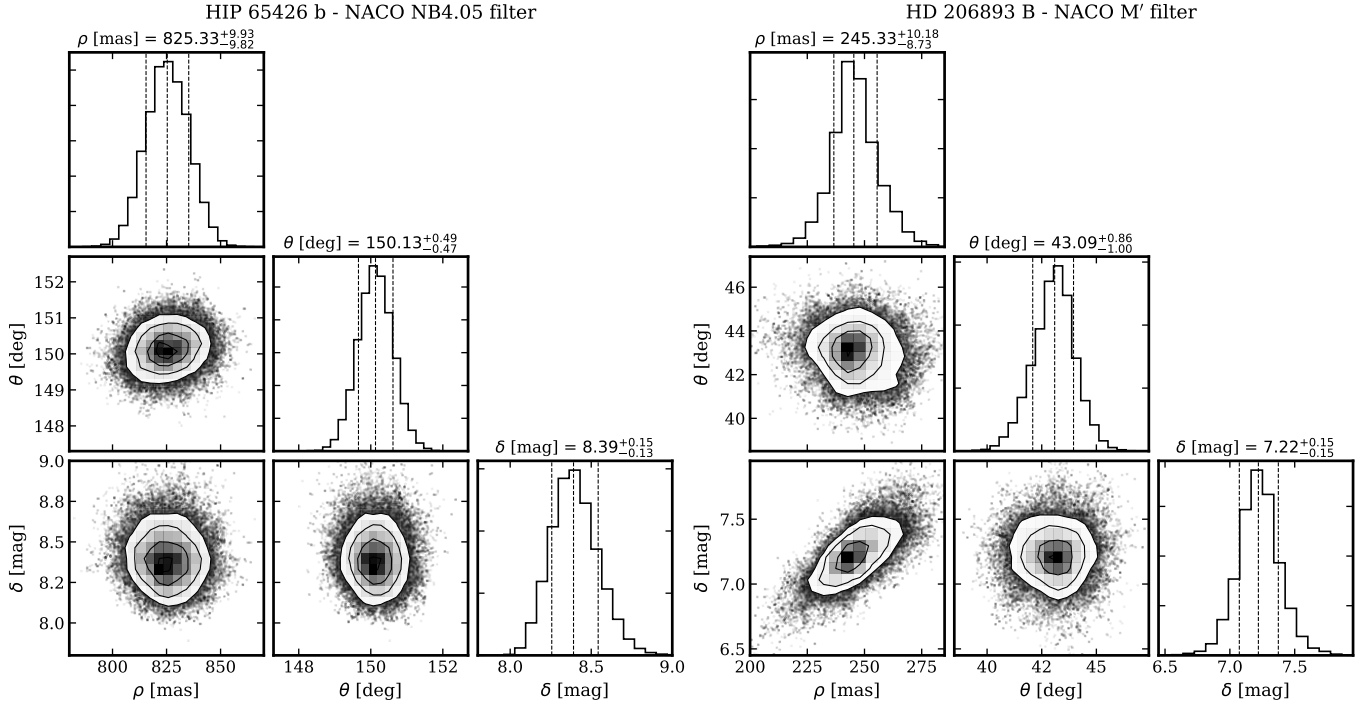
Fig. 4. Posterior distributions of the separation, ρ, position angle, θ, and flux contrast, δ, for HIP 65426 b in the NB4.05 filter (*left panel*) and HD 206893 in the M' filter (*right panel*). The MCMC results of the remaining targets and filters are displayed in Fig. A.1. *Diagonal panels*: marginalized 1D distributions of the parameters and *off-axis panels*: map the 2D probability space for all parameter pairs. The listed values and uncertainties are the median, and the 16th and 84th percentiles of the parameter samples, which are also indicated by the vertically dashed lines in the 1D distributions. Contours overlaid on the 2D distributions correspond to 1σ, 2σ, and 3σ confidence levels for Gaussian statistics.
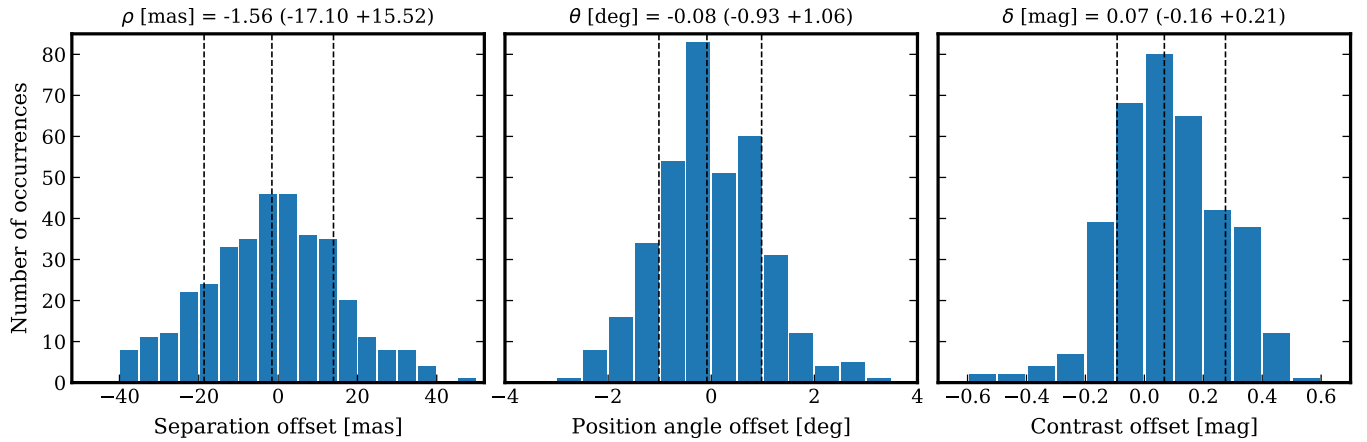
Fig. 5. Offset between the injected and retrieved values of the separation (*left panel*), ρ, position angle (*center panel*), θ, and flux contrast (*right panel*), δ, for the NB4.05 dataset of HIP 65426. The precision of the figure of merit was tested at 360 position angles with the best-fit separation and contrast from the MCMC analysis (see Fig. 4). The vertically dashed lines indicate the 16th, 50th, and 84th percentiles of the samples from which the uncertainties in the title above each panel have been derived.

uncertainty. Interestingly, Fig. 2 shows that in these two cases there is a prominent, negative residual present at the separation of the companion, likely related to an imperfect subtraction of the Airy pattern. Indeed, the L' image in Fig. 3 shows that the position of PZ Tel B coincides with a bright diffraction ring.

The stellar flux remained unsaturated throughout the observing sequences of the M' data and also some of the narrowband data. In these cases, we applied a one-on-one injection for each image such that the calibration error related to the PSF template can be excluded from the error analysis. For the other datasets, we measured the stellar flux with a circular aperture (diameter of $2\,FWHM$) in each of the calibration exposures (see Sect. 2). The

standard deviation on these flux measurements was adopted as a measure for the variation in the stellar flux which degrades the precision with which the companion signal is removed in each individual image. For the L' dataset of PZ Tel, we also considered the uncertainty in the transmission of the neutral density filter that was used for the unsaturated exposures. We adopted a filter transmission of $(2.33 \pm 0.10)\%$ from Bonnefoy et al. (2013) and included the uncertainty in the error budget (see Table 3).

The final contrast values and error components are listed in Table 3. Since the MIRACLES program focuses on atmospheric characterization with 4–5 μm photometry, we do not analyze the astrometry of the companions. Nonetheless, the uncertainty

Table 3. Photometry and error budget.

Target	Filter	MCMC contrast [a] (mag)	Bias offset [a] (mag)	Calib. error [b] (mag)	ND filter error (mag)	Final contrast [c] (mag)	Star magnitude (mag)	Apparent magnitude (mag)	Absolute magnitude (mag)	Flux (W m^{-2} μm^{-1})
β Pic b	NB3.74	8.13 ± 0.11	−0.33 ± 0.24	-	-	7.80 ± 0.26	3.45 ± 0.01 [d]	11.25 ± 0.26	9.77 ± 0.26	1.70 ± 0.42 × 10^{-15}
β Pic b	NB4.05	7.54 ± 0.02	−0.01 ± 0.04	-	-	7.53 ± 0.04	3.45 ± 0.01 [d]	10.98 ± 0.04	9.50 ± 0.05	1.61 ± 0.04 × 10^{-15}
HIP 65426 b	L′	8.56 ± 0.06	0.01 ± 0.08	0.07	-	8.56 ± 0.12	6.76 ± 0.04 [e]	15.32 ± 0.13	10.13 ± 0.13	3.90 ± 0.46 × 10^{-17}
HIP 65426 b	NB4.05	8.39 ± 0.15	0.07 ± 0.21	0.03	-	8.46 ± 0.26	6.77 ± 0.04 [f]	15.23 ± 0.26	10.04 ± 0.26	3.22 ± 0.78 × 10^{-17}
HIP 65426 b	M′	7.83 ± 0.16	0.06 ± 0.24	-	-	7.89 ± 0.29	6.76 ± 0.04 [g]	14.65 ± 0.29	9.46 ± 0.29	2.99 ± 0.82 × 10^{-17}
PZ Tel B	L′	4.94 ± 0.13	−0.16 ± 0.12	0.11	0.03	4.78 ± 0.21	6.26 ± 0.05 [e]	11.04 ± 0.22	7.68 ± 0.22	2.01 ± 0.41 × 10^{-15}
PZ Tel B	NB4.05	4.69 ± 0.01	−0.00 ± 0.02	0.04	-	4.69 ± 0.05	6.25 ± 0.05 [f]	10.94 ± 0.07	7.57 ± 0.07	1.67 ± 0.11 × 10^{-15}
PZ Tel B	M′	4.63 ± 0.01	−0.00 ± 0.02	-	-	4.63 ± 0.03	6.30 ± 0.02 [g]	10.93 ± 0.03	7.56 ± 0.03	9.18 ± 0.28 × 10^{-16}
HD 206893 B	L′	8.28 ± 0.12	−0.02 ± 0.28	0.02	-	8.27 ± 0.30	5.53 ± 0.07 [e]	13.80 ± 0.31	10.74 ± 0.31	1.59 ± 0.46 × 10^{-16}
HD 206893 B	NB4.05	7.62 ± 0.21	0.01 ± 0.32	0.08	-	7.63 ± 0.39	5.53 ± 0.07 [f]	13.16 ± 0.40	10.11 ± 0.40	2.16 ± 0.81 × 10^{-16}
HD 206893 B	M′	7.22 ± 0.15	0.01 ± 0.27	-	-	7.23 ± 0.31	5.54 ± 0.07 [g]	12.77 ± 0.32	9.72 ± 0.32	1.68 ± 0.50 × 10^{-16}

Notes. [a] For asymmetric uncertainties, we conservatively use the largest of the two values that were derived from the MCMC and bias analysis (see Figs. 4, 5, and A.1). [b] The uncertainty related to the variable observing conditions which may bias the relative flux calibration in cases where the unsaturated PSF was only sampled at the start and/or end of the observations. [c] The final contrast is calculated by adding the bias offset and combining the error components in quadrature. [d] The L′ and M′ photometry (3.454 ± 0.003 and 3.458 ± 0.009 mag, respectively) in the ESO photometric system have been adopted from Bouchet et al. (1991). The errors have been inflated to 0.01 mag to account for a minor color offset with the VLT/NACO filters. [e] The L′ magnitude has been adopted from the WISE W1 photometry (see top left panel of Fig. C.1). [f] The NB4.05 magnitude has been adopted from the WISE W1 photometry with a minor color correction of 0.01 mag for HIP 65426 and −0.01 mag for PZ Tel (see top right panel of Fig. C.1). [g] The M′ magnitude has been adopted from the WISE W1 magnitude with a minor color correction of 0.04 mag for PZ Tel and 0.01 mag for HD 206893 (see bottom left panel of Fig. C.1).

on the separation and position angle were simultaneously computed with the uncertainty on the contrast. We have therefore listed the retrieved position values and error components in Table B.1.

3.3.3. Absolute photometric calibration

With the results from the contrast measurements at hand, we determined the fluxes of the companions by considering the apparent magnitudes of their host star. However, apart from β Pic (see Table 1), we were not able to find L′, NB4.05, or M′ photometry for any of the stars in our sample. Since the stars are all young and therefore expected to be variable at optical and NIR wavelengths, we derived the stellar magnitudes directly from the available WISE photometry without considering shorter wavelengths. The uncertainty from the adopted WISE photometry has been included in the error budget.

The stellar fluxes for the NACO filters were all derived from the WISE W1 filter ($\lambda_0 = 3.4$ μm), which is most comparable to the NACO L′ filter (see Fig. 1). Although the WISE W2 filter ($\lambda_0 = 4.6$ μm) more closely resembles the NACO M′ filter than W1, it was shown by Avenhaus et al. (2012) that for targets with an apparent magnitude brighter than ∼6 mag, the recovered (because the WISE detector saturates at 6.7 mag) W2 photometry may not be reliable. Given the differences in wavelength coverage between the WISE and NACO filters, we used synthetic spectra to quantify the colors between these two filter systems.

We first computed synthetic colors from the BT-NextGen atmospheric models (Allard et al. 2012) for effective temperatures in the range of 4000–10 000 K (see Fig. C.1). Surface gravity and metallicity effects are negligible for the stellar temperatures of our sample. The results show that the color of the WISE W1 and NACO L′ is negligible for temperatures greater than ∼4500 K and so we adopted the WISE W1 magnitudes as L′ photometry for the stars in our sample (except for β Pic). The color of the WISE W1 and NACO NB4.05 filters is ≲0.02 mag in the considered temperature range above ∼4500 K. Therefore, the W1 magnitudes are also used for the NB4.05 photometry but we applied a color correction of +0.01 mag and −0.01 mag for

HIP 65426 and PZ Tel, respectively, as determined from their effective temperatures.

Figure C.1 shows that the deviation of the W1–W2 colors of our sample from the model predictions is largest for the brightest of the targets (HD 206893), which may indeed point to a problem with the W2 photometry. We therefore also adopt the W1 photometry for the M′ calibration of the host stars and apply a color correction of +0.04 mag and +0.01 mag for PZ Tel and HD 206893, respectively, while this is not required for HIP 65426 due to its earlier spectral type.

4. Results

4.1. Companion fluxes and colors

The extracted flux contrasts, calibrated magnitudes, and uncertainties in L′, NB4.05, and M′ are listed in Table 3 together with the new NB3.74 photometry of β Pic b. In addition, for the analysis of β Pic b, we adopted the L′ and M′ magnitudes from Stolker et al. (2019) because we applied a similar procedure for the photometric extraction and uncertainty estimation in that study.

The *Gaia* DR2 parallax of each target was used to calculate its distance (see Table 1), which we then used to convert the apparent magnitude into an absolute magnitude while taking into account the uncertainty on the parallax (and hence the distance) in the error propagation:

$$\sigma_M^2 = \sigma_m^2 + \sigma_d^2 \left(\frac{5}{d \log 10} \right)^2, \qquad (3)$$

where σ_M is the uncertainty on the absolute magnitude, σ_m the uncertainty on the apparent magnitude, σ_d the uncertainty on the distance, and d the distance in parsecs. The absolute magnitudes are listed in Table 3, revealing some diversity in the sample since the values provide a measure for the temperature of the atmospheres.

Apparent magnitudes were converted to physical fluxes by calculating the zero point of the filters with a flux-calibrated spectrum of Vega (Bohlin 2007) and setting its magnitude to 0.03 mag for all filters. The zero point of each filter is

Table 4. Companion colors at 3–5 μm.

Target	L'–NB4.05 (mag)	L'–M' (mag)	NB4.05–M' (mag)
β Pic b	0.32 ± 0.08	0.20 ± 0.13	-0.12 ± 0.13
HIP 65426 b	0.10 ± 0.25	0.68 ± 0.31	0.58 ± 0.36
PZ Tel B	0.10 ± 0.23	0.11 ± 0.22	0.01 ± 0.08
HD 206893 B	0.63 ± 0.46	1.02 ± 0.41	0.39 ± 0.43

computed as

$$\langle f_\lambda \rangle = \frac{\int f_\lambda(\lambda) R(\lambda) \mathrm{d}\lambda}{\int R(\lambda) \mathrm{d}\lambda}, \tag{4}$$

where $\langle f_\lambda \rangle$ is the mean flux for a given filter, $f_\lambda(\lambda)$ the wavelength-dependent flux, and $R(\lambda)$ the filter transmission. From this we obtained a zero point of 5.21×10^{-11} W m^{-2} μm^{-1} in NB3.74, 5.12×10^{-11} W m^{-2} μm^{-1} in L', 3.86×10^{-11} W m^{-2} μm^{-1} in NB4.05, and 2.10×10^{-11} W m^{-2} μm^{-1} in M'. We neglected additional components in the system response function such as telluric transmission, mirror reflectivity, optics transmission, and quantum efficiency, thereby assuming that these have a similar impact on the zero point and the companion's photometry, even though there will be dissimilarities between the spectral morphology of Vega and a substellar atmosphere. The calibrated fluxes are listed in Table 3.

From the derived magnitudes, we calculated the colors of the companions. These are listed in Table 4 with the errors calculated by adding the uncertainties of the individual filters in quadrature. The overview shows that many of the colors are within their uncertainties consistent with being blue, gray, or red. One exception is the L'–NB4.05 color of β Pic b, which is red at a 4σ confidence level. Furthermore, the red L'–NB4.05 color of HD 206893 B, the red NB4.05–M' color of HIP 65426 b, and the red L'–M' colors of β Pic b, HIP 65426 b, and HD 206893 B are all 1–2.5σ results. A more detailed comparison and analysis of the magnitudes and colors will be provided below.

4.2. Color and magnitude comparison: evolution, chemical composition, and clouds

The evolution of substellar-mass objects is marked by gravitational contraction which is only counteracted by electron degeneracy pressure. Consequently, brown dwarfs and gas-giant planets cool over time such that their luminosity has a strong dependence on age (e.g., Burrows et al. 2001). The spectral sequence of substellar objects extends from M-type stars to L, T, and Y dwarfs, which at a given age corresponds to an isochrone of decreasing mass and temperature. This implies that for a given spectral type, substellar objects can have a different surface gravity while having a similar temperature as a result of having a different age. Therefore, we investigate the L', NB4.05, and M' color and magnitude characteristics of our sample of young directly imaged low-mass companions by comparing them with synthetic photometry from atmospheric models, photometry from field and young, low-gravity brown dwarfs, and other directly imaged companions.

4.2.1. Selected empirical data and model spectra

The color and magnitude comparison was done with `species`, which is a toolkit that we developed for analyzing spectral and

photometric data of planetary and substellar atmospheres. The software provides a coherent, easy-to-use framework to store, inspect, analyze, and plot observational data and models. It benefits from a wide variety of publicly available data such as atmospheric model spectra, photometric libraries, spectral libraries, evolutionary tracks, photometry of directly imaged companions, and filter transmission profiles. The Python package is publicly available in the PyPI repository[3] and maintained on Github[4]. The online documentation[5] contains additional information on the workflow, implemented features, supported data, and several tutorials. Before presenting the results, we provide a brief summary of the empirical data and the model spectra that were selected.

The colors and magnitudes of the observed sample are empirically compared with the photometry from the Database of Ultracool Parallaxes (Dupuy & Liu 2012; Dupuy & Kraus 2013; Liu et al. 2016). This is an inventory of late-M, L, T, and Y dwarfs with measured parallaxes and photometry taken in the MKO filter system (including L' and M' magnitudes for a subset of the data). We extracted the field objects, which are old and are expected to have a high surface gravity, and also objects that were flagged as being young and/or having a low surface gravity.

While the NACO L' and M' filters are comparable to respective filters from the MKO system, the NB4.05 filter transmission deviates significantly from any of the available broadband photometry. We therefore used the IRTF spectral library (Cushing et al. 2005), which contains a selection of M and early L dwarf spectra in the 3–4 μm range. Specifically, these spectra extend up to 4.1 μm, where the NB4.05 and L' filter transmission has dropped to 10 and 65%, respectively. Synthetic fluxes for the L' and NB4.05 filters were therefore computed from these spectra while ignoring the slight deficit in the spectral range.

Spectra in the L' and M' band regime of late L and T dwarfs are more sparsely available due to their low luminosity and the high thermal background emission. We adopted the synthetic L' and NB4.05 photometry from Currie et al. (2014) which were calculated from spectra of brown dwarfs obtained with the AKARI infrared space telescope. In this sample, we included eleven L dwarfs and two early T dwarfs.

The photometry is also compared with predictions from evolutionary and atmospheric models. For this, we used the isochrone data from Baraffe et al. (2003), who coupled interior models to the non-gray atmospheric models by Allard et al. (2001). The AMES-Cond and AMES-Dusty models present two limiting cases for the treatment of the dust. The first model includes efficient gravitational settling of the dust with no effect on the thermal structure and the emitted spectrum. The second model considers inefficient gravitational settling with the grains remaining at the altitude where they form, following chemical equilibrium conditions. Isochrones were extracted at ages of 20 Myr, 100 Myr, and 1 Gyr and were interpolated for logarithmically spaced masses. From this, we obtained the effective temperature and surface gravity for each mass–age pair, and calculated the corresponding radius. We then used the spectra of the AMES-Cond and AMES-Dusty models to compute the synthetic photometry for the NACO filters.

Finally, we also gathered photometry of any directly imaged companions with magnitudes available in at least two of the H, L', NB4.05, and M' filters. The comparison includes photometry
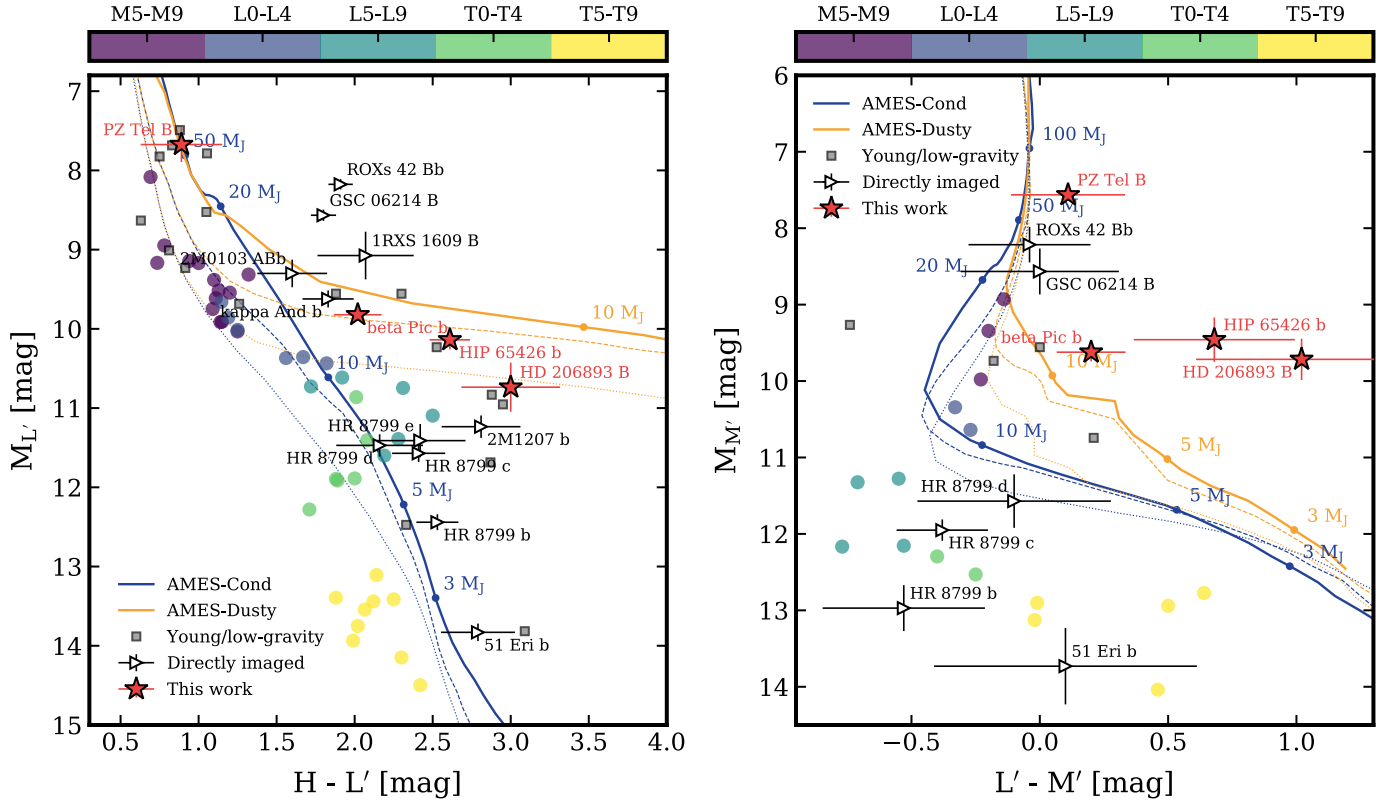
---

[3] https://pypi.org/project/species

[4] https://github.com/tomasstolker/species

[5] https://species.readthedocs.io

Fig. 6. Color–magnitude diagrams of *H–L′* vs. $M_{L'}$ (*left panel*) and *L′–M′* vs. $M_{M'}$ (*right panel*). The field objects are color coded by M, L, and T spectral types (see discrete color bar), the young and/or low-gravity dwarf objects are indicated with a gray square, and the directly imaged companions are labeled individually. The companions from this study are highlighted with a red star. The blue and orange lines show the synthetic colors and magnitudes computed from the AMES-Cond and AMES-Dusty evolutionary tracks for ages of 20 Myr (solid), 100 Myr (dashed), and 1000 Myr (dotted).

from β Pic b (Currie et al. 2013), HIP 65426 b (Chauvin et al. 2017), PZ Tel B (Biller et al. 2010), HD 206893 B (Milli et al. 2017), HR 8799 bcde (Galicher et al. 2011; Currie et al. 2012, 2014; Zurlo et al. 2016), GSC 06214 B (Ireland et al. 2011; Bailey et al. 2013), ROXs 42 Bb (Daemgen et al. 2017), 51 Eri b (Rajan et al. 2017), κ And b (Bonnefoy et al. 2014), 2M1207 b (Chauvin et al. 2004), 2M0103 ABb (Delorme et al. 2013), and 1RXS 1609 B (Lafrenière et al. 2008, 2010). Here, we loosely combined broadband filters from different instruments, as well as several magnitudes from the narrowband VLT/SPHERE *H2* filter. Small color offsets up to a few tenths of a magnitude are therefore to be expected but this will have no major impact on the overall picture emerging from these diagrams.

4.2.2. Color–magnitude diagrams

Color–magnitude diagrams illustrate the luminosity evolution of substellar atmospheres through the correlation between spectral type and absolute flux. Figure 6 presents two such diagrams based on the *H*, *L′*, and *M′* photometry of the companions in comparison with field and young/low-gravity dwarfs and a sample of directly imaged companions.

The absolute *L′* magnitudes and *H–L′* colors of the directly imaged companions appear to follow a similar track as the young/low-gravity objects (see left panel in Fig. 6). While the field and young/low-gravity dwarfs share the same characteristics at the high-mass end of the diagram, the young/low-gravity objects are redder compared to their older counterparts within the L and T-type regime of the sequence. A similar trend can be

identified in the photometric characteristics of the studied sample. PZ Tel B coincides with the photometry and colors from the isolated brown dwarfs while β Pic b, HIP 65426 b, and HD 206893 B are red in *H–L′* compared to the field dwarfs but follow the line of expectation with respect to the young/low-gravity objects. In contrast to the colors, the absolute *L′* fluxes of the sample follow approximately those of the field dwarfs if we consider the following spectral types: M7±1 for PZ Tel B (Maire et al. 2016), L2±1 for β Pic b (Chilcote et al. 2017), L5–L7 for HIP 65426 b (Chauvin et al. 2017), and late L for HD 206893 B (Delorme et al. 2017).

The atmospheric models predict comparable values at the high end of the spectral sequence where the temperatures are too high for cloud species to condense. Towards lower mass planets, the *H–L′* color becomes increasingly red due to the broad H_2O absorption features at 1.4 and 1.8 μm that partially covers the *H* band regime. The CH_4 absorption at 3.3 μm in the AMES-Cond (clear atmosphere) spectra starts to affect the *L′* photometry in the warm temperature regime. The combined effect causes the *H–L′* color to remain constant while the absolute *L′* flux decreases until the absorption by H_2O takes the overhand for the lowest planet masses. The AMES-Dusty (cloudy atmosphere) spectra show a weak CH_4 feature around 3.3 μm but the color in the warm and cold temperature regime is mostly affected by the strongly mixed dust grains. This continuum source shifts the overall flux towards longer wavelengths, hence the increasingly red color.

The right panel of Fig. 6 shows the *L′–M′* versus $M_{M'}$ color–magnitude diagram. In this case, there is a smaller sample of

directly imaged companions and isolated objects available. The colors of β Pic b, HIP 65426 b, and HD 206893 B are redder than both the field and young/low-gravity dwarfs (although with marginal significance), as well as the synthetic colors from the atmospheric models. Interestingly, the absolute M' magnitudes of β Pic b, HIP 65426 b, and HD 206893 B are all comparable ($\sim$9.5 mag) and similar to the magnitudes of the late M type field objects. That is, these three directly imaged objects are overluminous in M' given their spectral types. The color of PZ Tel B on the other hand is within its uncertainty consistent with the model predictions, but its absolute M' flux is potentially also slightly enhanced.

Absorption by CO in the M' band causes the color from the AMES-Cond models to become blue for the high-temperature objects, after which the CH_4 absorption in the L' band increases in its relative strength, resulting in a red color. The models overpredict the M' flux of the late L and T dwarf sequence because the abundances of CO and CH_4 in their atmospheres are likely driven by nonequilibrium chemistry due to vertical mixing of CO from the deeper convective layers (e.g., Griffith & Yelle 1999; Yamamura et al. 2010). Enhanced CO absorption in the M' band will lead to a bluer color and a lower flux in M'. The AMES-Dusty models predict colors that are mostly gray at high temperatures and increasingly red towards lower temperatures due the condensation of dust grains. The M' flux and color predicted by these models reaches closer to the observed values from β Pic b, HIP 65426 b, and HD 206893 but a significant discrepancy remains.

The L'–NB4.05 color in Fig. 7 varies moderately with values ranging up to only 1 mag for the directly imaged objects close to the L/T transition. PZ Tel B and HIP 65426 b have a color that is consistent with the field dwarfs and model predictions. HD 206893 B and β Pic b on the other hand are redder across these wavelengths, although the uncertainty on the color of HD 206893 B is large. Its color is comparable with the HR 8799 planets but these objects are intrinsically fainter at 4 μm. Similar to the M' magnitudes, all objects in our sample are brighter in NB4.05 compared to field dwarfs with similar spectral types.

The NB4.05 filter covers a regime with limited molecular absorption while CH_4 absorption at 3.3 μm can be present in the blue end of the L' filter. As a result, the AMES-Cond spectra show a color that changes from gray to red. In the AMES-Dusty spectra, the CH_4 absorption feature is much weaker and it is mostly the dust opacity that causes a reddening of the spectra, albeit weaker than the reddening seen in the H–L' colors.

4.2.3. Color–color diagrams

Color–color diagrams reveal correlations between the spectral slopes of two wavelength regimes. The left panel of Fig. 8 shows the relation between the H–K and K–L' colors. At high temperatures, the field and young/low-gravity dwarfs show a similar correlation, which is also followed by PZ Tel B and β Pic b. For late L spectral types on the other hand, the colors of the young/low-gravity objects start to deviate from the field dwarfs. Regarding our sample, HIP 65426 b lies in the color regime of the young/low-gravity objects while HD 206893 B has more distinctive color characteristics.

The synthetic colors computed from the AMES-Cond and AMES-Dusty models show a divergence at masses below 20 M_{Jup}. A clear atmosphere causes a blue color towards lower temperatures because of the decreasing K band flux while at higher temperatures the H–K color remains approximately constant because of the H_2O absorption at 1.4, 1.8, and 2.6 μm. The
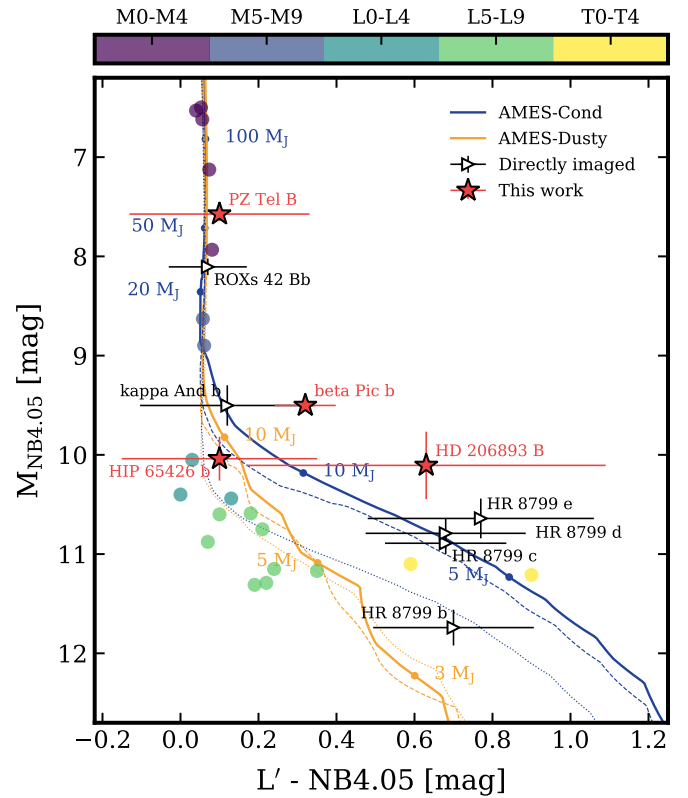
Fig. 7. Color–magnitude diagram L'–NB4.05 vs. $M_{NB4.05}$. The photometry and colors of the M, L, and T dwarfs were derived from IRTF (Cushing et al. 2005) and AKARI spectra (Currie et al. 2014). The field objects are color coded according to their spectral type (see discrete color bar) and the directly imaged companions are labeled individually. The companions from this study are highlighted with a red star. The blue and orange lines show the synthetic colors and magnitudes computed from the AMES-Cond and AMES-Dusty evolutionary tracks for ages of 20 Myr (solid), 100 Myr (dashed), and 1000 Myr (dotted).

K–L' color is close to gray at high temperatures and becomes redder at lower temperatures due to strong H_2O absorption in the K band even though the CH_4 absorption also increases in the L' band. For the AMES-Dusty model, the absorption features are largely muted and the increasingly red H–K color is mainly caused by the dust continuum opacity.

The color–color relation between K–L' and L'–M' shows a larger dispersion when comparing the field and young/low-gravity dwarfs with the directly imaged companions. While the K–L' colors are comparable for these two samples (see left panel of Fig. 8), the L'–M' colors of the young/low-gravity objects appear red compared to the field dwarfs. Interestingly, the four directly imaged companions from our sample are even redder compared to the isolated objects except for PZ Tel B. The photometric characteristics of PZ Tel B are within their uncertainties comparable to those of field dwarfs and model predictions. The offset between the model colors and the field dwarfs is caused by an under- and overprediction of the L' and M' flux, respectively. This is related to the nonequilibrium CH_4 and CO abundances, and similar to the discrepancy in the color–magnitude diagram of L'–M' versus $M_{M'}$ (see Fig. 6).

4.3. Modeling of β Pic b with a dynamical mass prior

The surface gravity of an atmosphere is an important parameter because it influences the vertical distribution of both the gas
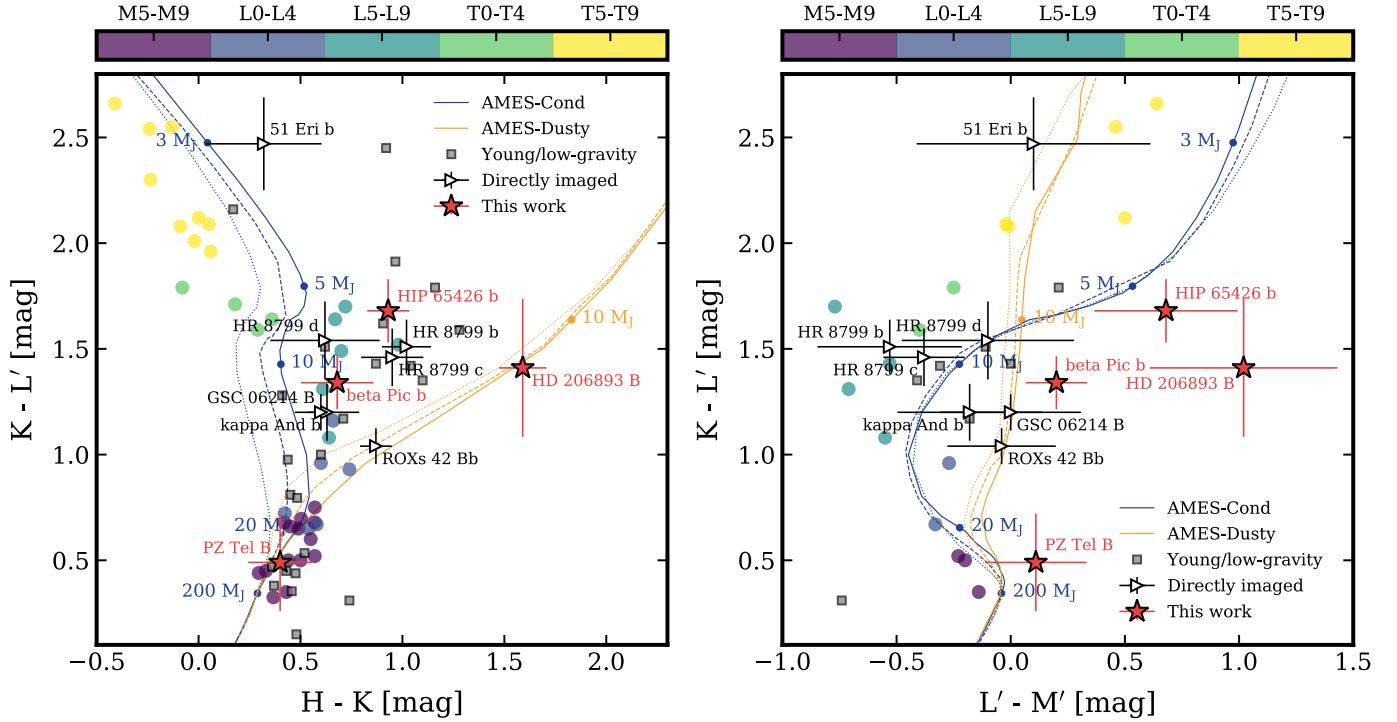
Fig. 8. Color–color diagrams for *H–K* vs. *K–L'* (*left panel*) and *L'–M'* vs. *K–L'* (*right panel*). The field objects are color coded into M, L, and T spectral types (see discrete color bar), the young and/or low-gravity dwarf objects are indicated with a gray square, and the directly imaged companions are labeled individually. The companion colors derived in this study are highlighted with a red star. The blue and orange lines are the synthetic colors that have been calculated from the AMES-Cond and AMES-Dusty evolutionary tracks for ages of 20 Myr (solid), 100 Myr (dashed), and 1000 Myr (dotted).

and the cloud condensates. A lower surface gravity will shift the photosphere to lower pressures and temperatures. Consequently, pressure broadening of the absorption lines becomes weaker and the atmosphere more transparent. However, quantifying the surface gravity from photometry alone is challenging and typically requires high-precision spectra.

Several studies of β Pic b have carried out a detailed atmospheric characterization to constrain the planet's main physical properties (e.g., Morzinski et al. 2015; Chilcote et al. 2017). Since the dynamical mass of β Pic b was recently determined (Snellen & Brown 2018; Dupuy et al. 2019), we carried out a simplified fitting procedure, only considering photometry of the planet but also applying a prior on its mass. The analysis was done with species with which we used a grid of synthetic spectra from the DRIFT-PHOENIX model (Helling et al. 2008). This is a radiative-convective equilibrium atmosphere model which includes a detailed cloud model for predicting the composition and size distribution of the condensates.

The posterior distributions of the effective temperature, $T_{\rm eff}$, surface gravity, log g, metallicity, [Fe/H], and radius, R, were computed with the affine-invariant sampler emcee (Foreman-Mackey et al. 2013). For each sample, the grid of spectra was linearly interpolated in the multidimensional space and synthetic photometry was computed with Eq. (4). The log-likelihood function was then calculated as

$$\log \mathcal{L} \propto -\frac{1}{2} \sum_i^N \left(\frac{f_i - m_i}{\sigma_i} \right)^2, \tag{5}$$

where f_i is the flux in filter i, m_i is the synthetic flux computed from the model spectra in filter i, and σ_i is the uncertainty on the observed flux. We used 200 walkers, each making a 1000 steps,

which were initialized at random positions close to the approximate temperature and radius, while the surface gravity and metallicity were uniformly initialized at random values between the grid boundaries.

The prior on the surface gravity was calculated by assuming a Gaussian prior distribution for the mass, which was centered at the model-independent value of 13 ± 3 $M_{\rm Jup}$ from Dupuy et al. (2019). A uniform prior was chosen for all other parameters. Although β Pic b has been routinely observed by various instruments, we only considered the following photometry without duplicating filter bandpasses: Magellan/VisAO Y_s filter (Males et al. 2014), VLT/NACO J filter (Currie et al. 2013), Gemini/NICI CH$_{4S,1\%}$ (Males et al. 2014), VLT/NACO H filter (Currie et al. 2013), VLT/NACO K_s filter (Bonnefoy et al. 2011), VLT/NACO L' and M' (Stolker et al. 2019), and VLT/NACO with NB3.74 and NB4.05 filters from this work.

Half of the walker's steps were discarded as burn-in, leaving a total of 10^5 samples. From these samples, we computed 30 randomly selected spectra which are shown in Fig. 9 together with the best-fit spectra for the two different metallicity values (see below). The residuals between the observed fluxes and the synthetic photometry are all within 2σ from the expected values. A dispersion of the spectra is mostly visible at the shorter wavelengths given the uncertainties on the photometry in the 1–2 μm region.

The sampled posterior distributions are shown in Fig. 10 (Foreman-Mackey 2016) with the median value of the samples indicated above each of the 1D distributions. Without the prior on the mass of the planet, the surface gravity was constraint to only log $g = 4.29^{+0.67}_{-0.76}$ dex while showing a strong degeneracy with the metallicity of the atmosphere. With the prior on the mass, we constrained the surface gravity to a value of
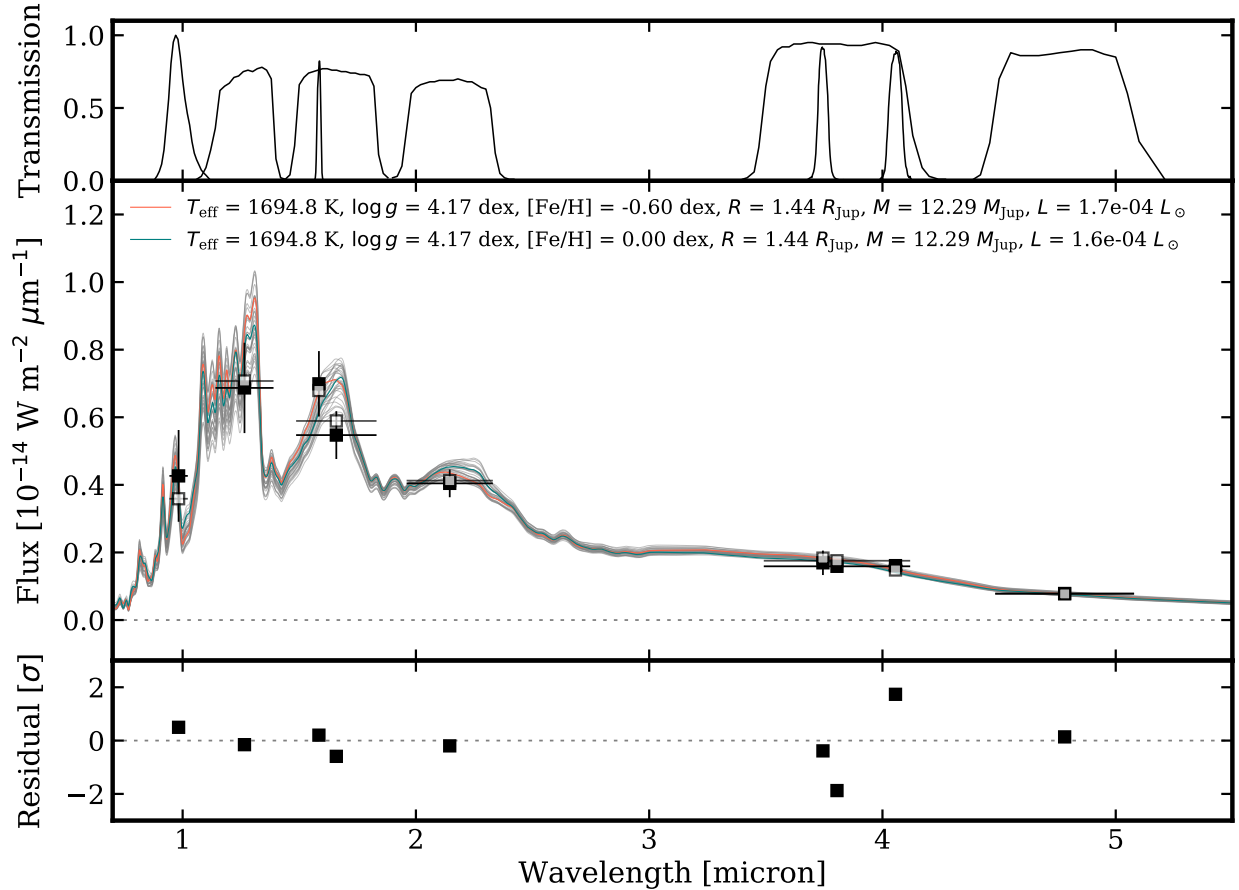
Fig. 9. Synthetic spectra ($\lambda/\Delta\lambda = 50$) of the DRIFT-PHOENIX atmospheric models that best describe the photometry of β Pic b with a prior on its mass. The two colored lines show the interpolated spectra for the median values of the posterior distributions, except for the metallicity, which is set to [Fe/H] = -0.06 (red) and [Fe/H] = 0.0 (teal). Additionally, 30 sets of parameter values were randomly drawn from the posterior distributions (see Fig. 10) and shown as gray spectra. The black solid squares are the data and the black open squares are the synthetic photometry for the best-fit model. The differences between the data and the synthetic photometry are shown in multiples of the 1σ uncertainties. The horizontal error bars indicate the FWHM of the filter transmission profiles, which are also plotted above the spectra. The mass and luminosity in the legend have been computed from the fit results.

$\log g = 4.17^{+0.10}_{-0.13}$ dex. The metallicity remained uncertain since its effect on the spectral appearance is relatively small compared to the uncertainties on the photometry. The posterior distribution of the metallicity has a peak both at -0.6 dex (at the lower edge of the grid) and a smaller peak close to solar metallicity. The effective temperature and radius are not significantly affected by the mass prior and were already constrained with the photometry alone. The best-fit values derived from the posterior distributions are $T_{\rm eff} = 1694 \pm 40$ K and $R = 1.44 \pm 0.05$ $R_{\rm Jup}$.

4.4. Detection limits in NB4.05 and M′ filters

The detection limits of our data were calculated using a method that is similar to the PSF subtraction described in Sect. 3.1. First, the azimuthal noise level was computed for a given position that was tested. An artificial planet was then injected with a S/N of 100 after which the self-subtraction inherent to PCA was determined by measuring the signal of the artificial source with a 1 *FWHM* aperture before and after the PSF subtraction. The contrast was then scaled to a false positive fraction (FPF) of 2.86×10^{-7}, which corresponds to a 5σ detection in the limit of Gaussian statistics. Since the noise measurement is limited by the small number of samples (i.e., the number of nonoverlapping apertures at a given separation), we assume a Student's

t-distribution for the calculation of the FPF (Mawet et al. 2014). The positions at which the detection limits were calculated started at an inner separation of 200 mas and increased in steps of 10 mas while the position angle was uniformly sampled with six positions. The average of the six azimuthal positions was stored and the procedure was repeated for a range of 5–30 PCs in steps of 5 PCs.

The results are presented in Fig. 11 for the NB4.05 and M' filters. The dispersion due to the different numbers of PCs that were used is largest at small separations where the impact of self-subtraction effects varies more strongly as function of PCs. The detection limits follow a similar steepness in the speckle-limited regime where field rotation causes only a small amount of movement of a signal while the curves quickly flatten off to the background-limited regime. This regime appears to start approximately at $\sim 8\lambda/D$ for the NB4.05 data (except for β Pic due to its brightness) and already at $\sim 5\lambda/D$ for the M' data. The limiting apparent magnitude in NB4.05 that was reached in the background-limited regime is approximately 12.5 mag for PZ Tel (16.8 min), 15.2 mag for HIP 65426 (2.6 h), 14.3 mag for HD 206893 (48.9 min), and 14.3 mag for β Pic (29.7 min). For the M' data, the limiting magnitude is about 13.3 mag for PZ Tel (24.4 min), 14.2 mag for HIP 65426 (54.9 min), and 14.3 mag for HD 206893 (73.9 min), where values between
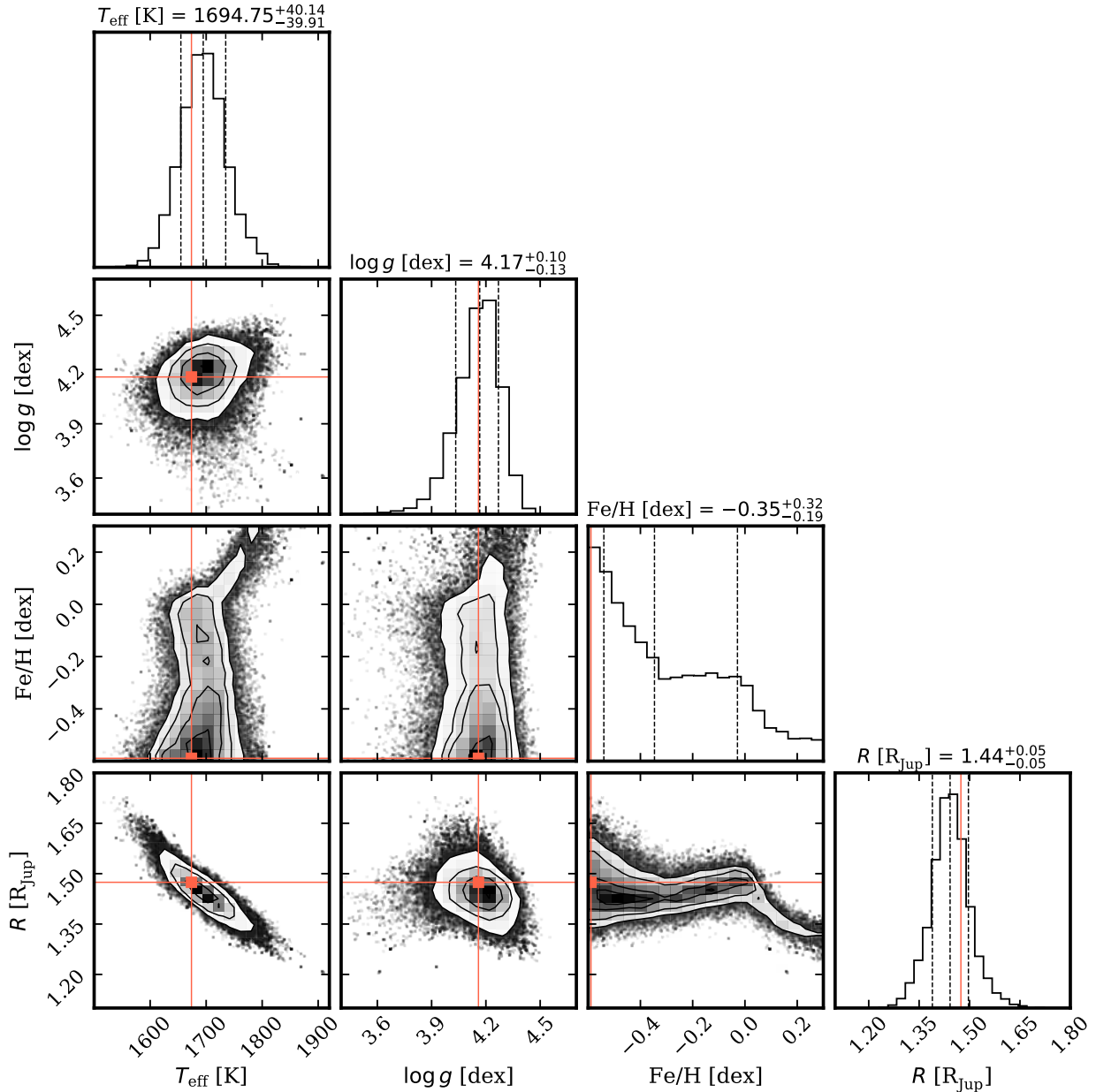
Fig. 10. Posterior distributions of the atmospheric model parameters that were fitted to the photometry of β Pic b with a prior on its mass. A comparison of the photometry and the sampled model spectra is shown in Fig. 9. The marginalized 1D and 2D distributions are shown together with the median, and 16th and 84th percentiles of the samples. The red horizontal and vertical lines indicate the sample position with the maximum posterior probability.

parentheses indicate the effective integration time after application of the frame selection.

5. Discussion and future outlook

5.1. Atmospheric insights from 3–5 μm photometry

Before discussing the extracted photometry and colors of our sample, we qualitatively describe the impact of molecular species and clouds on the appearance of gas giant and brown dwarf atmospheres at 3–5 μm. The photometric characteristics are shaped by the line opacities of the gas and continuum opacity of the cloud condensates. The composition of the gas depends on the temperature and pressure, which is most straightforwardly computed for a set of elemental abundances by assuming that

the reactions occur in chemical equilibrium. As an example, we show in the top panel of Fig. 12 how the equilibrium abundances of the most relevant species change with temperature while fixing the pressure. For the same species, we display in the bottom panel of Fig. 12 the molecular opacities at a spectral resolution of 1000.

The spectral window of the L' filter covers the fall and rise of two H_2O bands and also part of the CH_4 band at 3.3 μm. However, given the high temperatures ($T_{\rm eff} \gtrsim 1300$ K) and low surface gravity of the studied sample, the abundance of CH_4 is expected to be low. Specifically, there is evidence that the CO/CH_4 chemistry is not in equilibrium in the atmospheres of low-gravity objects (e.g., Konopacky et al. 2013; Moses et al. 2016). Therefore, only absorption by H_2O is expected to affect the L' photometry of our sample. The NB4.05 filter does not
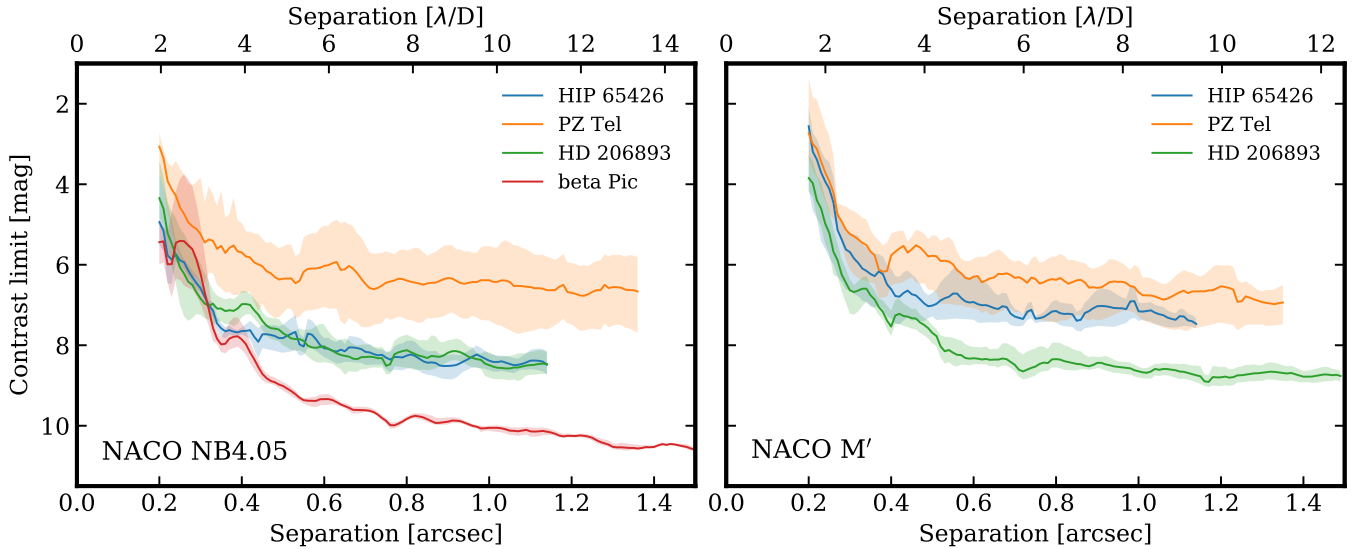
Fig. 11. Detection limits derived for the observed targets in the NACO NB4.05 (*left panel*) and *M′* (*right panel*) filters. The limits for the archival dataset of β Pic b in NB4.05 are also included as a reference. The solid lines show the mean contrast from six values of principal components (PCs) that were tested (5, 10, 15, 20, 25, and 30). The shaded area covers the range between the minimum and maximum detection limit from the different number of PCs. The upper horizontal axis is shown in units of λ/D, that is, 102 mas in NB4.05 and 120 mas in *M′* (at the central wavelengths of the filters).

cross any significant molecular absorption and is therefore a good tracer of the continuum. The opacity of H_2O increases between *L′* and NB4.05, which is continued across the *M′* filter. However, the opacity of CO is more than an order of magnitude larger than H_2O in the *M′* range. Therefore, CO will mostly affect the *M′* photometry, assuming that the CO and H_2O abundances are similar. The CO_2 fundamental stretching-mode band at 4.2 μm partially covers the *M′* band but its equilibrium abundance is about four orders of magnitude smaller than CO in the most optimistic case.

Since the NB4.05 filter is mostly sensitive to continuum emission, both the *L′*–NB4.05 and NB4.05–*M′* colors contain some information about the chemical composition of the atmosphere. Specifically, the *L′*–NB4.05 colors of our sample are expected to be affected by the H_2O opacity. The synthetic colors in Fig. 7 show that H_2O may indeed cause a red color at high temperatures which is further enhanced with the presence of clouds. The color has typical values of only a few tens of a magnitude, except for the low-temperature predictions for which CH_4 absorption occurs in *L′*. Similar effects of composition and clouds can be seen at 4–5 μm. At high temperature, both the predictions of the *L′*–*M′* color (see right panel in Fig. 6) and NB4.05–*M′* color (not shown) indicate blue values in the range of −0.6–0.0 mag. This occurs due to absorption by CO in *M′* if the effect of clouds is minimal. With clouds, the AMES-Dusty models predict NB4.05–*M′* colors that are close to gray while the *L′*–*M′* colors show a strong reddening towards lower-mass objects.

5.2. Photometric characteristics of the individual objects

The magnitudes and colors in the 3–5 μm range of β Pic b, HIP 65426 b, PZ Tel B, and HD 206893 B are presented in Sect. 4.1. Here, we discuss their photometric characteristics in the context of previous studies and the color predictions from atmospheric models. A color–color comparison based on 3–5 μm photometry of the directly imaged objects and various atmospheric models is therefore shown in Fig. 13.

5.2.1. β Pictoris b

The apparent magnitude of β Pic b that we derived from the NB4.05 data is 10.98 ± 0.04 mag. This value is consistent with 11.20 ± 0.23 mag from Quanz et al. (2010) and 11.04 ± 0.08 mag from Currie et al. (2013), within the respective uncertainties. In NB3.74, the photometry of the object is 11.25 ± 0.26 mag, which implies that the NB3.74–*L′* color is −0.05 ± 0.27 mag (*L′* = 11.30 ± 0.06 mag; Stolker et al. 2019). A (close to) gray color might indeed be expected since the central wavelength at NB3.74 lies in the center of the *L′* band, which covers a continuous spectral slope (see Fig. 9). In Sect. 4.3, we constrained the surface gravity of the planet to $\log g = 4.17^{+0.10}_{-0.13}$ dex by combining the photometry from *Y* to *M′* band and using prior on its mass. This value confirms the low-gravity nature of this object and is consistent with the $\log g = 4.18 ± 0.01$ dex that was derived by Chilcote et al. (2017) from the bolometric luminosity and age of the planet.

Figure 13 shows that the NB4.05–*M′* color of β Pic b (−0.12 ± 0.13 mag) is in line with the predictions by the DRIFT-PHOENIX, AMES-Dusty, BT-Settl, and cloudy petit-CODE models (i.e., all cloudy atmospheres). This blue color is likely caused by the CO fundamental band in *M′* regime. The red *L′*–NB4.05 color (0.32 ± 0.08 mag) might be a result of clouds, possibly combined with H_2O absorption in *L′*. This color is redder than what is predicted by these same cloudy atmosphere models and is more comparable to a ∼1000 K atmosphere. This may indicate that β Pic b has thicker clouds and/or stronger H_2O absorption. Such clouds cause an overall reddening of the spectrum, as is the case for the *L′*–*M′* color (see Fig. 6).

The sampled spectra from the MCMC posteriors in Sect. 4.3 point indeed towards a solution with some absorption by H_2O in the 3–4 μm region and CO absorption across the *M′* band (see Fig. 9). The strength of the absorption features that affect the precise values of the colors will also depend on the elemental abundances of carbon and oxygen. These are included in the combined metallicity parameter of the DRIFT-PHOENIX models which therefore restricts the fit to abundances only relative to solar values.
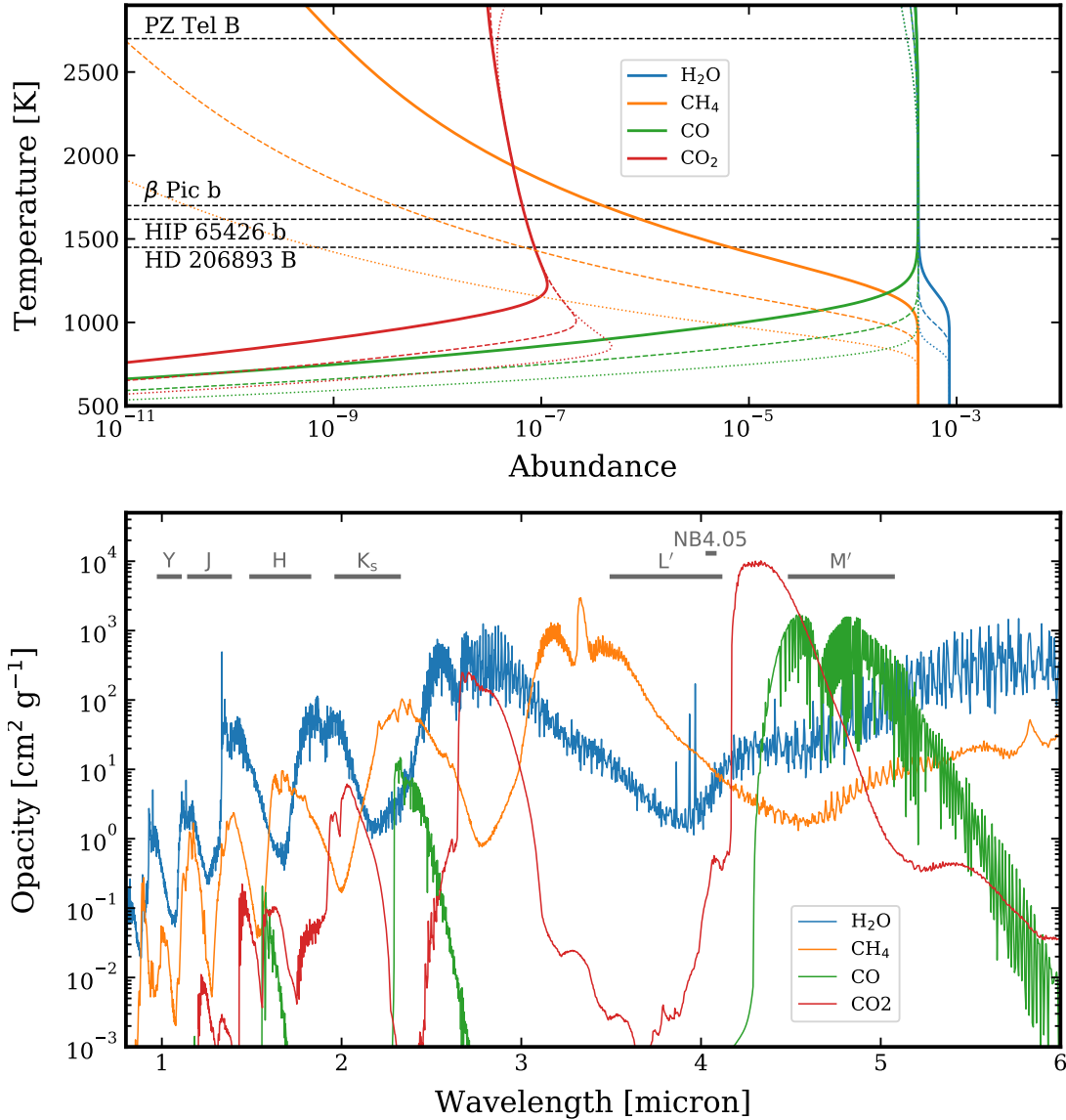
Fig. 12. *Top panel*: chemical equilibrium abundances calculated with `rate` at 1 bar (solid lines), 100 mbar (dashed lines), and 10 mbar (dotted lines) (Cubillos et al. 2019). For reference, model-dependent effective temperatures of β Pic b (derived in Sect. 4.3) and of the rest of the sample (Maire et al. 2016; Delorme et al. 2017; Cheetham et al. 2019) are indicated by horizontally dashed lines. *Bottom panel*: molecular opacities calculated with `petitRADTRANS` at 1 bar and 1500 K (Mollière et al. 2019). Filter bandpasses are indicated with horizontal bars.

5.2.2. HIP 65426 b

The L' and M' photometry of HIP 65426 b was previously reported by Cheetham et al. (2019). The authors determined a contrast of 8.47 ± 0.14 mag and 8.2 ± 0.4 mag in L' and M', respectively, and an apparent magnitude of 15.26 ± 0.15 mag and 15.1 ± 0.5 mag in the two filters. The contrast that we obtained in L' is 0.09 mag larger while the M' contrast is 0.31 mag smaller. A potential difference for the M' data could be related to the PSF template that is used. Specifically, we applied a one-on-one injection of the negative PSF for each image (see Sect. 3.3.2), thereby accounting for changes in the observing conditions. The derived apparent magnitude is 0.06 mag larger in L' and 0.45 mag smaller in M' compared to the results from Cheetham et al. (2019). Therefore, the flux calibration approach introduced an additional, minor difference in the final magnitudes.

The L'–NB4.05 color of HIP 65426 b is close to gray (0.1 ± 0.25 mag). Figure 13 shows that the color is consistent with all model predictions at the approximate $T_{\rm eff}$ of the planet ($\sim$1600 K), except for the clear petitCODE model (due to deeper CH_4 absorption in the L' band). The L'–NB4.05 color of HIP 65426 b is less red compared to β Pic b (see also Fig. 7), which may indicate that absorption by H_2O in the L' band is dampened more strongly by the atmosphere's dust content. Similarly, the red NB4.05–M' color (0.58 ± 0.36 mag) hints at a lack of CO absorption in the M' band due to enhanced cloud densities close to the photosphere. Indeed, the M' flux is relatively large and comparable to a late M dwarf (see Fig. 6). Considering the uncertainty on the 4–5 μm color, it is most similar (within 1–2σ) to the synthetic colors from the BT-Settl, DRIFT-PHOENIX, and cloudy petitCODE models.

5.2.3. PZ Telescopii B

The highly variable observing conditions of the L' dataset resulted in flux variations up to 40% (see Sect. 2). Although we applied a stringent frame selection to ensure that the science
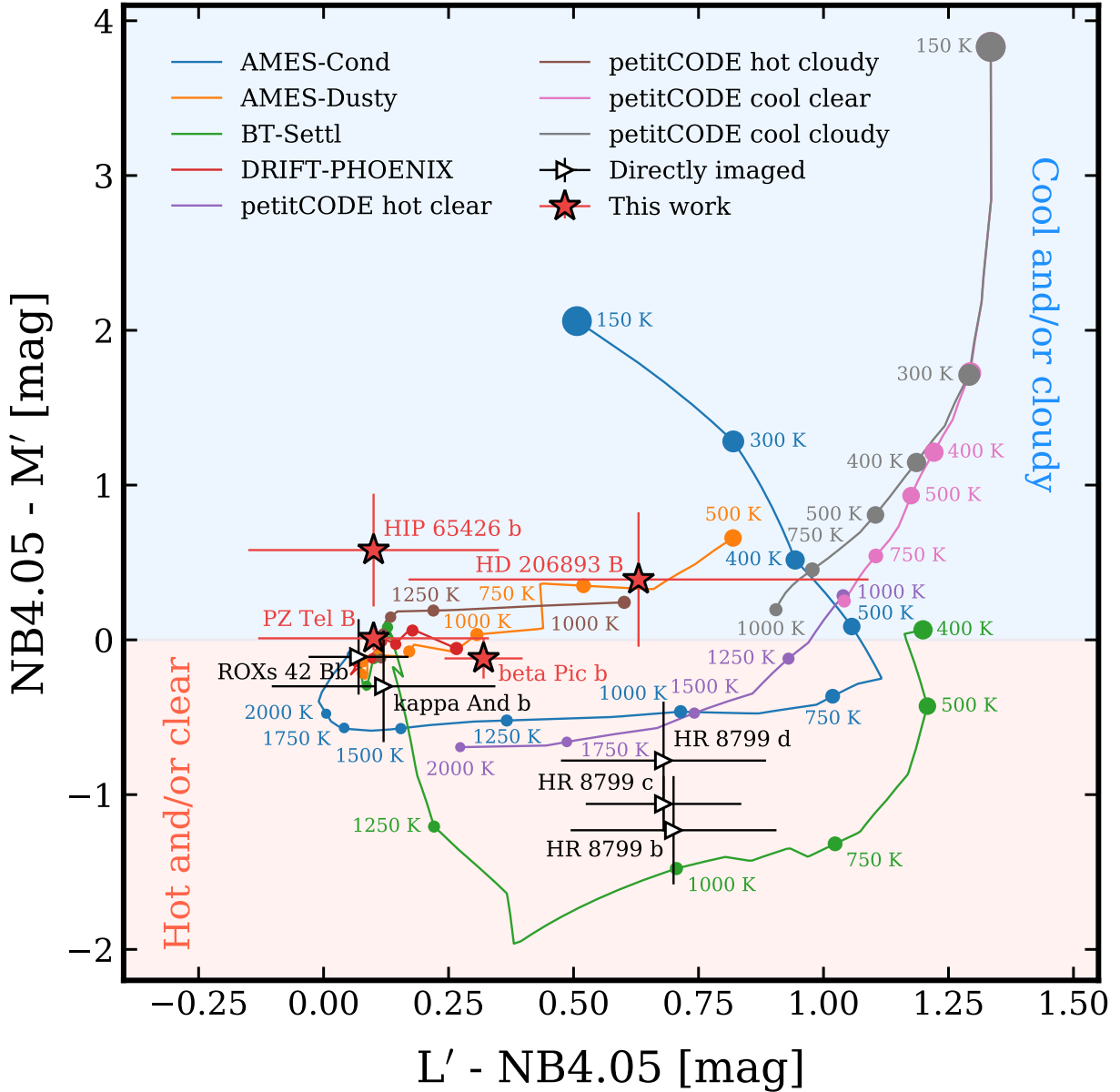
Fig. 13. Color–color diagram of L'–NB4.05 vs. NB4.05–M'. Synthetic colors are calculated from the AMES-Cond ($T_{eff} = 150$–3000 K), AMES-Dusty ($T_{eff} = 500$–3000 K), BT-Settl ($T_{eff} = 400$–3000 K), DRIFT-PHOENIX ($T_{eff} = 1000$–3000 K), petitCODE cool ($T_{eff} = 150$–1000 K), and petitCODE hot ($T_{eff} = 1000$–2000 K) atmospheric models (colored lines) across the indicated temperature range. The other parameters have been fixed to $\log g = 4.0$ dex, [Fe/H] = 0.0 dex, C/O = 0.55 (only petitCODE), and $f_{sed} = 2.0$ (only petitCODE cloudy). The dots decrease in size towards higher temperatures and correspond to 150, 300, 400, 500, 750, 1000, 1250, 1500, 1750, 2000, and 3000 K. The directly imaged companions for which photometry is available in the three filters are shown for comparison. The blue and red shaded regions indicate approximately two different NB4.05–M' regimes.

and flux exposures were obtained in comparable conditions, a remaining variation in the stellar flux may have caused a bias in the flux calibration which is difficult to correct or account for in the error budget. The derived contrast in L' of 4.78 ± 0.21 mag is 0.37 mag smaller than the value reported by Beust et al. (2016) (5.15 ± 0.15 mag), which could be related to a difference in frame selection. The contrast from Beust et al. (2016) was calibrated by Maire et al. (2016) to an apparent magnitude of 11.05 ± 0.18 mag. Therefore, our final magnitude is only 0.01 mag smaller due to our different approach with the flux calibration.

The L'–NB4.05 color of PZ Tel B is slightly red (0.10 ± 0.23 mag), but also consistent with being gray or blue, while the NB4.05–M' is gray (0.01 ± 0.08 mag). At 2700 K, the

L'–NB4.05 and NB4.05–M' colors are expected to be red and blue (by a tenth of a magnitude), respectively, due to absorption by H_2O and CO. Figure 13 shows that all model predictions converge at the highest temperatures because cloud condensation will not occur. Consequently, the differences in the sample of high- and low-gravity objects is more limited (see Figs. 6 and 7). The color–color diagram of Fig. 13 shows a clustering of synthetic colors close to gray, which approximately coincides with the colors of PZ Tel B (see also Fig. 8). The slight reddening (although not significant) of the L'–M' color could be caused by a thin layer of small cloud particles which may have formed at higher altitudes than predicted by the considered models.

5.2.4. HD 206893 B

The low-mass companion of HD 206893 was previously identified as an extremely red, late L dwarf (Milli et al. 2017; Delorme et al. 2017). The precise nature of the object is not clear because of the uncertain age of the system, yielding mass estimates in the range of 12–50 $M_{\rm Jup}$ (Delorme et al. 2017). The L' brightness of HD 206893 B that we derived in the present study is 13.80 ± 0.31 mag, which implies that the object is 0.37 mag fainter compared to results from Milli et al. (2017) (a similar stellar magnitude was used for the calibration). However, the values are comparable within their uncertainties, which are relatively large due to residual speckle noise at the small separation of HD 206893 B.

With a lower L' brightness and less red color, the object's photometric characteristics overlap with the low-gravity objects in the H–L' versus $M_{L'}$ diagram (see left panel in Fig. 6). The high absolute flux in M', on the other hand, results in a very red L'–M' color, which deviates by $\sim2\sigma$ from the field dwarfs (see Fig. 6). The L'–NB4.05 color of HD 206893B is redder than both β Pic b and HIP 65426 b, although the error bar is large. At an approximate temperature of 1450 K, methane absorption is expected to occur in the L' band in cases where abundances are in chemical equilibrium (see AMES-Cond and clear petitCODE predictions in Fig. 13). However, in a low-gravity environment, quenching of CH_4 and CO is expected to occur at the considered temperature, which would increase the L' flux. Furthermore, the NB4.05–M' color is red, but is approximately in line with predictions by the AMES-Dusty, petitCODE cloudy, and DRIFT-PHOENIX models at those temperatures. The combined photometric characteristics may hint at an extended amount of dust in the atmosphere of HD 206893 B or the presence of circumplanetary material which partially reprocesses its photospheric emission.

5.3. Upcoming opportunities for MIR imaging observations of planetary and substellar companions

As discussed above, adding MIR information to the SEDs of low-mass objects puts additional constraints on their atmospheric properties. In addition, the MIR allows in principle to search for cooler – and hence older or lower mass – objects than possible at shorter wavelengths. For cooler planets, there will be an increasing amount of information available from photometry in the 3–5 μm regime because the peak flux shifts towards the MIR and molecular absorption features increase in strength. Color effects between the L', NB4.05, and M' fluxes are therefore expected to become stronger towards lower temperatures (see Fig. 13).

From the ground, we expect the ERIS instrument for the VLT (Davies et al. 2018) to be an important step forward and to open up new discovery space in both new detections and better characterization potential. The ERIS instrument will sit behind the adaptive secondary mirror of the Adaptive Optics Facility and provide better AO performance than NACO. In combination with an overall reduced thermal background (due to fewer warm surfaces in the light-path compared to NACO) and state-of-the-art coronagraphs, ERIS will be significantly more sensitive. Furthermore, ERIS will have spectroscopic capabilities in the 3–5 μm range meaning that photometry can be complemented with medium-resolution spectra.

In terms of sensitivity and wavelength coverage in the MIR, JWST is expected to set new benchmarks for years to come. Depending on the final contrast performance, some of the known low-mass and planetary companions may be accessible to JWST instrumentation and their atmospheres can be probed for a variety of constituents including nitrogen- and phosphorus-bearing species such as NH_3 and PH_3 (Danielski et al. 2018). In addition, the broad spectral coverage will help determine the bolometric luminosity of these young objects, thereby putting additional constraints on evolutionary models (Marleau et al. 2019). Motivated by the expected significant increase in sensitivity, new evolutionary models were recently introduced (Linder et al. 2019) to make predictions for objects down to masses comparable to those of the ice giants in the Solar System. Whether or not JWST will indeed find such objects around young nearby stars remains to be seen.

Finally, as one of the first-generation instruments, the 39 m Extremely Large Telescope will feature the Mid-IR ELT Imager and Spectrograph (METIS; Brandl et al. 2018). While not being able to compete with JWST in terms of sensitivity, the unparalleled spatial resolution of METIS and its ability to point to the brightest stars in the sky will allow METIS to search for and potentially image small, terrestrial exoplanets around stars in the immediate vicinity of the Sun (Quanz et al. 2015b). The first experiment demonstrating the general feasibility of such observations was recently carried out in the context of the NEAR campaign with the upgraded VISIR instrument at the VLT (Kasper et al. 2017).

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. London Ser. A, 370, 2765
Amara, A., & Quanz, S. P. 2012, MNRAS, 427, 948
Avenhaus, H., Schmid, H. M., & Meyer, M. R. 2012, A&A, 548, A105
Bailey, V., Hinz, P. M., Currie, T., et al. 2013, ApJ, 767, 31
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
Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011, ApJ, 733, 65
Beust, H., Bonnefoy, M., Maire, A. L., et al. 2016, A&A, 587, A89
Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155
Biller, B. A., Liu, M. C., Wahhaj, Z., et al. 2010, ApJ, 720, L82
Boccaletti, A., Lagage, P. O., Baudoz, P., et al. 2015, PASP, 127, 633
Bohlin, R. C. 2007, ASP Conf. Ser., 364, 315
Bonnefoy, M., Lagrange, A. M., Boccaletti, A., et al. 2011, A&A, 528, L15
Bonnefoy, M., Boccaletti, A., Lagrange, A.-M., et al. 2013, A&A, 555, A107
Bonnefoy, M., Currie, T., Marleau, G. D., et al. 2014, A&A, 562, A111
Bonnefoy, M., Perraut, K., Lagrange, A. M., et al. 2018, A&A, 618, A63
Bouchet, P., Manfroid, J., & Schmider, F. X. 1991, A&AS, 91, 409
Bowler, B. P. 2016, PASP, 128, 102001
Brandl, B. R., Absil, O., Agócs, T., et al. 2018, SPIE Conf. Ser., 10702, 107021U
Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, Rev. Mod. Phys., 73, 719
Burrows, A., Sudarsky, D., & Hubeny, I. 2006, ApJ, 640, 1063
Chauvin, G., Lagrange, A. M., Dumas, C., et al. 2004, A&A, 425, L29
Chauvin, G., Desidera, S., Lagrange, A. M., et al. 2017, A&A, 605, L9

Cheetham, A. C., Samland, M., Brems, S. S., et al. 2019, A&A, 622, A80
Chilcote, J., Pueyo, L., De Rosa, R. J., et al. 2017, AJ, 153, 182
Cubillos, P. E., Blecic, J., & Dobbs-Dixon, I. 2019, ApJ, 872, 111
Currie, T., Burrows, A., Itoh, Y., et al. 2011, ApJ, 729, 128
Currie, T., Fukagawa, M., Thalmann, C., Matsumura, S., & Plavchan, P. 2012, ApJ, 755, L34
Currie, T., Burrows, A., Madhusudhan, N., et al. 2013, ApJ, 776, 15
Currie, T., Burrows, A., Girard, J. H., et al. 2014, ApJ, 795, 133
Cushing, M. C., Rayner, J. T., & Vacca, W. D. 2005, ApJ, 623, 1115
Cushing, M. C., Marley, M. S., Saumon, D., et al. 2008, ApJ, 678, 1372
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2012, VizieR Online Data Catalog: II/311
Daemgen, S., Todorov, K., Silva, J., et al. 2017, A&A, 601, A65
Danielski, C., Baudino, J.-L., Lagage, P.-O., et al. 2018, AJ, 156, 276
Davies, R., Esposito, S., Schmid, H. M., et al. 2018, SPIE Conf. Ser., 10702, 1070209
De Rosa, R. J., Rameau, J., Patience, J., et al. 2016, ApJ, 824, 121
Delorme, P., Gagné, J., Girard, J. H., et al. 2013, A&A, 553, L5
Delorme, P., Schmidt, T., Bonnefoy, M., et al. 2017, A&A, 608, A79
Dupuy, T. J., & Kraus, A. L. 2013, Science, 341, 1492
Dupuy, T. J., & Liu, M. C. 2012, ApJS, 201, 19
Dupuy, T. J., Brandt, T. D., Kratter, K. M., & Bowler, B. P. 2019, ApJ, 871, L4
Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M. 2018, A&A, 618, A138
Foreman-Mackey, D. 2016, J. Open Source Softw., 1
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Galicher, R., Marois, C., Macintosh, B., Barman, T., & Konopacky, Q. 2011, ApJ, 739, L41
Geißler, K., Chauvin, G., & Sterzik, M. F. 2008, A&A, 480, 193
Golimowski, D. A., Leggett, S. K., Marley, M. S., et al. 2004, AJ, 127, 3516
Gomez Gonzalez, C. A., Absil, O., & Van Droogenbroeck, M. 2018, A&A, 613, A71
Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, AJ, 132, 161
Greenbaum, A. Z., Pueyo, L., Ruffio, J.-B., et al. 2018, AJ, 155, 226
Griffith, C. A., & Yelle, R. V. 1999, ApJ, 519, L85
Helling, C., Dehn, M., Woitke, P., & Hauschildt, P. H. 2008, ApJ, 675, L105
Houk, N. 1978, Michigan catalogue of two-dimensional spectral types for the HD stars
Ireland, M. J., Kraus, A., Martinache, F., Law, N., & Hillenbrand, L. A. 2011, ApJ, 726, 113
Jenkins, J. S., Pavlenko, Y. V., Ivanyuk, O., et al. 2012, MNRAS, 420, 3587
Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890
Kasper, M., Arsenault, R., Käufl, H. U., et al. 2017, The Messenger, 169, 16
Käufl, H.-U., Kasper, M., Arsenault, R., et al. 2018, SPIE Conf. Ser., 10702, 107020D
Kenworthy, M. A., Snik, F., Keller, C. U., et al. 2018, SPIE Conf. Ser., 10702, 1070246
Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013, Science, 339, 1398
Lafrenière, D., Jayawardhana, R., & van Kerkwijk, M. H. 2008, ApJ, 689, L153
Lafrenière, D., Jayawardhana, R., & van Kerkwijk, M. H. 2010, ApJ, 719, 497
Lagrange, A.-M., Gratadour, D., Chauvin, G., et al. 2009, A&A, 493, L21

Leggett, S. K., Burningham, B., Saumon, D., et al. 2010, ApJ, 710, 1627
Lenzen, R., Hartung, M., Brandner, W., et al. 2003, Proc. SPIE, 4841, 944
Linder, E. F., Mordasini, C., Mollière, P., et al. 2019, A&A, 623, A85
Liu, M. C., Dupuy, T. J., & Allers, K. N. 2016, ApJ, 833, 96
Luhman, K. L. 2013, ApJ, 767, L1
Macintosh, B. A., Graham, J. R., Palmer, D. W., et al. 2008, SPIE Conf. Ser., 7015, 18
Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64
Maire, A. L., Bonnefoy, M., Ginski, C., et al. 2016, A&A, 587, A56
Males, J. R., Close, L. M., Morzinski, K. M., et al. 2014, ApJ, 786, 32
Marleau, G.-D., Coleman, G. A. L., Leleu, A., & Mordasini, C. 2019, A&A, 624, A20
Marley, M. S., & Robinson, T. D. 2015, ARA&A, 53, 279
Marley, M. S., Saumon, D., Cushing, M., et al. 2012, ApJ, 754, 135
Marois, C., Macintosh, B., Barman, T., et al. 2008, Science, 322, 1348
Mawet, D., Milli, J., Wahhaj, Z., et al. 2014, ApJ, 792, 97
Mawet, D., Wizinowich, P., Dekany, R., et al. 2016, SPIE Conf. Ser., 9909, 99090D
Meshkat, T., Kenworthy, M. A., Quanz, S. P., & Amara, A. 2014, ApJ, 780, 17
Metchev, S. A., & Hillenbrand, L. A. 2006, ApJ, 651, 1166
Milli, J., Hibon, P., Christiaens, V., et al. 2017, A&A, 597, L2
Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, A&A, 627, A67
Morley, C. V., Skemer, A. J., Allers, K. N., et al. 2018, ApJ, 858, 97
Morzinski, K. M., Males, J. R., Skemer, A. J., et al. 2015, ApJ, 815, 108
Moses, J. I., Marley, M. S., Zahnle, K., et al. 2016, ApJ, 829, 66
Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, AJ, 158, 13
Noll, S., Kausch, W., Barden, M., et al. 2012, A&A, 543, A92
Quanz, S. P., Meyer, M. R., Kenworthy, M. A., et al. 2010, ApJ, 722, L49
Quanz, S. P., Amara, A., Meyer, M. R., et al. 2015a, ApJ, 807, 64
Quanz, S. P., Crossfield, I., Meyer, M. R., Schmalzl, E., & Held, J. 2015b, Int. J. Astrobiol., 14, 279
Rajan, A., Rameau, J., De Rosa, R. J., et al. 2017, AJ, 154, 10
Rousset, G., Lacombe, F., Puget, P., et al. 2003, Proc. SPIE, 4839, 140
Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, A&A, 603, A57
Saumon, D., & Marley, M. S. 2008, ApJ, 689, 1327
Schmidt, T. O. B., Mugrauer, M., Neuhäuser, R., et al. 2014, A&A, 566, A85
Shkolnik, E. L., Allers, K. N., Kraus, A. L., Liu, M. C., & Flagg, L. 2017, AJ, 154, 69
Skemer, A. J., Marley, M. S., Hinz, P. M., et al. 2014, ApJ, 792, 17
Skemer, A. J., Morley, C. V., Allers, K. N., et al. 2016a, ApJ, 826, L17
Skemer, A. J., Morley, C. V., Zimmerman, N. T., et al. 2016b, ApJ, 817, 166
Snellen, I. A. G., & Brown, A. G. A. 2018, Nat. Astron., 2, 883
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
Stolker, T., Bonse, M. J., Quanz, S. P., et al. 2019, A&A, 621, A59
Stone, J. M., Eisner, J., Skemer, A., et al. 2016, ApJ, 829, 39
Stone, J. M., Skemer, A. J., Hinz, P. M., et al. 2018, AJ, 156, 286
Torres, C. A. O., Quast, G. R., da Silva, L., et al. 2006, A&A, 460, 695
Vigan, A., Bonnefoy, M., Ginski, C., et al. 2016, A&A, 587, A55
Wang, J. J., Graham, J. R., Dawson, R., et al. 2018, AJ, 156, 192
Wertz, O., Absil, O., Gómez González, C. A., et al. 2017, A&A, 598, A83
Yamamura, I., Tsuji, T., & Tanabé, T. 2010, ApJ, 722, 682
Zahnle, K. J., & Marley, M. S. 2014, ApJ, 797, 41
Zurlo, A., Vigan, A., Galicher, R., et al. 2016, A&A, 587, A57

Appendix A: Posterior probability distributions

In addition to Fig. 4, we present the posterior distributions for the remaining companions and filters in Fig. A.1.
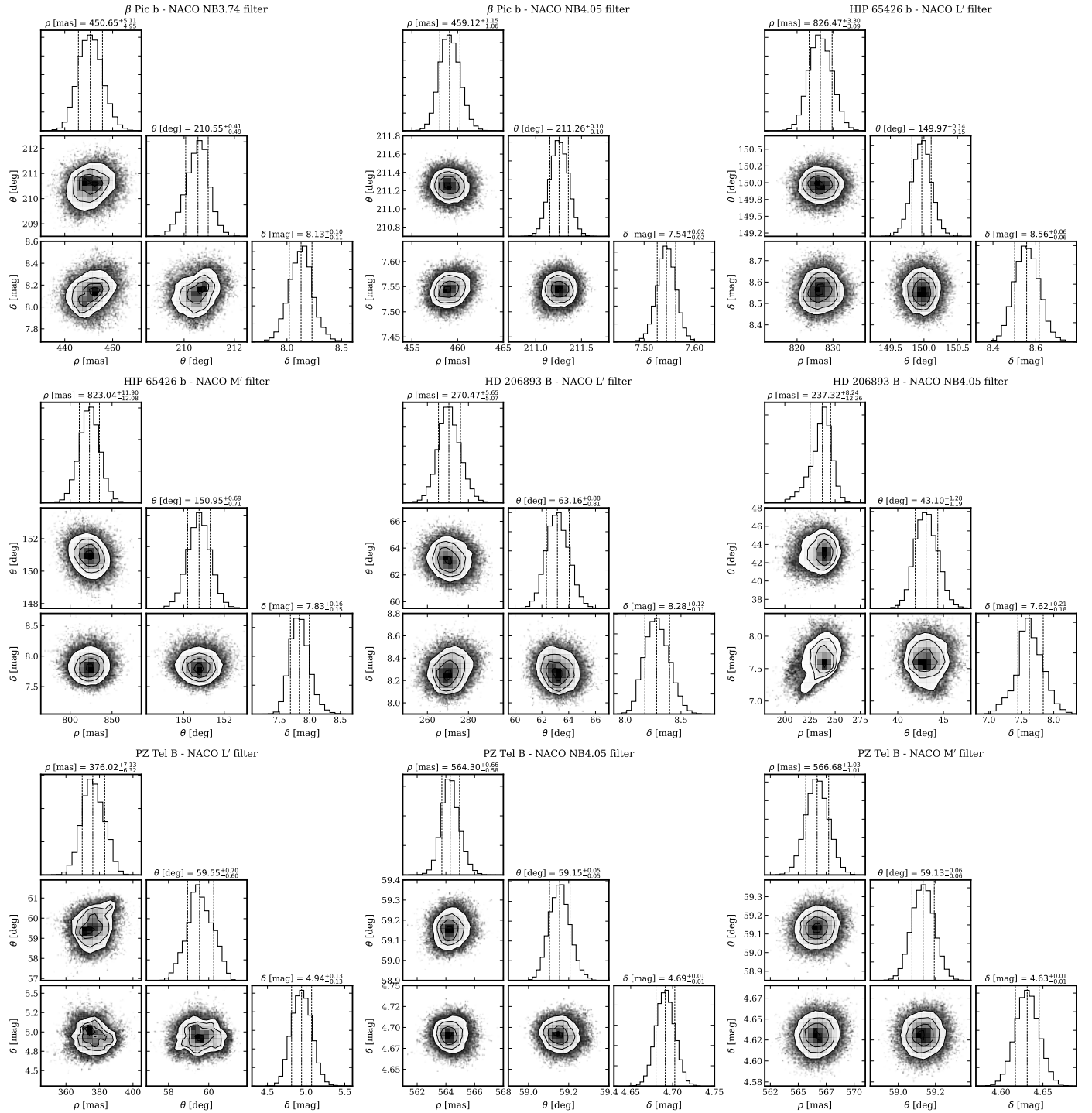

Fig. A.1. Posterior distributions of the separation, ρ, position angle, θ, and planet-to-star flux contrast, δ, for the remaining companions and filters. The companion name and filter is provided above each panel. Additional information can be found in the caption of Fig. 4.

Appendix B: Astrometry of the companions

The astrometry of the companions is listed in Table B.1. The overview includes the retrieved separation and position angle (measured east of north) from the MCMC analysis described in Sect. 3.3.1 and related posterior distributions presented in Figs. 4 and A.1. The measurement bias and related uncertainties are also listed, as described in Sect. 3.3.2 and illustrated in Fig. 5.

Table B.1. Astrometry and error budget.

Target	Filter	Separation MCMC [a] (mas)	Separation bias [a] (mas)	PA MCMC [a] (deg)	PA bias [a] (deg)	True north correction (deg)	Final separation [b] (mas)	Final PA [c] (deg)
β Pic b	NB3.74	450.65 ± 5.11	0.66 ± 12.58	210.55 ± 0.49	-0.02 ± 1.30	-0.40 ± 0.10	451.30 ± 13.58	210.13 ± 1.44
β Pic b	NB4.05	459.12 ± 1.15	0.03 ± 2.58	211.26 ± 0.10	-0.01 ± 0.20	-0.40 ± 0.10	459.15 ± 2.83	210.84 ± 0.46
HIP 65426 b	L'	826.47 ± 3.30	0.79 ± 7.57	149.97 ± 0.15	-0.00 ± 0.35	-0.40 ± 0.10	827.26 ± 8.26	149.56 ± 0.55
HIP 65426 b	NB4.05	825.33 ± 9.93	-1.56 ± 17.10	150.13 ± 0.49	-0.08 ± 1.06	-0.40 ± 0.10	823.77 ± 19.77	149.65 ± 1.23
HIP 65426 b	M'	823.04 ± 12.08	1.79 ± 18.75	150.95 ± 0.71	0.01 ± 1.09	-0.40 ± 0.10	824.83 ± 22.30	150.56 ± 1.36
PZ Tel B	L'	376.02 ± 7.13	-10.01 ± 10.74	59.55 ± 0.70	0.01 ± 0.47	-0.40 ± 0.10	366.01 ± 12.89	59.16 ± 0.94
PZ Tel B	NB4.05	564.30 ± 0.66	-0.08 ± 1.43	59.15 ± 0.05	0.00 ± 0.11	-0.40 ± 0.10	564.22 ± 1.57	58.76 ± 0.42
PZ Tel B	M'	566.68 ± 1.03	0.15 ± 1.78	59.13 ± 0.06	-0.01 ± 0.15	-0.40 ± 0.10	566.84 ± 2.06	58.72 ± 0.43
HD 206893 B	L'	270.47 ± 5.65	-0.95 ± 10.76	63.16 ± 0.88	0.00 ± 1.93	-0.40 ± 0.10	269.53 ± 12.15	62.76 ± 2.16
HD 206893 B	NB4.05	237.32 ± 12.26	1.80 ± 12.55	43.10 ± 1.28	-0.17 ± 1.71	-0.40 ± 0.10	239.12 ± 17.55	42.53 ± 2.17
HD 206893 B	M'	245.33 ± 10.18	1.18 ± 18.76	43.09 ± 1.00	0.11 ± 1.96	-0.40 ± 0.10	246.51 ± 21.34	42.80 ± 2.24

Notes. [a]For asymmetric uncertainties, we conservatively use the largest of the two values that were derived from both the statistical and systematic error analysis (see Figs. 4, 5, and A.1). [b]The final separation is calculated by adding the bias offset and combining the two error components in quadrature. [c]The final position angle (PA) is calculated by adding the bias and assumed true north offset, and combining the three error components in quadrature.

Appendix C: Color corrections for WISE and NACO

The photometry of HIP 65426, PZ Tel, and HD 206893 was calibrated from the available WISE photometry. We computed colors of the WISE and NACO filters from the BT-NextGen atmospheric models (Allard et al. 2012) in order to determine the color corrections that are required as a function of effective temperature. The color–color diagrams in Fig. C.1 show the comparison of the synthetic colors with the $W1-W2$ colors of our sample.
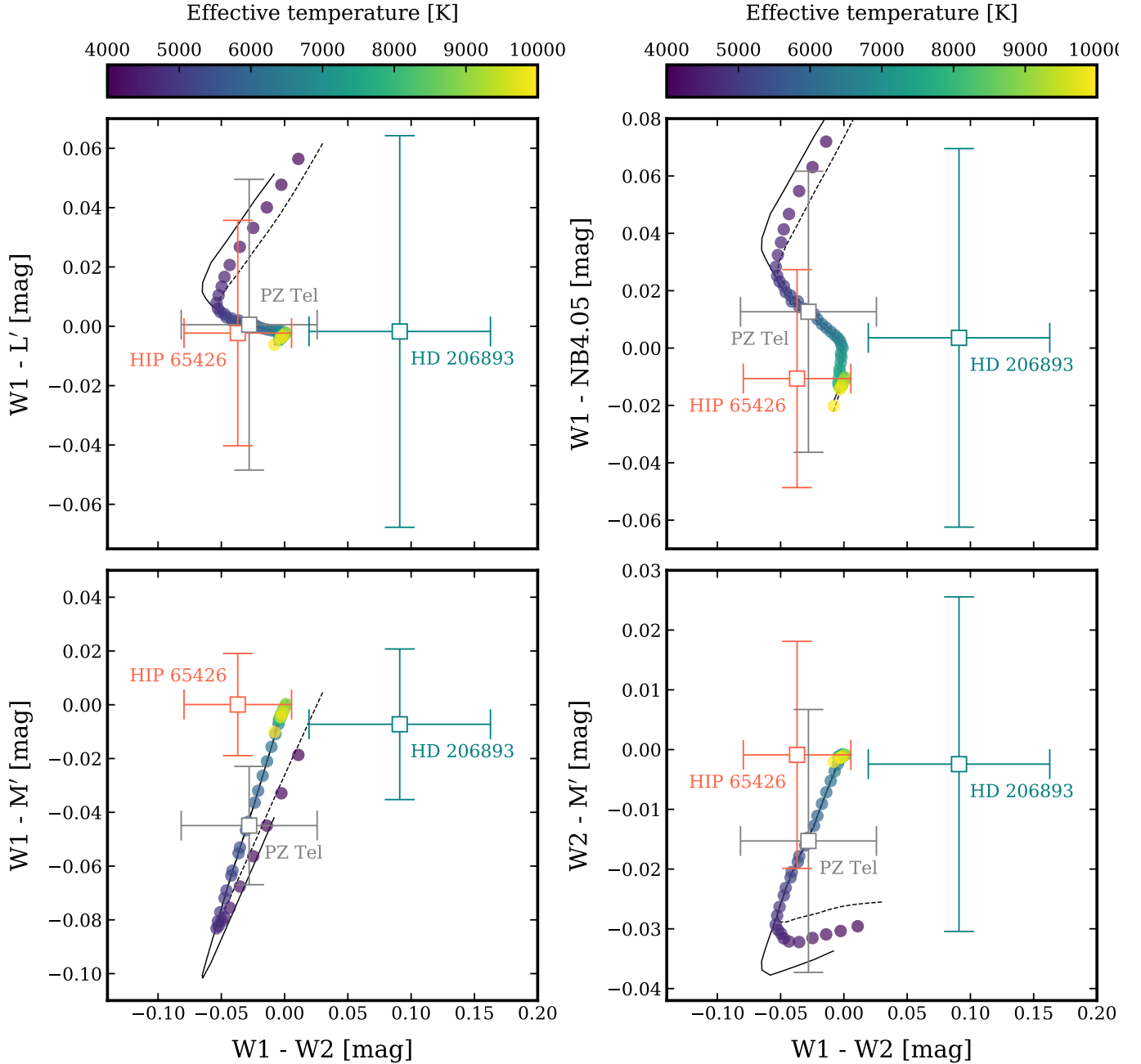

Fig. C.1. Color–color diagrams of $W1-W2$ vs. $W1-L'$ (*top left panel*), $W1-NB4.05$ (*top right panel*), $W1-M'$ (*bottom left panel*), and $W2-M'$ (*bottom right panel*). The synthetic photometry (colored dots) is calculated for 50 logarithmically spaced temperatures between 4000 and 10 000 K and log g = 4.5 dex. For comparison, the same calculation was done for log g = 4.0 dex (black solid line) and log g = 5.0 dex (black dashed line). The individual data points show the $W1-W2$ color and uncertainty of HIP 65426, PZ Tel, and HD 206893 along the horizontal axis (see Table 1). For the vertical axis, the color was computed from the BT-NextGen models at an effective temperature of 8840 K (HIP 65426; Chauvin et al. 2017), 5665 K (PZ Tel; Schmidt et al. 2014), and 6500 K (HD 206893; Delorme et al. 2017) with the error bars showing the uncertainty on either the $W1$ or $W2$ photometry.