

ACCEPTED TO AJ
 Preprint typeset using L^AT_EX style emulatej v. 5/2/11

PROSPECTING IN ULTRACOOOL DWARFS: MEASURING THE METALLICITIES OF MID- AND LATE-M DWARFS

ANDREW W. MANN^{1,2,3}, NIAL R. DEACON⁴, ERIC GAIDOS⁵, MEGAN ANSDELL³, JOHN M. BREWER⁶, MICHAEL C. LIU³,
 EUGENE A. MAGNIER³, KIMBERLY M. ALLER³

Accepted to AJ

ABSTRACT

Metallicity is a fundamental parameter that contributes to the physical characteristics of a star. However, the low temperatures and complex molecules present in M dwarf atmospheres make it difficult to measure their metallicities using techniques that have been commonly used for Sun-like stars. Although there has been significant progress in developing empirical methods to measure M dwarf metallicities over the last few years, these techniques have been developed primarily for early- to mid-M dwarfs. We present a method to measure the metallicity of mid- to late-M dwarfs from moderate resolution ($R \sim 2000$) K -band ($\simeq 2.2 \mu\text{m}$) spectra. We calibrate our formula using 44 wide binaries containing an F, G, K, or early M primary of known metallicity and a mid- to late-M dwarf companion. We show that similar features and techniques used for early M dwarfs are still effective for late-M dwarfs. Our revised calibration is accurate to ~ 0.07 dex for M4.5–M9.5 dwarfs with $-0.58 < [\text{Fe}/\text{H}] < +0.56$ and shows no systematic trends with spectral type, metallicity, or the method used to determine the primary star metallicity. We show that our method gives consistent metallicities for the components of M+M wide binaries. We verify that our new formula works for unresolved binaries by combining spectra of single stars. Lastly, we show that our calibration gives consistent metallicities with the Mann et al. (2013a) study for overlapping (M4–M5) stars, establishing that the two calibrations can be used in combination to determine metallicities across the entire M dwarf sequence.

Subject headings: stars: abundances — stars: binaries: visual — stars: fundamental parameters — stars: late-type — techniques: spectroscopic

1. INTRODUCTION

M dwarfs have become attractive targets for exoplanet searches (e.g., Fischer et al. 2012). M dwarfs represent $\sim 75\%$ of stars in the solar neighborhood (Henry et al. 2006) so their planets weigh heavily on any Galactic planet occurrence calculations. Stellar companions, which can impede giant planet formation (Kraus et al. 2012), dilute transits detections, and make Doppler detections more difficult, are less common around M dwarfs than for solar-type stars (Figueira et al. 2012). M dwarf’s low masses and small radii enhance Doppler and transit signals, thereby increasing the feasibility of detecting of Earth-sized planets in their habitable zones. These enhancements strengthen considerably from early- to late- M dwarfs. Early M-type dwarfs have masses and radii about half that of the Sun, while late-M dwarfs can have masses and radii $\sim 10\%$ that of the Sun (Dupuy et al. 2010; Boyajian et al. 2012) resulting in deeper transit depths and stronger transit signals for an equal size/mass planet. Further, the habitable zone for a late M-type dwarf is 5-10 times closer to the star than for an early M-type dwarf (Kopparapu et al. 2013),

resulting in a larger Doppler signal and more likely and frequent transits.

Studies of M dwarfs have already advanced the study of planet occurrence with stellar mass (e.g., Johnson et al. 2010; Gaidos et al. 2013) and metallicity (Mann et al. 2012, 2013c). Their low masses give additional leverage on any correlation between stellar mass and planet properties, and their large convective zones dilute any metallicity changes from pollution of the photosphere (Gonzalez 1997; Pinsonneault et al. 2001). However, fully exploiting M dwarfs to advance our knowledge of planet occurrence requires accurate metallicities for the entire sequence of M dwarfs, which are currently unavailable.

The advantages discussed above, among others, have motivated a number of planet surveys specifically targeting mid- to late-M dwarfs, most of which are coming online in the next few years. This includes near-infrared radial velocity surveys like CARMENES (Quirrenbach et al. 2012) and the Habitable-Zone Planet Finder (Mahadevan et al. 2012), transit surveys like APACHE (Sozzetti et al. 2013) and MEarth (Nutzman & Charbonneau 2008; Charbonneau et al. 2009), and direct imaging searches like PALMS (Bowler et al. 2012). Some of these surveys are directed at M dwarfs generally, but many are aimed at mid- to late-M dwarfs specifically. The Habitable-Zone Planet Finder, for example, is targeting M4-M9 dwarfs (Mahadevan et al. 2012).

Our knowledge of planet parameters are directly linked to our understanding of their host stars. Thus, these surveys require reliable stellar masses, radii, and metallic-

¹ Department of Astronomy, University of Texas at Austin

² Harlan J. Smith Fellow

³ Institute for Astronomy, University of Hawai‘i, 2680 Woodlawn Drive, Honolulu, HI 96822

⁴ Max Planck Institute for Astronomy, Konigstuhl 17, Heidelberg, 69117, Germany

⁵ Department of Geology & Geophysics, University of Hawai‘i, 1680 East-West Road, Honolulu, HI 96822

⁶ Department of Astronomy, Yale University, New Haven, CT 06511, USA

ities to properly characterize orbiting planets that are discovered. The *Gaia* spacecraft (de Bruijne 2012) is expected to measure parallaxes for the majority of the M dwarfs targeted by these surveys (Bailer-Jones et al. 2013), and distances can be used to derive luminosities and infer masses and radii (e.g., Delfosse et al. 2000; Bayless & Orosz 2006), but not metallicities. M dwarfs have sufficiently cool atmospheres to enable the formation of molecules with complex absorption bands. These bands are difficult to model but dominate the visible spectrum, making continuum identification difficult and creating line confusion. The result is that model-dependent methods such as spectral synthesis and curve of growth analysis that work well on solar-type stars are ineffective for M dwarfs (although improvements are ongoing, e.g., Önehag et al. 2012).

An alternative approach is to measure the metallicity of M dwarfs using empirical techniques. Such methods include measuring the position on a color-magnitude diagram (e.g., Schlafman & Laughlin 2010; Neves et al. 2012), the strength of molecular lines in the optical (e.g., Woolf & Wallerstein 2006; Dhital et al. 2012), or atomic lines in the optical or near-infrared (e.g., Rojas-Ayala et al. 2010; Mann et al. 2013a). Colors such as $g - r$ (e.g., West et al. 2004; Bochanski et al. 2013) or JHK (e.g., Johnson et al. 2012; Newton et al. 2014) can be used for predicting metallicity, but the errors are higher than other approaches and are subject to additional systematic errors (Mann et al. 2012). These methods are typically calibrated using wide binaries containing a solar-type primary (e.g., Bonfils et al. 2005). This assumes that wide binaries formed from the same molecular cloud and therefore have the same metallicity, which is well-established for solar-mass binaries (Desidera et al. 2004, 2006).

To date, calibrations of these empirical techniques have been created only for early and mid M-type dwarfs. The calibration from Mann et al. (2013a, henceforth M13) utilized the largest sample, but it contained only one M5, one M6, and nothing later. Newton et al. (2014) focused on cooler M dwarfs for the M_{Earth} survey, but included no M6 dwarfs, a single M7, and nothing cooler. As a result, the effectiveness of these calibrations for the coolest M dwarfs remains untested.

The faintness at optical wavelengths of the coolest M dwarfs makes them less likely to show up in long-time baseline proper motion surveys, because they generally rely on detections in the optical. Thus, until recently, it was difficult to locate wide, common proper motion pairs containing a late-type M dwarf. However, the problem has been mitigated significantly thanks to wide-field digital sky surveys such as the Sloan Digital Sky Survey (Aihara et al. 2011; West et al. 2011), the Two-Micron All-Sky Survey (2MASS Skrutskie et al. 2006), and the Panoramic Survey Telescope and Rapid Response System (PAN-STARRS Kaiser et al. 2010), which have provided proper motions and photometry for cooler and fainter objects than were previously accessible. Furthermore, now that methods to estimate the metallicity of early-M types are established, we can use pairs of early-M and late-M type dwarfs to extend the calibration to cooler temperatures.

In this paper we investigate methods to measure the

metallicities of M4.5–M9.5 dwarfs. We use 44 wide binaries containing an F, G, K, or early M dwarf primary and an M4.5–M9.5 companion. We determine metallicities for the primary stars by combining those from the literature with our own observations. Following the techniques of M13, we derive empirical calibrations between features in K -band spectra and the metallicity of ultracool dwarfs. In Section 2 we present our wide binary sample. In Section 3 we describe our observations of the primary and companion stars. In Section 4 we detail our calculations of metallicities for the primary stars and spectral classifications of the companions. We test how prior metallicity calibrations work on our ultracool dwarf sample in Section 5 then derive a new calibration in Section 6. We investigate the reliability of the calibration by employing a number of tests in Section 7. We conclude with a brief summary of our work in Section 8.

All wavelengths used in this work are stated as vacuum values.

2. SAMPLE

We constructed our wide binary sample from literature sources and from our own analysis of proper motion catalogs. Our sample contains M dwarfs as early as M4.5 for overlap with prior studies, but also includes stars out to the end of the M dwarf sequence at M9.5 (see Section 4.2 for details on spectral types assigned). We selected pairs with $-30^\circ < \delta < 68^\circ$ (easily reachable from Mauna Kea telescopes), companions with $K < 13$ (reasonable integration times), and pairs with primary-companion separation $> 5''$ (tighter binaries are difficult to observe and may have contaminated photometry). We also required that the primaries have spectral types later than F6, as stars earlier than this often have very few lines useful for metallicity determination. It is possible to determine the metallicities of M dwarfs as late as \simeq M5 using empirical techniques, although these methods are best tested for stars M4 and earlier, so we conservatively restrict our primary stars to F6 to M3.

In total we identified 61 pairs meeting the above criteria.

- 52 are benchmark systems previously identified in the literature (see Table 1 for list of references).
- Twelve pairs were taken from an ongoing search for ultracool and brown dwarf companions to *Hipparcos* stars using astrometry from the 2MASS and Pan-STARRS1 (Deacon et al. 2012, in review).
- Thirteen more targets were found by looking for co-moving pairs in Lépine & Shara (2005) and Lépine & Gaidos (2011, LG11) proper motion catalogs following the methods outlined in Lépine & Bongiorno (2007). Although this matching returned many targets already identified in the literature, this included seven early M + late-M pairs which were previously overlooked.
- A single final target was discovered in the finder while observing the primary (J19074+5905) for a separate project. We obtained a low-S/N spectrum of the companion, which enabled us to measure the spectral type and confirm the primary and companion have consistent distances based on the M_K -spectral type relation from Lépine et al. (2013).

Note that many targets are found in multiple sources.

We observed all 61 targets (Section 3). However, 17 of them were rejected from the final sample because: (1) the companion was earlier than M4.5 or an L dwarf, (2) the S/N of the observations was too low ($\lesssim 60$) to be useful, (3) the primary was too hot or cold to derive an accurate metallicity and had no reliable metallicity from the literature, or (4) the primary was a double-lined spectroscopic binary, which complicates the analysis (see Section 4.1 for more details).

In Figure 1 we show the distribution of spectral types for companion stars in this study as well as those from prior studies of M dwarf metallicities. Although there is significant overlap in the M4.5–M5.5 range, we have greatly expanded the number of companions with spectral types M6 and later. The full sample of binaries used for calibration is listed Table 1, including the spectral types (Section 4.2), primary star metallicities (Section 4.1) and references establishing the binarity of the pair. Because naming conventions for these stars vary throughout the literature, we also provide the coordinates of each target.

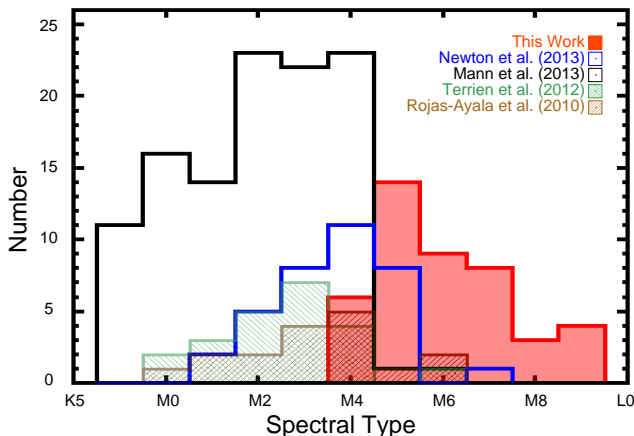


Figure 1. Distribution of spectral types covered by our calibration sample (red) and calibrators used in prior studies of M dwarf metallicities (Rojas-Ayala et al. 2012; Terrien et al. 2012; Mann et al. 2013a; Newton et al. 2014). We consider only reported spectral types, and do not account for systematic differences between spectral typing methods. Note that many sources contain overlapping targets.

3. OBSERVATIONS AND REDUCTION

3.1. *ESPaDOnS/CFHT*

We observed 15 F-, G-, and K-type primary stars with the Canada-France-Hawaii Telescope (CFHT) Echelle SpectroPolarimetric Device for the Observation of Stars (ESPaDOnS; Donati 2003). Observations were taken in queued service mode, in the star+sky setting on ESPaDOnS. This yielded a resolution of $\lambda/\Delta\lambda \simeq 65000$ and wavelength coverage from $0.37 \mu\text{m}$ to $1.05 \mu\text{m}$. Because of cosmic rays and atmospheric variations we did not use exposure times higher than 2400s. For fainter sources, we took multiple exposures and stacked them after reduction. All final spectra had $S/N > 100$ at $0.67 \mu\text{m}$, and typical S/N was > 150 per resolving element. The data were reduced automatically by the *Libre-ESpRIT*

pipeline described in Donati et al. (1997). Four of the 15 stars with ESPaDOnS spectra were later rejected because of complications with the primary (see Section 2 for more details).

3.2. *SpeX/IRTF*

We obtained near-infrared spectra of the 61 companions and 19 M dwarf primaries with the SpeX spectrograph (Rayner et al. 2003) attached to the NASA Infrared Telescope Facility (IRTF) on Mauna Kea. We used the cross-dispersed mode with the $0.3''$ slit. This provided simultaneous coverage from 0.8 to $2.4 \mu\text{m}$ at a resolution of $R \simeq 2000$. Targets were placed at two positions along the slit (A and B). We took exposures following an ABBA slit-nodding pattern, with at least six exposures per target. Although spectral features used for calculating metallicities (see Section 6) are free of telluric and OH lines, we still choose to minimize the effect of H_2O and other atmospheric variation by capping individual exposure times were at 120 s.

Resulting S/N in the *H* and *K* bands for all spectra was > 60 (typically > 90) for the companions, and > 100 (typically > 120) for the early-M dwarf primaries. To avoid effects of flexure in the optical path, we obtained flat-field and argon lamp calibration data at or near the same pointing as the target. We observed an A0V-type star within 1 hr and 0.1 airmasses of each target to remove telluric lines. The faintest targets took more than 1 hr of time (including overhead), and many targets moved by more than 0.1 airmasses over the course of an observation sequence. In these cases we took two A0V stars, sometimes slewing away from the target between exposures to observe one of the A0V stars, then slewing back to take additional target exposures.

We extracted and reduced spectra using the *Spextool* IDL package (Cushing et al. 2004). *Spextool* performed flat-field correction, wavelength calibration, sky subtraction, as well as extraction of the 1D spectrum. We stacked multiple exposures with the IDL routine *xcombspec* (part of *Spextool*). While running *xcombspec* we checked (by eye) to see if exposures were consistent with each other and to remove outliers. However, we removed only three images, and all were taken through thick clouds (> 2 magnitudes of extinction) and had relatively low S/N. After the spectra were stacked, we performed telluric corrections and flux calibration using the A0V stars with the *xtellcor* package (Vacca et al. 2003). Separate orders were combined with the *xmergeorders* IDL routine.

As a test, we tried performing telluric correction using different A0V stars for a given target taken in the same night. This generated small color differences amounting to $J-K_s < 0.04$, which is consistent with those found by Mann et al. (2013b). The change in overall shape is likely due to seeing changes between the target and standard star (Rayner et al. 2009). Such color terms were only significant when measured across the whole *JHK* spectrum, and thus were unlikely to effect our results. Within each echellette order, we found spectra were consistent within errors, provided the A0V standard was taken within 0.15 airmasses.

Reduced spectra were put in vacuum wavelengths using the formula from Ciddor (1996). We put spectra in the stars' rest frame by cross-correlating them with the

Table 1

Companion				Primary				
Name	R.A.	δ	SpT	Name	SpT	[Fe/H] ^a	[Fe/H] Ref ^b	Binary Ref ^c
HIP 70623 B	14 ^h 26 ^m 45 ^s .74	-5°10'20"	M4.5	HIP 70623	K0V	+0.56 ± 0.03	SPOCS	15,18
LSPM J0212+1249W	2 ^h 12 ^m 19 ^s .76	+12°49'25"	M4.5	NLTT 7300	M3V	+0.46 ± 0.10	SpeX	17
HIP 98535 B	20 ^h 1 ^m 2 ^s .17	+48°16'27"	M4.5	HIP 98535	G5V	-0.18 ± 0.03	SME	15,18
2M 1743+2136	17 ^h 43 ^m 15 ^s .32	+21°36'4"	M4.5	HIP 86722	K0V	-0.39 ± 0.05	F08	11,15
LSPM J2047+1051N	20 ^h 47 ^m 16 ^s .75	+10°51'45"	M4.5	HIP 102582	K2V	-0.53 ± 0.03	SME	15
NLTT 15867	5 ^h 58 ^m 17 ^s .18	-4°38'1"	M4.5	HIP 28267	G7V	-0.10 ± 0.03	SPOCS	5
PM I14254+2035	14 ^h 25 ^m 25 ^s .89	+20°35'45"	M5.0	HIP 70520	F9V	-0.57 ± 0.05	R07	9
NLTT 8870	2 ^h 45 ^m 41 ^s .23	+44°57'2"	M5.0	HIP 12886	M1V	+0.11 ± 0.10	SpeX	5
HIP 114424 B	23 ^h 10 ^m 22 ^s .08	-7°48'54"	M5.0	HIP 114424	K0V	+0.10 ± 0.03	SPOCS	15
HIP 114456 B	23 ^h 10 ^m 54 ^s .78	+45°30'43"	M5.0	HIP 114456	K0V	+0.21 ± 0.03	SPOCS	15,18
LSPM J0253+6321	2 ^h 53 ^m 15 ^s .55	+63°21'6"	M5.0	HIP 13394	G0V	-0.14 ± 0.08	C11	9
LSPM J1841+2447N	18 ^h 41 ^m 9 ^s .81	+24°47'19"	M5.0	GJ 1230A	M3V	+0.18 ± 0.10	SpeX	17
HIP 106551 B	21 ^h 34 ^m 45 ^s .17	+38°31'0"	M5.0	HIP 106551	K1III	+0.05 ± 0.06	C01	15
LSPM J0932+2659E	9 ^h 32 ^m 48 ^s .25	+26°59'43"	M5.5	HIP 46843	G9V	-0.09 ± 0.05	F08	4,5,9,13
LSPM J1659+0635	16 ^h 59 ^m 5 ^s .58	+6°35'32"	M5.5	HIP 83120	K0V	+0.25 ± 0.03	SME	9,5
LSPM J1207+1302	12 ^h 7 ^m 24 ^s .01	+13°2'13"	M5.5	HIP 59126	K0V	-0.02 ± 0.03	SME	4,5,9
PM I19074+5905 B	19 ^h 7 ^m 24 ^s .83	+59°5'9"	M5.5	I19074+5905	M2V	+0.30 ± 0.10	SpeX	16
I10005+2717	10 ^h 0 ^m 35 ^s .71	+27°17'6"	M5.5	HIP 49046	M1V	+0.26 ± 0.10	SpeX	13,15,18
LSPM J1748+1143	17 ^h 48 ^m 44 ^s .32	+11°43'47"	M5.5	HIP 87182	K4V	+0.02 ± 0.03	SME	9
LSPM J0731+1958	7 ^h 31 ^m 38 ^s .88	+19°58'32"	M5.5	HIP 36607	K0V	+0.05 ± 0.03	SME	4,5,9,13
LSPM J1302+3227	13 ^h 2 ^m 20 ^s .81	+32°27'10"	M5.5	HIP 63636	G8IV	+0.05 ± 0.04	T05	9
LSPM J1124+2330E	11 ^h 24 ^m 40 ^s .17	+23°30'57"	M5.5	NLTT 27298	M2V	+0.08 ± 0.10	SpeX	17
LSPM J2049+3216W	20 ^h 49 ^m 13 ^s .75	+32°16'51"	M6.0	HIP 102766	K2V	-0.02 ± 0.03	SME	5,9
PM I10008+3155	10 ^h 0 ^m 50 ^s .19	+31°55'44"	M6.0	HIP 49081	G3V	+0.20 ± 0.03	SPOCS	14
NLTT19472	8 ^h 24 ^m 52 ^s .44	-3°41'1"	M6.0	HIP 41211	F8V	-0.28 ± 0.08	C11	5
NLTT28453	11 ^h 45 ^m 35 ^s .39	-20°21'4"	M6.0	HIP 57361	M2V	-0.05 ± 0.10	SpeX	5
LSPM J1210+1858E	12 ^h 10 ^m 9 ^s .79	+18°58'7"	M6.5	HIP 59310	K3V	+0.30 ± 0.03	SME	13,15,18
HIP 81910 B	16 ^h 43 ^m 49 ^s .50	-26°48'40"	M6.5	HIP 81910	G3V	+0.24 ± 0.03	SPOCS	15
LSPM J0942+2351	9 ^h 42 ^m 57 ^s .18	+23°51'19"	M6.5	NLTT 22411	M1V	+0.05 ± 0.10	SpeX	17
PM I11055+4331	11 ^h 5 ^m 30 ^s .90	+43°31'17"	M6.5	HIP 54211	M2V	-0.32 ± 0.10	SpeX	5
2M 0318+0828	3 ^h 18 ^m 42 ^s .14	+8°28'0"	M7.0	NLTT 10534	M2V	+0.19 ± 0.10	SpeX	12
PM I16555-0823	16 ^h 55 ^m 35 ^s .29	-8°23'40"	M7.0	HIP 82817	M3V	-0.08 ± 0.10	SpeX	13
2M 1320+0957	13 ^h 20 ^m 41 ^s .59	+9°57'50"	M7.0	HIP 65133	K4V	+0.07 ± 0.03	SME	10
NLTT 36549	14 ^h 12 ^m 12 ^s .13	-0°35'16"	M7.5	NLTT 36548	M3V	-0.26 ± 0.10	SpeX	4
2M 1200+2048	12 ^h 0 ^m 32 ^s .92	+20°48'51"	M7.5	G 121-42	M2V	-0.15 ± 0.10	SpeX	10
GJ 569B	14 ^h 54 ^m 29 ^s .36	+16°6'8"	M7.5	HIP 72944	M2V	-0.08 ± 0.10	SpeX	3
2M 1916+0509	19 ^h 16 ^m 57 ^s .60	+5°9'1"	M7.5	HIP 94761	M2V	+0.11 ± 0.10	SpeX	1
2M 2331-0406	23 ^h 31 ^m 1 ^s .64	-4°6'19"	M8.0	HIP 116106	F8V	-0.26 ± 0.03	SPOCS	7
2M 0003-2822	0 ^h 3 ^m 42 ^s .28	-28°22'41"	M8.0	HIP 296	G8V	+0.31 ± 0.03	SME	3,7
2M 0430-0849	4 ^h 30 ^m 51 ^s .57	-8°49'0"	M8.5	LP 655-23	M3V	+0.03 ± 0.10	SpeX	7
HIP 78184 B	15 ^h 57 ^m 55 ^s .32	+59°14'25"	M9.0	HIP 78184	M0V	+0.08 ± 0.10	SpeX	6,15,18
2M 2010+0634	20 ^h 10 ^m 35 ^s .39	+6°34'36"	M9.0	NLTT 48838	M3V	-0.01 ± 0.10	SpeX	12
2M 0739+1305	7 ^h 39 ^m 43 ^s .85	+13°5'6"	M9.0	BD+13 1727	K2V	+0.15 ± 0.03	SME	5
2M 2237+3922	22 ^h 37 ^m 32 ^s .55	+39°22'39"	M9.5	HIP 111685	M1V	+0.03 ± 0.10	SpeX	2

^a [Fe/H] values shown here include our applied corrections (see Section 4).

^b Metallicity References – C01 = Cayrel de Strobel et al. (2001), SPOCS = Valenti & Fischer (2005), T05 = Takeda et al. (2005), R07 = Ramírez et al. (2007), F08 = Fuhrmann (2008), C11 = Casagrande et al. (2011), SME = Spectroscopy Made Easy analysis of CFHT/ESPaDONs data, SpeX = empirical calibrations of M13 applied to IRTF/SpeX data

^c Binary Reference – 1 = van Biesbroeck (1944), 2 = Kirkpatrick et al. (2001), 3 = Martín et al. (2000), 4 = Chanamé & Gould (2004), 5 = Gould & Chanamé (2004), 6 = Pinfield et al. (2006), 7 = Caballero (2007), 8 = Cruz et al. (2007), 9 = Lépine & Bongiorno (2007), 10 = Faherty et al. (2010), 11 = Allen et al. (2012), 12 = Luhman et al. (2012), 13 = Tokovinin & Lépine (2012), 14 = Mann et al. (2013a), 15 = Deacon et al. (in review), 16 = SpeX Finder, 17 = Lépine & Shara (2005) + Lépine & Gaidos (2011), 18 = Washington Double Star Catalog.

spectra of template stars from the IRTF spectral library (Cushing et al. 2005; Rayner et al. 2009). For stars of M4.5 to M6.5 we used the M5V template Gl 51, and for later-type stars we used the M9V template LHS 2065. As a test we tried cross-correlating our spectra with different templates from the IRTF library and found that differences in the derived radial velocity offset were small (~ 1 resolving element) provided the template was within $\simeq 3$ spectral subtypes of the target star.

Newton et al. (2014) noted that for high S/N spectra `Spextool` underestimates the error, because of the presence of correlated noise. However, Newton et al. (2014) targets are significantly brighter and have higher S/N than our targets (S/N > 200 versus > 90). At S/N > 200 Poisson errors are almost negligible, but are likely the dominant source of noise for our spectra. As a test, we took the individual 1D spectra (just prior to stacking with `xcombpec`) of 15 random stars from our sample and then re-stacked them into two spectra of each star (using just half the unstacked spectra in each case). We did the same for the corresponding A0V star. We find that differences in equivalent widths between the stacks of the same star are consistent within Poisson or photon noise-based errors except the five stars where the S/N was $\gg 150$. For these targets the errors are scaled according to the scatter in the stacking.

Reduced spectra of all companions are included with this manuscript.

4. ANALYSIS

4.1. Primary Star Metallicities

As with M13, we drew metallicities for the primaries both from the literature and from our own observations. Seven primaries have metallicities in the Spectroscopic Properties of Cool Stars catalog (SPOCS, Valenti & Fischer 2005), which is based on analysis of high-resolution spectra with the Spectroscopy Made Easy (SME, Valenti & Piskunov 1996) software package. Another seven primaries have metallicities from various literature sources, most of which make use of high-resolution spectra and MOOG (Snedden 1973).

Different literature sources use slightly different techniques and thus may have small systematic inconsistencies. As in M13, we corrected for this by using stars common to both the given literature source and the SPOCS sample. Literature sources with less than 30 stars of overlap with SPOCS were not utilized. Generally the overlap sample is $\ll 100$ stars, and the corrections are $\ll 0.1$ dex. This method enabled us to put all metallicities on the same scale (in this case the SPOCS scale) and estimate the error from residual scatter after applying the correction. More details on the corrections, number of overlapping stars, and derived errors can be found in M13.

We observed 15 of the FGK primaries with CFHT/ESPaDOnS, although four of these stars were removed because of complications with their primary (see Section 2). To determine stellar properties for these stars we modeled each spectrum with the SME software (Valenti & Piskunov 1996), fitting to the set of lines tuned for the SPOCS catalog (Valenti & Fischer 2005). We simultaneously solved for surface gravity, effective temperature, projected rotational velocity, and individ-

ual abundances of Na, Si, Ti, Fe, and Ni as in the SPOCS analysis. Solar values were assumed for all of the initial models and after obtaining an initial fit, we perturbed T_{eff} by $\pm 100\text{K}$ and fit again. Corrections based on Vesta and stellar binary observations as detailed in Valenti & Fischer (2005) were then applied. The SME-determined $[\text{Si}/\text{Fe}]$ was used as a proxy for alpha-element enhancement.

Torres et al. (2012) showed that different analysis methods of high-resolution spectra can produce systematically different results for the same stars and that fitting for all parameters simultaneously can create strong correlations between $[\text{Fe}/\text{H}]$, $\log g$, and T_{eff} . We mitigated this effect by utilizing *Hipparcos* parallaxes (van Leeuwen & Fantino 2005; van Leeuwen 2007, where available) and the Yonsei–Yale evolutionary models (Demarque et al. 2004) to independently constrain $\log g$, following the method from Valenti et al. (2009). We first used the distance and color to derive a bolometric luminosity. We combined this with the T_{eff} , $[\text{Fe}/\text{H}]$, and $[\text{Si}/\text{Fe}]$ from an initial fit using SME, which we interpolated onto the Yonsei–Yale grid to get $\log g$. This $\log g$ was compared to the value determined by SME, and if the two did not match, the SME analysis was run again with the gravity fixed to the isochrone value. The process was repeated until the $\log g$ values agreed. The iterative process did not converge for a single star (HIP 102582), which is probably due to an erroneous parallax, color, unresolved companion (tertiary), and/or the relatively low temperature of this star ($T_{\text{eff}} \simeq 4600\text{K}$). However, because the $[\text{Fe}/\text{H}]$ did not significantly change during the iteration, the metallicity for this object is likely reliable and included in our sample. Final stellar parameters (T_{eff} , $\log g$, $[\text{M}/\text{H}]$, etc.) for stars observed with ESPaDOnS are listed in Table 2.

The remaining 19 primaries are late-K or early-M dwarfs. For these targets we calculated their metallicities by applying the empirical methods from M13 to SpeX spectra. M13 provides empirical calibrations between observed atomic and molecular line strengths and the metallicities of stars from K5 to M5 for visible, J –, H –, and K –band spectra. We calculated the weighted mean of the H – and K –band metallicities accounting for measurement (mostly Poisson) and calibration errors. For stars with visible wavelength spectra available from Lépine et al. (2013) we included the visible calibration metallicities. Calibration errors (typically 0.08 dex) were *not* assumed to be uncorrelated because they are based on the same underlying sample and technique. Thus the calibration errors represent the error floor on the metallicities of these stars. Adopted metallicities for these stars are reported in Table 1.

4.2. Spectral Types

We used a custom by-eye matching routine to determine the spectral type of each companion or M dwarf primary. The routine uses NIR spectra and is based on the by-eye matching routine from the HAMMER spectral typing suite (Covey et al. 2007). Our routine separately displays the normalized J –, H –, and K –band spectrum from the target alongside a NIR template spectrum from Rayner et al. (2003) or Cushing et al. (2005). An initial guess template is shown based on a χ^2 comparison of the target and NIR templates. The user is allowed

Table 2
Parameters of Primary Stars Observed at CFHT

Name	$T_{\text{eff}}^{\text{a}}$	$\log g^{\text{a}}$	[Fe/H] ^a	[M/H] $\pm \sigma$	[Na/H] $\pm \sigma$	Run Type ^b
HIP 296	5561	4.50	+0.31	+0.24 \pm 0.03	+0.32 \pm 0.03	ITER
HIP 36607	5077	4.55	+0.05	+0.02 \pm 0.07	+0.10 \pm 0.04	ITER
BD+13 1727	5098	4.70	+0.15	+0.10 \pm 0.04	+0.11 \pm 0.03	SME VESTA
HIP 59126	4862	4.61	-0.02	-0.05 \pm 0.04	-0.00 \pm 0.04	ITER
HIP 59310	4739	4.60	+0.30	+0.25 \pm 0.03	+0.43 \pm 0.04	ITER
HIP 65133	4604	4.64	+0.07	+0.00 \pm 0.04	-0.03 \pm 0.05	ITER
HIP 83120	5072	4.51	+0.25	+0.23 \pm 0.05	+0.33 \pm 0.05	ITER
HIP 87182	4687	4.63	+0.02	-0.02 \pm 0.03	-0.04 \pm 0.03	ITER
HIP 98535	5181	3.89	-0.18	-0.17 \pm 0.03	-0.23 \pm 0.03	ITER
HIP 102766	4988	4.60	-0.02	-0.04 \pm 0.05	-0.05 \pm 0.03	ITER
HIP 102582 ^c	4574	4.72	-0.53	SME VESTA

^a Errors on T_{eff} , $\log g$, and [Fe/H] stars are 44 K, 0.06 dex, and 0.03 dex for all stars.

^b ITER: parameters determined using *Hipparcos* parallaxes and Y^2 isochrones. VESTA: parameters determined using classical SME fitting (no parallax information included) with a correction using Vesta as described in Valenti & Fischer (2005).

^c The fit for HIP 102582 was run in ITER mode, but failed to converge. The initial fit suggests that the derived [Fe/H] is reliable, but other derived parameters are discarded for this star.

to switch templates (both spectral type and luminosity class) manually to get a better by-eye match. Half subtype templates were constructed (if they are not already included) from normalized, linear combinations of spectral pairs (e.g., an M5.5 is constructed by adding together an M5 and an M6). We also repeated this method using the entire *JHK* spectrum as a crosscheck, and found differences were ≤ 1 subtype in all cases.

To test the reliability of our spectral-typing method we analyzed a sample of stars with both visible and NIR data from Reid et al. (1995), West et al. (2011), Lépine et al. (2013), or this program. Spectral types for these targets were determined from their visible-wavelength spectra (based on the system of Kirkpatrick et al. 1991). We found no significant systematic offset between the optical spectral types and those determined from our by-eye matching of NIR spectra. Based on the scatter we estimated the errors on assigned spectral types to be $\simeq 0.5$ for M4.5–M7.5 and $\simeq 0.8$ for M7.5–M9. The higher errors for the latest spectral types may be in part due to systematic discrepancies between assigned spectral types in the optical from different surveys.

Rojas-Ayala et al. (2012) and Newton et al. (2014) presented relations between the empirical H₂O-K2 index and the spectral type of the star. The Newton et al. (2014) is more relevant, as it includes more late-type stars. Interestingly we found that the Newton et al. (2014) relation predicts spectral types systematically 0.5 subtypes later than those from our by-eye analysis. Since the spectral types from the H₂O-K2 index are based on the continuum shape, while our matching is based on the more traditional method of matching indices (albeit by eye), we only report our spectral types. This also kept the spectral types more consistent with the M13 study, which are based on optical spectra.

5. APPLICABILITY OF PRIOR CALIBRATIONS TO LATE-M DWARFS

We examined the performance of previous M dwarf metallicity calibrations using our ultracool dwarf companion sample. Our sample covers a different range of

spectral types than those used for previous calibrations, which focused on early- to mid-M dwarfs. The goal is to determine how these calibrations work (or fail) for the latest M dwarfs.

We followed the procedures for measuring metallicities given in Terrien et al. (2012), M13, and Newton et al. (2014) to determine the metallicities of each companion. Each of these calibrations uses SpeX spectra to measure the equivalent widths of strong lines in the *H*- or *K*-band, although the set of lines varies. Thus applying their methods required using different feature (wavelength) definitions, as well as different procedures to estimate the (pseudo-)continuum for a given feature. The calibration of Rojas-Ayala et al. (2010, 2012) based on the TripleSpec spectrograph was not tested, because there are small but significant systematic offsets between equivalent widths using the different instruments (Newton et al. 2014) and because the targets from Rojas-Ayala et al. (2010) were already folded into the Newton et al. (2014) analysis.

We show the derived companion metallicity using each of the literature calibrations versus the primary star metallicity in Figure 2. For comparison, we calculate the adjusted coefficient of determination (R_{ap}^2) for each relation applied to our wide binary sample. R_{ap}^2 is defined as:

$$R_{\text{ap}}^2 = 1 - \frac{(n-1) \sum (y_{i,\text{model}} - y_i)^2}{(n-p) \sum (y_i - \bar{y})^2}, \quad (1)$$

where p is the number of changeable parameters, n is the number of data points in the fit, y_i primary star metallicity of the i th star, $y_{i,\text{model}}$ is the metallicity of the i th star predicted by the fit, and \bar{y} is the average of y . A R_{ap}^2 closer to 1 implies that the model accurately explains the variance of the sample whereas $R_{\text{ap}}^2=0$ implies that it can explain none. We report R_{ap}^2 and the standard deviation (σ) for each relation in Table 3.

All prior calibrations show significant systematics with spectral type, resulting in inaccurate metallicities for the latest-type dwarfs. This is expected, since these cali-

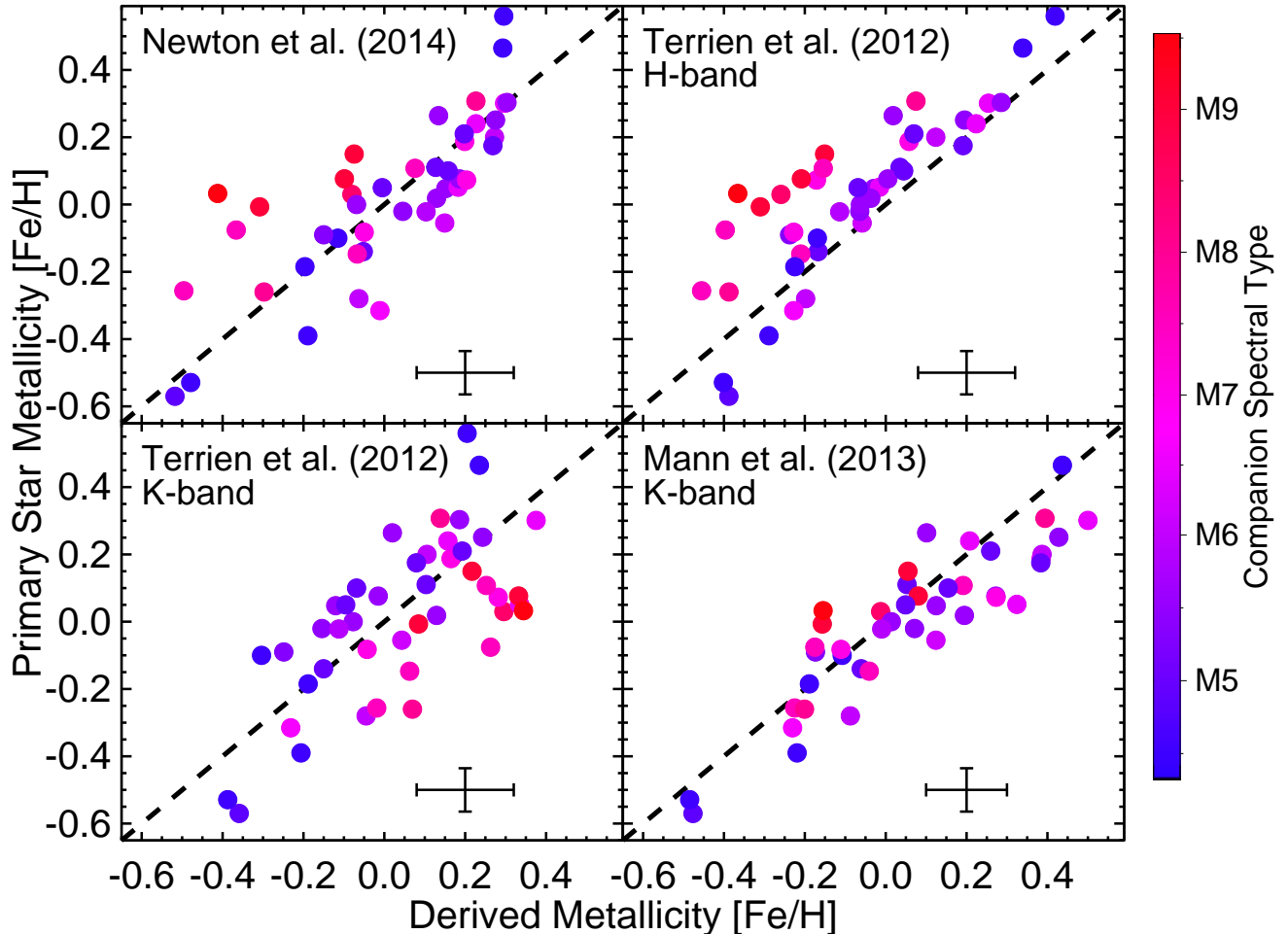


Figure 2. Primary star (system) metallicity as a function of the metallicity determined for our ultracool companion sample based on the calibrations from Newton et al. (2014), Terrien et al. (2012) *H*-band, Terrien et al. (2012) *K*-band, and Mann et al. (2013a) *K*-band. Data points are colored by spectral type. The error bar in the bottom right denotes a typical (median) error from the primary stars and the error in the calibration reported from the relevant reference. The dashed line has slope unity and is added for reference.

Table 3
Tests of Each Calibration on Mid to Late-M Dwarf Sample

Reference	Band	R_{ap}^2	σ
Newton et al. (2014)	<i>K</i>	0.56	0.15
Terrien et al. (2012)	<i>H</i>	0.47	0.13
Terrien et al. (2012)	<i>K</i>	0.37	0.18
Mann et al. (2013a)	<i>K</i>	0.62	0.12
This work	<i>K</i>	0.89	0.07

brations were based almost entirely on stars \simeq M5 and earlier. If we remove the stars later than M6 from the sample, all calibrations yield results (as determined by R_{ap}^2) consistent with those reported in the respective paper.

The M13 calibration shows the least systematics with spectral type, in that this calibration accurately predicts the metallicity of M7-M9 dwarfs but on average underestimates the metallicities of the whole sample. The calibration from Newton et al. (2014) performs reasonably well on our sample, most likely because their sample in-

cludes more mid-M dwarfs than other analyses. However Newton et al. (2014) assigns incorrect metallicities for stars with $[\text{Fe}/\text{H}] > +0.3$ (already noted by Newton et al. 2014) and underestimates the metallicities of stars M7 and later. The two calibrations from Terrien et al. (2012) have a similar issue with M7 and later stars, although the *K*-band calibration under- rather than overestimates the metallicity of the latest M dwarfs.

6. MEASURING THE METALLICITIES OF LATE-M DWARFS

The random and systematic errors present when we applied prior calibrations to the ultracool dwarf sample motivated a new metallicity calibration tuned for late-M dwarfs. To do this, we first used the list of metal-sensitive spectroscopic features and pseudo-continuum definitions from M13. We followed the method of M13 to determine which combination of features from this list gave the best calibration (as determined by χ^2_p) for each wavelength regime (*J*, *H*, and *K* band). We fit for calibrations of the form:

$$[\text{Fe}/\text{H}] = \sum_n (A_n(F_n) + B_n(F_n)^2) + \quad (2)$$

Table 4
Spectral features used

Name	Feature (μm)	Blue continuum (μm)	Red continuum (μm)
Na I	2.2045 – 2.2113	2.1940 – 2.1985	2.2130 – 2.2190
Ca I	2.2610 – 2.2670	2.2450 – 2.2520	2.2717 – 2.2781

$$C(\text{H}_2\text{O} - \text{K2}) + D,$$

where $(\text{H}_2\text{O} - \text{K2})$ is a temperature-sensitive H_2O index defined by Rojas-Ayala et al. (2012), F_n is the equivalent width of the n -th feature from M13, $[\text{Fe}/\text{H}]$ is the metallicity of the system (derived from the primary star), and the other variables (A , B , C , D) were determined by Levenberg-Marquart least-squares minimization (Markwardt 2009).

Starting with one feature and no square term ($B_n = 0$ and $n=1$) we tried all features from M13 for a given wavelength regime. After the best single feature was found we tried adding an additional feature. The total number of included features (n) was determined by an F -test, which measures whether the coefficient for the new term is consistent with zero. Additional terms are only added if the probability that the new coefficient is significant exceeds 95.5% (2σ). The same test was applied to determine if squared terms should be included. Higher-order (≥ 3 rd) terms were not explored, but as we explain below, even second-order terms were not required.

Errors in equivalent widths and the H_2O -K2 index were calculated via Monte Carlo (MC). Noise was added to the spectrum equal to the measurement error computed by `Spectool`. Equivalent widths were then recalculated on the perturbed spectrum. This process was repeated 1000 times, and the error in a given F_n and H_2O -K2 index was taken as the standard deviation of these 1000 values.

We found the best fit in the K -band required just the Na I and Ca I lines:

$$[\text{Fe}/\text{H}] = 0.131(\text{EW}_{\text{Na}}) + 0.210(\text{EW}_{\text{Ca}}) - 3.07(\text{H}_2\text{O} - \text{K2}) + 1.341, \quad (3)$$

where equivalent widths (EW_{Na} and EW_{Ca}) are given in \AA^7 . We show the binary system metallicities as a function of the derived metallicity in Figure 3. Equation 3 yielded a scatter (σ) of 0.07 dex, a R_{ap}^2 of 0.89, and a χ_ν^2 of 2.1 ($\nu = 39$). The quality of the calibration is similar to that of M13 for K5–M5 dwarfs, and significantly better than applying previous calibrations developed for early–mid M dwarfs (Section 5). Feature and continuum regions for all measured spectral lines are identical to those in M13, but the two main features used in this work (Ca I and Na I) are also listed in Table 4.

We attempted to find a metallicity relation useful for J - and H -band spectra following an identical prescription for the K -band. However, we found that calibrations with comparable performance (similar χ_ν^2 , σ , and R_{ap}^2) to Equation 3 required using three or more features and still showed significant systematics with metallicity. We were able to find a formula with a relatively low σ (0.12 dex), but such calibrations systematically underes-

timated the metallicity of the most metal-rich stars and systematically overestimated the metallicity of the most metal-poor stars. The primary cause is that the most effective features for measuring metallicity for K5–M5 become weak and difficult to measure past M4 (Figure 4). The situation is even worse in the J -band where there were similar issues measuring features but we were also plagued by much lower S/N.

Newton et al. (2014) found that the best K -band calibration was achieved using just the Na I index, including a square term, and without the H_2O -K2 index. In contrast, we found that the inclusion of any squared term (B_n) was not justified by an F -test, but that the inclusion of the H_2O -K2 index was well justified by the same test. The Newton et al. (2014) sample is calibrated on a relatively small range of spectral types (mostly M3–M5), thus it was probably not necessary to include the H_2O -K2 index, which was designed to help adjust for spectral changes with T_{eff} (Rojas-Ayala et al. 2012). Like Newton et al. (2014) we find a second-order term is justified if we use just one feature (just Na I). However, we find we get better results using Na I and Ca I than using higher order terms.

M13 found a best fit relation using multiple Na I and CO bands in the K -band, although they also identified the Ca I triplet as a strong metallicity indicator. However, the Na I line at $\simeq 2.335 \mu\text{m}$ weakens past M5, while the Na I doublet at $2.208 \mu\text{m}$ is only mildly sensitive to spectral type (Figure 5). Here we found that the CO bands past $2.28 \mu\text{m}$, despite being very strong, are poor indicators of metallicity for M5–M9, and thus are not useful in this calibration. The Ca I triplet at $2.265 \mu\text{m}$ does become weaker for the coolest stars, but is still measurable even in the M9 dwarfs (Figure 5) and remains a reliable predictor of metallicity based on our analysis.

Interestingly, the Na I and Ca I lines used in our calibration are the same two features identified by Rojas-Ayala et al. (2010) to determine the metallicity of early M dwarfs. Because our method of finding the best calibration considers all metal-sensitive features identified by M13 and has no preference for these particular lines, this is a strong verification for the power of these atomic lines to measure the metallicity of M dwarfs.

7. THE ROBUSTNESS OF OUR CALIBRATION

7.1. Chance Calibration Probability

Feeding a large number (30) of features to Equation 2 increases the possibility of getting a reasonable calibration simply by chance, since each feature adds several potential degrees of freedom. Finding a calibration of quality similar to that of Equation 3 is unlikely. However, we want to better quantify *how unlikely* this is.

As a test, we reassigned the metallicities of each primary star randomly to another binary in the sample. We then reapplied our method as described above: feeding in metal-sensitive features from M13 into Equation 2, finding the fit by least squares, and adding in more features until the change in χ_ν^2 is minimal. We then repeated this process 10,000 times, each time with re-randomized metallicities (although binaries are forbidden from having their original metallicity) and recording the R_{ap}^2 and rms values of the final fit. The rms is often a poor metric when comparing multiple fits, because it

⁷ An IDL program for applying Equation 3 can be found at <http://github.com/awmann/metal>

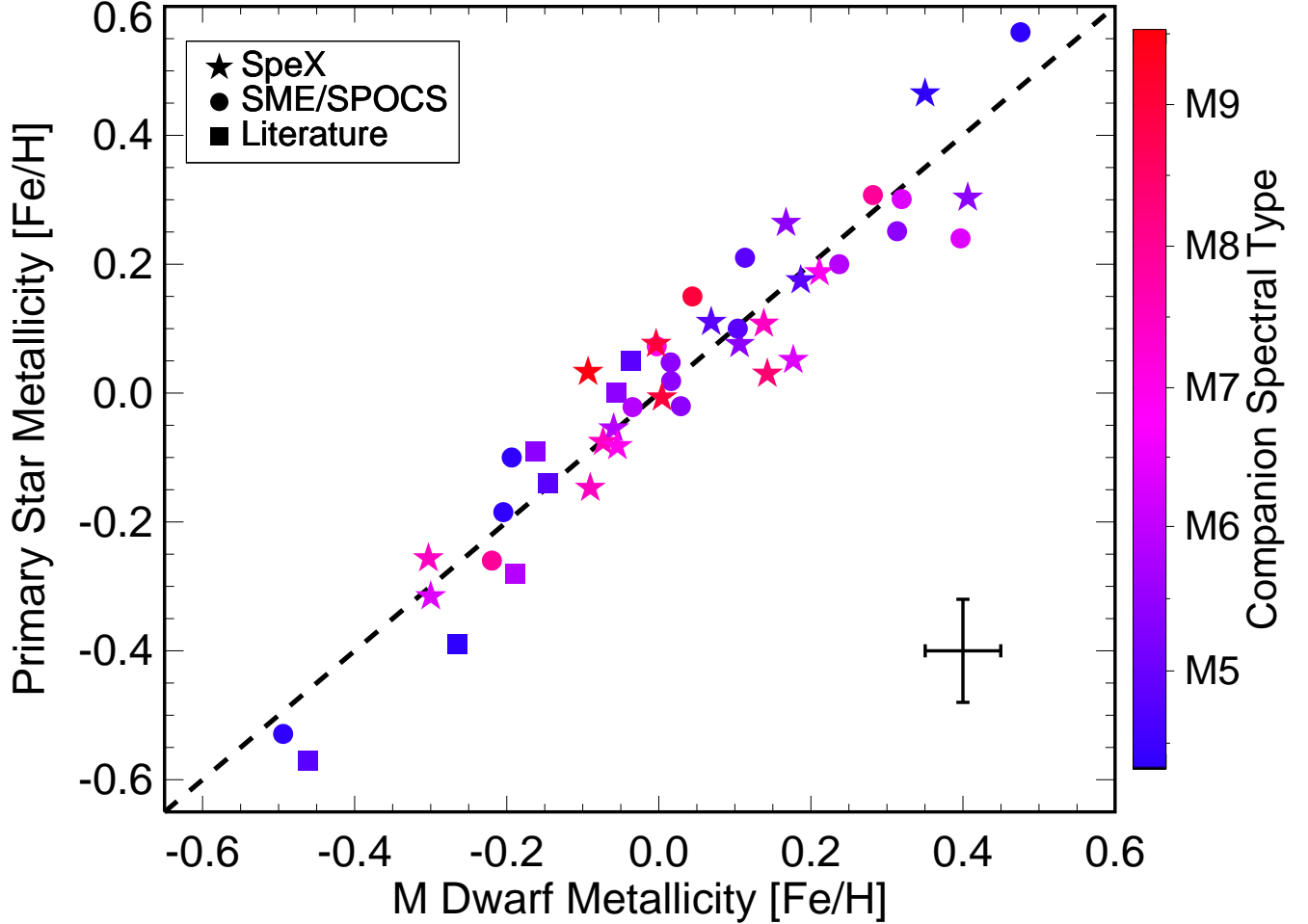


Figure 3. Primary star (system) metallicity as a function of the metallicity derived for the ultracool companion sample based on Equation 3. Points are colored according to spectral type, and symbols indicate the source for the primary star metallicity. The error bar in the bottom right denotes the median error on primary star metallicity (y-axis) and the median measurement (mostly Poisson) error from estimating the M dwarf metallicity. The calibration error (≈ 0.07 dex) should be considered separately and is not shown.

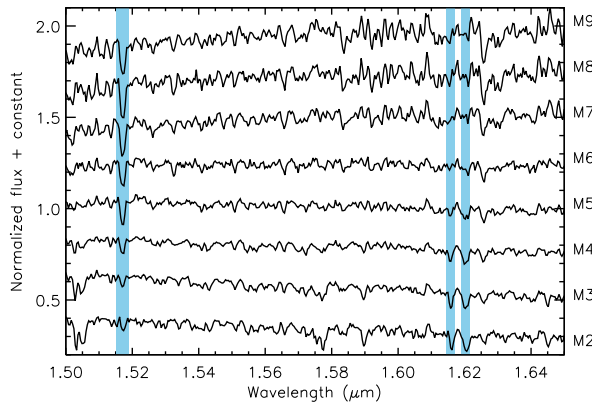


Figure 4. *H*-band spectra of M dwarfs with a range of spectral types. The metal-sensitive features identified by Terrien et al. (2012) are shown in teal. The Ca I lines near $1.62 \mu\text{m}$ are strong and easily measured out to M4, but become impractical weak for later spectral types. The K I line blueward of $1.52 \mu\text{m}$ is easy to measure for the full range, but has a strong spectral type dependence that is difficult to remove.

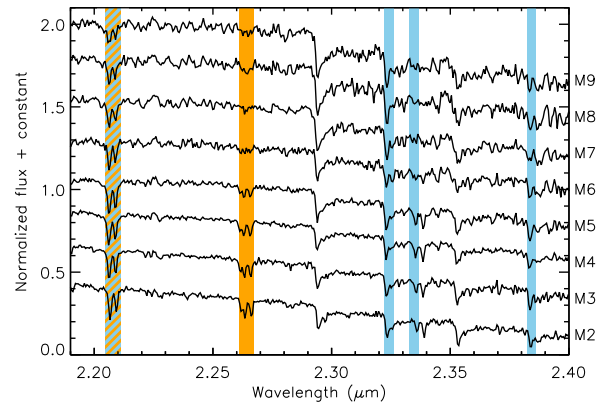


Figure 5. *K*-band spectra of M dwarfs with a range of spectral types. Features used in Equation 3 are shown in orange, with features used in M13 marked in teal. The labeled features at $2.345 \mu\text{m}$ and $2.385 \mu\text{m}$ are useful metallicity indicator to $\approx M5$, after which they becomes difficult to measure.

is sensitive to the underlying distribution. For example, if our primary star metallicities were clustered around $-0.2 < [\text{Fe}/\text{H}] < +0.2$, a calibration that assigns all stars

solar metallicity would have an rms $\ll 0.2$, even though such a calibration would not be useful. This is why we reassigned, rather than randomized metallicities: to preserve the sample’s metallicity distribution.

We found that the 99.7% highest (3σ) R_{ap}^2 value amongst our random sample is 0.41, which is less than half of that from our fit. Similarly, the 99.7% lowest rms is 0.19, more than twice as high as the rms of Equation 3. Thus the probability of getting an R_{ap}^2 of 0.89 and an rms of 0.07 is $\ll 0.1\%$, demonstrating that our method is statistically significant despite the introduction of a large number of variables into the fit.

7.2. Systematics

We searched for systematic issues in our calibration by comparing the fit residuals with spectral type, metallicity, and source of metallicity for the primary using a Spearman rank test. For all three cases we found no statistically significant correlation ($P=0.31, 0.13, 0.06$, respectively), suggesting that our calibration is robust over the range of metallicities and spectral types covered in our sample. This was slightly complicated by the distribution of points for these three parameters. For example, our sample includes only four dwarfs M9 or later, all of which have near solar metallicity ($-0.05 < [\text{Fe}/\text{H}] < +0.20$).

To test the limits of our calibration we applied Equation 3 to two additional stars outside the range of our calibrators. HIP 114962B is an M3.5 subdwarf companion to an F8 subgiant, and GJ 1048B is an early-L dwarf companion to a K2 dwarf (Gizis et al. 2001). HIP 114962 has $[\text{Fe}/\text{H}] = -1.40 \pm 0.08$ (Cayrel de Strobel et al. 2001; Casagrande et al. 2011; Lee et al. 2011), and GJ 1048 has $[\text{Fe}/\text{H}] = +0.06 \pm 0.03$ (Adibekyan et al. 2012). These two pairs were not included in the initial calibration sample because they are outside the range of spectral types considered (see Section 2). However, they are still useful tests because if our calibration fails just outside the range of companion star spectral types it suggests a problem.

Applying Equation 3 to spectra of these two companions yielded metallicities of $[\text{Fe}/\text{H}] = -1.26$ and 0.13 for HIP 114962B and GJ 1048B, respectively. Accounting for measurement and calibration errors, the differences between derived and primary star metallicities were 1.2σ and 0.7σ , respectively. The agreement suggests that the calibration may be effective slightly outside the range of spectral types of the calibrators.

7.3. Unresolved Binaries

Approximately 45% of wide M dwarf binaries contain at least one more star (Law et al. 2010), typically a close (< 30 AU separation) companion to one of the components (or a close companion to each of the components for quadruple systems). At the median distance to our targets (32 pc) any such close-in companion would be unresolvable (separations $\lesssim 1$ arc second) in the SpeX slit-viewing camera. Unresolved companions to the primary star can be identified as a spectroscopic binary in our ES-PaDOnS spectra. However, for both M dwarf primaries and all companions, the spectra was not high enough resolution to detect multiple lines. The presence of an unresolved star may change the $\text{H}_2\text{O-K2}$ index and continuum measurements, which will in turn add scatter to the calibration.

We tested the effect of binarity on our calibration and the M13 calibration using a sample of bright late-K and M dwarfs from LG11 or this program. More than 400 of these targets have NIR SpeX spectra, primarily from a program aimed at studying the properties of M dwarfs and their planets (Gaidos et al. in preparation). NIR data for these stars were taken, reduced, and analyzed with the exact same techniques used for this work (see Section 3). We selected the 253 K7-M8 dwarfs with parallaxes from *Hipparcos* (van Leeuwen 2007) or ground-based surveys (Costa et al. 2005; Henry et al. 2006; Lépine et al. 2009; Jao et al. 2011; Dittmann et al. 2014). We calculated the expected distance to each star using the mean of the $\text{H}_2\text{O-K2} - M_K$ relation from Newton et al. (2014) and the M_J -spectral type relation from Lépine et al. (2013). We removed 65 dwarfs with spectroscopically determined distances $> 3\sigma$ larger than those based on trigonometric parallax, as these are likely unresolved binaries.

We used this sample to construct 200 unique artificial binary spectra. Specifically, we randomly combined two stars with metallicity differences < 0.07 dex (similar to the measurement error), which we determined from each spectrum using the calibration from M13 for K7-M4, and Equation 3 for M4.5–M8. To accurately place the stars at the same distance, we normalized and scaled each spectrum according to their M_K . We calculated the masses of each star using the empirical relation from Delfosse et al. (2000). We assigned a random orbital period following the log-normal distribution from Raghavan et al. (2010), but with a cutoff at semi-major axes of 32 AU. This cutoff corresponds to a $1''$ separation (resolvable in the SpeX guider) at the median distance to our targets. We then calculated the radial velocity shift assuming a random inclination and circular orbit, which we applied to the fainter of the two stars. We combined the two resulting spectra to form an empirical binary spectrum. We analyzed the resulting spectrum just as we would our other observations, determining the spectral type, moving the star to its rest frame, and recalculating the metallicity following the method as was applied for the single stars.

Figure 6 shows the difference between the metallicity derived for the artificial binaries and the mean metallicity of their components as a function of ΔM_K . We found that there is a small bias in favor lower metallicities, but the median difference is only -0.01 dex. This offset arises due to large radial velocity variations smearing out the features resulting in slightly lower equivalent widths. However, even in cases of tight, similar mass binaries, this effect is small compared to calibration and measurement errors. The scatter in metallicities is only 0.02 dex and no pairs changed in metallicity by more than 0.06 dex. Most of the scatter can be explained by small (< 0.07 dex) differences in the metallicity of the two components and the additional measurement noise.

7.4. M+M Wide Binaries

The technique of calibrating metallicity diagnostics using wide binaries relies on the assumption that wide binary components have the same metallicity. This assumption must be at least partially valid, since we would not be able to derive such a precise calibration if the metallicities of the primary and companion were uncorrelated (see Section 7.1). However, if instead there was

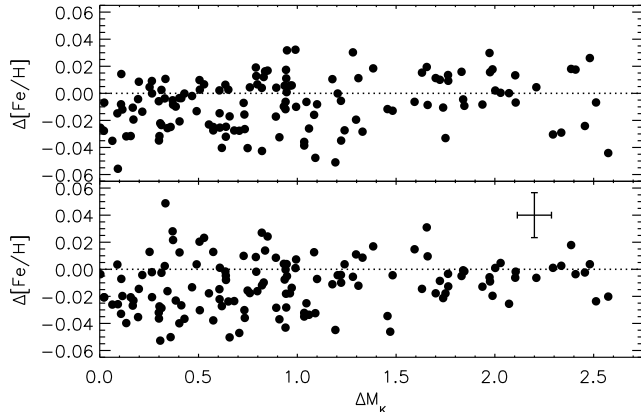


Figure 6. Difference in metallicity (as determined by Equation 3 or calibration of M13) of synthetic binary formed from combining the NIR spectra of two single K7-M8 dwarfs of similar metallicity. The top plot shows difference in metallicity between the synthetic binary and the mean of the two components, while the bottom plot shows the difference compared to the brighter of the two components. The error bar shows the median error in $[\text{Fe}/\text{H}]$ (measurement error only) and M_K . The dotted line indicates zero difference.

a dispersion in the metallicity of the star-forming cloud, or one of the stars accreted metal-rich material over its lifetime, the binary elements could have similar but not identical metallicities.

Examination of FGK+FGK wide binary systems have found consistent metallicities to within expected measurement errors (e.g., Desidera et al. 2004), suggesting that metallicity differences between binary components, if present, are very small. Rojas-Ayala et al. (2012) showed that, within errors, their method to measure M dwarf metallicities gave consistent results for both components of five M+M pairs. However, Rojas-Ayala et al. (2012) found that the color-magnitude metallicity estimates (Johnson & Apps 2009; Schlaufman & Laughlin 2010) did not show the same consistency. Thus it is prudent to apply a similar test using our calibration.

We selected a sample of seven M+M wide binaries following the same methods as explained in Section 2, with the restrictions that both components are earlier than M4.5, or both components are M4.5 or later (so the same calibration can be used) and that the components have separations $> 5''$ (so that each star can be studied separately). The sample is listed in Table 5.

We measured the metallicities of each component of these wide pairs, using the M13 calibration for M0–M4 dwarfs, and Equation 3 for M4.5–M9 dwarfs. We found a median difference between primary and companion metallicity of 0.01 dex and a maximum difference of 0.06 dex. Most of these differences were similar in size to the measurement errors, and all differences were less than the calibration errors.

7.5. Bridging the Calibrations

Combining our work with that of M13 it should be possible to measure the metallicities of dwarfs from K5 to M9.5. Although since these methods are calibrated on a different set of stars there is a possibility there will be systematic differences between the two calibrations. To

investigate this, we applied Equation 3 and the K -band calibration from M13 to a sample of 15 M4–M5 stars from Lépine et al. (2013). The metallicities between the two calibrations for 14 of these stars were consistent within 1σ , with the remaining star showing a difference of 1.6σ . We found no evidence of a systematic offset between the two sets of derived metallicities for these stars (median difference = 0.02 dex).

8. SUMMARY

We have used wide binaries containing an F, G, K, or early M dwarf primary with a M4.5–M9.5 companion to calibrate spectroscopic metallicity diagnostics for the coolest M dwarfs. Although many calibrations already exist, based either on spectroscopy or absolute magnitude, none have been calibrated with the latest M-type dwarfs. We showed that these prior spectroscopic calibrations yield systematically inaccurate metallicities for the coolest M dwarfs (Figure 2). We derived a new calibration for late-M (M4.5–M9.5) dwarfs and found that the Na I doublet and Ca I triplet were the most effective metallicity indicators for late-M dwarfs. We found that our calibration (Equation 3) predicts metallicities accurate to $\simeq 0.07$ dex for $-0.58 < [\text{Fe}/\text{H}] < +0.56$. The error is comparable to that reported by M13 for the K5–M5 sample. By combining this work with that of M13, it is possible to measure metallicities of stars across the entire M dwarf sequence.

For the F, G, and K dwarf primaries $[\text{Fe}/\text{H}]$ is generally measured directly using the plethora of Fe lines present in their spectra. However, measuring metallicities of M dwarfs generally relies on Ca and Na. We would therefore expect to get a smaller scatter relating the strength of these features to $[\alpha/\text{H}]$, $[\text{M}/\text{H}]$, and $[\text{Na}/\text{H}]$. Unsurprisingly, this was seen in previous studies using similar features (e.g. Rojas-Ayala et al. 2012; Mann et al. 2013a). This also may be the source of the higher scatter between primary and companion metallicity seen at lower metallicity (M13). However, most of the literature sources we use only report $[\text{Fe}/\text{H}]$. Although calibrations exist to determine $[\text{M}/\text{H}]$ for early M dwarfs, these calibrations use almost identical lines to the $[\text{Fe}/\text{H}]$ calibrations, which may complicate the result. Thus we would be left with only 18 stars, which is not enough for a meaningful investigation. Future analysis of this issue would require a more homogenous analysis of the primary stars.

We performed a number of tests to assess the quality and applicability of the calibration. We verified that:

- Despite the use of a large line list and many free parameters, the precision of the calibration cannot be due to chance ($P \ll 0.001$).
- The metallicities predicted by Equation 3 are free of significant trends as a function of spectral type, metallicity, or source of metallicity for the primary.
- The calibration (and that of M13) is unaffected by unresolved binaries (triples) in the calibration sample.
- Both this calibration and that of M13 yield consistent metallicities for each component of M+M wide binaries.

Table 5
M+M Binaries

Primary			Companion			
Name	SpT	[Fe/H]	Name	SpT	[Fe/H]	$\Delta[\text{Fe}/\text{H}] \pm \sigma^a$
GJ 1245A	M6.0	+0.02	GJ 1245C	M5.5	-0.01	0.03 ± 0.05
GJ 896A	M4.5	-0.07	GJ 896B	M4.5	-0.04	0.03 ± 0.03
GJ 118.2B	M0.0	+0.20	GJ 118.2C	M3.5	+0.25	0.05 ± 0.03
GJ 617A	M0.5	+0.14	GJ 617B	M2.5	+0.20	0.06 ± 0.02
GJ 4049A	M3.0	-0.17	GJ 4049B	M3.5	-0.18	0.01 ± 0.03
GJ 725A	M3.0	-0.29	GJ 725B	M3.5	-0.33	0.04 ± 0.01
LP 213-67	M7.0	-0.01	LP 213-68	M8.0	-0.02	0.01 ± 0.06

^a Includes measurement error only.

- The calibration from M13 and this work predict metallicities for M4-M5 dwarfs (where the calibration samples overlap) that are in agreement.

Another potential source of error is the presence of false common-proper-motion companions (chance alignment) in the calibration sample. Presumably two unassociated stars will have random metallicities, and therefore appear as outliers in our relation. The lack of outliers in Figure 3 suggests our sample is relatively free of false binaries. Based on the published proper motions and statistical arguments from Lépine & Bongiorno (2007) and Tokovinin & Lépine (2012) we expect the false-binary rate to be $\ll 8\%$. The true number is probably significantly lower than this, as many pairs from the literature are identified using distance and radial velocity information in conjunction with proper motions.

Although the Na I and Ca I lines are strong metallicity indicators for M dwarfs, the Na I doublet becomes significantly weaker and the Ca I triplet is essentially not detectable in L dwarfs at this S/N and resolution. It is promising that our current calibration works for a single L dwarf, however it is hard to draw conclusions from a single star. Extending this calibration further into the L dwarf regime may require fine-tuning the calibration and/or using an entirely different set of lines. We will investigate measuring the metallicities of L dwarfs in a future paper.

We thank the anonymous referee for their useful comments on this paper. This work was supported by the Harlan J. Smith Fellowship from the University of Texas at Austin to AWM, NASA grants NNX10AI90G (Astrobiology: Exobiology & Evolutionary Biology) and NNX11AC33G (Origins of Solar Systems) to EG; and NSF grant AST09-09222 to MCL, 0822443 to KMA, and AST-0709460 to EAM. Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. Also based on observations obtained with the Infrared Telescope Facility, which is operated by the University of Hawaii under Cooperative Agreement no. NNX-08AE38A with the National Aeronautics and Space Administration, Science Mission Directorate, Planetary Astronomy Program. This research has made use of the Washington Double Star Catalog maintained at the U.S. Naval Observatory. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna

Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

Facility: IRTF:SpeX, CFHT:ESPaDO nS

REFERENCES

- Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, *A&A*, 545, A32
- Aihara, H., Allende Prieto, C., An, D., et al. 2011, *ApJS*, 193, 29
- Allen, P. R., Burgasser, A. J., Faherty, J. K., & Kirkpatrick, J. D. 2012, *AJ*, 144, 62
- Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, *A&A*, 559, A74
- Bayless, A. J., & Orosz, J. A. 2006, *ApJ*, 651, 1155
- Bochanski, J. J., Savcheva, A., West, A. A., & Hawley, S. L. 2013, *AJ*, 145, 40
- Bonfils, X., Delfosse, X., Udry, S., et al. 2005, *A&A*, 442, 635
- Bowler, B. P., Liu, M. C., Shkolnik, E. L., et al. 2012, *ApJ*, 753, 142
- Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012, *ApJ*, 757, 112
- Caballero, J. A. 2007, *ApJ*, 667, 520
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, 530, A138
- Cayrel de Strobel, G., Soubiran, C., & Ralite, N. 2001, *A&A*, 373, 159
- Chanamé, J., & Gould, A. 2004, *ApJ*, 601, 289
- Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, *Nature*, 462, 891
- Ciddor, P. E. 1996, *Appl. Opt.*, 35, 1566
- Costa, E., Méndez, R. A., Jao, W.-C., et al. 2005, *AJ*, 130, 337
- Covey, K. R., Ivezić, Ž., Schlegel, D., et al. 2007, *AJ*, 134, 2398
- Cruz, K. L., Reid, I. N., Kirkpatrick, J. D., et al. 2007, *AJ*, 133, 439
- Cushing, M. C., Rayner, J. T., & Vacca, W. D. 2005, *ApJ*, 623, 1115
- Cushing, M. C., Vacca, W. D., & Rayner, J. T. 2004, *PASP*, 116, 362
- de Bruijne, J. H. J. 2012, *Ap&SS*, 341, 31
- Deacon, N. R., Liu, M. C., Magnier, E. A., et al. 2012, *ApJ*, 757, 100
- , in review, *ApJ*
- Delfosse, X., Forveille, T., Ségransan, D., et al. 2000, *A&A*, 364, 217
- Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, *ApJS*, 155, 667
- Desidera, S., Gratton, R. G., Lucatello, S., & Claudi, R. U. 2006, *A&A*, 454, 581
- Desidera, S., Gratton, R. G., Scuderi, S., et al. 2004, *A&A*, 420, 683
- Dhital, S., West, A. A., Stassun, K. G., et al. 2012, *AJ*, 143, 67
- Dittmann, J. A., Irwin, J. M., Charbonneau, D., & Berta-Thompson, Z. K. 2014, *ApJ*, 784, 156
- Donati, J.-F. 2003, in *Astronomical Society of the Pacific Conference Series*, Vol. 307, *Astronomical Society of the Pacific Conference Series*, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 41
- Donati, J.-F., Semel, M., Carter, B. D., Rees, D. E., & Collier Cameron, A. 1997, *MNRAS*, 291, 658

- Dupuy, T. J., Liu, M. C., Bowler, B. P., et al. 2010, *ApJ*, 721, 1725
- Faherty, J. K., Burgasser, A. J., West, A. A., et al. 2010, *AJ*, 139, 176
- Figueira, P., Marmier, M., Boué, G., et al. 2012, *A&A*, 541, A139
- Fischer, D. A., Gaidos, E., Howard, A. W., et al. 2012, *ApJ*, 745, 21
- Fuhrmann, K. 2008, *MNRAS*, 384, 173
- Gaidos, E., Fischer, D. A., Mann, A. W., & Howard, A. W. 2013, *ApJ*, 771, 18
- Gizis, J. E., Kirkpatrick, J. D., & Wilson, J. C. 2001, *AJ*, 121, 2185
- Gonzalez, G. 1997, *MNRAS*, 285, 403
- Gould, A., & Chanamé, J. 2004, *ApJS*, 150, 455
- Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, *AJ*, 132, 2360
- Jao, W.-C., Henry, T. J., Subasavage, J. P., et al. 2011, *AJ*, 141, 117
- Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, *PASP*, 122, 905
- Johnson, J. A., & Apps, K. 2009, *ApJ*, 699, 933
- Johnson, J. A., Gazak, J. Z., Apps, K., et al. 2012, *AJ*, 143, 111
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7733, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Kirkpatrick, J. D., Henry, T. J., & McCarthy, Jr., D. W. 1991, *ApJS*, 77, 417
- Kirkpatrick, J. D., Liebert, J., Cruz, K. L., Gizis, J. E., & Reid, I. N. 2001, *PASP*, 113, 814
- Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, 765, 131
- Kraus, A. L., Ireland, M. J., Hillenbrand, L. A., & Martinache, F. 2012, *ApJ*, 745, 19
- Law, N. M., Dhital, S., Kraus, A., Stassun, K. G., & West, A. A. 2010, *ApJ*, 720, 1727
- Lee, Y. S., Beers, T. C., Allende Prieto, C., et al. 2011, *AJ*, 141, 90
- Lépine, S., & Bongiorno, B. 2007, *AJ*, 133, 889
- Lépine, S., & Gaidos, E. 2011, *AJ*, 142, 138
- Lépine, S., Hilton, E. J., Mann, A. W., et al. 2013, *AJ*, 145, 102
- Lépine, S., & Shara, M. M. 2005, *AJ*, 129, 1483
- Lépine, S., Thorstensen, J. R., Shara, M. M., & Rich, R. M. 2009, *AJ*, 137, 4109
- Luhman, K. L., Loutrel, N. P., McCurdy, N. S., et al. 2012, *ApJ*, 760, 152
- Mahadevan, S., Ramsey, L., Bender, C., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S., & Hilton, E. J. 2013a, *AJ*, 145, 52
- Mann, A. W., Gaidos, E., & Ansdell, M. 2013b, *ApJ*, 779, 188
- Mann, A. W., Gaidos, E., Kraus, A., & Hilton, E. J. 2013c, *ApJ*, 770, 43
- Mann, A. W., Gaidos, E., Lépine, S., & Hilton, E. J. 2012, *ApJ*, 753, 90
- Markwardt, C. B. 2009, in *Astronomical Society of the Pacific Conference Series*, Vol. 411, *Astronomical Data Analysis Software and Systems XVIII*, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251
- Martín, E. L., Koresko, C. D., Kulkarni, S. R., Lane, B. F., & Wizinowich, P. L. 2000, *ApJ*, 529, L37
- Neves, V., Bonfils, X., Santos, N. C., et al. 2012, *A&A*, 538, A25
- Newton, E. R., Charbonneau, D., Irwin, J., et al. 2014, *AJ*, 147, 20
- Nutzman, P., & Charbonneau, D. 2008, *PASP*, 120, 317
- Önehag, A., Heiter, U., Gustafsson, B., et al. 2012, *A&A*, 542, A33
- Pinfield, D. J., Jones, H. R. A., Lucas, P. W., et al. 2006, *MNRAS*, 368, 1281
- Pinsonneault, M. H., DePoy, D. L., & Coffee, M. 2001, *ApJ*, 556, L59
- Quirrenbach, A., Amado, P. J., Seifert, W., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJS*, 190, 1
- Ramírez, I., Allende Prieto, C., & Lambert, D. L. 2007, *A&A*, 465, 271
- Rayner, J. T., Cushing, M. C., & Vacca, W. D. 2009, *ApJS*, 185, 289
- Rayner, J. T., Toomey, D. W., Onaka, P. M., et al. 2003, *PASP*, 115, 362
- Reid, I. N., Hawley, S. L., & Gizis, J. E. 1995, *AJ*, 110, 1838
- Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2010, *ApJ*, 720, L113
- . 2012, *ApJ*, 748, 93
- Schlaufman, K. C., & Laughlin, G. 2010, *A&A*, 519, A105
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Snedden, C. A. 1973, PhD thesis, the University of Texas at Austin
- Sozzetti, A., Bernagozzi, A., Bertolini, E., et al. 2013, in *European Physical Journal Web of Conferences*, Vol. 47, European Physical Journal Web of Conferences, 3006
- Takeda, Y., Ohkubo, M., Sato, B., Kambe, E., & Sadakane, K. 2005, *PASJ*, 57, 27
- Terrien, R. C., Mahadevan, S., Bender, C. F., et al. 2012, *ApJ*, 747, L38
- Tokovinin, A., & Lépine, S. 2012, *AJ*, 144, 102
- Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, *ApJ*, 757, 161
- Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, *PASP*, 115, 389
- Valenti, J. A., & Fischer, D. A. 2005, *ApJS*, 159, 141
- Valenti, J. A., & Piskunov, N. 1996, *A&AS*, 118, 595
- Valenti, J. A., Fischer, D., Marcy, G. W., et al. 2009, *ApJ*, 702, 989
- van Biesbroeck, G. 1944, *AJ*, 51, 61
- van Leeuwen, F. 2007, *A&A*, 474, 653
- van Leeuwen, F., & Fantino, E. 2005, *A&A*, 439, 791
- West, A. A., Hawley, S. L., Walkowicz, L. M., et al. 2004, *AJ*, 128, 426
- West, A. A., Morgan, D. P., Bochanski, J. J., et al. 2011, *AJ*, 141, 97
- Woolf, V. M., & Wallerstein, G. 2006, *PASP*, 118, 218