

THE SURFACE NITROGEN ABUNDANCE OF A MASSIVE STAR IN RELATION TO ITS OSCILLATIONS, ROTATION, AND MAGNETIC FIELD

C. AERTS^{1,2,3}, G. MOLENBERGHS^{3,4}, M. G. KENWARD⁵, AND C. NEINER⁶

¹ Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium; conny@ster.kuleuven.be

² Department of Astrophysics, IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

³ Faculty of Science, Hasselt University, Martelarenlaan 42, B-3500 Hasselt, Belgium

⁴ I-BioStat, KU Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium

⁵ Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E7HT, UK

⁶ LESIA, UMR 8109 du CNRS, Observatoire de Paris, UPMC, Paris Diderot, 5 Place Jules Janssen, F-92195 Meudon Cedex, France

Received 2013 October 30; accepted 2013 December 12; published 2014 January 13

ABSTRACT

We have composed a sample of 68 massive stars in our galaxy whose projected rotational velocity, effective temperature, and gravity are available from high-precision spectroscopic measurements. The additional seven observed variables considered here are their surface nitrogen abundance, rotational frequency, magnetic field strength, and the amplitude and frequency of their dominant acoustic and gravity modes of oscillation. A multiple linear regression to estimate the nitrogen abundance combined with principal component analysis, after addressing the incomplete and truncated nature of the data, reveals that the effective temperature and the frequency of the dominant acoustic oscillation mode are the only two significant predictors for the nitrogen abundance, while the projected rotational velocity and the rotational frequency have no predictive power. The dominant gravity mode and the magnetic field strength are correlated with the effective temperature but have no predictive power for the nitrogen abundance. Our findings are completely based on observations and their proper statistical treatment and call for a new strategy in evaluating the outcome of stellar evolution computations.

Key words: methods: statistical – stars: abundances – stars: evolution – stars: magnetic field – stars: massive – stars: oscillations

Online-only material: color figure

1. INTRODUCTION

The mixing of chemical species inside stars is poorly understood. Yet, its effect on stellar evolution and supernova explosions, and by implication on the chemical enrichment of galaxies is of prime importance. It was realized quite some time ago that stellar rotation and the mixing it induces must play a major role in stellar evolution theory (e.g., Herrero et al. 1992) but its inclusion in models is far from trivial (see, e.g., Langer 1992; Talon et al. 1997 for early attempts and discussions), while Maeder (2009) is a recent extensive monograph in this topic.

Ways to test the theoretical descriptions used to represent rotational mixing are scarce and are mainly limited to studies of the chemical composition and projected rotational velocity of stellar atmospheres (e.g., Venn & Lambert 2005; Hunter et al. 2008; Przybilla et al. 2010, among many others). A particularly strong observational diagnostic for mixing is the surface nitrogen abundance of stars that are undergoing hydrogen fusion through the carbon–nitrogen–oxygen cycle in their interior. It turns out that the current theoretical concepts of rotational mixing seemingly fail to explain observations of various massive stars in the Milky Way and in the Magellanic Clouds (Hunter et al. 2008; Brott et al. 2011; Rivero González et al. 2012; Bouret et al. 2012, 2013), leading to intense debates on yet unknown causes of the discrepancies and on the way to identify other physical ingredients lacking in the theoretical models (e.g., Meynet et al. 2011; Potter et al. 2012; Mathis et al. 2013).

Here, we shed new light on the matter by considering a variety of *observational* data and by subjecting them to careful statistical analysis. In particular, we investigate the relationship between observed stellar oscillations, rotational frequency, magnetic field strength, and surface nitrogen abundance in a sample of galactic massive stars for which this multitude of

data has recently become available. A direct comparison between observed rotational properties with evolutionary models would ideally be based on a value for the equatorial rotation velocity. This would require having either precise measurements of both the rotational frequency and the stellar radius, or else of the inclination angle i between the rotational axis of the star and the line-of-sight of the observer, in addition to $v \sin i$. While the combination $(v \sin i, i)$ is available for some pulsating stars from an analysis of time-series spectroscopy and for a few magnetic stars from spectro-polarimetry, this concerns very few stars. Similarly, direct measurements of the radii in general are not available for massive stars. Hence, a specific point of attention in this work is whether the use of the rotational frequency, rather than the projected rotational velocity, leads to an improved diagnostic.

2. SAMPLE SELECTION

Sample selection was restricted to stars whose effective temperature, gravity, and projected rotational velocity are available from high signal-to-noise, high-resolution spectroscopy. From this sample, we kept the stars for which at most one of the following four additional properties has missing data: nitrogen abundance, rotational frequency, magnetic field strength, and oscillations. This rather strict criterion of missingness was adopted to achieve a good starting point for the statistical analysis outlined in the following section and implied that only Galactic OB stars were retained in the sample.

Regarding the oscillations detected, we considered the frequency and the amplitude of the dominant acoustic mode and of the dominant gravity mode of the star (see Aerts et al. 2010, for a definition and extensive description of such heat-driven oscillations in massive stars) whenever available in the

Table 1
Characteristics of the Observed Data Set

Variable	Quantity	Physical Meaning	Unit	Observed Range	Tr?	% Obs.
X_1	$v \sin i$	Projected rotational velocity	km s^{-1}	[0,310]	No	100%
X_2	f_{rot}	Rotational frequency	per day	[0,2.505]	No	94%
X_3	f_g	Gravity-mode frequency	per day	[0.039,1.157]	No	32%
X_4	A_g	Amplitude of f_g	mmag	[0.11,24.90]	Yes	96%
X_5	f_p	Acoustic mode frequency	per day	[3.258,10.935]	No	32%
X_6	A_p	Amplitude of f_p	mmag	[0.05,38.10]	Yes	96%
X_7	$\log(1 + B_{\text{pol}})$	Magnetic field strength	Gauss	[0.00,4.20]	Yes	96%
X_8	$\log T_{\text{eff}}$	Effective temperature	Kelvin	[4.061,4.633]	No	100%
X_9	$\log g$	Gravity	cm s^{-2}	[2.70,4.43]	No	100%
X_{10}	$12 + \log[\text{N}/\text{H}]$	Nitrogen abundance	dex	[7.42,8.95]	Yes	59%

Notes. The magnetic field is assumed to be an oblique dipole and B_{pol} is the polar field strength. “Tr?” stands for “Truncated with upper limit” or not. All variables are positive quantities and are thus truncated with zero as the lower limit. The last column gives the percentage of observed values expressing the level of completeness within the data set, where we considered both measured values and measured limits to determine the percentage.

literature. We downloaded the Hipparcos data from Perryman (1997) and derived upper limits for the oscillation amplitudes as four times the standard deviation of these data. These upper limits were adopted in the case of stars for which one or both types of oscillations have not yet been detected.

The magnetic field strength under the assumption of an oblique dipole configuration is either available as a measured value (in case a time series of spectro-polarimetry was observed) or as a lower limit (because the projection factor between the line-of-sight and the magnetic axis at the time of measurement is unknown). We adopted the most recent published results derived from high-resolution polarimeters such as ESPaDOnS, NARVAL, or HARPSpol (data sources are listed along with the data in Table 2). In absence of those, we took results based on the lower-resolution spectrographs such as, e.g., FORS1/2. For a comparison between the interpretations based on these two types of measurements, we refer to Shultz et al. (2012). We added to those data all unpublished null detections obtained by the MiMeS (Magnetism in Massive Stars) collaboration.⁷

The rotational frequency was deduced from time series of spectro-polarimetric data or from rotational splitting of oscillation frequencies. It is noteworthy that, whenever values from these two independent methods are available for a star, they are in excellent agreement (see Briquet et al. 2013 for a recent example). Whenever available, we took the $v \sin i$ values deduced from time-series spectroscopy analyzed in such a way that line-broadening due to the oscillations was carefully taken into account (e.g., Aerts & De Cat 2003; Aerts et al. 2010, Chapter 6). In the absence of time-series spectroscopy or spectro-polarimetry, we relied on the $v \sin i$ values derived from high-resolution single snapshot spectra.

For the values of $\log T_{\text{eff}}$, $\log g$, and the nitrogen abundance, various sources are available in the literature, based on different methods and analysis codes. We relied on recent NLTE analyses of high-resolution, high signal-to-noise spectroscopy, mostly done by Nieva & Przybilla (2012), Martins et al. (2012a, 2012b), and Morel et al. (2008) as well as on detailed asteroseismic or spectro-polarimetric studies of individual targets which also included NLTE line-fitting of the high-precision spectroscopy. For stars not covered in this way, we took the most recent high-precision spectroscopic data source available.

Table 1 gives a summary of the observed properties of the 68 sample stars, with an indication of the level of completeness

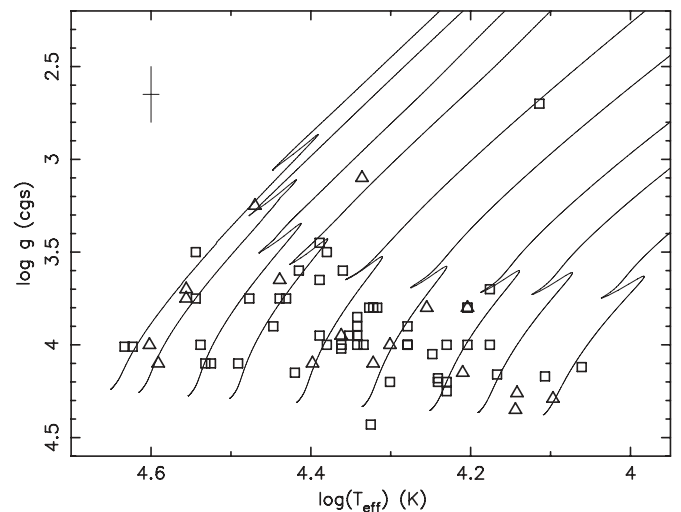


Figure 1. ($\log T_{\text{eff}}$, $\log g$) diagram of the measured stars (squares for single stars, triangles for spectroscopic binaries) and of evolutionary tracks (full lines, see the text) for masses of 3, 4, 5, 7, 10, 15, 20, 30, and 40 M_{\odot} , from right to left. A typical error box is indicated in the upper left corner.

of the sampling for each of the 10 variables X_1, \dots, X_{10} , while Table 2 contains the actual data of all stars we used in our analyses along with all of the data sources. All stars are plotted in a ($\log T_{\text{eff}}$, $\log g$) diagram in Figure 1, together with some evolutionary tracks starting from the zero-age main sequence (ZAMS). These were computed with the MESA stellar evolution code (Paxton et al. 2011, 2013), taking the solar composition given by Asplund et al. (2009) and ignoring rotational mixing. We adopted the mixing-length theory of convection with a mixing-length value of 1.8 local pressure scale heights. The Schwarzschild criterion for convection was used, along with a fully mixed core-overshoot region extending over 0.15 local pressure scale heights, which is a good average of all seismically determined values for OB pulsators (Aerts 2014). Among the 68 stars are 17 unevolved spectroscopic binaries whose orbits are known and for which the binarity has appropriately been taken into account in the primary’s parameters listed in Table 2. These are indicated with triangles in Figure 1.

Some comments on Figure 1 are warranted. Hubrig et al. (2006) already pointed out the high $\log g$ value with large uncertainty for the star HD 46005, placing it below the ZAMS.

⁷ <http://www.physics.queensu.ca/~wade/mimes/>

Table 2
Data Set Used in This Study Composed as per 2013 September 1

Object	SpT	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	SB
108	O4-8f?p	0 (1)	0.000 (1)	... (...)	<4.80 (2)	... (...)	<4.80 (2)	3.10 (1)	4.544 (1)	3.50 (1)	8.77 (1)	0
46223	O5e	100 (34)	0.249 (35)	0.755 (35)	0.46 (35)	7.554 (35)	0.17 (35)	0.00 (33)	4.633 (34)	4.01 (34)	8.85 (34)	0
46150	O5Vf	100 (34)	0.144 (35)	0.055 (35)	0.32 (35)	... (...)	<0.01 (35)	0.00 (33)	4.623 (34)	4.01 (34)	8.48 (34)	0
148937	O5.5-6f?p	45 (65)	0.142 (65)	... (...)	<10.00 (2)	... (...)	<10.00 (2)	3.01 (65)	4.602 (1)	4.00 (1)	8.48 (1)	1
37022	O7V	24 (1)	0.065 (20)	... (...)	<22.80 (2)	... (...)	<22.80 (2)	3.04 (1)	4.591 (1)	4.10 (1)	7.82 (1)	1
191612	O8fpe	1 (1)	0.002 (81)	... (...)	<8.00 (2)	... (...)	<8.00 (2)	3.39 (81)	4.556 (1)	3.75 (1)	8.43 (1)	1
46966	O8V	50 (34)	0.084 (35)	0.039 (35)	0.89 (35)	8.226 (35)	0.06 (35)	0.00 (33)	4.544 (34)	3.75 (34)	8.08 (34)	0
46149	O8.5V	30	0.084	0.175	0.20	3.844	0.05	0.00	4.556	3.70	7.90	1
57682	O9IV	15 (1)	0.016 (14)	... (...)	<3.20 (2)	... (...)	<3.20 (2)	2.94 (48)	4.538 (48)	4.00 (48)	8.11 (1)	0
214680	O9V	16 (85)	0.147 (23)	... (...)	<2.40 (2)	3.258 (31)	5.50 (31)	0.00 (33)	4.532 (31)	4.10 (31)	7.81 (86)	0
46202	O9V	25 (36)	... (...)	0.510 (36)	0.11 (36)	4.856 (36)	0.09 (36)	0.00 (33)	4.525 (34)	4.10 (34)	8.00 (34)	0
37742	O9Iab:	100 (1)	0.167 (23)	... (...)	<4.40 (2)	... (...)	<4.40 (2)	1.88 (1)	4.470 (1)	3.25 (1)	7.52 (1)	1
149438	B0.2V	6 (66)	0.024 (66)	... (...)	<1.60 (2)	... (...)	<1.60 (2)	2.70 (66)	4.491 (1)	4.10 (1)	8.15 (1)	0
51756	B0.5IV	28 (45)	0.527 (45)	... (...)	<0.01 (2)	... (...)	<0.01 (2)	0.00 (33)	4.477 (45)	3.75 (45)	... (...)	0
111123	B0.5IV	18 (55)	... (...)	... (...)	<4.00 (2)	5.231 (56)	3.10 (56)	0.00 (33)	4.439 (28)	3.65 (28)	7.61 (28)	1
46328	B0.7IV	9 (28)	0.235 (37)	... (...)	<8.40 (2)	4.772 (38)	16.20 (39)	>3.08 (39)	4.439 (28)	3.75 (28)	8.00 (28)	0
44743	B1III	23 (28)	0.054 (29)	... (...)	<5.60 (2)	3.979 (30)	21.00 (30)	0.00 (4)	4.380 (28)	3.50 (28)	7.59 (28)	0
50707	B1Ib	34 (43)	0.107 (44)	... (...)	<2.40 (2)	5.419 (44)	4.6 (44)	0.00 (9)	4.415 (28)	3.60 (28)	8.03 (28)	0
66665	B1V	10 (52)	0.048 (52)	... (...)	<6.00 (2)	... (...)	<6.00 (2)	2.83 (52)	4.447 (52)	3.90 (52)	... (...)	0
187879	B1III+	97 (79)	0.129 (79)	... (...)	<0.01 (79)	... (...)	<0.01 (79)	... (...)	4.336 (79)	3.10 (79)	7.55 (79)	1
37017	B1.5V	90 (19)	1.110 (19)	... (...)	<5.60 (2)	... (...)	<5.60 (2)	>3.78 (14)	4.322 (19)	4.10 (19)	... (...)	1
64740	B1.5Vp	160 (50)	0.752 (50)	... (...)	<2.00 (2)	... (...)	<2.00 (2)	4.20 (14)	4.380 (14)	4.00 (14)	7.89 (51)	0
74575	B1.5III	11 (5)	... (...)	... (...)	<2.00 (2)	... (...)	<2.00 (2)	0.00 (9)	4.360 (5)	3.60 (5)	7.92 (5)	0
180642	B1.5III	25 (75)	0.075 (75)	0.308 (76)	1.60 (76)	5.487 (76)	38.10 (76)	0.00 (9)	4.389 (76)	3.45 (76)	8.00 (76)	0
886	B2IV	3 (3)	0.005 (3)	0.682 (3)	1.99 (3)	6.590 (3)	6.59 (3)	0.00 (4)	4.342 (5)	3.95 (5)	7.76 (5)	0
3360	B2IV	17 (6)	0.186 (6)	0.640 (6)	1.30 (2)	... (...)	<1.00 (2)	2.53 (6)	4.317 (5)	3.80 (5)	8.23 (5)	0
16582	B2IV	1 (7)	0.075 (7)	0.318 (7)	0.43 (7)	6.206 (7)	11.62 (7)	0.00 (8)	4.327 (5)	3.80 (5)	8.23 (5)	0
29248	B2III	6 (11)	0.017 (11)	0.432 (12)	3.20 (12)	5.763 (12)	36.90 (12)	0.00 (4)	4.342 (5)	3.85 (5)	7.93 (5)	0
36485	B2V	32 (18)	0.677 (18)	... (...)	... (...)	... (...)	... (...)	4.00 (18)	4.301 (18)	4.00 (18)	... (...)	1
37479	B2Vp	170 (14)	0.840 (21)	... (...)	... (...)	... (...)	... (...)	3.98 (22)	4.362 (14)	4.00 (14)	... (...)	0
37776	B2IV	95 (14)	0.650 (24)	... (...)	<7.20 (2)	... (...)	<7.20 (2)	4.18 (25)	4.342 (14)	4.00 (14)	... (...)	0
55522	B2V	70 (46)	0.366 (46)	... (...)	<7.20 (2)	... (...)	<7.20 (2)	>3.42 (47)	4.241 (47)	4.20 (47)	... (...)	0
61068	B2II	10 (39)	... (...)	... (...)	<3.60 (2)	6.010 (30)	19.50 (30)	0.00 (39)	4.420 (5)	4.15 (5)	8.00 (5)	0

Table 2
(Continued)

Object	SpT	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	SB
67621	B2IV	20	0.279	...	<3.20	...	<3.20	>2.95	4.279	4.00	...	0
		(14)	(14)	...	(2)	...	(2)	(14)	(14)	(14)	...	
85953	B2III	18	0.020	0.266	11.20	...	<1.00	0.00	4.322	3.80	7.66	0
		(39)	(10)	(10)	(10)	...	(10)	(33)	(28)	(28)	(28)	
96446	B2IIIp	3	0.175	1.149	<4.40	10.763	<4.40	3.89	4.334	4.00	7.42	0
127381	B2III	68	0.331	...	<2.00	10.935	3.10	2.70	4.362	4.02	8.26	0
		(60)	(60)	...	(2)	(61)	(61)	(60)	(60)	(60)	(60)	
142184	B2V	290	1.967	...	<4.80	...	<4.80	4.00	4.230	4.25	...	0
		(48)	(48)	...	(2)	...	(2)	(48)	(48)	(48)	...	
157056	B2IV	31	0.107	...	<1.00	7.116	9.40	...	4.398	4.10	7.78	1
		(67)	(68)	...	(69)	(69)	(69)	(58)	(28)	(28)	(28)	
163472	B2V	63	0.275	...	<1.00	7.148	13.22	2.60	4.352	3.95	7.99	0
		(70)	(71)	...	(72)	(72)	(72)	(71)	(28)	(28)	(28)	
182180	B2Vn	310	1.918	...	<10.00	...	<10.00	4.06	4.248	4.05	...	0
		(77)	(77)	...	(2)	...	(2)	(77)	(77)	(77)	...	
184927	B2V	14	0.105	...	<4.40	...	<4.40	3.59	4.342	3.90	...	0
		(78)	(78)	...	(2)	...	(2)	(14)	(14)	(14)	...	
205021	B2IIIev	25	0.083	...	<1.00	5.250	37.00	2.48	4.431	3.75	8.11	0
		(82)	(83)	...	(2)	(30)	(30)	(83)	(5)	(5)	(5)	
214993	B2III	36	0.120	0.355	5.00	5.179	38.10	0.00	4.389	3.65	7.64	0
		(87)	(87)	(88)	(88)	(88)	(88)	(33)	(28)	(28)	(28)	
216916	B2IV	20	0.275	...	<2.00	5.911	2.55	0.00	4.362	3.95	7.78	1
		(89)	(90)	...	(91)	(91)	(91)	(33)	(5)	(5)	(5)	
48977	B2.5V	29	0.637	0.517	2.23	...	<0.01	0.00	4.301	4.20	7.53	0
		(41)	(41)	(41)	(41)	...	(41)	(33)	(41)	(41)	(41)	
208057	B3Ve	104	0.694	0.802	9.00	...	<7.60	>2.70	4.279	3.90	...	0
		(58)	(84)	(31)	(31)	...	(2)	(15)	(15)	(15)	...	
129929	B3V	2	0.012	...	<1.00	6.462	11.80	0.00	4.389	3.95	7.73	0
		(62)	(63)	...	(62)	(62)	(62)	(33)	(28)	(28)	(28)	
50230	B3V	7	0.044	0.684	1.84	4.922	1.24	...	4.255	3.80	...	1
		(42)	(42)	(42)	(42)	(42)	(42)	...	(42)	(42)	...	
74560	B3IV	13	0.010	0.645	14.30	...	<1.00	0.00	4.210	4.15	...	1
		(39)	(10)	(10)	(10)	...	(2)	(9)	(39)	(39)	...	
43317	B3IV	110	1.115	1.101	1.44	4.331	0.56	3.08	4.230	4.00	7.66	0
		(26)	(26)	(26)	(26)	(26)	(26)	(27)	(26)	(26)	(26)	
35298	B3Vw	260	0.540	...	<9.60	...	<9.60	>3.95	4.204	3.80	...	0
		(14)	(15)	...	(2)	...	(2)	(16)	(14)	(14)	...	
24587	B5V	32	0.426	1.157	7.90	...	<1.00	0.00	4.142	4.26	...	1
		(8)	(8)	(8)	(8)	...	(2)	(9)	(10)	(10)	...	
189775	B5III	85	0.384	...	<10.40	...	<10.40	>3.65	4.204	3.80	...	0
		(80)	(80)	...	(2)	...	(2)	(14)	(80)	(80)	...	
176582	B5IV	105	0.632	...	<7.20	...	<7.20	3.85	4.204	4.00	...	0
		(74)	(74)	...	(2)	...	(2)	(74)	(74)	(74)	...	
142990	B5V	125	1.021	...	<4.80	...	<4.80	>3.88	4.230	4.20	...	0
		(64)	(15)	...	(2)	...	(2)	(15)	(64)	(64)	...	
61556	B5IVn	70	0.524	3.60	4.176	4.00	...	0
		(49)	(49)	(14)	(14)	(14)	...	
46769	B5II	72	0.103	...	<0.01	...	<0.01	0.00	4.114	2.70	8.08	0
		(40)	(40)	...	(40)	...	(40)	(40)	(40)	(40)	(40)	
35502	B5V	80	1.176	...	<5.60	...	<5.60	>3.83	4.204	3.80	...	1
		(14)	(14)	...	(2)	...	(2)	(17)	(14)	(14)	...	
105382	B6IIIe	73	0.772	...	<7.20	...	<7.20	3.36	4.241	4.18	...	0
		(46)	(46)	...	(2)	...	(2)	(54)	(46)	(46)	...	
125823	B7IIIpv	15	0.113	...	<18.80	...	<18.80	>4.15	4.279	4.00	8.10	0
		(59)	(59)	...	(2)	...	(2)	(59)	(59)	(59)	(59)	
46005	B8V	150	2.505	...	<26.00	...	<26.00	0.00	4.325	4.43	...	0
		(4)	(31)	...	(2)	...	(2)	(4)	(31)	(31)	...	
175362	B8IVs	35	0.272	...	<18.00	...	<18.00	>4.32	4.176	3.70	8.95	0
		(73)	(15)	...	(2)	...	(2)	(15)	(73)	(73)	(73)	
140873	B8III	69	1.226	1.152	13.40	...	<1.00	0.00	4.144	4.35	...	1
		(8)	(8)	(10)	(10)	...	(10)	(33)	(57)	(57)	...	
32633	B9p	19	0.156	...	<5.20	...	<5.20	>3.93	4.107	4.17	...	0
		(13)	(13)	...	(2)	...	(2)	(13)	(13)	(13)	...	

Table 2
(Continued)

Object	SpT	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	SB
123515	B9IV	15 (57)	0.038 (10)	0.685 (10)	20.90 (10)	... (10)	<1.00 (10)	0.00 (58)	4.097 (10)	4.29 (10)	... (10)	1
181558	A0	12 (8)	0.127 (8)	0.808 (8)	24.90 (8)	... (8)	<1.00 (8)	0.00 (4)	4.167 (10)	4.16 (10)	... (10)	0
112413	A0spe	17 (13)	0.183 (13)	... (2)	<40.80 (2)	... (2)	<40.80 (2)	>3.20 (13)	4.061 (13)	4.12 (13)	... (13)	0

Notes. The object number corresponds to the Henry-Draper (HD) number. The spectral type of the star (SpT) was retrieved from the Simbad database and was used to order the stars. The rightmost column is a flag to indicate spectroscopic binarity (1) or not (0). These three entries were not used in the statistical modeling. The meaning of X_i is given in Table 1. The four stars without acceptable imputed values for their missing data are indicated in *italic*. The data source for each entry is listed in brackets beneath the value as an index which corresponds to the same index in the reference list except for (33), which stands for an unpublished null detection by the MiMeS collaboration.

References. (1) Martins et al. 2012b; (2) Perryman 1997; (3) Handler et al. 2009; (4) Silvester et al. 2009; (5) Nieva & Przybilla 2012; (6) Neiner et al. 2003; (7) Aerts et al. 2006; (8) De Cat et al. 2005; (9) Shultz et al. 2012; (10) De Cat & Aerts 2002; (11) Pamyatnykh et al. 2004; (12) Handler et al. 2004; (13) Silvester et al. 2012; (14) Petit et al. 2013; (15) Bychkov et al. 2005; (16) Yakunin et al. 2011; (17) Romanyuk & Kudryavtsev 2008; (18) Leone et al. 2010; (19) Bolton et al. 1998; (20) Petit et al. 2008; (21) Townsend et al. 2010; (22) Oksala et al. 2012; (23) Kaper et al. 1996; (24) Mikulášek et al. 2008; (25) Kochukhov et al. 2011; (26) Pápics et al. 2012; (27) Briquet et al. 2013; (28) Morel et al. 2008; (29) Mazumdar et al. 2006; (30) Stankov & Handler 2005; (31) De Cat et al. 2007; (32) Degroote et al. 2010; (33) MiMeS collaboration, unpublished null detection; (34) Martins et al. 2012a; (35) Blomme et al. 2011; (36) Briquet et al. 2011; (37) Fournet-Ravard et al. 2011; (38) Saesen et al. 2006; (39) Hubrig et al. 2009; (40) Aerts et al. 2013a; (41) Thoul et al. 2013; (42) Degroote et al. 2012; (43) Lefever et al. 2010; (44) Shobbrook et al. 2006; (45) Pápics et al. 2011; (46) Briquet et al. 2004; (47) Briquet et al. 2007a; (48) Grunhut et al. 2012; (49) Rivinius et al. 2003; (50) Bohlender & Landstreet 1990; (51) Fraser et al. 2010; (52) Petit et al. 2011; (53) Neiner et al. 2012c; (54) Alecian et al. 2011; (55) Aerts et al. 1998; (56) Cuypers et al. 2002; (57) Aerts et al. 1999; (58) Hubrig et al. 2006; (59) Bohlender et al. 2010; (60) Henrichs et al. 2012; (61) Koen & Eyer 2002; (62) Aerts et al. 2004; (63) Aerts et al. 2003b; (64) Cidale et al. 2007; (65) Wade et al. 2012; (66) Donati et al. 2006; (67) Briquet et al. 2005; (68) Briquet et al. 2007b; (69) Handler et al. 2005; (70) Briquet et al. 2012; (71) Neiner et al. 2012a; (72) Handler et al. 2012; (73) Leone & Manfre 1997; (74) Bohlender & Monin 2011; (75) Aerts et al. 2011; (76) Briquet et al. 2009; (77) Rivinius et al. 2013; (78) Wade et al. 1997; (79) Tkachenko et al. 2013; (80) Lyubimkov et al. 2002; (81) Wade et al. 2011; (82) Telting et al. 1997; (83) Henrichs et al. 2013; (84) Henrichs et al. 2009; (85) Simón-Díaz et al. 2010; (86) Villamariz et al. 2002; (87) Desmet et al. 2009; (88) Handler et al. 2006; (89) Aerts et al. 2003a; (90) Thoul et al. 2003; (91) Chapellier et al. 1995.

On the other side of low $\log g$ values, we see that only one star in the sample is evolved beyond core-hydrogen burning. This is the rotational variable B5II star HD 46769 studied from time-resolved *CoRoT* space photometry and high-precision spectroscopy (Aerts et al. 2013a). The additional seemingly evolved object is the double-lined spectroscopic eclipsing binary V380 Cyg, for which a core-overshoot value of 0.6 local pressure scale heights was reported to bring the model-independent measured dynamical mass deduced for the primary in agreement with single-star evolutionary models (Guinan et al. 2000). This binary is the brightest star observed by the *Kepler* spacecraft and these high-precision photometric space data along with an extensive new set of time-resolved, high-resolution spectroscopy led to dynamical masses with a relative precision near 1% and revealed the primary to be a low-amplitude rotational variable with mild silicon spots and additional stochastic low-frequency photometric variability. These latest data confirmed the earlier findings that a high core-overshoot parameter is necessary to bring the primary's mass into agreement with single-star models and that the secondary does not fit the isochrones for the measured equatorial rotational velocity and metallicity of both stars (Tkachenko et al. 2013). This binary indicates that the current stellar models have too limited near-core mixing. As can further be seen in Figure 1, all other stars in our sample cover the main-sequence phase of evolution.

Finally, the missing data due to lack of measurements were assumed to be missing at random (Rubin 1976; Molenberghs & Kenward 2007), in line with common statistical practice. As is often done in astronomy, we consider logarithmic quantities for the variables X_4 , X_6 , X_7 , X_8 , X_9 , and X_{10} , given their wide ranges of possible values among OB stars, where X_7 was transformed in such a way that a null detection for the magnetic field translates into zero value.

3. STATISTICAL METHODOLOGY AND ANALYSIS

Handling of the incomplete and truncated data was done by multiple imputation combined with acceptance–rejection sampling (Rubin 1987; Schafer 1997; Little & Rubin 2002; Carpenter & Kenward 2013). These are established techniques in medical statistics but less so in astrophysics. We thus explain first why we used that methodology and how we tuned it to our application before outlining the various steps involved in the analysis.

The great strength of multiple imputation is that it is a principled statistical method that takes proper account of the loss of information in the incomplete data. As with any statistical technique it does of course rest on assumptions. The role of these assumptions becomes more critical with increasing amounts of missing information. However, the connection between missing information and the amount of missing data is far from straightforward. For example, missing information also depends on the type of outcome variables, the patterns of missingness, and the dependence structure between what is observed and what is missing, and the model fitted. For these reasons, we have limited the data set to the variables listed in Table 1, where the poorest level of completeness occurs for the oscillation frequencies (X_3 and X_5 , which each have a coverage of 32%) but we have well-determined upper limits for their amplitudes from the Hipparcos data (X_4 and X_6).

Because data come from various sources, there is heterogeneity in the error with which they are measured. Moreover, it is well known that good error propagation for spectroscopic quantities (X_1 , X_7 , X_8 , X_9 , and X_{10}) is a difficult problem due to the possible dominance of systematic uncertainties over statistical errors. These systematic uncertainties result from a combination of instrument calibrations, varying atmospheric

conditions, spectrum normalization uncertainties, and limitations in the theory of spectral line predictions. Given that our analysis is based on multiple regression, in which one works conditionally on the values observed for the explanatory variables, such heterogeneity is not explicitly accommodated. The implication of this is that estimated relationships may be somewhat attenuated (Carroll et al. 2010), but the logic of the analysis is not undermined. In particular, ignorance of the measurement errors will not create artifacts such as non-existing relationships.

3.1. Incomplete Data: Missingness and Truncation

Let $X_i = (X_{i1}, \dots, X_{ip})^T$ be the column vector of p measurements for star $i = 1, \dots, n$. The data are commonly organized in a data set $X = (X_1^T, \dots, X_n^T)^T$ that takes the form of a rectangular matrix of dimension $n \times p$. When the data are complete, a wide variety of statistical models can be fitted to them. As an example, consider a regression model

$$X_{i1} = \beta_1 + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad (1)$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is a vector of unknown regression coefficients and the error term ε_i is assumed to follow a distribution with mean zero and variance σ^2 . Using conventional methodology and standard statistical software, the parameters β and σ^2 can then be estimated, along with measures of precision, i.e., confidence intervals.

In the current problem, we are faced with two types of incompleteness. First, some values X_{ij} are entirely missing. Second, for some values X_{ij} bounds are available per individual star, but not the actual value, i.e., the information is restricted to $\ell_{ij} \leq X_{ij} \leq u_{ij}$, with ℓ_{ij} (u_{ij}) a star-specific lower bound (upper bound). Note that, if only one bound is available for all stars simultaneously, then we use an additional superscript r to indicate that, i.e., either $\ell_{ij} = \ell_{ij}^r$ for one lower bound only or $u_{ij} = u_{ij}^r$ for one upper bound only, where $[\ell_{ij}^r, u_{ij}^r]$ is then the total range of the random variable X_{ij} . This occurs, e.g., due to the requirement for positive frequency values, leading to zero as lower bound for all stars for variables X_2 , X_3 , and X_5 . Similarly, we required the values for X_{10} to be bound by physically meaningful values as explained in the following section. In summary, it is possible to unify missingness and truncation by simultaneously setting both bounds to their range limits. Even a properly observed value can be placed within this setting: $\ell_{ij} = X_{ij} = u_{ij}$. Thus, the vectors ℓ_i and u_i fully describe the data available on star i .

Modeling such data has received a large amount of attention. Here, we opted for so-called multiple imputation (Rubin 1987; Schafer 1997; Little & Rubin 2002; Carpenter & Kenward 2013) combined with acceptance–rejection sampling. Multiple imputation consists of three steps. In the first or imputation step, the principle is to replace each missing value with M copies or so-called imputations. These are drawn from the predictive distribution of what is missing, given what is observed. Because values are drawn multiple times, rather than filled in once, the phenomenon that incomplete data lead to reduced statistical information is maintained, in contrast to single imputation. Let us write the model to be fitted symbolically as

$$f(\mathbf{x}_i | \boldsymbol{\theta}), \quad (2)$$

where \mathbf{x}_i is the realized value of X_i and $\boldsymbol{\theta}$ groups all model parameters. The modeler thus obtains M completed data sets. In the second or modeling step, each of these is analyzed

separately, as if the data were complete. Thus, M estimates of $\boldsymbol{\theta}$, are obtained. We denote these by $\hat{\boldsymbol{\theta}}_m$, with $m = 1, \dots, M$. The same is true for the corresponding measures of precision. Let the estimated variance-covariance matrix for $\hat{\boldsymbol{\theta}}_m$ be \hat{U}_m . In the third or analysis step, these M estimates are combined into a single set of parameter and precision estimates, using Rubin’s rules (Rubin 1987; Schafer 1997; Little & Rubin 2002; Carpenter & Kenward 2013):

$$\hat{\boldsymbol{\theta}} = \frac{1}{M} \sum_{m=1}^M \hat{\boldsymbol{\theta}}_m, \quad (3)$$

$$\hat{U} = \frac{1}{M} \sum_{m=1}^M U_m + \frac{M+1}{M(M-1)} \sum_{m=1}^M (\hat{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}})^T. \quad (4)$$

Step 2 requires fitting a model to a complete set of data, and to repeat this exercise M times; in step 3, Rubin’s rules (3) and (4) are applied. The most involved step is the first one. We turn to it next.

For our purposes in particular, the constraints make the use of multiple imputation non-standard, in contrast to when the only complication is missingness. Consider for our purposes, the general setting where X_i is a vector with three subvectors $X_i = (X_{i1}^T, X_{i2}^T, X_{i3}^T)^T$, where the notation implies that we first transpose the individual subvectors from column to row, place them all next to each other, and then turn it into a column once again, in such a way that X_{i1} is observed, X_{i2} is truncated with conditions $\ell_{i2} \leq X_{i2} \leq u_{i2}$, and X_{i3} is fully missing. Sampling is then needed from

$$f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \mathbf{c}_{i2}, \boldsymbol{\theta}) = f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \boldsymbol{\theta}) \cdot \frac{1}{f(\mathbf{c}_{i2} | \mathbf{x}_{i1}, \boldsymbol{\theta})}. \quad (5)$$

By contrast, should X_{i2} as well as X_{i3} be fully unobserved, then it would be sufficient to sample from $f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \boldsymbol{\theta})$, the first factor of Equation (5).

Assuming a multivariate normal distribution for the variables in the imputation model implies that every conditional distribution of one subset given another is still multivariate normal (Johnson & Wichern 2000; Carpenter & Kenward 2013, Appendix B). Hence, relying on this convenient property, the conditional density is also multivariate normal and takes the following form:

$$f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \boldsymbol{\theta}) = \phi_{p_2+p_3}(\mathbf{x}_{i,23} | \boldsymbol{\mu}_{23} + \Sigma_{23,1} \Sigma_{1,1}^{-1}(\mathbf{x}_{i1} - \boldsymbol{\mu}_1); \Sigma_{23,23} - \Sigma_{23,1} \Sigma_{1,1}^{-1} \Sigma_{1,23}), \quad (6)$$

where $\phi_{p_2+p_3}(\cdot; \cdot)$ is the multivariate normal density with dimension $p_2 + p_3$. In this notation, an index “23” indicates selection of the appropriate components of the full vector or matrix pertaining to the second and third subvector combined. Not only is this predictive distribution much simpler than Equation (5), expressions of this form are part of standard implementations of multiple imputation (Carpenter & Kenward 2013).

The above considerations suggest combining multiple imputation with so-called acceptance–rejection sampling (von Neumann 1951; Gilks et al. 1996; Robert & Casella 2004). Generally, when sampling from $h(\mathbf{x})$ is required, but sampling from $g(\mathbf{x})$ is much easier, then one can proceed by sampling from the latter density, provided that there is a value $M > 1$ such that $h(\mathbf{x})/g(\mathbf{x}) \leq M$. In our case, $h(\mathbf{x}_i) = f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \mathbf{c}_{i2}, \boldsymbol{\theta})$

while $g(\mathbf{x}_i) = f(\mathbf{x}_{i2}, \mathbf{x}_{i3} | \mathbf{x}_{i1}, \boldsymbol{\theta})$, and the ratio between both is $M = f(\mathbf{c}_{i2} | \mathbf{x}_{i1}, \boldsymbol{\theta})^{-1} \geq 1$.

Acceptance–rejection sampling operates in the following way. Draw \mathbf{X} from $g(\mathbf{x})$ and $U \sim U(0, 1)$, a uniform variable on the unit interval. Then, accept the draw if $U \leq h(\mathbf{x})/[Mg(\mathbf{x})]$ and reject it otherwise. Given the form of Equation (5), $h(\mathbf{x})/[Mg(\mathbf{x})] = 1$ when \mathbf{C}_{i2} is satisfied and 0 otherwise. Hence, thanks to the special situation posed by truncation, every draw that satisfies the constraint is always accepted, otherwise, it is always rejected.

Thus, in practice, multiple imputations are drawn from a multivariate distribution assuming that the missing values and the truncated values are all missing, but draws are accepted only if they satisfy the constraints \mathbf{C}_{i2} . The draws that are accepted then form the appropriate predictive distribution. Acceptance–rejection sampling can be very inefficient when $g(\cdot)$ and $h(\cdot)$ are very different, resulting in small proportions of acceptable draws. To improve the efficiency of our imputation procedure, we have additionally used transformations that automatically respect range restrictions whenever such restrictions are applicable to a variable for all stars simultaneously, rather than to values for some particular stars only. For example, an additional square root transformation was adopted to ensure non-negativity.

3.2. Principal Components

Principal component analysis (Krzanowski 1988; Johnson & Wichern 2000), abbreviated as PCA, is a classical exploratory method, based on rotating outcome vectors $\mathbf{Y}_i = L\mathbf{X}_i$ with the requirement that the components of the transformed vectors are uncorrelated, that the first one has maximal variance, the second one maximal variance given the first, etc. Ordinarily, PCA is conducted on standardized, hence unitless, input variables, \mathbf{X}_i . Technically, the transformation matrix L is the set of eigenvectors of the correlation matrix of the input variables \mathbf{X}_i . The variances of the transformed variables are found as the eigenvalues of the said correlation matrix. For example, the first principal component

$$\mathbf{Y}_{i1} = \ell_{11}\mathbf{X}_{i1} + \dots + \ell_{1p}\mathbf{X}_{ip} \quad (7)$$

is determined by the first eigenvector $\boldsymbol{\ell}_1 = (\ell_{11}, \dots, \ell_{1p})^T$ and its variance is λ_1 , the leading eigenvalue. The coefficients of $\boldsymbol{\ell}_1$ indicate the importance with which the observed variables \mathbf{X}_{ij} contribute to the first principal component, with the same logic applying to the other principal components. Further, λ_1/p is the fraction of the total variance in the original, standardized variables that is captured by the first principal components.

3.3. Application to the Selected Sample

For our application, variables X_1 , X_8 , and X_9 ($v \sin i$, $\log T_{\text{eff}}$, and $\log g$) are observed for all 68 stars, while all others have missing values (see Table 1). By definition, all variables must be positive and are thus truncated at zero as their lower limit, but we use the measured lower limits for X_7 (the magnetic field strength) when available. Furthermore, there is truncation as an upper limit on variables X_4 , X_6 , and X_{10} (the oscillation mode amplitudes and the nitrogen abundance). For X_{10} , we required physically meaningful results in that imputed values must be contained in [6.8, 9.0] following results achieved for the Milky Way and the Magellanic Clouds (Trundle et al. 2007; Rivero González et al. 2012).

We fit a linear regression model to the nitrogen abundance (X_{10}), with X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 , X_8 , and X_9 as potential predictors. Although obviously from a physical perspective a linear model does not necessarily reflect the underlying theory, it does have specific advantages that are appropriate in this setting. First, a linear regression does not rest on a priori theory and hence is neutral with respect to the specific relationships uncovered. Second, it is not the aim to establish an entire physical theory for the processes being modeled, but rather to bring to light important relationships that would be hard to unravel when considering only one variate at a time and that can then be explored further.

Twenty imputations were drawn, using the Monte Carlo Markov Chain method (Neter et al. 1996). The advantage of this method is that it can easily handle missingness of a non-monotone type, i.e., where arbitrary patterns of missingness occur among the stars. It is well-known that small numbers of imputations typically produce valid results, and Little & Rubin (2002) even quote five as a sufficient choice. This is relevant for very large databases, where drawing imputations is computationally costly, while here, the number of imputations is not really of concern, given that there are only 68 stars. On the other hand, acceptance–rejection sampling with bounds applying to the variables, necessitates drawing large numbers of imputations. For some stars, 5,000 imputations were needed to obtain the requested number of 20 non-rejected imputations. Even then, for 4 of the 68 stars, no valid draws could be found. These stars are indicated in Table 2. A sensitivity analysis was conducted to address this issue (see Section 3.4), as is common practice in statistical modeling.

A regression model was then fitted to the data by first fitting it to all 20 imputations and then combining the results using the rules outlined in Section 3.1. To select a well-fitting, parsimonious model, two routes were followed. In the first, backward selection (Neter et al. 1996), all nine predictor variables were included. Then, the least significant one was deleted, the model re-fitted, and the process repeated until only significant predictors remained. Significance is defined at the conventional five percent level (indicated as $p < 0.05$). It turns out that only two predictors (X_5 and X_8 , i.e., f_p and $\log T_{\text{eff}}$) are significant (top part of Table 3). In the second, forward selection starts from including only the most significant predictor at first, which is X_8 as can be seen from the top left-hand columns in Table 3. Thereafter, the model is re-fitted with X_8 included by default, paired with each one of the remaining predictors in turn. The most significant of these is again X_5 . The process was iterated but no further significant predictors could be added. While generally, backward and forward selection can lead to very different models, here the same two predictors for the nitrogen abundance (X_{10}) are selected with both approaches, i.e., the effective temperature (X_8) and the frequency of the dominant acoustic oscillation mode (X_5). The top part of Table 3 describes the results of the model selection processes.

Returning to the left-hand side of Table 3, these are the coefficients (with standard errors in brackets), and significance levels when the two predictors are considered one at a time. Qualitatively, X_5 , i.e., the frequency of the dominant acoustic mode, is a significant predictor on its own for the nitrogen abundance (X_{10}). The fraction of the variance explained by the model (denoted as R^2) for this single predictor ranges up to 36% for the nitrogen abundance. The effective temperature (X_8) explains up to 37% of the variance of X_{10} and is also a significant predictor on its own.

Table 3
Parameter Estimates (Standard Errors), p -values, and the Fraction of the Variance Explained by the Model (R^2)

Pred.	Separate Models			Joint Model		
	Estimate (s.e.)	p	R^2 Range	Estimate (s.e.)	p	R^2 Range
Models for X_{10}						
Intercept	0.6744(1.9285)	...	
X_5	0.0622(0.0225)	0.0068	[0.05;0.36]	0.0595(0.0217)	0.0079	
X_8	1.5848(0.5009)	0.0018	[0.08;0.37]	1.5515(0.4428)	0.0006	[0.21;0.54]
Models for X_3						
Intercept	6.0323(2.2180)	...	
X_7	0.1801(0.0412)	<0.0001	[0.26;0.56]	0.1546(0.0386)	0.0001	
X_8	-1.8711(0.5888)	0.0018	[0.11;0.31]	-1.2723(0.5087)	0.0136	[0.34;0.65]

Notes. Left-hand columns: separate models with a single predictor; right-hand columns: joint model with significant predictors simultaneously included. The top part concerns models for the nitrogen abundance (X_{10}) and the bottom part for the frequency of the dominant gravity-mode oscillation (X_3). (Pred.: predictor variable).

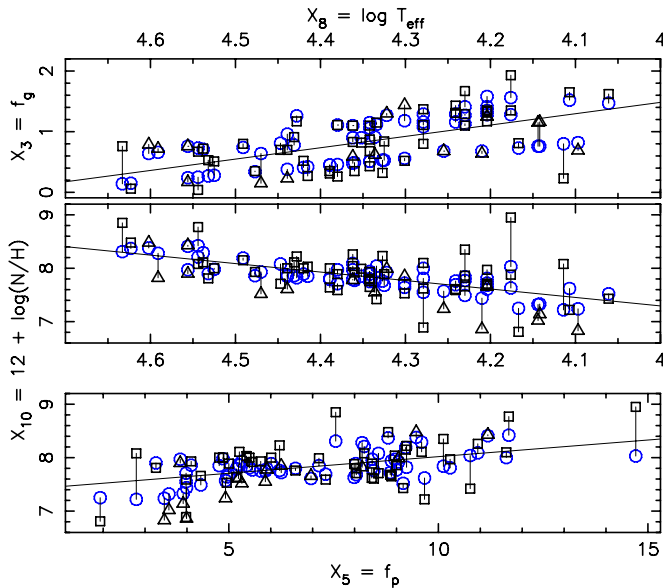


Figure 2. Observed or averaged imputed values (squares for single stars, triangles for spectroscopic binaries) connected by lines to the values predicted by the joint models described in Table 3 (circles, blue) for the gravity-mode frequency (X_3 , upper panel) and for the nitrogen abundance (X_{10} , lower two panels). The full lines represent the univariate model fits listed in the left part of Table 3.

(A color version of this figure is available in the online journal.)

The lower two panels of Figure 2 present an illustration of the quality of the joint model. The observed or averaged imputed values of X_{10} (squares for single stars, triangles for spectroscopic binaries) are connected by lines to those predicted by the joint model for X_{10} given in the upper part of Table 3 (circles), as a function of X_8 (middle panel) and of X_5 (lower panel). Although the joint model is satisfactory, the corresponding fraction of the variance explained by the model across the 20 imputations is at best 54%.

The results of our statistical modeling lead to the following hypotheses. The acoustic mode frequencies are a measure of the mean density of the star (e.g., Aerts et al. 2010). Following the mass–radius relation that holds during core-hydrogen burning, the acoustic frequencies decrease as the mass of the star increases. On the other hand, the effective temperature increases

as the mass increases. From the coefficients of the joint model for the nitrogen abundance in Table 3 and the range of values for X_5 and X_8 as given in Table 1, we find that the effect of the temperature is dominant over the one of the acoustic mode. Thus, among the sample of OB stars considered here, the higher-mass O stars have more nitrogen enrichment than the lower-mass B stars, as also found in the study by Rivero González et al. (2012), but the presence of acoustic modes further seems to increase the nitrogen abundance and this increase due to the oscillations gets larger as we move from the early O stars to the late B stars.

To investigate inter-relationships among the variables as well as the variance in the data set as a whole, without focusing solely on the nitrogen abundance, we performed a PCA (Krzanowski 1988; Johnson & Wichern 2000), as discussed in Section 3.2. This revealed evidence of two relationships: a first one between the dominant gravity-mode frequency (X_3), the magnetic field strength (X_7), the effective temperature (X_8), and the rotational variables (X_1 and X_2) explaining 29% of the variance in the data set and a second one between the nitrogen abundance (X_{10}), some of the mode properties (X_4 , X_5) and the effective temperature (X_8) explaining 19% of the variance. It is reassuring that the joint model for the nitrogen abundance is recovered in this way from the second principal component. Following the first principal component, a joint model analysis was repeated for variable X_3 (f_g) and led to the results in bottom part of Table 3: the lower the effective temperature and the stronger the magnetic field, the higher the gravity-mode frequency, while the rotational frequency and projected rotational velocity were found to be insignificant as predictors for X_3 .

The upper panel of Figure 2 is an illustration of the quality of this joint model for X_3 . Again, the observed or averaged imputed values of X_3 are connected to those predicted by the joint model given in the lower part of Table 3 (circles), as a function of X_8 . In this case, the joint model across the 20 imputations explains up to 65% of the variance in the frequency of the dominant gravity mode. The increase of the gravity-mode frequencies with a decrease of the effective temperature was already reported by De Cat & Aerts (2002, their Figure 18) from their much smaller sample of slowly pulsating B stars, which covers only a very narrow range in effective temperature compared to the sample we composed here. In addition, we find a correlation between the gravity-mode frequency and the magnetic field

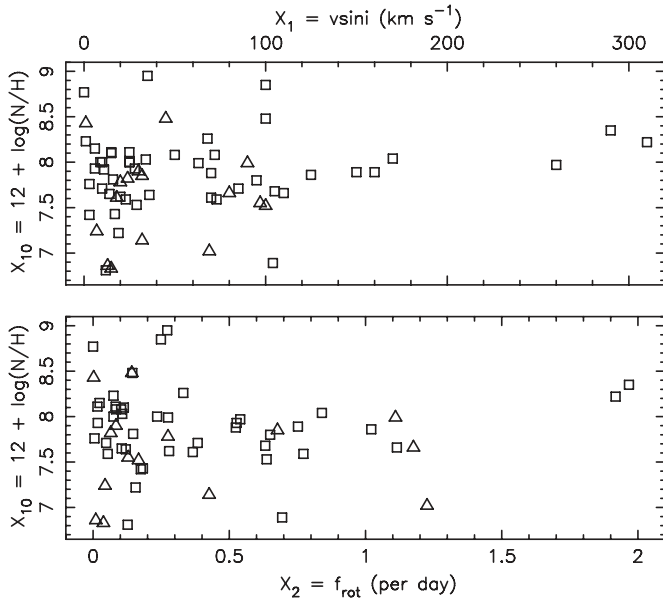


Figure 3. Observed or averaged imputed values (squares for single stars, triangles for spectroscopic binaries) for the nitrogen abundance (X_{10}) as a function of $v \sin i$ and the rotation frequency (X_1 and X_2).

strength. As far as we are aware, such a connection has not been reported before from observational diagnostics, although tight relationships between low-frequency gravity modes and the magnetic field strength have been presented in theoretical work (e.g., Hermans et al. 1990). Of course, we must keep in mind that the spectro-polarimetric measurements upon which we relied in this work reveal only the topology and strength of the magnetic field in the outer layers of the star while the internal properties may be quite different. Moreover, the relationship we find results directly from the imputed values of the magnetic field strength and gravity-mode frequency. Indeed, it cannot be revealed without imputation, because very few stars with both gravity-mode oscillations and a positive magnetic field strength have been detected. In our sample, e.g., 20 stars have both magnetic field measurements and detected gravity modes, but for 18 of those a null detection for the magnetic field strength occurs. Hence few gravity-mode oscillators have a magnetic field, but the imputed values for X_7 for all stars suggest a correlation between X_3 and X_7 which remains to be checked in the future from new magnetic gravity-mode pulsators yet to be discovered.

3.4. Sensitivity Analysis

The fact that four stars were removed from the joint analyses out of necessity needs to be addressed. We performed a so-called sensitivity analysis by also considering multiple imputation and acceptance–rejection without requiring zero lower bounds and without transforming the data to their square-root value. This allowed all 68 stars to be included and led to very similar results, although the frequency of rotation (X_2) or of the gravity mode (X_3) for some of these stars were assigned negative imputed values.

4. DISCUSSION

The theory of single-star evolution relying on rotational mixing predicts that the surface nitrogen abundance of a star (X_{10}) should strongly depend on its equatorial rotation velocity. Hence, recent observational studies aimed to test the theoretical

predictions mainly considered the observed projected rotational velocity (X_1) and corrected it by assuming equal probability of all possible inclination angles to take away the dependence on the unknown factor, $\sin i$ (e.g., Hunter et al. 2008; Brott et al. 2011; Rivero González et al. 2012; Bouret et al. 2012, 2013). Those studies led to the conclusion that the theoretical models fail to explain the observed values of the nitrogen abundance. Here, we come to a similar conclusion, by relying on a multitude of observed stellar properties of 64 galactic OB stars. In particular, our sample contains a majority of stars with $v \sin i$ below 100 km s^{−1} while their nitrogen abundance ranges from about 7 to almost 9 dex. We deduce that neither the projected rotational velocity (X_1) nor the rotational frequency (X_2) has predictive power for the measured nitrogen abundance. This is illustrated graphically in Figure 3. Comparison of the middle and lower panels of Figure 2 and both panels of Figure 3 visually shows that the rotational parameters are indeed less suitable as predictors of the nitrogen abundance compared to the effective temperature and the dominant acoustic oscillation mode frequency.

In fact, our results imply that none of the nine variables considered here is by itself able to predict more than a third of the variance in the nitrogen abundance. The joint bivariate model developed for the nitrogen abundance listed in Table 3, i.e.,

$$12 + \log[N/H] = 0.6744(1.9285) + 0.0595(0.0217) \cdot f_p + 1.5515(0.4428) \cdot \log T_{\text{eff}} \quad (8)$$

offers an appropriate *yet still incomplete* comparison between theory and observations, for the range of stellar parameters in the considered galactic sample.

We come to the conclusion that mixing must occur already early on in the core-hydrogen burning phase for a considerable fraction of massive stars and that it cannot be due to rotational mixing alone. The same conclusion was reached by an independent study of LMC O stars by Rivero González et al. (2012). Other dynamical phenomena causing mixing must thus be considered. From the statistical modeling based on the observed quantities treated in this work, heat-driven oscillation mode frequencies are a more suitable predictor for mixing compared to the magnetic field strength and the rotation in the sample of stars we studied here. Hence we suggest including the oscillation properties into future evaluations of theoretical model predictions, e.g., by means of Equation (8), rather than or in addition to the rotational properties.

Whether our conclusions hold for evolved massive stars and/or for metal-poor stars cannot be checked at present, because the asteroseismic data needed for such a study is lacking. Unfortunately, relatively few massive stars have been monitored sufficiently long in high-precision space photometry because such objects hardly occurred in the fields-of-view of the *CoRoT* and *Kepler* space missions, given that these satellites focused on exoplanet hunting around faint cool stars in rather crowded fields, for which the presence of bright massive stars is a nuisance. It will thus require large dedicated and coordinated ground-based efforts to remedy this situation, because long-term monitoring is necessary to reveal and unravel the oscillation properties of evolved OB stars. The potential of such studies from space photometry is clear and it could be tackled with the *Kepler2.0* mission concept (see, e.g., Aerts et al. 2013b) and with the future PLATO2.0 mission (Rauer et al. 2013).

Apart from the observed acoustic and gravity modes excited by a heat-mechanism in the stellar envelope upon which we

relied in this work, large-scale, low-frequency internal gravity waves excited by the convective core, as in the studies by Browning et al. (2004), Neiner et al. (2012b), Rogers et al. (2013), and Mathis et al. (2013) may contribute considerably to the surface nitrogen abundance if these waves are able to propagate through the acoustic mode cavity and to transport chemical species in addition to angular momentum. So far, theoretical or simulation studies focused on the angular momentum transport of such internal gravity waves but it should be well possible to extend those studies to the chemical transport they may induce. The frequency spectrum of internal gravity waves computed from two-dimensional simulations for a $3 M_{\odot}$ star by Rogers et al. (2013) coincides with the frequency range of running waves expected for OB stars and leads to the conclusion that the convective penetration depth *decreases* as a function of the rotation frequency. This result is in disagreement with the one based on three-dimensional simulations for a $2 M_{\odot}$ star by Browning et al. (2004), who finds the penetration depth to be *increasing* with rotation frequency (see Mathis 2013 for a discussion of the origin of this discrepancy). Clearly, a better understanding of such type of simulations for stars with a convective core is needed. Nevertheless, both simulation studies do find a range of the penetration depth and of core overshooting in agreement with the findings from asteroseismology based on standing waves (Aerts 2014). Moreover, these simulation studies show that *internal gravity waves transport angular momentum, thereby decreasing the level of differential rotation and hence counteract the effect of rotational mixing as the stellar evolution progresses*. We therefore consider the simultaneous act of internal gravity waves and differential rotation to be the best explanation for the measured surface nitrogen abundances in massive stars, despite the fact that the chemical transport that such waves may induce remains to be computed. At present, direct observational velocity diagnostics resulting from internal gravity waves are at best speculative in terms of macroturbulent spectral line broadening (e.g., Aerts et al. 2009; Simón-Díaz et al. 2010). As soon as velocity signatures due to internal gravity waves can be established with certainty, they can be added as variables for joint statistical modeling, in addition to the signatures of heat-driven acoustic and gravity-mode oscillations that were used in the current work.

The authors are grateful to the MiMeS collaboration for having provided magnetic field null detections prior to publication. They also acknowledge the numerous constructive comments by the referee, Dr. Joachim Puls, which helped to improve the presentation of this work appreciably. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement No. 227224 (PROSPERITY) and from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy). Conny Aerts is grateful to the Lorentz Center in Leiden, the Netherlands, for providing a stimulating research environment by means of its workshop program; the idea for this study found its origin during the workshop “Steps Towards a New Generation of Stellar Models.”

REFERENCES

- Aerts, C. 2014, in *Setting a New Standard in the Analysis of Binary Stars*, EAS Publication Series, in press (arXiv:1311.6242)
- Aerts, C., Briquet, M., Degroote, P., Thoul, A., & van Hoolst, T. 2011, *A&A*, **534**, A98
- Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, *Asteroseismology* (Astronomy and Astrophysics Library; Berlin: Springer)
- Aerts, C., & De Cat, P. 2003, *SSRv*, **105**, 453
- Aerts, C., De Cat, P., Cuypers, J., et al. 1998, *A&A*, **329**, 137
- Aerts, C., De Cat, P., Kuschnig, R., et al. 2006, *ApJL*, **642**, L165
- Aerts, C., De Cat, P., Peeters, E., et al. 1999, *A&A*, **343**, 872
- Aerts, C., Lehmann, H., Briquet, M., et al. 2003a, *A&A*, **399**, 639
- Aerts, C., Puls, J., Godart, M., & Dupret, M.-A. 2009, *A&A*, **508**, 409
- Aerts, C., Simón-Díaz, S., Catala, C., et al. 2013a, *A&A*, **557**, A114
- Aerts, C., Thoul, A., Daszyńska, J., et al. 2003b, *Sci*, **300**, 1926
- Aerts, C., Waelkens, C., Daszyńska-Daszkiewicz, J., et al. 2004, *A&A*, **415**, 241
- Aerts, C., Zwintz, K., Marcos-Arenal, P., et al. 2013b, arXiv:1309.3042
- Alecian, E., Kochukhov, O., Neiner, C., et al. 2011, *A&A*, **536**, L6
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
- Blomme, R., Mahy, L., Catala, C., et al. 2011, *A&A*, **533**, A4
- Bohlender, D. A., & Landstreet, J. D. 1990, *ApJ*, **358**, 274
- Bohlender, D. A., & Monin, D. 2011, *AJ*, **141**, 169
- Bohlender, D. A., Rice, J. B., & Hechler, P. 2010, *A&A*, **520**, A44
- Bolton, C. T., Harmanec, P., Lyons, R. W., Odell, A. P., & Pyper, D. M. 1998, *A&A*, **337**, 183
- Bouret, J.-C., Hillier, D. J., Lanz, T., & Fullerton, A. W. 2012, *A&A*, **544**, A67
- Bouret, J.-C., Lanz, T., Martins, F., et al. 2013, *A&A*, **555**, A1
- Briquet, M., Aerts, C., Baglin, A., et al. 2011, *A&A*, **527**, A112
- Briquet, M., Aerts, C., Lüftinger, T., et al. 2004, *A&A*, **413**, 273
- Briquet, M., Hubrig, S., Schöller, M., & De Cat, P. 2007a, *AN*, **328**, 41
- Briquet, M., Lefever, K., Uytterhoeven, K., & Aerts, C. 2005, *MNRAS*, **362**, 619
- Briquet, M., Morel, T., Thoul, A., et al. 2007b, *MNRAS*, **381**, 1482
- Briquet, M., Neiner, C., Aerts, C., et al. 2012, *MNRAS*, **427**, 483
- Briquet, M., Neiner, C., Leroy, B., & Pápics, P. I. 2013, *A&A*, **557**, L16
- Briquet, M., Uytterhoeven, K., Morel, T., et al. 2009, *A&A*, **506**, 269
- Brott, I., Evans, C. J., Hunter, I., et al. 2011, *A&A*, **530**, A116
- Browning, M. K., Brun, A. S., & Toomre, J. 2004, *ApJ*, **601**, 512
- Bychkov, V. D., Bychkova, L. V., & Madej, J. 2005, *A&A*, **430**, 1143
- Carpenter, J. R., & Kenward, M. G. 2013, *Multiple Imputation and its Application* (Chichester: John Wiley & Sons)
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. 2010, *Measurement Error in Nonlinear Models* (2nd ed.; Boca Raton, FL: CRC/Chapman & Hall)
- Chapellier, E., Le Contel, J. M., Le Contel, D., Sareyan, J. P., & Valtier, J. C. 1995, *A&A*, **304**, 406
- Cidale, L. S., Arias, M. L., Torres, A. F., et al. 2007, *A&A*, **468**, 263
- Cuypers, J., Aerts, C., Buzasi, D., et al. 2002, *A&A*, **392**, 599
- De Cat, P., & Aerts, C. 2002, *A&A*, **393**, 965
- De Cat, P., Briquet, M., Aerts, C., et al. 2007, *A&A*, **463**, 243
- De Cat, P., Briquet, M., Daszyńska-Daszkiewicz, J., et al. 2005, *A&A*, **432**, 1013
- Degroote, P., Aerts, C., Michel, E., et al. 2012, *A&A*, **542**, A88
- Degroote, P., Briquet, M., Auvergne, M., et al. 2010, *A&A*, **519**, A38
- Desmet, M., Briquet, M., Thoul, A., et al. 2009, *MNRAS*, **396**, 1460
- Donati, J.-F., Howarth, I. D., Jardine, M. M., et al. 2006, *MNRAS*, **370**, 629
- Fourtune-Ravard, C., Wade, G. A., Marcolino, W. L. F., et al. 2011, in *IAU Symp. 272*, ed. C. Neiner, G. Wade, G. Meynet, & G. Peters (Cambridge: Cambridge Univ. Press), **180**
- Fraser, M., Dufton, P. L., Hunter, I., & Ryans, R. S. I. 2010, *MNRAS*, **404**, 1306
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. 1996, *Markov Chain Monte Carlo in Practice* (London: Chapman & Hall)
- Grunhut, J. H., Rivinius, T., Wade, G. A., et al. 2012, *MNRAS*, **419**, 1610
- Guinan, E. F., Ribas, I., Fitzpatrick, E. L., et al. 2000, *ApJ*, **544**, 409
- Handler, G., Jerzykiewicz, M., Rodríguez, E., et al. 2006, *MNRAS*, **365**, 327
- Handler, G., Matthews, J. M., Eaton, J. A., et al. 2009, *ApJL*, **698**, L56
- Handler, G., Shobbrook, R. R., Jerzykiewicz, M., et al. 2004, *MNRAS*, **347**, 454
- Handler, G., Shobbrook, R. R., & Mokgwetsi, T. 2005, *MNRAS*, **362**, 612
- Handler, G., Shobbrook, R. R., Uytterhoeven, K., et al. 2012, *MNRAS*, **424**, 2380
- Henrichs, H. F., de Jong, J. A., Verdugo, E., et al. 2013, *A&A*, **555**, A46
- Henrichs, H. F., Kolenberg, K., Plaggenborg, B., et al. 2012, *A&A*, **545**, A119
- Henrichs, H. F., Neiner, C., Schnerr, R. S., et al. 2009, in *IAU Symp. 259, Cosmic Magnetic Fields: From Planets, to Stars and Galaxies*, ed. K. G. Strassmeier, A. G. Kosovichev, & J. E. Beckman (Cambridge: Cambridge Univ. Press), **393**
- Hermans, D., Goossens, M., & Kerner, W. 1990, *A&A*, **231**, 259
- Herrero, A., Kudritzki, R. P., Vilchez, J. M., et al. 1992, *A&A*, **261**, 209
- Hubrig, S., Briquet, M., De Cat, P., et al. 2009, *AN*, **330**, 317
- Hubrig, S., Briquet, M., Schöller, M., et al. 2006, *MNRAS*, **369**, L61

- Hunter, I., Brott, I., Lennon, D. J., et al. 2008, *ApJL*, **676**, L29
- Johnson, R. A., & Wichern, D. W. 2000, *Applied Multivariate Statistical Analysis* (4th ed.; Englewood Cliffs, NJ: Prentice-Hall)
- Kaper, L., Henrichs, H. F., Nichols, J. S., et al. 1996, *A&AS*, **116**, 257
- Kochukhov, O., Lundin, A., Romanyuk, I., & Kudryavtsev, D. 2011, *ApJ*, **726**, 24
- Koen, C., & Eyer, L. 2002, *MNRAS*, **331**, 45
- Krzanowski, W. J. 1988, *Principles of Multivariate Analysis* (Oxford: Clarendon)
- Langer, N. 1992, *A&A*, **265**, L17
- Lefever, K., Puls, J., Morel, T., et al. 2010, *A&A*, **515**, A74
- Leone, F., Bohlender, D. A., Bolton, C. T., et al. 2010, *MNRAS*, **401**, 2739
- Leone, F., & Manfre, M. 1997, *A&A*, **320**, 257
- Little, R. J. A., & Rubin, D. B. 2002, *Statistical Analysis with Missing Data* (New York: Wiley)
- Lyubimkov, L. S., Rachkovskaya, T. M., Rostopchin, S. I., & Lambert, D. L. 2002, *MNRAS*, **333**, 9
- Maeder, A. 2009, *Physics, Formation and Evolution of Rotating Stars* (Astronomy and Astrophysics Library, Berlin: Springer)
- Martins, F., Escolano, C., Wade, G. A., et al. 2012a, *A&A*, **538**, A29
- Martins, F., Mahy, L., Hillier, D. J., & Rauw, G. 2012b, *A&A*, **538**, A39
- Mathis, S. 2013, *Habilitation thesis*, Université Paris XI Orsay, France
- Mathis, S., Decressin, T., Eggenberger, P., & Charbonnel, C. 2013, *A&A*, **558**, A11
- Mazumdar, A., Briquet, M., Desmet, M., & Aerts, C. 2006, *A&A*, **459**, 589
- Meynet, G., Eggenberger, P., & Maeder, A. 2011, *A&A*, **525**, L11
- Mikulášek, Z., Krčička, J., Henry, G. W., et al. 2008, *A&A*, **485**, 585
- Molenberghs, G., & Kenward, M. G. 2007, *Missing Data in Clinical Studies* (Chichester: John Wiley & Sons)
- Morel, T., Hubrig, S., & Briquet, M. 2008, *A&A*, **481**, 453
- Neiner, C., Alecian, E., Briquet, M., et al. 2012a, *A&A*, **537**, A148
- Neiner, C., Floquet, M., Samadi, R., et al. 2012b, *A&A*, **546**, A47
- Neiner, C., Geers, V. C., Henrichs, H. F., et al. 2003, *A&A*, **406**, 1019
- Neiner, C., Landstreet, J. D., Alecian, E., et al. 2012c, *A&A*, **546**, A44
- Neter, J., Nachtsheim, C. J., Wasserman, W., & Kutner, M. H. 1996, *Applied Linear Statistical Models* (Homewood, IL: Richard D. Irwin Inc.)
- Nieva, M.-F., & Przybilla, N. 2012, *A&A*, **539**, A143
- Oksala, M. E., Wade, G. A., Townsend, R. H. D., et al. 2012, *MNRAS*, **419**, 959
- Pamyatnykh, A. A., Handler, G., & Dziembowski, W. A. 2004, *MNRAS*, **350**, 1022
- Pápics, P. I., Briquet, M., Auvergne, M., et al. 2011, *A&A*, **528**, A123
- Pápics, P. I., Briquet, M., Baglin, A., et al. 2012, *A&A*, **542**, A55
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, **192**, 3
- Paxton, B., Cantiello, M., Arras, P., et al. 2013, *ApJS*, **208**, 4
- Perryman, M. A. C. (ed.) 1997, in *The HIPPARCOS and TYCHO Catalogues. Astrometric and Photometric Star Catalogues Derived from the ESA HIPPARCOS Space Astrometry Mission* (ESA SP-1200), 1
- Petit, V., Massa, D. L., Marcolino, W. L. F., et al. 2011, *MNRAS*, **412**, L45
- Petit, V., Owocki, S. P., Wade, G. A., et al. 2013, *MNRAS*, **429**, 398
- Petit, V., Wade, G. A., Drissen, L., Montmerle, T., & Alecian, E. 2008, *MNRAS*, **387**, L23
- Potter, A. T., Chitre, S. M., & Tout, C. A. 2012, *MNRAS*, **424**, 2358
- Przybilla, N., Farnstein, M., Nieva, M. F., Meynet, G., & Maeder, A. 2010, *A&A*, **517**, A38
- Rauer, H., Catala, C., Aerts, C., et al. 2013, *ExA*, submitted (arXiv:1310.0696)
- Rivero González, J. G., Puls, J., Najarro, F., & Brott, I. 2012, *A&A*, **537**, A79
- Rivinius, T., Stahl, O., Baade, D., & Kaufer, A. 2003, *IBVS*, **5397**, 1
- Rivinius, T., Townsend, R. H. D., Kochukhov, O., et al. 2013, *MNRAS*, **429**, 177
- Robert, C. P., & Casella, G. 2004, *Monte Carlo Statistical Methods* (New York: Springer)
- Rogers, T. M., Lin, D. N. C., McElwaine, J. N., & Lau, H. H. B. 2013, *ApJ*, **772**, 21
- Romanyuk, I. I., & Kudryavtsev, D. O. 2008, *AstBu*, **63**, 139
- Rubin, D. B. 1976, *Biometrika*, **63**, 581
- Rubin, D. B. 1987, *Multiple Imputation for Nonresponse in Surveys* (New York: Wiley)
- Saesen, S., Briquet, M., & Aerts, C. 2006, *CoAst*, **147**, 109
- Schafer, J. L. 1997, *Analysis of Incomplete Multivariate Data* (London: Chapman & Hall)
- Shobbrook, R. R., Handler, G., Lorenz, D., & Mogorosi, D. 2006, *MNRAS*, **369**, 171
- Shultz, M., Wade, G. A., Grunhut, J., et al. 2012, *ApJ*, **750**, 2
- Silvester, J., Neiner, C., Henrichs, H. F., et al. 2009, *MNRAS*, **398**, 1505
- Silvester, J., Wade, G. A., Kochukhov, O., et al. 2012, *MNRAS*, **426**, 1003
- Simón-Díaz, S., Herrero, A., Uytterhoeven, K., et al. 2010, *ApJL*, **720**, L174
- Stankov, A., & Handler, G. 2005, *ApJS*, **158**, 193
- Talon, S., Zahn, J.-P., Maeder, A., & Meynet, G. 1997, *A&A*, **322**, 209
- Telting, J. H., Aerts, C., & Mathias, P. 1997, *A&A*, **322**, 493
- Thoul, A., Aerts, C., Dupret, M. A., et al. 2003, *A&A*, **406**, 287
- Thoul, A., Degroote, P., Catala, C., et al. 2013, *A&A*, **551**, A12
- Tkachenko, A., Degroote, P., Aerts, C., et al. 2013, *MNRAS*, in press (arXiv:1312.3601)
- Townsend, R. H. D., Oksala, M. E., Cohen, D. H., Owocki, S. P., & ud-Doula, A. 2010, *ApJL*, **714**, L318
- Trundle, C., Dufton, P. L., Hunter, I., et al. 2007, *A&A*, **471**, 625
- Venn, K. A., & Lambert, D. L. 2005, in *ASP Conf. Ser. 336, Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis*, ed. T. G. Barnes, III & F. N. Bash (San Francisco, CA: ASP), 93
- Villamariz, M. R., Herrero, A., Becker, S. R., & Butler, K. 2002, *A&A*, **388**, 940
- von Neumann, J. 1951, *Nat. Bureau Standards*, **12**, 36
- Wade, G. A., Bohlender, D. A., Brown, D. N., et al. 1997, *A&A*, **320**, 172
- Wade, G. A., Grunhut, J., Gräfener, G., et al. 2012, *MNRAS*, **419**, 2459
- Wade, G. A., Howarth, I. D., Townsend, R. H. D., et al. 2011, *MNRAS*, **416**, 3160
- Yakunin, I., Romanyuk, I., Kudryavtsev, D., & Semenko, E. 2011, *AN*, **332**, 974