

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325290931>

An empirical examination of WISE /NEOWISE asteroid analysis and results

Article in Icarus · May 2018

DOI: 10.1016/j.icarus.2018.05.004

CITATIONS

11

READS

532

1 author:



Nathan Myhrvold

Intellectual Ventures

163 PUBLICATIONS 1,495 CITATIONS

[SEE PROFILE](#)

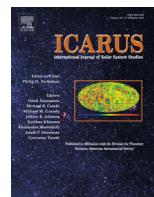
Some of the authors of this publication are also working on these related projects:



The Overdrawn Risks of Artificial Intelligence [View project](#)



Dinosaur growth and metabolism [View project](#)



An empirical examination of WISE/NEOWISE asteroid analysis and results

Nathan Myhrvold

Intellectual Ventures, Bellevue, WA 98005 USA



ARTICLE INFO

Article history:

Received 21 November 2016

Revised 2 May 2018

Accepted 8 May 2018

Available online 22 May 2018

Keywords:

Asteroids

Near-Earth objects

NEATM

WISE

NEOWISE

ABSTRACT

Asteroid observations by the *WISE* space telescope and the analysis of those observations by the NEOWISE project have provided more information about the diameter, albedo, and other properties of approximately 164,000 asteroids, more than all other sources combined. The raw data set from this mission will likely be the largest and most important such data on asteroids available for many years. To put this trove of data to productive use, we must understand its strengths and weaknesses, and we need clear and reproducible methods for analyzing the data set. This study critically examines the *WISE* observational data and the NEOWISE results published in both the original papers and the NASA Planetary Data System (PDS). There seem to be multiple areas where the analysis might benefit from improvement or independent verification. The NEOWISE results were obtained by the application of 10 different modeling methods, many of which are not adequately explained or even defined, to 12 different combinations of *WISE* band data. More than half of NEOWISE results are based on a single band of data. The majority of curve fits to the data in the NEOWISE results are of poor quality, frequently missing most or all of the data points on which they are based. Complete misses occur for about 30% of single-band results, and among the results derived from the most common multiple-band combinations, about 43% miss all data points in at least one band. The NEOWISE data analysis relies on assumptions that are in many cases inconsistent with each other. A substantial fraction of *WISE* data was systematically excluded from the NEOWISE analysis. Building on methods developed by Hanuš et al. (2015), I show that error estimates for the *WISE* observational data were not well characterized, and all observations have true uncertainty at least a factor of 1.3–2.5 times larger than previously described, depending on the band. I also show that the error distribution is not well fit by a normal distribution. These findings are important because the Monte Carlo error-analysis method used by the NEOWISE project depends on both the observational errors and the normal distribution. An empirical comparison of published NEOWISE diameters to those in the literature that were estimated by using radar, occultation, or spacecraft (ROS) measurements shows that, for 129 results involving 105 asteroids, the NEOWISE diameters presented in tables of thermal-modeling results exactly match prior ROS results from the literature. While these are only a tiny fraction (0.06%) of the asteroids analyzed, they are important because they represent the only independent check on NEOWISE diameter accuracy. After removing the exact matches and adding additional ROS results, I find that the accuracy of diameter estimates for NEOWISE results depends strongly on the choice of data bands and on which of the 10 models was used. I show that systematic errors in the diameter estimates are much larger than previously described and range from –5% to +23%. In addition, random errors range from –15% to +19% when all four *WISE* bands were used, and from –39% to +57% in cases employing only the W2 band. The empirical results presented here show that much work remains to be done in analyzing data from the *WISE*/NEOWISE mission and interpreting it for asteroid science.

© 2018 The Author. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license.
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Infrared (IR) observations by space telescopes have generated information about asteroids that is unique or hard to obtain via other means, including estimates of their diameters and, when

E-mail address: nathan@nathanmyhrvold.com

Table 1

Summary of published NEOWISE papers and numbers of asteroids analyzed in each. FC: primary papers that analyze *WISE* observations from the full-cryo (FC) mission phase. 3B+PC: primary papers analyzing observations from the three-band (3B) and post-cryo (PC) phases. Re: primary papers from the reactivation mission. Other: non-NEOWISE papers that estimate diameters by thermal modeling. For the FC papers, the NEOWISE group reported model fits that estimate the parameters D , H , p_v , p_{IR} , η ; where noted, the studies may determine separate W1 and W2 albedos p_{IR1} , p_{IR2} . Note that many asteroids appear in multiple NEOWISE studies or other source studies. The count column contains the number of asteroids for which thermal modeling results were presented in the paper.

Source Studies				
	Reference	Abbreviation here (abbreviation in PDS)	Topic	Count
FC	(Mainzer et al., 2014b)	Mainzer/NEO:Tiny (Mai14)	Tiny near-Earth objects (NEO)	106
	(Grav et al., 2011a)	Grav/JT:Pre (Gr11)	Jovian Trojans	1742
	(Grav et al., 2011b)	Grav/Hilda (Gr12a)	Hildas	1023
	(Grav et al., 2012)	Grav/JT:Tax (Gr12b)	Jovian Trojans p_{IR1} , p_{IR2}	478
	(Mainzer et al., 2011c)	Mainzer/TMC	Thermal-model parameters	0
	(Mainzer et al., 2011b)	Mainzer/NEO:Pre (Mai11)	NEO	428
	(Masiero et al., 2011)	Masiero/MB:Pre (Mas11)	Main-belt asteroids (MBA)	129,478
	(Masiero et al., 2014)	Masiero/MB:NIR (Mas14)	MBA p_{IR1} , p_{IR2}	2835
3B + PC	(Mainzer et al., 2012a)	Mainzer/PP:3	MBA, NEO	116
	(Masiero et al., 2012)	Masiero/MB:3 (Mas12)	MBA	13,669
Re	(Nugent et al. 2015)	Nugent/Re1 (Nug15)	1st year	7956
	(Nugent et al., 2016)	Nugent/Re2	2nd year	9092
Other	(Tedesco et al., 2002)	IRAS	IRAS	2228
	(Ryan and Woodward, 2010)	RW	STM, NEATM	118
	(Usui et al., 2014, 2012, 2011a, 2011b)	AKARI	AKARI	4844

coupled with optical observations, of their visible albedos, as recently reviewed by Mainzer et al. (2015b).

WISE, a medium-class space telescope deployed by NASA, observed in four bands: W1, W2, W3, and W4, centered on the wavelengths 3.4, 4.6, 12, and 22 μ m, respectively (Wright et al., 2010). NASA's NEOWISE project added to *WISE* the post-processing capability necessary to identify and observe asteroids and other small bodies in the solar system (Mainzer et al., 2011a). *WISE* and NEOWISE data are available for download from the Infrared Science Archive (IRSA)/*WISE* image archive (NASA/IPAC Infrared Science Archive, 2017).

The NEOWISE project used asteroid thermal modeling to estimate the diameters and albedos of about 164,000 asteroids, far more than all previous studies combined by more than a factor of 16. The enormous *WISE*/NEOWISE data (both raw observations and modeled results) is a treasure trove of data for planetary science and will be critically important for understanding asteroids for many years to come. No current or planned space mission will produce more four-band IR observations of asteroids.

This study was undertaken to independently examine the results and methodology published as part of the NEOWISE project in order to assess how best to use this incredible resource. It is vitally important that the astronomical community understands which aspects of the NEOWISE analysis represent the best or final word on extracting astronomical information from the data, as well as where more research work remains to be done.

Table 1 lists the publications by the NEOWISE group that include fits to thermal models, notably the Near-Earth Asteroid Thermal Model (NEATM), as well as resulting asteroid physical properties—including diameter D , visible-band albedo p_v , and near-infrared albedo covering the W1 and/or W2 bands (p_{IR} , p_{IR1} , p_{IR2}). These results are available from the electronic archives of the journals in which the papers were published. For convenience, I refer to the NEOWISE studies listed in Table 1 collectively as the “NEOWISE papers.” Although the papers were published by collaborations that include members of the NEOWISE group, some collaborators and some aspects of the work may have been performed outside the official scope of the NASA/JPL NEOWISE project. Table 1 also includes other papers that report asteroid physical parameter estimates used in this study but that did not make use of *WISE*/NEOWISE data.

Together, the suite of NEOWISE papers asserts that their estimates of asteroid diameter, visible albedo, and infrared albedo are relatively precise in the majority of cases. Some of the NEOWISE papers include caveats that errors can be higher in certain cases and that the results must be interpreted with caution. But the assertion by Masiero et al. (2011) that “Using a NEATM thermal model fitting routine, we compute diameters for over 100,000 main belt asteroids from their IR thermal flux, with errors better than 10%” (Masiero et al., 2011) is a claim typical of the papers. The claim is repeated in the NASA Planetary Data System (PDS) archive of NEOWISE results (Mainzer et al., 2016) documentation and is explained in some detail in a 2015 review that stated:

NEATM-derived diameters generally reproduce measurements from radar, stellar occultations, and *in situ* spacecraft visits to within $\pm 10\%$, given multiple thermally dominated IR measurements that adequately sample an asteroid’s rotational light curve with good signal-to-noise ratio (SNR) and an accurate determination of distance from knowledge of its orbit (Mainzer et al., 2011c). It is worth noting that the accuracy of the diameters of objects used to confirm the performance of radiometric thermal models (such as radar or stellar occultations) is typically $\sim 10\%$. (Mainzer et al., 2015b)

These assertions imply that analysis of the *WISE*/NEOWISE data is complete, at least with respect to diameter, because it is already within or close to the tolerance of the best available comparison data (measurements from radar, occultation and spacecraft observations, denoted here as ROS).

Work outside the NEOWISE group to fully understand and exploit the data is still in its early stages, however. Seven years after the initial publication of the NEOWISE calculations, the results have yet to be replicated (i.e., physical properties obtained by model fits performed on the observational data) by any independent group. The question of replicability is important because numerous astronomers have relied on the NEOWISE results to draw conclusions about many salient topics in solar-system science (Bauer et al., 2013; Faherty et al., 2015; Mainzer et al., 2012b, 2012c, 2011e; Masiero et al., 2015a; 2015b, 2013; Nugent et al., 2012; Sonnett et al., 2015). The initial paper by Mainzer et al. (2011a) has been cited at least 270 times, and at

least 1400 papers mention or reference NEOWISE, according to a recent query of Google Scholar.

NEOWISE results have been applied to study the distribution of asteroid diameter and/or visible-band albedo values (Wright et al., 2016), including the distributions of the size and albedo of near-Earth objects (NEO). Those distributions in turn have strongly influenced discussion about a search for potentially hazardous Earth-impacting asteroids (Grav et al., 2015; Mainzer et al., 2015a; 2015b; Myhrvold, 2016), underscoring the importance of both minimizing and properly characterizing the inevitable error in such estimates.

The analysis of broad frequency distributions of asteroid sizes or albedos can tolerate a degree of uncertainty in individual estimates, but systematic biases pose greater problems. Studies of asteroid families or individual asteroids—such as those that apply more detailed and sophisticated thermophysical modeling than the simple thermal models considered here (Ali-Lagoa et al., 2013; Hanuš et al., 2015; Koren et al., 2015)—require higher accuracy and thoroughly characterized errors.

In the course of a project to find asteroids within the NEOWISE results which might be fast-tumbling asteroids and thus have a different spectrum than that predicted by the NEATM (Myhrvold, 2016), I attempted to understand and replicate the NEOWISE results. It soon became apparent that the NEOWISE data analysis procedure has not yet been documented in sufficient detail to allow replication. Other inconsistencies were also noticed, leading to the empirical examination presented here.

This study is organized into nine principal sections. Section 2 presents a summary of the WISE and NEOWISE data sources used in this study. Section 3 gives an overview of the NEOWISE results, which the NEOWISE group obtained by applying 10 very different modeling methods to 12 distinct combinations of WISE data bands. The accuracy of these methods varies markedly, both among methods and within each method as it is applied to the various band combinations. The NEOWISE results thus must be viewed as a collection of results from disparate methods applied to disparate data resources.

Of the 165,794 asteroid physical property results summarized here, more than half of them are based on data from a single WISE band; just 2% make use of all four bands of data, in large part because the NEOWISE analysis systematically discarded large amounts of data.

Section 4 reports the discovery that, in 129 cases involving 105 asteroids, the diameter estimates presented in NEOWISE papers as the result of thermal-model fitting exactly match diameter estimates from previous ROS sources. Most of these cases have been omitted from the more recent PDS archive. However, nine such cases, involving six asteroids, remain in the PDS and bear a model code describing them as having a diameter not determined by fitting (i.e., a diameter set as an input to the modeling rather than being an output).

Section 5 further develops work by Hanuš et al., (2015) that examined the small percentage of NEOWISE asteroid observations that occur in the regions where successive WISE frames overlap. The successively imaged asteroids provide an ideal data set for measuring the quality of estimated standard error in the resulting flux observations. I expand the analysis of Hanuš et al., which was limited to two WISE bands and 400 cases, to include all four bands and more than 100,000 cases. I find that the WISE pipeline systematically underestimated observational error by a large factor, which varies in magnitude by band. Moreover, the flux errors do not follow a Gaussian (normal) distribution. Examination of WISE observations of ≈ 1 million stars finds that this effect is not confined to asteroids and is not related to detector-edge effects.

Section 6 is an analysis of the accuracy of NEOWISE model fits to the WISE data. Surprisingly, I find that nearly a third of single-band NEOWISE curves miss every data point they are supposed to

fit in the sense that the residuals for the curve are all the same sign, so the curve passes either above all of the data points or below all of the data points. Of the multiband modeling results, only 48% to 58% pass among data points in each band used (i.e., have residuals that include positive and negative signs or are zero). Those fits that do pass among the data points include many that appear to be of poor quality.

Section 7 shows that nearly 15,000 NEOWISE results violate the mathematical definition relating diameter and visible-band geometric albedo to absolute visible-band magnitude H , to a degree greater than can be explained by numerical issues such as rounding. Some of the violations would cause more than a ten-fold change in p_v or as much as a fourfold change in D —evidence of serious mathematical inconsistency among the NEOWISE models that fit p_v . I also show that updating the values of H used in the NEOWISE analysis (presumably from 2010) to 2017 values from the Minor Planet Center (MPC) alters the estimate of p_v by 10% or more for over half of the NEOWISE results. Considering that values of H from MPC have a dispersion of ~ 0.3 magnitude when compared to precisely measured values (Pravec et al., 2012), this may be the largest source of uncertainty in derived albedo values.

Section 8 examines the error analysis used as the basis for the claim in the NEOWISE papers that results are accurate and finds systematic flaws at multiple levels of the analysis.

Section 9 compares the accuracy of NEOWISE diameter estimates to ROS results and two prior sets of thermal-modeling results from IRAS and Ryan and Woodward (2010) (abbreviated here as RW). Concordance among the estimates is found to vary strongly with the data bands analyzed and the modeling methods used. Across the various combinations of models and bands included in the PDS, the systematic error ranges from -7% to $+23\%$. In addition, the 68.27% confidence interval that accounts for random errors is calculated to span -13.8% to $+18.8\%$ when all four WISE bands were used, with wider confidence intervals (-23% to $+27\%$) when fewer bands were employed.

Section 10 discusses these findings and their implications on work that has been based on the NEOWISE results. Details of selected points and methods are presented in the Appendix and in the Supplementary Information (SI) document.

2. Sources of WISE/NEOWISE observations and results

The WISE mission operated in four distinct phases (Wright et al., 2010). The first, or full-cryo (FC), phase involved all four bands and resulted in observations of 147,667 asteroids. Low cryogen stores triggered a shift briefly to a three-band (3B) phase that collected observations in W1, W2, and W3 for 7850 asteroids. The third, post-cryo (PC), phase recorded observations from only the W1 and W2 bands. Recently, asteroid observations by the WISE satellite restarted after a hiatus as the NEOWISE reactivation mission (Mainzer et al., 2014a; Nugent et al., 2016; 2015). Together these two phases contributed two-band observations of 18,885 asteroids. Some asteroids were observed in multiple or even all of the mission phases. Also, within each study there may be multiple thermal modeling results for a given asteroid, due to the fact that the NEOWISE methodology separately analyzes 3- to 10-day epochs (see Section 3.6 and Section S4 of SI.)

WISE/NEOWISE observational data were obtained in March 2017 from a search of the MPC observations database (MPCAT-OBS) for observatory code C51. Observation times and positions were then used as search inputs to the IRSAs WISE catalogs (Using Data Tags in Journals, 2018) (All-Sky, 3-Band, Post-Cryo, and Reactivation data releases), with a tolerance of 2 s in time and 20 arcsec in position. The searches returned ~ 2.5 million observations: 1.8 million from the FC mission, 78,456 from 3B, 69,288 from PC, and 484,410 from the ongoing reactivation mission. This is substantially

the same process documented in the NEOWISE papers (for example Masiero/MB:Pre; see SI Section S4). JPL Horizons was used to obtain solar phase angle and distances from each asteroid to the telescope and to the Sun, by using the observation time.

The NEOWISE project processed selected observational data to produce model fits that generate estimates for four or five physical properties of each asteroid: D , p_V , p_{IR} (or individually p_{IR1} and p_{IR2}), and the NEATM beaming parameter η . Details on the differences in data-processing methods among NEOWISE papers can be found in SI Section 4.

Results for some asteroids appear in multiple NEOWISE papers. Masiero/MB:NIR, for example, reports on a subset of the asteroids that were studied earlier (with different model parameterization) in Masiero/MB:Pre. The 3B, PC, and Re mission phases revisited many asteroids observed in the FC mission. In addition, 15 asteroids appear in both Masiero/MB:Pre and Mainzer/NEO, and 3 asteroids appear in both Masiero/MB:Pre and Grav/Hilda. Because these groups are presumably mutually exclusive, these cases of overlap may reflect mis-categorization.

Ryan and Woodward (2010) published estimates derived using the NEATM and the Standard Thermal Model (STM) for 118 asteroids, of which 92 were found to have WISE/NEOWISE data in IRSA and to have FC + 3B + PC data. Prior to NEOWISE, the largest thermal-modeling efforts were IRAS, which has 1728 asteroids in common to both studies, and AKARI (Usui et al., 2014, 2012; 2011a, 2011b), which has 4844 asteroids in common with NEOWISE.

ROS asteroid diameters, estimated from radar, occultation, and spacecraft observations, are critical in determining the accuracy of the thermal-modeling estimates because they offer the only diameter estimates obtained by completely independent methods and observations. A list of ROS references used in this study appears in SI Table S5. For concision, I use “ROS asteroids” or “ROS diameters” to refer to cases where ROS diameter estimates are available.

Many of the papers published in 2011 with the initial large release of results from WISE/NEOWISE include “Preliminary” in their titles (Table 1). The publications of the NEOWISE group since that time, as well as those of the larger astronomical community, have focused primarily on applying these preliminary results or on the 3B and PC mission phases. Some papers in Table 1 have reanalyzed a small subset of asteroids using a different model (i.e., Masiero/MB:NIR, Grav/JT:Tax). For the majority of asteroids, however, the results from the preliminary papers have not been revised or superseded by any more recent reanalysis.

In 2016, the NEOWISE group compiled and archived data from some of the papers in Table 1 in the PDS as “NEOWISE diameters and albedos V1.0” (Mainzer et al., 2016). This data archive (available at <https://sbn.psi.edu/pds/resource/neowisediam.html>) and its associated documentation thus serves as an alternative source for many of the NEOWISE model fits. Note that Nugent/Re2 was too recent to be incorporated in the PDS archive, and as a consequence will not be further considered in the current study.

Given the large number of papers that relied on NEOWISE results as published in the original papers, it is important to check whether this most recent source contains the same results as the papers or whether it constitutes a new release with different results. The descriptions published with the data files in the PDS archive (Mainzer et al., 2016) clearly state that they are the same. The overall description of the data set states that it represents “a compilation of published diameters” and other properties from the earlier NEOWISE papers (see file dataset.cat). The documentation further clarifies that the PDS results “were collected entirely by referencing results published in the literature” (see file publithost.cat). Individual result records are also tagged with codes referencing the previous NEOWISE papers from which the results were compiled. These codes are documented as “Reference to the original publication of fitted parameters” in the .lbl file that accom-

panies each data file (for example, neowise_mainbelt.lbl for the main belt asteroids).

Model-fit results from Masiero/MB:Pre, the paper that reports the bulk of the NEOWISE results, and from Masiero/MB:NIR, which reanalyzes a subset, were compared with model fits in the PDS that cite that one of those papers as the source. The model results in the PDS attributed to Masiero/MB:Pre match the original publication and data archive of *Astrophysical Journal* (ApJ) for 125,706 asteroids, after allowing for different numerical precision in the PDS version. While this is the majority of the results, a comparison of PDS archive results to the ApJ published sources reveals numerous discrepancies. Those involving Masiero/MB:Pre and Masiero/MB:NIR are shown in Table 2. Since a given asteroid can have multiple results in a single study, the number of results (fits) is different than the asteroid count.

There are 3772 asteroids that have model fits were published in Masiero/MB:Pre and appear in ApJ, yet results attributed to that paper are missing from the PDS. Conversely, the PDS includes results for 1211 asteroids with Masiero/MB:Pre cited as the source, yet data on those asteroids did not appear in that paper in ApJ.

Because all of these heretofore unpublished results report being based on data from the W3 and W4 bands, the observations must have been made during the FC mission phase. MPC data sets were examined to check whether alternate asteroid names could be a source of the discrepancies, but that does not appear to be the case.

A similar comparison of Masiero/MB:NIR to the PDS found that 213 asteroids that have results in the ApJ version are missing from the PDS, and 7 asteroids that did not appear in the ApJ version have results in the PDS attributed to that study. Of these, 6 of the newly appearing asteroids have a zero observation count for band W4. This is significant because the text of Masiero/MB:NIR in ApJ states that a requirement for inclusion in that study is to have WISE observations in all four bands. Consistent with this, all of the results published in ApJ for that study have a non-zero observation count in each band. The 6 asteroids that were retroactively attributed to Masiero/MB:NIR in the PDS archive would thus appear to be ineligible for that study.

When an asteroid does appear in both the ApJ and PDS, it may not have the same results in both. Most fit results do match; 134,181 result fits are the same between ApJ and PDS to within quoted numerical precision. But among results for asteroids in both sources there also are 1311 model fits in Masiero/MB:Pre that have no counterpart in the PDS; conversely, 1219 fits in the PDS have no counterpart in ApJ. These mismatching fits for asteroids in both sources occur because some results published in ApJ have been omitted from the PDS, or because new results were included in the PDS and retroactively attributed to the paper. Note that a modified result (one or more parameters differ) is not distinguished here from an entirely new result.

The reason that the mismatching fit counts differ (i.e., there are 1311 results from ApJ with no match, while the PDS has 1219 results with no match) is that there is a net omission of 92 results. Asteroids can have multiple results, so some results may be omitted while other results for the same asteroid occur in the PDS.

Similarly, for Masiero/MB:NIR, 2632 of the fits match within quoted numerical precision, 229 of the results that appear in the ApJ version have no counterpart in the PDS and, in the opposite direction, 189 results in the PDS have no counterpart in the original ApJ paper but are attributed to it.

Section 9 below compares NEOWISE results to those from IRAS and finds that the original ApJ published fits have a relatively poor accuracy compared to IRAS (36.6% systematic difference) when diameters that exactly match prior ROS diameters are removed. However, the accuracy improves considerably with the modifications introduced in the PDS (to a 6.6% difference).

Table 2

Comparison of NEOWISE results reported in the original paper and in the NASA Planetary Data System (PDS). The NEOWISE V1.0 archive in the PDS contains model fits attributed to previous NEOWISE publications. A comparison reveals numerous differences, however. The table enumerates these discrepancies for the two largest sources: Masiero/MB:Pre and Masiero/MB:NIR, both of which were originally published in *Astrophysical Journal (ApJ)* (see Table 1 for details on these papers). Many of the model fits in the PDS are incorrectly attributed; they are for asteroids that do not appear in the referenced publication and data files (*upper half of table*). In many other cases, different fits than those archived in PDS were published for corresponding asteroids in the referenced publication (*lower half*). In addition, 3988 model fits for 3772 asteroids described in Masiero/MB:Pre, and 219 fits for 213 asteroids described in Masiero/MB:NIR, are missing from the PDS. Discrepancies of these kinds also occur for other NEOWISE papers referenced by the PDS.

	Category	Masiero/MB:Pre		Masiero/MB:NIR	
		In ApJ	In PDS	In ApJ	In PDS
Asteroid count	Total	129,478	126,917	2835	2629
	Common to PDS and ApJ	125,706	125,706	2622	2622
	In PDS but not ApJ	0	1211	0	7
	In ApJ but not PDS	3772	0	213	0
Model fit count	Total	139,480	136,626	3080	2829
	Common to PDS and ApJ	135,492	135,400	2861	2821
	Matching fits	134,181	134,181	2632	2632
	Mismatching fits	1311	1219	229	189
	For asteroids in PDS but not ApJ	0	1226	0	8
	For asteroids in ApJ but not PDS	3988	0	219	0

Table 3

Examples of model fits altered between publication Masiero/MB:Pre in ApJ and the PDS attributed to that study. Estimates of physical parameters were examined for a selection of the more than 1000 asteroids that appear in the PDS with model fits attributed to Masiero/MB:Pre that are given different fits in the PDS. For the five objects detailed here, diameter estimates varied between the two sources by up to 35%. Diameters are in km; the rightmost columns list the count of observations in each band used in the estimates.

Asteroid	Source	D	p_v	η	p_{IR}	W1	W2	W3	W4
53	PDS	90.893	0.031	0.909	0.047	12	13	13	13
	ApJ	115.000	0.04	1.063	0.0376	12	13	13	13
4525	PDS	10.118	0.034	0.896	0.052	14	14	16	16
	ApJ	9.800	0.0776	0.921	0.1164	14	14	16	16
5742	PDS	12.773	0.226	1.089	0.339	0	0	7	5
	ApJ	17.205	0.1243	1.826	1.0	4	5	7	5
13561	PDS	21.26	0.052	0.923	1.0	0	11	11	11
	ApJ	20.216	0.0306	0.856	0.0458	8	11	11	11
56283	PDS	4.138	0.306	1.115	0.456	0	15	15	15
	ApJ	3.815	0.5309	0.987	0.3445	14	15	15	15

New results and revised versions of old results are certainly welcome, but they should be identified clearly, both in the overall description of the data archive and in reference tags within individual data records. The consequences of these additions and subtractions for research based on these data sources could be substantial, depending on what role the asteroids that were omitted, added, or modified play in the analysis.

To better understand why asteroid property estimates were altered between the original publication in ApJ and archiving in PDS, selected asteroids were examined in greater detail (Table 3). In the case of asteroid 53, the ApJ diameter estimate of 115 km exactly matches a previous ROS source, as described in Section 4. The diameter in the PDS for this asteroid is an apparent model-fit estimate of 90.893 km that was retroactively attributed to the ApJ paper, but did not appear in it. For asteroid 4525, the diameter value published in ApJ does not appear to correspond to an ROS source that I could find, but it may be from an uncited source because it is rounded to the nearest 100 m, which is typical of ROS results; NEOWISE-calculated diameters, in contrast, are typically given to the nearest meter or 10 m. The fact that, for both of these objects, the number of observations in each WISE band is the same in both ApJ and the PDS suggests that the estimates result from dif-

ferent analyses of the same observations. The new diameter value is retroactively attributed to the ApJ paper.

For the other three example objects in Table 3, observations in band W1—and for asteroid 5742, band W2—have been omitted from the PDS results, based on the counts of observations in each band. In the case of 5742, the original ApJ fit had the rather high value of $p_{IR} = 1.0$, which was revised down to $p_{IR} = 0.339$, but since neither the W1 or W2 band were used, this is an assumed value (model code DVB-). In the case of 13561, the ApJ fit has $p_{IR} = 0.0458$, but the revised PDS result has a fit value (model code DVBI) $p_{IR} = 1.0$, despite omitting the W1 band. Although about 90% of the data in the PDS is attributed to Masiero/MB:Pre, similar cases of omission of results from the ApJ data sets or inclusion of results not in the ApJ data sets in the PDS occur as well for the other NEOWISE papers listed in Table 1.

All of the asteroids modeled in Masiero/MB:NIR had been previously modeled in Masiero/MB:Pre. In many cases, the Masiero/MB:NIR result is found in the PDS, but the Masiero/MB:Pre result has been omitted. However this is not always true: for 103 asteroids, results in the PDS are attributed to both Masiero/MB:Pre and Masiero/MB:NIR. I attempted to find a consistent objective rationale or rule for these cases. As an example, I tested whether

Table 4

Taxonomy of NEOWISE results in the PDS. Counts of model fits in the PDS are given according to the set of data bands used in the fits and a code (e.g., DVBI) indicating the type of model being fit (see Sections 3.1–3.4). Although the PDS codes do not distinguish between models using one or two albedo parameters, in this study I denote the cases where p_{IR1} , p_{IR2} were used by DV2, D-B2, and DV-2. Band combinations are classified as fully thermal (incorporating two or more thermal bands), partially or possibly thermal, not thermal, and single band. Shaded cells indicate unlikely or inconsistent combinations (see Section 3.5).

		NEATM-like					STM-like & other					Rows & groups		
Data bands available		DVBI	DVB2	DVB-	D-B-	D-BI	DV-I	DV-2	DV--	D---	-VB-	Totals	%	%
Fully thermal	W1,W2,W3,W4	237	2883	160	1	1	9	2	0	0	1	3294	2.0%	44.3%
	W2,W3,W4	998	93	3	0	0	1	0	0	0	0	1095	0.7%	
	W1,W3,W4	23	0	6201	9	0	15	0	1	0	5	6254	3.8%	
	W3,W4	0	0	62,184	367	0	0	0	231	0	0	62,782	37.9%	
Partial	W1,W2,W3	349	0	81	0	0	262	28	4	0	0	724	0.4%	1.2%
	W1,W3	0	0	0	0	0	68	23	3	0	0	94	0.1%	
	W2,W3	1	0	819	7	0	2	0	346	0	0	1175	0.7%	
Not	W1,W2	5	0	3	0	0	2454	0	3913	0	0	6375	3.8%	3.8%
	W1	0	0	0	0	0	11	0	2	0	0	13	0.0%	
Single-band	W2	0	0	0	0	0	2	0	9074	0	0	9076	5.5%	50.7%
	W3	0	0	0	1	0	0	0	73,788	1061	3	74,853	45.1%	
	W4	0	0	0	0	0	0	0	59	0	0	59	0.0%	
	Column totals	1613	2976	69,451	385	1	2824	53	87,421	1061	9	165,794	100%	
Column %		1.0%	1.8%	41.9%	0.2%	0%	1.7%	0%	52.7%	0.6%	0%	100%		
Grouped		44.9%					55.1%							

the cases where both results were retained were because the Masiero/MB:Pre case had lower estimated error in D or other parameters. That is not the case, nor could I find any other pattern.

Masiero/MB:NIR estimates near-IR albedos separately for bands W1 and W2; indeed that is the primary thrust of that study. The PDS nevertheless includes just one value for IR albedo; it thus fails to fully describe the model fits as originally published. The same is true for fits from Grav/JT:Tax. Adding the second IR parameter to the PDS would make the resource more useful.

The inconsistencies between the PDS and Masiero/MB:NIR in such cases are important to understand because one cannot simply use the PDS archive as a replacement for the original paper; the PDS omits a crucial datum from each result. Moreover, any use of values from the original ApJ paper raises the issue of how to cope with the omissions, additions, and changes tallied in Table 2.

Unlike the NEOWISE papers, the PDS includes both a code that describes the type of model fit performed and a count of the data points in each band used for the model fit. These codes are very useful additions to the repository, as discussed below.

For this study, I drew on results from both the original papers and the PDS. Although the latter is the most recent publication, most studies that rely on NEOWISE results are based on the original publications from ApJ, so it is important to include them as well.

3. Overview of NEOWISE results

Table 4 uses counts obtained from the PDS to provide an overview of the diverse combinations of bands and kinds of model fits that are reflected in the NEOWISE results. Strikingly, a majority (50.7%) of NEOWISE results are based on a single WISE band, a fact not directly evident from the NEOWISE papers themselves. True thermal modeling appears to have informed only 44.3% of the results (the sum of the first four data rows; see the rightmost column), almost all of which are based on two bands (W3, W4).

Codes introduced in the PDS indicate which parameters were fit by a model for each result. A “D” in the code indicates that di-

ameter was fit. A “V” refers to p_V , “B” to η , the so-called “beaming parameter” that is fit to each asteroid in the NEATM (Harris, 1998), and “I” to p_{IR} . The PDS uses “I” for any case where the fit is to one value of p_{IR} for W1, W2 bands and for cases where there are two values fit, one for each band (e.g., p_{IR1} , p_{IR2}). In this study, I denote the latter case with a “2” rather than “I” to avoid ambiguity. Hyphens in the code indicate parameters not obtained by fitting. The most common code among the NEOWISE results is DV--: a model that fits for diameter and p_V but uses assumed values (see below) for η and p_{IR} .

A close examination of how model fitting was performed in the NEOWISE papers raises several general issues with the statistical and numerical methodology that are independent of the models being fit. The first issue is a lack of clarity about the methods used. Mainzer et al. (2014b) provided the most detailed description:

The best-fit values for diameter, visible geometric albedo p_V , infrared albedo p_{IR} , and beaming parameter η were determined using a least-squares optimization that accounted for the measurement uncertainties for bands W1, W2, W3, W4, absolute magnitude H , and phase curve slope parameter G .

That overview omits crucial details necessary to replicate the fits. None of the NEOWISE papers indicate whether ordinary least-squares (OLS) fitting or a χ^2 or weighted least-squares (WLS) approach was used, for example. The PDS model codes reveal the parameters that were fit but also raise important questions, discussed below.

3.1. NEATM-like models

The PDS model code for the NEATM, as conventionally applied, is DVB- or D-B-. The NEATM and other thermal models were originally applied to IR observations in multiple bands of wavelengths long enough that reflected sunlight was negligible, such as W3 and W4 in the case of NEOWISE. When fitting D , p_V , p_{IR} , and η to input parameters H and G , one must respect the constraint that D , p_V , and H are related. The NEOWISE papers, like many previous asteroid studies, model asteroids as spheres. H is thus related to D and p_V

through the well-known definition for visible-band albedo (Bowell et al., 1988; Harris and Lagerros, 2002)

$$p_v \equiv \left(\frac{1329}{D} \right)^2 10^{-\frac{2H}{3}}, \quad (1)$$

where D is in km. This expression or its equivalent appears in each of the NEOWISE papers. Using “independent” variables that are not actually independent can lead to problems in nonlinear least-squares fitting (Feigelson and Babu, 2012; Ivezić et al., 2014; Wall and Jenkins, 2012). Standard statistical practice would use Eq. (1) to represent p_v in terms of D and H , thus eliminating p_v as an independent variable.

The passage quoted above implies that H is a required input for thermal modeling. Yet it also states that p_v is an independent variable, in which case H would not be required as a dependent variable. The PDS model codes further highlight this inconsistency. If H is available, then DVB- and D-B- ought to be identical. Instead it appears that asteroids coded as D-B- model fits are those for which H was not available, which would imply that NEOWISE results having “V” in their model codes might explicitly violate Eq. (1), as discussed further in Section 7. In cases where p_v was not varied as a free parameter, its value must have been fixed, or p_v must have been replaced in a re-parameterization. The NEOWISE papers do not explain the assumptions used in such cases.

To the extent that the NEOWISE papers used the NEATM, that implies using the NEATM formula for subsolar temperature T_{ss} ,

$$T_{ss}(r_{as}) = \left(\frac{S(1 - p_v q)}{0.9\sigma\eta r_{as}^2} \right)^{0.25} = \frac{T_1}{\sqrt{r_{as}}} \quad (2)$$

(or some alternate parametrization that is equivalent), where σ is the Stefan-Boltzmann constant, q is the H-G model phase integral, S is the solar constant, and r_{as} is the distance from the asteroid to the sun. Using p_v and η as independent parameters can be problematic because they appear as the quotient inside the expression for T_{ss} , and p_v is bounded. As a result, any change in p_v can be masked by a change in η , making it far more likely for the nonlinear least-squares algorithm to become trapped in a false minimum. A more reliable numerical approach is to reparametrize the relation by using the parameter T_1 , which is the subsolar temperature at $r_{as} = 1$ AU (Myhrvold, 2018). This latter approach eliminates the absolute magnitude H from the thermal modeling.

3.2. STM-like models

The Standard Thermal Model (STM) was a predecessor to the development of the NEATM by Harris (Harris, 1998). The NEATM differs from the STM principally in using a fixed value of η rather than allowing the beaming parameter to vary for each asteroid. The STM also lacks the phase-angle adjustment built into the NEATM, so phase angle is often set by using an empirical, linear adjustment of 0.01 mag deg⁻¹ (Lebofsky et al., 1986; Spencer et al., 1989; Veeder et al., 1989).

The NEOWISE papers do allow η to vary in cases that can make use of “two thermally dominated bands” (Masiero et al., 2011), defined as bands having $\geq 75\%$ of the flux from thermal emission rather than reflected sunlight. How to determine *a priori* whether that criterion applies is unclear, however, and would require accounting for Kirchhoff’s law of thermal radiation, which holds that the emissivity in each band is directly related to the reflectivity in the band. The NEOWISE approach violates Kirchhoff’s law in this case (Myhrvold, 2018, 2016) and thus may have incorrectly determined which bands are thermally dominated.

In many instances, the NEOWISE analysis does not find two thermally dominated bands and sets η to a fixed value, much as the STM does. That approach may be intuitive, but it should be justified with either a mathematical argument or test cases that show

how well such asteroids are handled with the fixed- η assumption or whether they might be better handled in other ways.

In any event, having adopted this approach, it should be applied consistently, but the NEOWISE papers instead vary details of the approach among papers and also make numerous exceptions within a single paper. Each paper states that it applies a value of η derived by statistical analysis of whatever group of asteroids are the focus of that study. Masiero/MB:Pre sets $\eta = 1.0$, for example, with the justification that this value is the mean for the main-belt cases. Nugent/Re1 and Nugent/Re2 instead use the mode of the distribution, given as $\eta = 0.99$.

However, if we take all of the full NEATM model fits in the PDS that are in the main belt and that used models DVBI or DVB2, the mean is $\bar{\eta} = 0.94$, across 4229 results. If instead we cast a wider net to include the model code DVB-, the mean is $\bar{\eta}_- = 1.03$ across 71,243 results. But the DVB- fits have assumed values for p_{IR} that might alter the results, so the fully modeled approach would seem the better choice. Note that each original NEOWISE paper was able to incorporate results only up to the time of its publication, so these articles could not use the statistics of the PDS, which span the entire collection of papers.

Table 5 shows the stated assumed value of η for each paper along with the actual mean values $\bar{\eta}$ for the NEATM fits in each of the NEOWISE papers of **Table 1**. The stated assumed value of η does not match the actual mean value for any of the papers, although some of the differences are small. If we assume that the asteroids requiring an assumed value of η are statistically similar to the NEATM-fit asteroids of the same category, then the difference between the stated value and the actual mean will introduce a systematic error in diameter estimates. As an example, Masiero/MB:NIR assumes a value of $\eta = 1.0$, whereas the actual mean for the best NEATM fits is $\bar{\eta} = 0.91$. Using $\eta = 1.0$ as the assumed value if, in reality, $\bar{\eta} = 0.91$ would systematically overestimate the diameter by $\approx 6\%$. The total range of η across NEOWISE models that fit it is $0.55 \leq \eta \leq 3.14$. If an asteroid of $\eta = 0.55$ is modeled as having $\eta = 1$, then the diameter estimate will be too large, at 1.44 times the actual diameter. If an asteroid of $\eta = 3.14$ is modeled as having $\eta = 1$, the estimated diameter will be too small, at 0.47 times the actual diameter.

Moreover, as shown in **Table 5**, the assumed values of η archived in the PDS sometimes contradict the assumed value reported in the papers to which they are attributed. Mainzer/NEO:tiny reports using a value of $\eta = 2.0$ for those asteroids that have an assigned value of η , for example, yet the majority of the fits in the PDS attributed to that paper instead use some other value, ranging from 0.9 to 3.0.

Because model codes were not used in the original NEOWISE papers published in ApJ, the only way to recognize cases that used an assumed value of η is to compare to the stated assumed value. However, as shown in **Table 5**, the stated assumed value was not always used, so one cannot reliably determine whether or not a model result published in the ApJ version of the data sets has an assigned η . The model codes that appear in the PDS version do provide a consistent way to determine this case, showing that this is was a useful addition to the data.

For other papers, the number of discordant values is small on a relative basis. Masiero/MB:Pre clearly states that $\eta = 1.0$ was used in such cases, yet three fits in the PDS that reference Masiero/MB:Pre instead used $\eta = 1.2$. It is important to understand these inconsistencies, including how and why these “fixed” values were adjusted.

The NEATM has largely superseded the STM because a variable beaming parameter greatly improves the accuracy of diameter estimates (Harris, 1998; Ryan and Woodward, 2010). A formal error analysis of the NEOWISE studies could process multiband data on asteroids with both fixed and variable η and then compare the re-

Table 5

Comparison of assumed values stated in original references to corresponding values in the PDS. In cases where fewer than two thermal bands were available, the NEOWISE papers assumed a fixed value of η rather than calculating a value from a model fit. This is found for model codes DV-, DV-I, and DV-2. NEOWISE studies that focused on individual categories of asteroids ("Single category") selected one value to use for each category (NEO, main belt, Jovian Trojan, Hildas). Two papers analyzed asteroids from multiple categories ("Multi") and assumed one value of η for each category. The leftmost columns list each original reference and the value of η reported in that paper as being used for all such cases, usually on the basis that the assumed value is the mean of values for asteroids where η is found by fitting. The middle columns show the actual mean values of η for the full NEATM fits within the same paper. The column DVBI,2 is the mean across fits having model code DVBI or DV2. The +DVB- column includes those results, plus results having the DVB- model code. The columns at right show the actual values for η for model codes DV-, DV-I, and DV-2. In some cases, these values are inconsistent with the model fits recorded in the PDS that cite these papers (shaded cells). Columns labeled N_{fits} list counts of fits that used the adjacent value of η . Additional values of η were used for just one or two fits in some papers ("Other values").

Assumed η stated in reference Paper		Mean $\bar{\eta}$ NEATM DVBI,2 +DVB-		Actual assumed η and counts found in the PDS				Other values
		η	N_{fits}	η	N_{fits}	η	N_{fits}	
Single category	Grav/JT:Pre	0.88	n/a	.90	0.88	193		
	Grav/Hilda	.85	0.84	0.90	0.85	294		
	Grav/JT:Tax	.77	0.80	0.86	0.77	79		
	Mainzer/NEO:Pre	1.4	1.58	1.60	1.35	97	1.4	19
	Mainzer/NEO:Tiny	2.0	1.93	1.96	2.0	27	1.4	19
	Masiero/MB:Pre	1.0	1.02	1.03	1.0	68,971	1.2	3
	Masiero/MB:3	1.0	0.98	1.00	1	11,468		
	Masiero/MB:NIR	1.0	0.91	0.91	1.2	9		
Multi	Mainzer/PP:3	1.0, 1.4	1.53	1.64	1.4	100	1.0	2
	Nugent/Re1	0.95, 1.4	n/a	0.89	0.95	8766	1.4	235
							0.7	1
							1.2	1.53; 2.0

sults to assess the impacts on accuracy. Such an analysis has not yet been reported, however, and unfortunately it cannot be fully performed by third parties because the NEOWISE models are not currently replicable.

3.3. Models involving reflected sunlight

An empirical analysis of how the NEOWISE analysis determined near-IR albedo reveals further issues that call into question the validity of many of the albedo results reported to date. The handling of p_{IR} is important because it can have a large impact on the other parameters in the model fit.

The NEATM and STM are not suitable, in their original forms, for handling bands such as W1 and W2 that can include a substantial, or even dominant, contribution from reflected sunlight. The NEOWISE studies thus used NEATM-like and STM-like models modified in various ways. Under some conditions, the NEOWISE analysis fits a model that includes p_{IR} as a parameter. The models denoted by PDS codes DVBI and D-BI take the NEATM approach, whereas the DV-I model uses fixed η values. In other cases, the NEOWISE analysis splits p_{IR} into two parameters $p_{\text{IR}1}$, $p_{\text{IR}2}$. Either way, p_{IR} and $p_{\text{IR}1}$, $p_{\text{IR}2}$ were handled by the NEOWISE papers in a way that violates Kirchhoff's law of thermal radiation (Myhrvold, 2018) by assuming that emissivity in the W1 and W2 bands is 0.9, regardless of the value of albedo in the same bands (i.e., p_{IR} and $p_{\text{IR}1}$, $p_{\text{IR}2}$).

When the NEOWISE analysis was unable to fit p_{IR} , a value for that parameter was assumed; such models are coded DVB- and D-B- in the NEATM-like case and DV-- in the STM-like case. The additional model types D--- and -VB- are quite distinct from either the NEATM or the STM, as discussed in Section 3.4.

Masiero/MB:Pre states that, in cases where p_{IR} was assumed rather than fit, the analysis set near-IR albedo to a fixed multiple of visible albedo: $p_{\text{IR}} = a p_v$, with $a = 1.5$ for that paper. Note that a value $a > 1$ implies that the asteroid spectrum is more "red" (really near-IR) than the incident sunlight. As shown in Table 6, the stated values of the adjustment factor a vary substantially among the NEOWISE papers; some papers do not report what value was used.

I used the PDS data to empirically examine the values of a selected for model fits of the type DVB-, D-B-, DV--, D---, and -VB-.

Histograms of the results are shown in SI Figure S1. The empirical analysis reveals that, contrary to the documented procedure, p_{IR} was not always set to a single multiple of p_v .

In both Masiero/MB:Pre and Masiero/MB:3, the value of IR albedo appears to have been clipped to 1.0, i.e., $p_{\text{IR}} = \min(1.0, 1.5 p_v)$ —see blue histograms in Figure S1—either during the analysis or after modeling. Note that while $p_{\text{IR}} = 1.0$ would be considered an extremely high value, it is not a hard upper bound on p_{IR} . The NEOWISE papers assume that p_{IR} follows the HG phase system, so its maximum value is $1/q$, where q is the HG phase integral. For the default slope parameter, $G = 0.15$ and $q = 0.384$, so $p_{\text{IR}} \leq 2.6$ (Myhrvold, 2018).

Even beyond the clipped cases, values of p_{IR} varied in a fairly broad distribution ($\pm 0.1 p_v$), typically peaking at a fixed value. For the papers Grav/JT:Pre and Grav/JT:Tax, the dispersion in a seems likely to be a result of numerical rounding of the reported results. But rounding is not a sufficient explanation for the distributions seen for the other papers. For several NEOWISE papers, fits marked in the PDS as having assumed p_{IR} (i.e., code ending in “-” rather than “I”) span wide ranges, with $0.2 p_v \leq p_{\text{IR}} \leq 3.0 p_v$. This is inconsistent with the stated policy of setting p_{IR} to a (possibly clipped) multiple of p_v .

A different empirical check is to see how well the assumption that $p_{\text{IR}} = a p_v$ is upheld in fully modeled results that derive p_{IR} from the data (i.e., results with model codes DVBI or DV2.) To investigate this issue, I fit the linear equations $p_{\text{IR}} = a p_v$ and $p_{\text{IR}} = a p_v + b$ to results for models that do fit p_{IR} (i.e., those having codes ending in “I”). Table 6 compares the resulting fit values to the values reported in the corresponding papers. These values can also be compared to those derived empirically (SI Figure S1). In all cases, the value of a obtained by fitting is smaller than either the value reported by NEOWISE in the text or the peak found empirically. A linear fit that allows for a constant offset b performs much better in most cases, as evaluated by the corrected Akaike Information Criterion (ΔAIC_c). Nevertheless, neither equation generates good fits in absolute terms, due to considerable scatter in the data. The assumption that p_{IR} is a multiple of p_v is empirically inaccurate for the cases where NEOWISE fits each of them (i.e., results having model codes DVBI or DV2).

As described elsewhere (Myhrvold, 2018), the NEOWISE studies approach all calculations involving p_{IR} in a manner that violates

Table 6

Comparison of linear adjustment factor a used in NEOWISE papers to estimate p_{IR} with values of a obtained by fitting linear equations, with and without offsets, to results in the PDS. In cases where the near-IR albedo p_{IR} is set by assumption, the NEOWISE papers set it to a multiple of visible band albedo as $p_{\text{IR}} = a p_v$, and the value of a is reported in the paper and reproduced here in the leftmost columns. The center group of columns show the minimum, maximum, and mean value of p_{IR}/p_v for results that fit these parameters. The right group of columns shows linear fits to the equations $p_{\text{IR}} = a p_v$, or $p_{\text{IR}} = a p_v + b$ for parameters a, b to NEOWISE results in the same papers that determined p_{IR} and p_v by fitting. The difference in the corrected Akaike Information Criterion, ΔAIC_c , is shown for the two fits in each row. A value $\Delta \text{AIC}_c \leq 2$ indicates strong statistical support; in most cases, the linear model with constant term b has strong support, and the linear model without it does not. That is a relative determination; in absolute terms neither linear relationship fits the data well. Note that no results in the PDS attributed to the paper Grav/JT:Pre fit both p_{IR} and p_v , and the paper Masiero/MB:NIR included no results that rely on an assumed value of a .

Stated assumed values of a , for $p_{\text{IR}} = a p_v$		Actual p_{IR}/p_v from DVBI, DVB2			Linear fits to p_v, p_{IR} from DVBI, DVB2			
Paper	a	Min	Max	Mean	$p_{\text{IR}} = a p_v$		$p_{\text{IR}} = a p_v + b$	
		a_{min}	a_{max}	\bar{a}	a	ΔAIC_c	a	b
Single category	Grav/JT:Pre	2.0	No results in PDS attributed to this paper fit both p_{IR} and p_v .					
	Grav/Hilda	2.0	1.07	3.32	1.94	1.76	24.00	1.17
	Grav/JT:Tax	2.0	1.04	4.22	2.07	1.96	62.50	0.97
	Mainzer/NEO:Pre	1.6	0.52	5.37	1.95	1.46	27.60	0.95
	Mainzer/NEO:Tiny	1.6	0.45	4.33	1.64	1.52	0.78	1.42
	Masiero/MB:Pre	1.5	0.58	5.30	1.58	1.35	359.00	0.85
	Masiero/MB:3	1.4	0.024	13.1	1.39	1.30	157.00	1.18
	Masiero/MB:NIR	n/a	0.30	10.9	1.40	1.31	144.00	1.16
Multi	Mainzer/PP:3	1.6	1.01	2.07	1.48	1.23	0	0.76
	Nugent/Re1	1.5,1.6	0.075	24.5	2.60	1.62	28.90	1.12

Kirchhoff's law. It is nonphysical to fit for p_{IR} (or to assume its value) in bands W1 and W2 without also changing the emissivity in those bands. The NEOWISE method assumes that emissivity in the W1 and W2 bands is always 0.9, but it allows p_{IR} to vary regardless. This violation of Kirchhoff's law leads to errors in diameter estimation that are especially problematic for certain combinations of models and bands (Myhrvold, 2018). While it is common in thermal-modeling studies to fix the value of emissivity, typically to $\epsilon = 0.9$, doing so is usually acceptable because the observations are made at longer wavelengths, where reflected sunlight is negligible. This is not the case in the WISE W1 and W2 bands.

3.4. Other models

Several enigmatic model codes, including D--- (occurs for 1061 results) and -VB- (occurs for 9 results), appear in the PDS but do not correspond to any definition of thermal modeling normally used in the field. The NEOWISE papers offer no explanation of such models. D--- is found only in Masiero/MB:Pre and Masiero/MB:3, only in cases where a single band of W3 data was analyzed and where H apparently was not available. These results offer no values for p_v or p_{IR} , and $\eta = 1$ for all of them. The papers do not discuss the method used to obtain the diameter D .

The code -VB- appears to indicate that the diameter was not obtained by using D as a model fit parameter, so it must have been fixed to a value prior to fitting a model on remaining variables p_v and η . The source of such diameter values has not been disclosed. As discussed further in Section 4, other evidence suggests that PDS results bearing -VB- model codes have diameters that exactly match prior ROS estimates in the literature.

3.5. Inconsistent data/model combinations

As detailed in Table 4, the NEOWISE results archived in the PDS make use of 47 different combinations of model codes and data bands. Of these, 19 combinations (shaded entries in Table 4), used to produce 16,794 results, seem unlikely or contradictory, based on the definition of the models.

In cases of full thermal modeling, W3 and W4 bands are both present. Those bands are always dominated by thermal emis-

sion and thus, according to rules described in the NEOWISE papers, should be analyzed using a full NEATM model: DVBI, DVB-, D-B-, or D-BI. Yet 265 asteroids for which W3 and W4 data are present are given non-NEATM-like model codes in the PDS.

For eight results in the PDS, only W1- and W2-band observations are available, yet the results bear NEATM-like codes DVBI and DVB-. It seems extremely unlikely that these bands could be "thermally dominated" in the sense used in the NEOWISE papers. Similarly, it is questionable that DV-- and DV-I models could operate with only W1- and W2-band data, but 6367 results are annotated with those codes and bands.

No established method exists to perform true thermal modeling using just a single band of IR observational data, so single-band cases are modeled non-thermally by using additional assumptions and the visible absolute magnitude H . Results obtained in this way are not true modeling results but should instead be considered rescaled results from a single idealized asteroid having the assumed η and a p_{IR} that is a fixed product of p_v and a . The majority of the NEOWISE fit results in the PDS (more than 84,000, or 50.7%) are based on single-band observations.

In the 74,853 cases of single-band results that use data only from the W3 band, the key assumption is the value of η . An assumption about p_{IR} is irrelevant in such cases because it only pertains to bands W1 and W2. As shown in Table 5, the NEOWISE papers state that they use specific assumed values of η for these cases, but in practice some PDS results use other values of η .

Across full NEATM fits (DVBI) in Masiero/MB:Pre, $0.604 \leq \eta \leq 3.14$, with a mean of 1.02, a median of 1.00, and a 68.27% confidence interval (CI) of $0.862 \leq \eta \leq 1.16$. There are 68,713 fits in this paper that are based solely on W3-band data, and the stated assumption is that $\eta = 1.0$. Assuming that the distribution of η found for the DVBI cases applies to these asteroids, the minimum and maximum diameter estimation due to variation in η alone is -27% to $+132\%$, a very wide range. The median error is zero in this case because the stated diameter is the same as the median of the DVBI cases. The 68.27% CI on the error is -9% to $+10.4\%$, but this only reflects the error due to making an assumption about η , not the total error.

Asteroids for which three or four bands of observations exist and models carry DVBI codes tend to be larger than the asteroids modeled using only W3-band data (Fig. 8 and SI Fig. S10). It is therefore unclear how much we can rely on the assumption that the single-band cases are statistically identical (with respect to η) to the three- and four-band cases.

In the 9076 cases where data from only the W2 band was used, assumptions about both η and $p_{IR} = a p_V$ determine the resulting diameter estimate. The NEOWISE analysis determined the amount of reflected sunlight in the band by assuming that $p_{IR} = a p_V$ (using values of a listed in Table 6). The value of a completely determines the emitted flux, which is then modeled by assuming a value of η . Errors in either assumption will lead to error in the estimated diameter.

The NEOWISE papers fail to address whether such methods are valid, and if so how accurate their results are. Both questions could be answered easily by performing single-band modeling (using all models represented in the PDS) on asteroids for which multiple bands exist, and then comparing single-band to multiband results. This will be important future work to establish confidence in the NEOWISE estimates.

3.6. Utilization of observational data

WISE single-exposure flux data must be filtered prior to use for thermal modeling. By downloading all of the asteroid observations and following the same data preparation and filtering steps as done with NEOWISE, one can estimate the total available data for each asteroid. The counts presented here regard as “available” observations for which all bands exist, have no problems indicated by the WISE pipeline quality and artifact flags, and have a WISE estimated $\sigma \leq 0.25$, corresponding to estimated SNR ≥ 4.3 . The data were obtained from IRSA using the same methods, and were filtered using the same criteria, as was employed in Masiero/MB:Pre. Section S4 of the SI provides a detailed account of NEOWISE data preparation and selection.

Table 7 categorizes the available observations for asteroids and band combinations in the three left columns; the four rightmost columns show the counts of those bands actually analyzed by the NEOWISE team for results from the FC mission.

The FC mission, for example, yielded data having at least one data point in each of the four bands on 31,516 asteroids, 23% of the total asteroids observed in that mission phase. Yet an examination of the PDS finds that just 4,114 asteroids are shown as having data from all four bands, a mere 3% of the total from the FC mission in the PDS. Of the asteroids that potentially have four bands of data, only 13% are actually analyzed with four bands of data; for all of the rest, entire bands of data were discarded.

This phenomena is not limited to the four-band cases. There are 56,969 asteroids that were observed in the W3 and W4 bands (at least one point in each), which is the largest group in that mission phase, comprising 41% of it. Yet fewer than half of these cases (46%) were analyzed using both bands. For the rest, only one band of data was analyzed.

The discarding of data was not done in response to poor data quality (i.e., low SNR, background objects, noise, artifacts, etc). In both the NEOWISE papers and the analysis here, these data-quality issues are handled by the WISE pipeline, which assigns quality and artifact flags to observations, as well as by the WISE pipeline estimated observational error. The quality criteria used for selecting data for the counts in Table 7 are the same criteria used by Masiero/MB:Pre and Masiero/MB:NIR (see SI Section S4). Instead, several aspects of the NEOWISE methodology that are unrelated to data quality led to observational data being discarded.

The first is the requirement that there be at least three data points in each band. While there are 31,516 asteroids for which at

least one point is available per band, the number drops to 17,146 for two points per band minimum, and to 11,860 for three points. Some minimum may be necessary for a given analytical technique, but these counts show that a method that could usefully utilize smaller data counts would greatly expand the number of asteroids included in a fuller data analysis.

The number of asteroids having at least three points in each of the four bands is still 2.9 times larger than the actual count of 4114 asteroids that NEOWISE analyzed with four bands. The next factor that causes data discarding is that the NEOWISE papers group WISE observations into “epochs,” which are periods of time when the WISE cadence and asteroid orbit produce a set of observations separated by a maximum time interval. Thermal models were fit separately on each epoch for each asteroid rather than on the complete set of observations available for each asteroid.

The maximum interval defining an epoch is 3 days in one NEOWISE paper (Mainzer et al., 2011d) but 10 days in another (Mainzer et al., 2011b); other papers do not specify the interval used. An epoch typically includes 8–12 observations over 36 h, but some asteroids were observed hundreds of times in a single epoch. Observations can also occur in multiple epochs. Among the asteroids sampled from the papers in Table 1, 66.6% were observed during a single epoch, 28.6% during two epochs, 4.3% in three epochs, and 0.5% in four epochs (all during the FC mission). Because the epochs are driven primarily by the interaction of the asteroid orbit with the WISE cadence, the epochs do not represent an intrinsic property of either the asteroid or the observation (i.e., the epoch is not directly related to solar phase angle). No rationale or justification is presented in the papers for why breaking the analysis into epochs for analysis is either valid or desirable in this circumstance.

The minimum point count per band is applied on a per-epoch basis. In cases where there are multiple epochs, this minimum-data-point requirement discards data that would not be discarded if the artificial separation into epochs had not been performed.

The next unconventional rule for discarding data could be called the “40% rule.” Within each epoch, counts of data points in each band are compared, and all data are discarded from any band that has fewer than 40% of the highest count. Grav/JT:Pre explains the 40% rule as follows:

In order to avoid having low-level noise detections and/or cosmic rays contaminating our thermal model fits, we required each object to have at least three uncontaminated observations in a band. Any band that did not have at least 40% of the observations of the band with the most numerous detections (W3 or W4 for the Trojans), even if it has 3 observations, was discarded.

While low-level noise and cosmic rays are important to reject, the method described would seem to be ill-suited to either task. Noise and cosmic rays are intended to be handled by the WISE pipeline quality and artifact flags, as well as by the pipeline error estimates. No discussion is offered in Grav/JT:Pre of why those mechanisms are insufficient.

Indeed, intrinsic sensitivity issues limit the number of observations in some bands. Taken across the entire FC mission, the W3 band includes the most observations of asteroids. The W1 band has only 21% as many observations as W3, and W2 has 28% as many (see Table A2). The disparities arise in large part from the fact that the WISE detectors and exposure times were not intended only for asteroid observation; instead WISE is a general-purpose survey mission, which by necessity requires compromises (Wright et al., 2010).

The strategy of discarding data that does not conform to the 40% rule would, if applied to the entire FC mission data set, eliminate both W1 and W2 entirely. Conversely, those asteroids for which data counts in W1 or W2 exceed 40% of the W3 count

Table 7

Utilization of NEOWISE observational data. The approach to data analysis employed by NEOWISE routinely discards enormous amounts of data. Observations: NEOWISE data available in the FC mission where $\sigma \leq 0.25$, equivalent to $\text{SNR} \geq 4.3$, applying the same filters on the artifact and quality flags as used by Masiero/MB:Pre. Fits in PDS: counts of data according to bands employed in NEOWISE fits, as documented in the PDS.

Available	Observations		Fits in the PDS			
	Count	% of Total	Actually used	Count	% Utilized	% of Total
W1 W2 W3 W4	31,516	23.0%	W1 W2 W3 W4	4114	13.1%	3.0%
			W1 W2 W3	8	0.0%	0.0%
			W1 W3 W4	724	2.3%	0.5%
			W2 W3 W4	3808	12.1%	2.8%
			W1 W3	4	0.0%	0.0%
			W2 W3	17	0.1%	0.0%
			W3 W4	16,694	53.0%	12.2%
			W3	6119	19.4%	4.5%
			W4	28	0.1%	0.0%
W1 W2 W3	1461	1.1%	W1 W2 W3	0	0.0%	0.0%
			W2 W3	13	0.5%	1.1%
			W3	1448	99.1%	0.1%
W1 W3 W4	19,513	14.2%	W1 W3 W4	107	0.5%	0.1%
			W1 W3	1	0.0%	0.0%
			W3 W4	10,284	52.7%	7.5%
			W3	9111	46.7%	6.7%
			W4	10	0.1%	0.0%
W2 W3 W4	11,264	8.2%	W2 W3 W4	1762	15.6%	1.3%
			W2 W3	12	0.1%	0.0%
			W3 W4	7195	63.9%	5.3%
			W3	2294	20.4%	1.7%
			W4	1	0.0%	0.0%
W1 W3	3012	2.2%	W1 W3	1	0.0%	0.0%
			W3	3011	100.0%	2.2%
W2 W3	517	0.4%	W2 W3	14	2.7%	0.0%
			W3	503	97.3%	0.4%
W3 W4	56,969	41.6%	W3 W4	26,198	46.0%	19.1%
			W3	30,746	54.0%	22.4%
			W4	25	0.0%	0.0%
W3	12,737	9.3%	W3	12,737	100.0%	9.3%

are by definition atypical. The use and implications of epochs and the 40% rule are described and discussed more fully in Appendix Section 12.1 and SI Section 4.

A further consequence of the decision to restrict analysis to WISE epochs is that the NEOWISE analysis does not benefit from pooling data across mission phases. An asteroid that is observed with a small number of counts in the W1 and W2 bands during the FC mission could, in principle, be supplemented with observations in those bands in the subsequent 3B, PC, and Re mission phases, the last of which continues to the present. The current NEOWISE analysis can only use PC and Re mission data only as new epochs that are analyzed separately.

The current state of WISE/NEOWISE data utilization can be viewed as an opportunity for planetary science. The application of new analytical techniques that are able to handle pooled data across epochs, and even across mission phases, would allow significant improvements to our knowledge of asteroids. It may be challenging to do this analysis, but should such techniques be developed, they could more than double the number of asteroids that have the highest-quality modeling approach with all four bands of data and greatly increase the number of asteroids analyzed by full thermal modeling in the sense of Table 4 (i.e., with both W3 and W4 present).

4. Comparison with previous ROS results

The accuracy of asteroid thermal modeling is best assessed by comparing diameter estimates derived from a model to estimates obtained by other means, such as from ROS observations made by radar, stellar occultations, or spacecraft close approaches.

Mainzer/TMC, one of the earliest NEOWISE papers, addressed the error and accuracy of NEOWISE thermal modeling. That paper used ROS diameters to model the NEATM flux from 50 objects. It then compared the modeled flux to the observed fluxes. This approach is, in effect, the opposite of the technique used to create the ~164,000 NEOWISE model fits. It was intended to be part of the mission's error and accuracy analysis, which is treated in detail below in Section 9.

Of the 53 results presented for 50 objects in Mainzer/TMC, two are moons rather than asteroids and are not considered here. Three asteroids (1866, 68216, and 164121) are either not found in the ROS references or are given a different diameter than that listed in Table 1 of Mainzer/TMC. It seems likely that these asteroids had ROS values from unpublished sources. Diameters given for all other objects exactly match a value in one of the previous studies, mostly but not entirely from the ROS references in that paper. The use of ROS diameters in Mainzer/TMC for test objects is not prob-

Table 8

Examples of NEOWISE published “fits” for diameter that exactly match reference ROS measurements. This table lists a small subset of diameter estimates published in the NEOWISE papers that exactly match those published in prior studies. See SI Table S1 for the complete list of all 129 matching estimates for 105 asteroids. Certain asteroids appear in multiple papers, some of which do not report an ROS-matched diameter; such cases are shaded.

Asteroid	NEOWISE published results			Prior radar, occultation, spacecraft references		
	D	σ_D	Paper	D	σ_D	Paper
2	544,000	42,916	Masiero/MB:Pre	544,000	N/A	Shevchenko et al., 2006
	544,000	60,714	Masiero/MB:Pre			
	610,818	49,581	PDS:Mas14	No match		
	642,959	13,759	PDS:Mas14	No match		
	No match			512,000	6000	Durech et al., 2010
	No match			515,000	N/A	Dunham & Herald, 2010
	No match			522,000	N/A	Shevchenko et al., 2006
	No match			523,156	1128	Dunham & Herald, 2010
	No match			531,800	34,100	Dunham & Herald, 2010
	No match			539,000	28,000	Durech et al., 2010
	No match			545,109	1693	Dunham & Herald, 2010
5	115,000	9353	Masiero/MB:Pre	115,000	6000	Durech et al., 2010
	106,699	3140	PDS:Mas12	No match		
	108,290	3700	Masiero/MB:NIR	No match		
	108,293	3703	PDS:Mas14	No match		
	No match			116,397	2845	Dunham & Herald, 2010
	No match			120,000	14,000	Magri et al., 1999
6	185,000	10,688	Masiero/MB:Pre	185,000	10,000	Magri et al., 1999
	195,639	5441	PDS:Mas14	No match		
	195,640	5440	Masiero/MB:NIR	No match		
	197,871	33,614	PDS:Nug15	No match		
	No match			165,000	21,000	Durech et al., 2010
	No match			180,000	40,000	Durech et al., 2010
	No match			207,500	1400	Dunham & Herald, 2010
	No match			220,805	N/A	Dunham & Herald, 2010
8	140,000	1160	Masiero/MB:Pre	140,000	7000	Durech et al., 2010
	147,490	1030	Masiero/MB:NIR	No match		
	147,491	1025	PDS:Mas14	No match		
	169,354	43,365	PDS:Nug15	No match		
	No match			136,000	N/A	Dunham & Herald, 2010
	No match			138,000	9000	Magri et al., 1999
	No match			141,000	10,000	Durech et al., 2010
	No match			160,800	N/A	Shevchenko et al., 2006
	No match			169,538	7803	Dunham & Herald, 2010

lematic because source references are cited and the purpose of the paper is error and accuracy analysis. The test objects are presented as such, not as final modeling results. This is corroborated by the PDS archive, which lists no results from that paper.

The primary result papers for main belt and NEO asteroids analyzed by NEOWISE are Masiero/MB:Pre and Mainzer/NEO:Pre. These papers present the results of thermal-model fitting, including diameters. The tables of fit results for these papers (both in the paper and in the online data repository for ApJ) also include diameter estimates that exactly match the values published in prior ROS references, intermingled in the table with results from thermal-model fitting. There is no datum to mark or denote which results have diameters that exactly match prior ROS references, in contrast to those that are the result of thermal modeling.

Among the asteroids for which the NEOWISE result diameter exactly matches a prior ROS diameter are the 48 asteroids used as test cases in Mainzer/TMC. However the set is not limited to those asteroids. A search revealed a total of 129 NEOWISE results for 105 asteroids that have diameters that exactly match ROS di-

ameters from the literature published prior to the NEOWISE papers (Table 8 and SI Table S1). This is a tiny fraction of NEOWISE results on an overall basis (0.064%), but it is significant because it covers the majority of ROS diameters available in the 2010–2011 time frame. Because ROS diameters are the only independent way to assess the accuracy of NEOWISE estimates, these asteroids are particularly important. Both papers reference Mainzer/TMC, but these references appear to focus on its error analysis results and do not mention the use of ROS diameters, nor are there other mentions or references to ROS diameters in the papers.

Both Masiero/MB:Pre and Mainzer/NEO:Pre give multiple results for some asteroids. In each case where an asteroid has a NEOWISE result diameter that exactly matches a prior ROS diameter, that is the only diameter presented for the asteroid.

Masiero/MB:Pre and Mainzer/NEO:Pre quote their result diameters to the nearest 1 m. In contrast, ROS sources typically round to the nearest 1 km or (typically for very small asteroids) to the nearest 100 m. The matches are exact to accuracy quoted by NEOWISE

(typically 5–6 digits)—a degree of coincidence extremely unlikely to occur by chance (odds roughly 1 in 10^{210} , according to a Monte Carlo analysis [see SI Section S8]).

The list of results with exactly matching diameters may be larger than noted above. There are 21 asteroids given a diameter estimate in Masiero/MB:Pre that are presented as rounded to the nearest 1 km or 100 m but have diameters that do not exactly match any prior ROS estimate I have found. The rounding could be coincidental or could indicate a ROS origin for the diameter. For asteroid 381, for example, $D = 129,100$ m in (Shevchenko and Tedesco, 2006), a reference that is the source for many other exact matches, but Masiero/MB:Pre reported $D = 129,000$ m for this asteroid. That is not included as an exact match in the current study.

Data in the PDS include nine results, for six asteroids, that bear the model code -VB-. All of these results are attributed to Mainzer/NEO:Pre. The leading hyphen in the model code indicates that the diameter was fixed to an assumed value rather than being determined by a model fit (Mainzer et al., 2016). For eight of the results, the diameter exactly matches a prior ROS diameter. The result for asteroid 230111 (2001 BE10) has model code -VB- and D rounded to nearest 100 m. Asteroid 230111 was the subject of radar observations prior to the WISE/NEOWISE mission (Benner et al., 2008), but that paper does not publish a diameter, and I have not found any other published reference for its diameter; it might have been obtained from an unpublished source. The existence of the -VB- model code is significant because it shows that NEOWISE used at least one model that set diameters by a means other than thermal-model fitting.

Because ROS estimates offer the only diameter measurements derived independently of the NEOWISE methods, they are crucial to determining the accuracy of the NEOWISE diameter estimates. Indeed, this is how the NEOWISE group judged IRAS and RW results for accuracy and systematic bias (Mainzer et al., 2011d). The group did not perform the same analysis on NEOWISE results; this is discussed further in Section 9.

Other NEOWISE papers that present asteroid results, such as Masiero/MB:NIR and Grav/Hilda from the FC mission phase, appear to have fit-derived diameters.

Thermal modeling of asteroids has historically focused on estimation of diameter as the most important result, and this is reflected in Masiero/MB:Pre: “Using a NEATM thermal model fitting routine we compute diameters for over 100,000 Main Belt asteroids from their IR thermal flux, with errors better than 10%.” The intermingling of ROS diameters with estimates computed by thermal modeling would appear to be inconsistent with that goal. If accurate diameter estimates are intended to be due to thermal modeling, then using ROS diameter (i.e., model code -VB-) is counterproductive. Conversely, the only way to assess the accuracy of the NEOWISE results is to compare them to an independent metric: the ROS diameters. That assessment obviously cannot be made if the ROS diameters have been used as the source of the diameter estimate.

4.1. Treatment of results matching ROS diameters in the PDS NEOWISE archive

Many existing studies have relied on the original NEOWISE data sets published in ApJ and available for download from that journal. The recent addition of NEOWISE results to the PDS archive offers an alternative source for these data. Although the PDS files purport to provide the same information as previously published, the NEOWISE data published in ApJ and the PDS differ in many ways, as shown in Section 2 and Tables 2 and 3. To avoid confusion and conflicting results, it is very important to understand how and why the NEOWISE group revised the results.

The differences between the ApJ and PDS sources of NEOWISE results particularly effect the results with diameters that exactly match prior ROS results. Aside from the nine cases with model code -VB- described above, all other results published in ApJ with NEOWISE diameter estimates that exactly match prior ROS diameters have been omitted altogether from the PDS. Table 9 presents excerpted examples of such cases; SI Table S2 provides a complete list.

The first case shown in Table 9 is asteroid 2 (Pallas), for which two results published in ApJ in Masiero/MB:Pre and matching a ROS diameter were omitted from the PDS. The PDS includes two new results for this asteroid that do not appear in the ApJ version of Masiero/MB:NIR; the reference fields for these new results retroactively attribute them to that paper. Note that Masiero/MB:NIR states that one selection criteria for its analysis required observations in each of the four WISE bands. Yet the PDS lists zero observations in the W4 band as contributing to the two results for Pallas. These observation counts for Pallas are consistent with the counts given for results in Masiero/MB:Pre that exactly match an ROS diameter.

For asteroid 5 (Astraea), the PDS omits the result from Masiero/MB:Pre; the result from Masiero/MB:NIR is included and differs only by rounding. This is the most common pattern seen in SI Table S2: for 82 asteroids, all results from Masiero/MB:Pre were omitted, while those from Masiero/MB:NIR were allowed to remain (some with rounding). Asteroids 334 (Chicago) and 1512 (Oulu) similarly appeared in both Masiero/MB:Pre (matching a ROS diameter) and Grav/Hilda (with a fit diameter). The PDS omits their entries from Masiero/MB:Pre but includes the Grav/Hilda diameter.

In the case of asteroid 13 (Egeria), the original papers include only a result from Masiero/MB:Pre; no result was published in the ApJ paper Masiero/MB:NIR. While that situation is similar to that for Pallas discussed above, the treatment in the PDS is different. A new result with a non-ROS diameter was included in the PDS and was retroactively attributed to Masiero/MB:Pre paper. This pattern of modification of the original result occurs for 12 asteroids: 13, 36, 46, 53, 54, 84, 105, 134, 313, 336, 345, and 2867.

Asteroid 10115 (1992 SK) originally appeared in both Masiero/MB:Pre and Mainzer/NEO:Pre, in both cases with a diameter that exactly matches a prior ROS diameter. The Masiero/MB:Pre entry has been omitted from the PDS, but the entry for Mainzer/NEO:Pre has been kept as is, with the ROS diameter and with model code -VB-. Nine results, involving six asteroids (10115, 68216, 164121, 230111, 341843, 363067), match this pattern.

Finally, in the case of asteroid 522 (Helga), the result from Masiero/MB:Pre was omitted from the PDS and not replaced, so no result appears in the PDS for this asteroid.

The text and data field definitions in the PDS archive clearly state that it is purely a compilation of results from the original papers rather than an updated analysis (see Section 2). There is no explanation of why results were omitted or retroactively attributed to prior papers. Ideally the PDS or another publication would include the results as originally published, updated results that are explicitly marked as such, and NEOWISE model-fit results for the asteroids that previously were assigned diameters that exactly match prior ROS diameters.

The cases described here are just a subset of the 3772 asteroids that appeared in Masiero/MB:Pre but have been omitted from the PDS; the 1311 model fits that appear in the PDS with a different diameter than in the original ApJ paper; and the 1211 asteroids that have been retroactively attributed to the paper (Table 2). Other cases show patterns like the example of asteroid 4525 (Jonbauer) listed in Table 3. The ApJ diameter for that asteroid appears to be from a ROS source because it is rounded to the nearest 100 m. In addition, the PDS entry for that asteroid gives a modified

Table 9

Examples of treatment in the PDS archive of NEOWISE results from ApJ that include a diameter that exactly matches a prior ROS diameter. These examples show how NEOWISE results having diameters that exactly match prior ROS diameters were treated in the PDS archive for six asteroids discussed in the text. SI Table S2 presents a complete list of such cases. For each asteroid (leftmost column), multiple rows list the diameter estimates published in prior ROS study (see SI Table S1), NEOWISE papers (Table 1), and/or the PDS. The rightmost column notes whether results present in original ApJ papers were omitted from the PDS (red), included as is or with rounding, modified (green), or added anew with retroactive attribution (yellow). One of the example results (blue) was included with the model code -VB-.

Asteroid	Original Source	D (km)		PDS Treatment	Attribution
		ApJ	PDS		
2	ROS	544			
	Masiero/MB:Pre	544	544	omitted	
	Masiero/MB:Pre	544	544	omitted	
			642.959	retroactive attribution	Mas 14
5	ROS	115			
	Masiero/MB:Pre	115	115	omitted	
	Masiero/MB:NIR	108.29	108.293	rounded	Mas 14
334	ROS	174.1			
	Masiero/MB:Pre	174.1	174.1	omitted	
	Grav/Hiida	198.77	198.77	exact	Grav11
13	ROS	227			
	Masiero/MB:Pre	227	222.792	modified	Mas11
10115	ROS	1			
	Masiero/MB:Pre	1	1	omitted	
	Masiero/NEO:Pre	1	1	exact: -VB-	Mai11
522	ROS	83.7			
	Masiero/MB:Pre	83.7	83.7	omitted	

diameter that is not rounded. In the absence of a prior ROS reference, I have not listed this asteroid here as an ROS-matching case because I could not find a prior ROS reference, but it could be unpublished or from a source I could not find.

Nonetheless, it seems implausible that all of the results tallied in Table 2 as omitted from the PDS matched ROS diameters; the literature in the 2010–2011 timeframe did not contain that many ROS-estimated diameters. Other, unknown factors must have prompted most of the omissions, modifications, and retroactive attributions.

5. Constraining systematic observational error

The NEOWISE papers used a Monte Carlo approach to estimate error in fitted parameters. Section 8 below discusses some of the issues with this approach. Here I focus on the inputs to the analysis: the observed fluxes and their errors.

The WISE data-analysis pipeline produces estimates of flux f_{obs} and the standard error σ_{obs} (both expressed in WISE magnitudes) for each observation in each band. There are many possible sources of error in the observations, as discussed at length in the WISE Explanatory Supplement (Cutri et al., 2011).

The NEOWISE Monte Carlo analysis makes a simplifying assumption that each flux may be modeled as a normal distribution based on the WISE pipeline. I will use the notation $x \sim \mathcal{N}(m, \sigma)$ to signify that the random variable x has a normal distribution with mean m and standard deviation σ . In those terms, the NEOWISE Monte Carlo error estimates assume that Monte Carlo trial fluxes $f_{\text{MC}} \sim \mathcal{N}(f_{\text{obs}}, \sigma_{\text{obs}})$. The error analysis depends crucially on that assumption being correct.

As an artifact of the WISE observing cadence, some asteroids are observed twice in quick succession, with a time separation of ≈ 11 s, when the asteroid appears in the overlapping region between consecutive WISE frames. Hanuš et al. (2015) realized that

these “double detections” provide an opportunity to independently check the accuracy of the assumption that WISE observational fluxes are $\sim \mathcal{N}(f_{\text{obs}}, \sigma_{\text{obs}})$. Their work assumes that no change to the physical fluxes from the asteroid occurs during the ≈ 11 s interval. It further assumes that any differences detected between consecutive flux measurements f_1 and f_2 , with corresponding uncertainty estimates σ_1 and σ_2 , must be due to random observational error. Hanuš et al. (2015) studied the statistic \hat{f} , which they termed the “normalized difference” in fluxes between the two observations in a double detection.

$$\hat{f} = \frac{f_1 - f_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (3)$$

They pointed out that if $f_1 \sim \mathcal{N}(f_1, \sigma_1)$ and $f_2 \sim \mathcal{N}(f_2, \sigma_2)$, then $\hat{f} \sim \mathcal{N}(0, 1)$. In classical statistics, the statistic \hat{f} is usually denoted Z , and this would be called a two-sample Z-test (Chihara and Hesterberg, 2011; Kanji, 2006). In that context, the distribution of Z is the sampling distribution of the null hypothesis that $f_1 \sim \mathcal{N}(f_1, \sigma_1)$ and $f_2 \sim \mathcal{N}(f_2, \sigma_2)$.

Instead, using data from ≈ 400 WISE double detections in the W3 and W4 bands, Hanuš et al. found that \hat{f} had a distribution of $\mathcal{N}(0, 1.4)$ in the W3 band and $\mathcal{N}(0, 1.3)$ in the W4 band. They concluded that the estimates of standard error in the observations σ_1, σ_2 are systematically too low and should be a factor of 1.4 and 1.3 higher, respectively, for W3 and W4. In addition, they found that a Kolmogorov-Smirnov test gave support at the 0.92 and 0.87 level to the hypothesis that \hat{f} follows a normal distribution for bands W3 and W4, respectively.

Here I extend this analysis to all double observations from the FC mission (Fig. 1). Data were obtained from IRSA in the same manner as described in Masiero/MB:Pre and Section 3.6. Suspect observations were filtered out by using the WISE pipeline quality and artifact flags and the same filtering criteria employed in Masiero/MB:Pre (see SI Section S4 and SI Tables S3 and S4). An

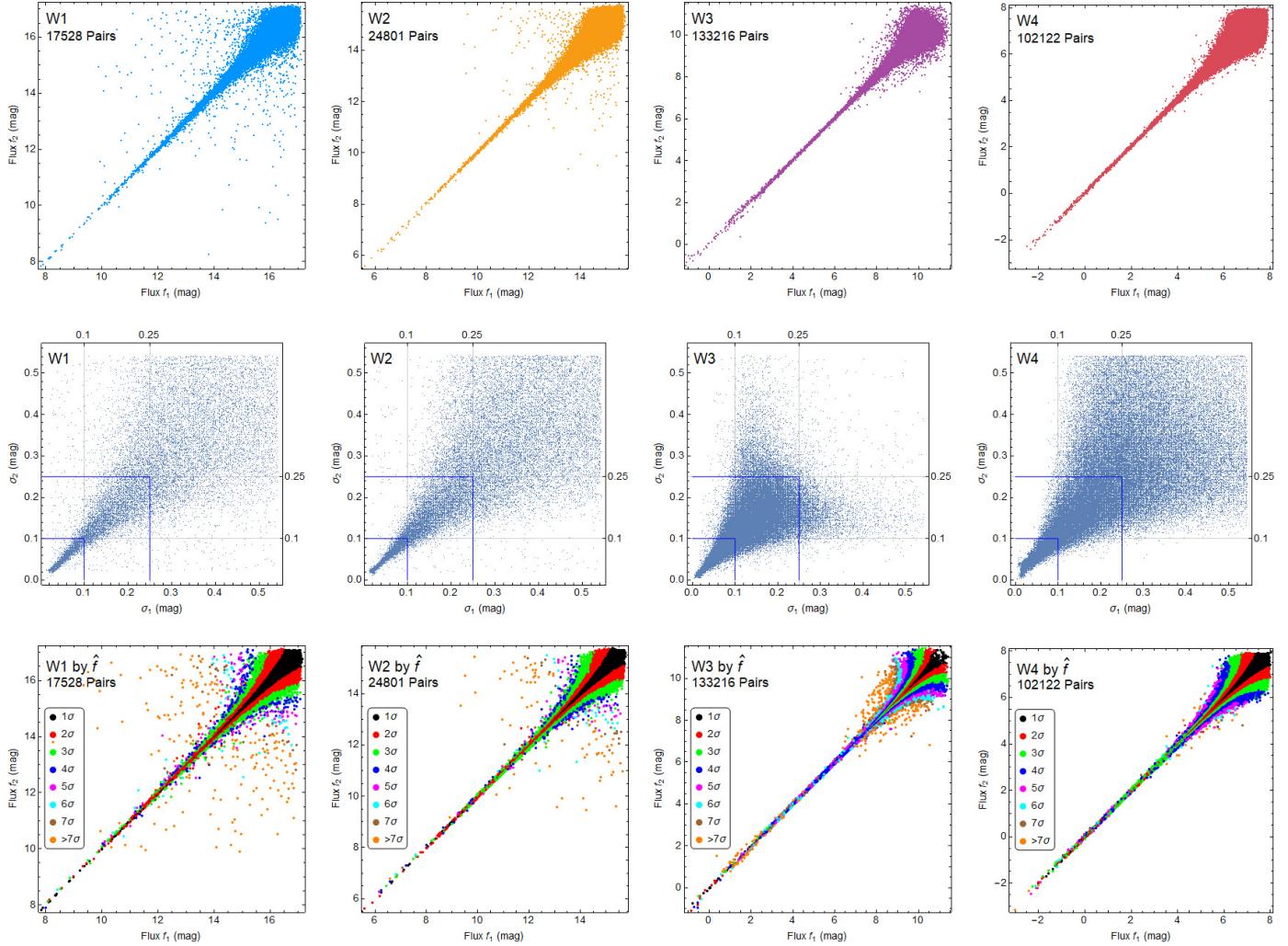


Fig. 1. Fluxes measured and observational errors estimated in double detections. About 10% of WISE/NEOWISE observations fall in a region that was captured in successive WISE frames. These “double detection” observations measure a flux f_1 (plotted here on the x-axis), followed by a second flux f_2 (plotted on the y-axis) about 11 s later. If the two measurements were completely consistent, the plots of points (f_1, f_2) for each band in the top row would all lie on the line $y = x$. Instead, these pairs of observations illustrate the variability in flux that occurs. The double detections also include estimated observational errors σ_1 and σ_2 , which are similarly plotted as points (σ_1, σ_2) for each band (middle row). Blue squares indicate the subsets $\sigma_1, \sigma_2 \leq 0.25$ and $\sigma_1, \sigma_2 \leq 0.1$ that are discussed in the text. If the estimated observational error were correct, then the statistic \hat{f} of Eq. (3) should have a normal distribution $\mathcal{N}(0, 1)$ with zero mean and unit standard deviation. The bottom row plots the flux points colored by integral values of \hat{f} , which corresponds to standard deviations of that distribution. The points in black are $\leq 1\sigma$; those in orange are $>7\sigma$. The many points lying several standard deviations out are evidence that it is extremely unlikely that the observational errors were correctly estimated.

SNR threshold was not applied on the data because the point of the analysis is to understand variations in the sigma estimated by the WISE pipeline. Double detections were found by examining the WISE time stamps for each observation.

The top row of graphs in Fig. 1 plot each double detection f_1, f_2 as a point. The estimated standard deviations σ_1, σ_2 for these double detections also vary (middle row of graphs). When flux values are the same for each observation in the double detections, the corresponding point falls on the diagonal line $y = x$. Blue squares in the middle row of charts demarcate the subsets $\sigma_1, \sigma_2 \leq 0.25$ (i.e., SNR ≥ 4.3) and $\sigma_1, \sigma_2 \leq 0.1$ (i.e., SNR ≥ 10.9), corresponding to the putatively “best” observations having the lowest estimated errors (highest SNR). Yet considerable scatter is evident even in these subsets.

The bottom row of graphs in Fig. 1 plot each pair of observations (f_1, f_2) colored by \hat{f} for the pair (from their associated σ_1, σ_2). If $|\hat{f}| < 1$, the points lie within one standard deviation of $\mathcal{N}(0, 1)$ and are colored black. If $1 \leq |\hat{f}| < 2$, the points fall between one and two standard deviations away and are colored red – and so forth up to the last color (orange), which repre-

sents points that would be seven or more standard deviations out.

It is immediately evident from Fig. 1 that the points are far from the assumed distribution of $\mathcal{N}(0, 1)$. Indeed, the maximum values of \hat{f} are 46.2, 34.8, 31.2, and 12.6 standard deviations for bands W1, W2, W3, and W4, respectively.

I tallied the same statistic \hat{f} for all near-simultaneous observations in all four WISE bands in the FC mission. Three SNR or σ thresholds were considered, including $\sigma_1 \leq 0.532$ and $\sigma_2 \leq 0.532$, which corresponds to $\text{SNR} \geq 2.0$; $\sigma_1 \leq 0.25$ and $\sigma_2 \leq 0.25$, which corresponds to $\text{SNR} \geq 4.3$; and $\sigma_1 \leq 0.1$ and $\sigma_2 \leq 0.1$, which corresponds to $\text{SNR} \geq 10.9$. Note that a threshold of $\text{SNR} \geq 4.3$ is the same threshold used for Masiero/MB:Pre (see SI Tables S3 and S4).

Fig. 2 shows the resulting histograms, which are poorly fit by a normal distribution (fit using maximum log-likelihood) and instead closely match the generalized form of the Student’s t-distribution $S(\mu_t, \sigma_t, \nu_t)$, which has a location parameter μ_t , a scale parameter σ_t , and generalized degrees of freedom ν_t . Statistical tests of the normal distribution were applied, including the Cramér-von Mises test, the Komolgorov-Smirnov test, and the Anderson-Darling test.

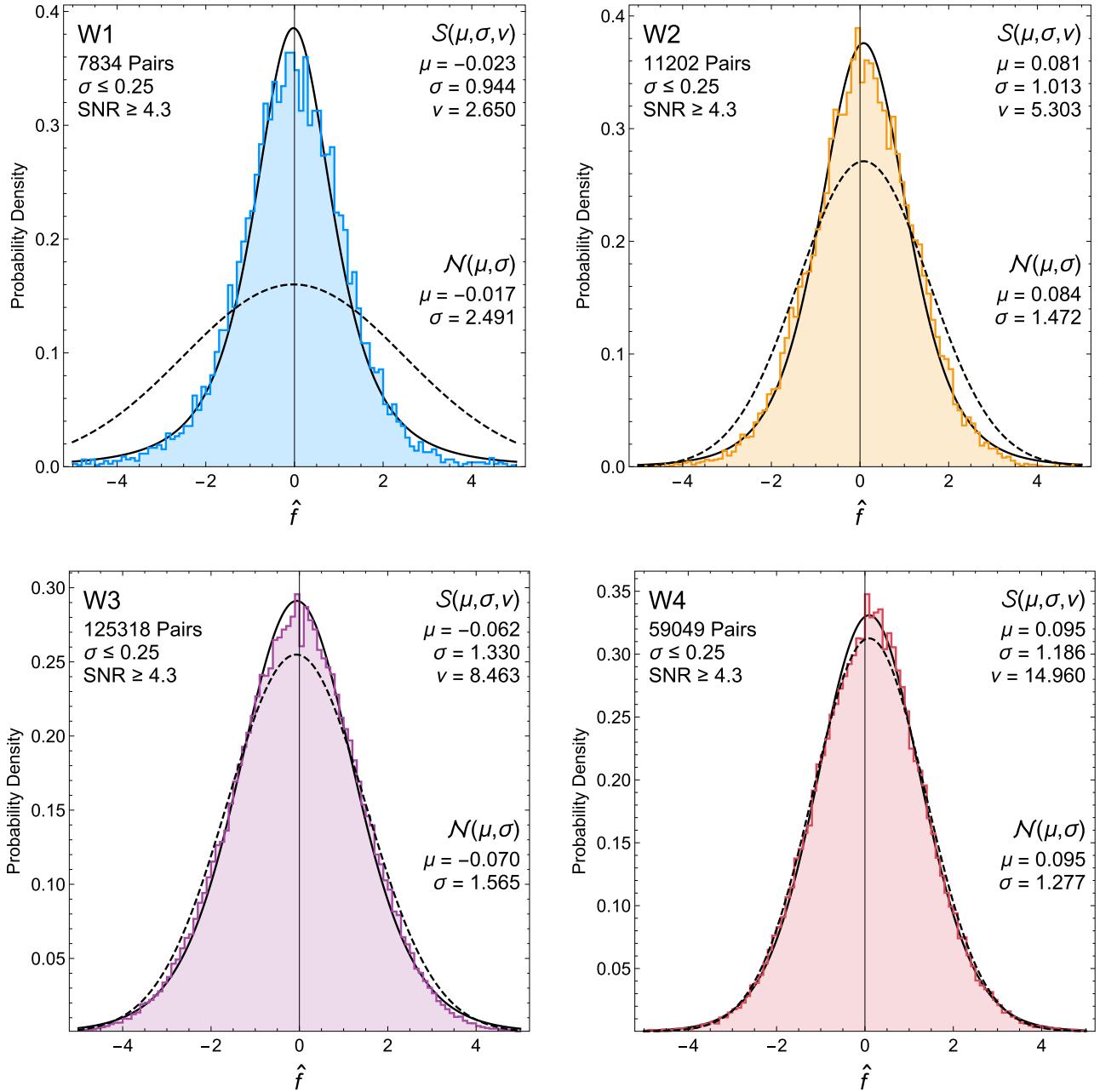


Fig. 2. Sampling distribution for near-simultaneous observations. The statistic \hat{f} of Eq. (3) is plotted as a histogram of double observations in each band, for observations having $\sigma_1, \sigma_2 \leq 0.25$; other values are shown in the appendix. The solid black curve is a best fit to the generalized Student's t-distribution $S(\mu, \sigma, v)$ with the parameters indicated. The dashed black curve is a best fit to a normal distribution $N(\mu, \sigma)$. The best fits are determined by maximum log-likelihood.

All gave a result $p \ll 10^{-12}$ in each band, indicating no support for the null hypothesis that \hat{f} has a normal distribution.

SI Figure S2 illustrates the cases of $\sigma_1, \sigma_2 \leq 0.1$ and of all observations ($\sigma_1, \sigma_2 \leq 0.532$). These histograms also are best fit by a Student's t-distribution, and the deviation from a normal distribution (i.e., smaller values of v_t) is greater for the lower error (i.e., smaller σ_1, σ_2) scenarios. This demonstrates that the effect is *not* due to faint or noisy observations. Fit parameters for each of these cases are listed in Appendix Table A2 and A3 and are discussed in Section 12.2.

The findings here are roughly consistent with those of Hanuš et al. because, at the high values of $v_t = 9.954$ for W3 and $v_t = 29.72$ for W4 (for all observations), the distributions are approximately normal, especially when using a much smaller data set (see dashed curves in Fig. 2). Hanuš et al. found $\sigma = 1.4$ for W3 and

$\sigma = 1.3$ for W4, roughly consistent with the σ_t values that I find (Fig. 2 and Appendix Table A2, A3).

Student's t-distribution (see Section 12.2) covers the behavior of a statistic like \hat{f} of Eq. (3) when the fluxes f_1 and f_2 are means and σ_1 and σ_2 are standard deviations of small samples drawn from a normal distribution. For integer values of v_t , the sample size is $v_t + 1$. The Student's t-distribution has much fatter tails than the normal distribution, as can be seen in the proliferation of points that would be $\geq 7\sigma$ in the bottom row of Fig. 1 if $\hat{f} \sim N(0, 1)$.

WISE estimates flux by combining data from multiple pixels using a point-spread function (Cutri et al., 2015), which is not a simple mean across n samples. The WISE pipeline estimate for observational uncertainty is similarly a statistical sample of the point-spread function and per-pixel variation, not a simple standard deviation. Nevertheless, the fact that \hat{f} is empirically well approxi-

mated by a Student's t-distribution suggests that the distribution of flux estimates in the W1 band behaves much like a mean of $n \approx 4$ (really $3.65 = v_t + 1$) samples drawn from $(\mu_{\text{true}}, \sigma_{\text{true}})$. Moreover, σ_1 and σ_2 are *estimates* of σ_{true} , not the standard error σ_{true} itself. This is why they can vary so much (middle row of Fig. 1) even absent a physical change to the system in the short interval between observations.

As a result, we should expect that the observational errors estimated for a single observation in a *WISE* band will *not* be normally distributed, even if one assumes that the true distribution is normal. Hanuš et al. were correct that the difference of two normal distributions ought to be normal—but that is applicable here only if σ_1 and σ_2 are the true parameters of two *different* normal distributions. Instead they are small sample-size estimates of the same distribution $\mathcal{N}(\mu_{\text{true}}, \sigma_{\text{true}})$, as illustrated by the considerable variation between successive observations in Fig. 1.

The differences observed in the values of v_t between bands probably arises from intrinsic properties of the *WISE* focal-plane sensor, which was designed for the primary *WISE* mission of mapping stars, galaxies, and other fixed sources. The NEOWISE search for moving objects such as asteroids stretches the capabilities of the array. Pixels on the *WISE* sensors are $18 \mu\text{m}$ square for each band; in the case of band W4, the pixels are ganged 2×2 and are thus effectively $36 \mu\text{m}$ square. The size of the point-spread function is set by the wavelength range: $2.8 - 3.8 \mu\text{m}$ for band W1, $4.1 - 5.2 \mu\text{m}$ for W2, $7.5 - 16.5 \mu\text{m}$ for W3, and $20 - 28 \mu\text{m}$ for W4. Variations among bands in the number of pixels activated, electron counts in the detector, and background noise for a given source would be expected and may drive differences in the estimated flux and flux uncertainty, leading in turn to variations in the effective sample size and thus the value of v_t in each band. The primary mission of *WISE* was to produce a map of the sky with approximately 8 to 11 co-added frames in each location, and these errors would be expected to converge in such a case. In contrast, NEOWISE uses single-exposure observations, and the errors in the *WISE* pipeline for those cases appear to be poorly characterized by a normal distribution.

Hanuš et al. noted that it is possible that the variations seen between successive observations arise from edge-dependent effects, such as vignetting or asymmetry in the detector. Such effects would not be relevant to the majority of observations that occur elsewhere on the sensor. To test this possibility, I obtained millions of *WISE* observations of stars from the IRSA archive of stellar observations with stable visible-band photometry, observations that were previously used for SDSS calibration (Ivezić et al., 2007). Despite the obvious temperature differences between stars and asteroids, the W1 and W2 bands yielded enough observations to provide meaningful statistics. In addition to the ≈ 11 s between overlapping successive observations, the *WISE* cadence repeats stellar observations at intervals of roughly 1.5 h, 3, 4, 6, and multiple days. These repeated observations do not require the source to appear in the overlap regions at the edges of the sensor.

As shown in SI Figures S4 and S5, I find that, just as for asteroids, the f -statistic histograms for stars are well fit by Student's t-distribution, both for double detections of stars and for pairwise comparisons of f across longer intervals that do not occur at the edge of the sensor. Values of the distribution parameters do differ from those for asteroids, probably as a result of much higher fluxes and different electron-count statistics across the point-spread function. This result is consistent with the intuitive model that both flux and flux errors behave like small- n samples taken from an underlying distribution. The specifics of the effective sample size depend on the observational statistics. As with asteroids, selecting subsets of the data with $\sigma_1, \sigma_2 \leq 0.25$ or $\sigma_1, \sigma_2 \leq 0.1$ increases the deviation from normality. The phenomenon is thus not specific to asteroids, nor is it confined to the edges of the sensors.

The importance of this finding is that the assumption in the NEOWISE Monte Carlo analysis that $f_{\text{MC}} \sim \mathcal{N}(f_{\text{obs}}, \sigma_{\text{obs}})$ is refuted by the empirical statistics of the double detections. The best-fitting normal distributions have standard deviations that are much larger (Fig. 2, Table A2 and A3), but the normal distribution is a poor choice. The distribution is better approximated by $f_{\text{MC}} \sim \mathcal{S}(f, \sigma_{\text{obs}} \sigma_t, v_t)$, where f is the *WISE* flux in a given band, σ_{obs} is the *WISE* standard error estimate for that band, and σ_t, v_t are given for the band in Table A2.

The implications of this finding for error analysis of the NEOWISE data set are substantial because the NEOWISE results are based on Monte Carlo simulations that draw modeled band magnitudes from $\mathcal{N}(f_{\text{obs}}, \sigma_{\text{obs}})$. These estimates are systematically too low, and they are affected by previously unrecognized errors because they are non-normally distributed. The assumption that errors are approximately normally distributed is common in science, and interpreting the *WISE* estimated σ_{obs} as the standard deviation of a normal distribution is a natural choice. Unfortunately, the double detections show that this conventional assumption does not apply here.

The same caveat applies beyond NEOWISE to any modeling method based on the single-exposure *WISE* fluxes and flux uncertainties, such as thermophysical modeling (Alí-Lagoa et al., 2014; Hanuš et al., 2015; Koren et al., 2015). Indeed that was the context for the study by Hanuš et al. (2015).

Although the analysis in this section is limited to the FC mission, double detections are a feature of the 3B and PC missions as well. The *WISE* pipeline appears to have changed for the Re mission; the majority of successive overlapping observations have identical fluxes ($f_1 = f_2$), although extreme values $f_1 \neq f_2$ remain.

6. Models that miss the data

The utility of asteroid thermal modeling hinges on its ability to fit observational data well, so that model parameters can be used to estimate physical properties accurately. In my empirical examination of NEOWISE fits, I discovered that many are a surprisingly poor fit to the data, to the extent that in a great many cases the fit curve misses every data point in an entire band—and sometimes in all bands (Figs. 3 and 4).

As a very crude metric of fit, I counted for each model fit in the FC mission whether the fit curve intersected the error bars of any data point in a band (in which case the fit residuals will be both positive and negative or zero across the data points) or whether it misses the band entirely (generating residuals that all have the same sign). Most statistical studies would use finer tools to compare how well competing models fit the data. But a starting assumption for such tools is that no model lies either entirely above or entirely below every point in the data set being fit.

Yet that is a common occurrence in the FC mission and associated NEOWISE papers. Fig. 3 shows one example—asteroid 25916, for which the NEOWISE model is systematically lower (brighter) in magnitude than the data points in each band. A best-fit NEATM model (red curves) passes through the clouds of data points in each band (blue curves), yielding an estimated diameter that is 8.8% smaller. As seen in Fig. 4 (blue bars), this case exemplifies a much larger issue.

The crude count of band misses fails to capture the full extent of the problem because many NEOWISE models that do intersect some points offer, upon inspection, obviously very poor fits that would be considered unacceptable in most scientific or statistical studies. Fig. 3 illustrates as an example asteroid 90367, for which the model systematically overestimates the magnitude of the object, as indicated by the bulk of the data. A best-fit NEATM model fits much better and yields a diameter estimate that is 10.3% larger.

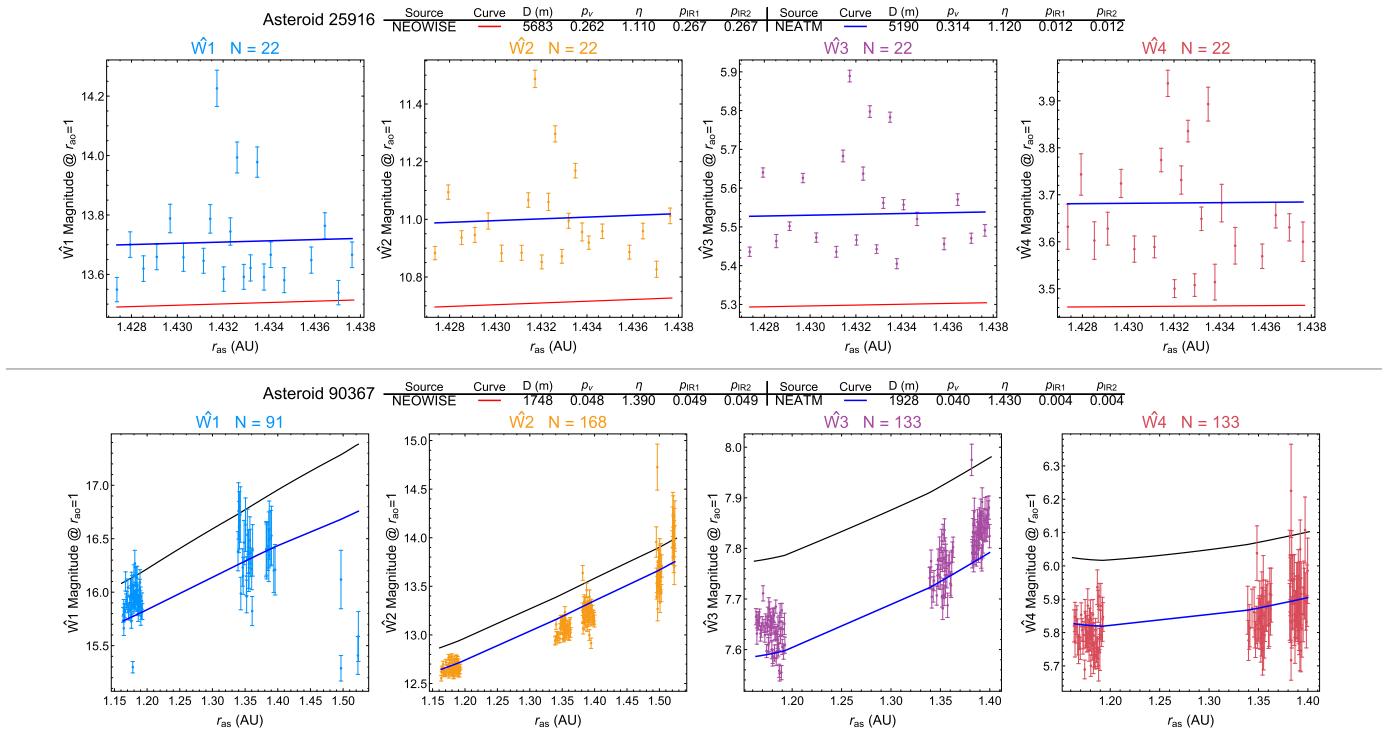


Fig. 3. Example models that show poor fit to data. Each row shows the four *WISE* bands observed for an asteroid during the FC mission phase, along with their observational errors σ estimated by the *WISE* pipeline (*error bars*), where r_{as} is the distance of the asteroid from the Sun. For asteroid 25916, NEOWISE model fits (red curves) miss every data point in all four bands. The best-fit NEATM model (blue curves) easily passes through the clouds of data points in each band. For asteroid 90367, NEOWISE model fits (black curves) do pass through the clouds of data points, but the quality of fits is quite poor. The best-fit NEATM model is also shown (blue curves). The plotting method is discussed in Section 5 of the SI. Only data points for which $\text{SNR} \geq 4.3$, as estimated by the *WISE* pipeline, are included, after filtering using the *WISE* pipeline data quality and artifact flags (see SI Section A4).

Fig. 4 summarizes the results of the census of hits and misses for the FC mission. (See SI Figure S8 for comparable counts of NEOWISE fits archived in the PDS.) Consistent with Table 4, the majority of NEOWISE results are based on a fit to a single band, with most of those being fits to W3.

In 19,122 (29%) of the cases that use W3 only, the model misses every data point. The next largest group of results was fit with both W3 and W4, yet only 57% of those model fits hit each band (in the sense here of having residuals that are mixed sign or zero.) Overall, the model fits that rely on at least two fully thermal bands (i.e., W3 and W4) have a combined success rate of 57% to 58% at hitting every band of data used.

These counts almost certainly underestimate the true prevalence of band missing, because they include in the assessment all data points from the FC mission, whereas the NEOWISE model fits are limited to the small subset of data taken during a single 3- to 10-day epoch. My inclusion of extra data points in some cases registers a “hit” in a band when the fit actually missed all of the data available to the NEOWISE model.

A model fit that fails to intercept *any* of the data under analysis cannot be deemed a “result” derived from observation; rather it is an artifact manufactured entirely by the analytical process. Some combinations of assumptions and constraints applied in the NEOWISE analysis prevents many curves from approaching or intersecting the clouds of data points, or, in less extreme cases, precludes a good fit.

One example of such a constraint would be an assumption that sets p_{IR} or η to assumed values, as discussed in Section 3. However this cannot be the only source of the poor fits because such fits also occur in cases (DVBI, DV2) where these particular assumptions are not made. Indeed only about 57% of fits using three or

four bands of data hit the data points (i.e., generate residuals of mixed sign) in all bands.

The direct consequence of the band missing is that $\approx 30\%$ of the single-band fits and $\approx 43\%$ of the multi-band fits should be considered suspect because they miss one or more bands of data. More generally, a fitting procedure that produces so many cases where the model fit misses entire bands of data is itself suspect, raising questions about the integrity of all of its results. A fitting procedure that is so constrained by assumptions that it miss entire bands may, in some cases, also miss best fits in cases where it does not miss the entire band. An example is shown in Fig. 3, where the model for asteroid 90367 does hit the clouds of data, but is not the best fit when evaluated by a least-squares criterion.

These results depend crucially on whether the fluxes that I calculate from the NEOWISE result parameters are the same as those calculated by the NEOWISE group. Since I could not make that comparison directly, the output of the NEATM flux code used for this analysis was validated by extensive comparison to several other NEATM implementations, including those by A. W. Harris (DLR), V. Ali-Lagoa, and J. Emery. Conversion from NEATM fluxes to *WISE* magnitudes was done using the formulas of Wright (2013), the *WISE* principle investigator.

Fig. 5 provides an additional validation check by plotting the residuals from the NEOWISE model fits. Any systematic problem in the NEATM implementation I used would be apparent as a shift in the distribution away from a median of zero. No such shift is seen, which also indicates that the cases where NEOWISE models miss a band are, on average, roughly balanced between those that are too high and those that are too low. In the case of the W3 band, the median of the residuals is -0.174 magnitude, possibly the result of a “linear correction” to the W3 band that is men-

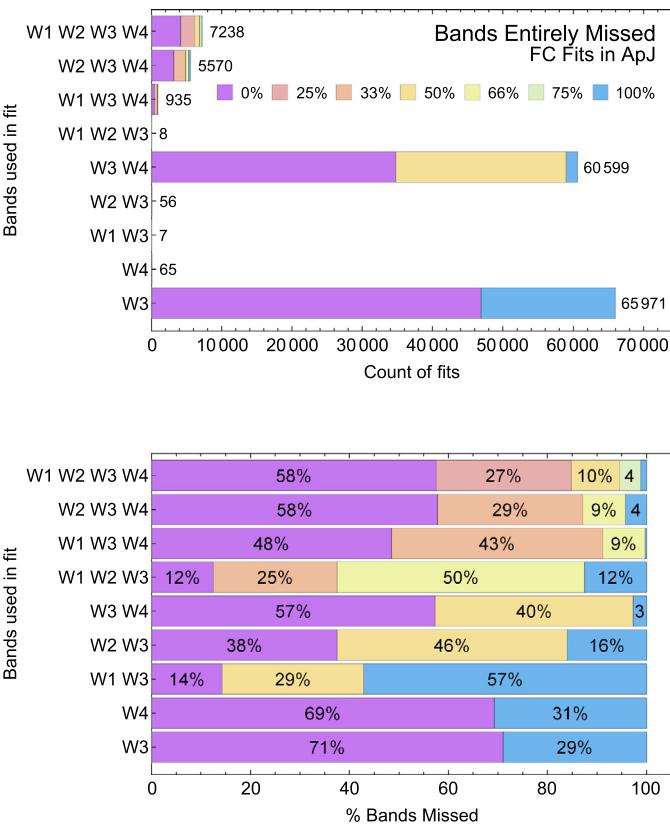


Fig. 4. Bands of data hit and missed by NEOWISE model fits. Top: the absolute number of fits that hit or miss bands of data, organized by the bands that the NEOWISE model fits purport to be based on, for model fits from the FC mission from papers published in ApJ (see Table 1). See SI Figure S8 for comparable plots of NEOWISE fits in the PDS. Bottom: the same data expressed as percentages. For this purpose, a “miss” of a data band means that the curve has residuals of the same sign (all positive or all negative) for all data points in a band. The curve thus lies entirely above or entirely below the data points. A hit means that the residuals have mixed signs or are zero.

tioned in some NEOWISE papers without being documented (see SI Section 4).

7. Relationship between albedo, diameter, and absolute magnitude

The relationship between geometric albedo p_v , diameter D , and absolute visible-band magnitude H is defined for a spherical asteroid by Eq. (1). This can be used to check the consistency of the NEOWISE fits by calculating, for each fit,

$$\Delta p_v = \left| p_v - \left(\frac{1329}{D} \right)^2 10^{-\frac{H}{5}} \right|, \quad (4)$$

$$\Delta D = \left| D - \frac{1329}{\sqrt{p_v}} 10^{-\frac{H}{5}} \right|,$$

where D is in km. The interpretation of these values is the difference between Eq. (1) and the published results. If the fit is perfectly in accord with Eq. (1), and if p_v and D were expressed with infinite numerical precision, then $\Delta p_v = \Delta D = 0$.

NEOWISE results quote D in km to a precision of 0.001 km and p_v to a precision of 0.001. Therefore we would expect that compliance with the rounded figures would occur at plus or minus half of that value; i.e., when either $\Delta p_v \leq 0.0005$ or $\Delta D \leq 0.0005$. Of 164,347 NEOWISE fits in the PDS which report p_v , D , and H , the vast majority yield delta values less than these limits. However, 14,953 fits ($\approx 9\%$ of the total) exceed these limits in Δp_v , ΔD , or both. Eq. (4) cannot be used to determine whether the departure

from Eq. (1) is explained entirely by a difference in p_v of Δp_v , a difference in D of ΔD , or by both p_v and D being off by amounts consistent with Eq. (1).

Fig. 6 plots NEOWISE fits from the PDS archive that deviate from Eq. (1) as points $(\Delta p_v, \Delta D)$ (left panel) and divided by the fit values $(\Delta p_v/p_v, \Delta D/D)$ (right panel).

Remarkably, the cloud of points depicted on the left becomes two curves when plotted on the right. Equally surprising is the size of the deviations ΔD , which can be much more than 10 km, while Δp_v approaches 1.0 (left). On a fractional basis, D can change by up to a factor of 4 (i.e., 400%) and p_v by up to a factor of 24 (i.e., 2400%).

These large factors are the extremes, but they are not isolated outliers. There are 4982 NEOWISE fits for which $\Delta p_v/p_v \geq 10\%$, and 98 fits where $\Delta p_v/p_v \geq 100\%$. These deviations are unlikely to be caused by numerical issues because Eqs. (1) and (4) use only the very simplest numerical functions: multiplication, division, square root and powers of ten (10^x), which should be implemented with high accuracy on any computer system.

The points in Fig. 6 are examples where the NEOWISE results systematically deviate from the definition of geometric albedo for spherical asteroids. This might have occurred if means or medians of values from the NEOWISE Monte Carlo error analysis were used without enforcing Eq. (1) (see Section 8), for example if results from different trials were simply averaged. Because the relationship between D , p_v , and H is nonlinear, whereas averaging is linear, it is possible that the average of D , p_v , and H across multiple trials would fail to satisfy Eq. (1) even when each individual trial does. Other explanations are also possible.

The NEOWISE papers state that the value of H varied in Monte Carlo trials, but they do not say whether the value of H that was reported is the original MPC value or whether it was sometimes modified by the Monte Carlo trials. The NEOWISE results in the PDS archive quote H values with two digits right of the decimal point (i.e., 0.00) but rounded to the nearest 0.1 magnitude, suggesting that they came from MPC values, presumably downloaded from MPC at some point prior to publication of the original papers. Unfortunately the MPC does not maintain a change log or version system that would allow a comparison to the database as of the original download date.

The NEOWISE papers also state that the H - G phase-law slope parameter G is varied independently of H in the Monte Carlo analysis. For most asteroids analyzed by NEOWISE, G has the default value of $G = 0.15$. This suggests that the MPC value of H for these asteroids is likely due to observations from a limited set of phase angles, resulting in too little data to estimate G . If the phase angles measured were near opposition, then the value of H would be largely independent of G , and the approach of varying the two independently to within typical observational error would be reasonable. However, if the value of H was based on limited observations at a phase angle far from opposition, then there is a strong relationship between the MPC-derived value of H and the assumption that $G = 0.15$. In that case, it may not be appropriate to vary H and G independently because uncertainty in H will be the product of an independent range of variation due to photometric uncertainty multiplied by a component due to uncertainty in G . Unfortunately, the necessary information to calculate the correct correlation is currently lacking from the MPC or other standard sources. Recent work on determining H and G from photometric surveys seems to be in the right direction for taking this into account in the future (Veres et al., 2015).

A separate issue with p_v , D , and H values in the NEOWISE analyses is that H values for many asteroids are significantly different than those currently in the Minor Planet Center database. Using 2017 values for the asteroids in Masiero/MB:Pre results in changes to the value of p_v by 10% or more for 55% of the NEOWISE results.

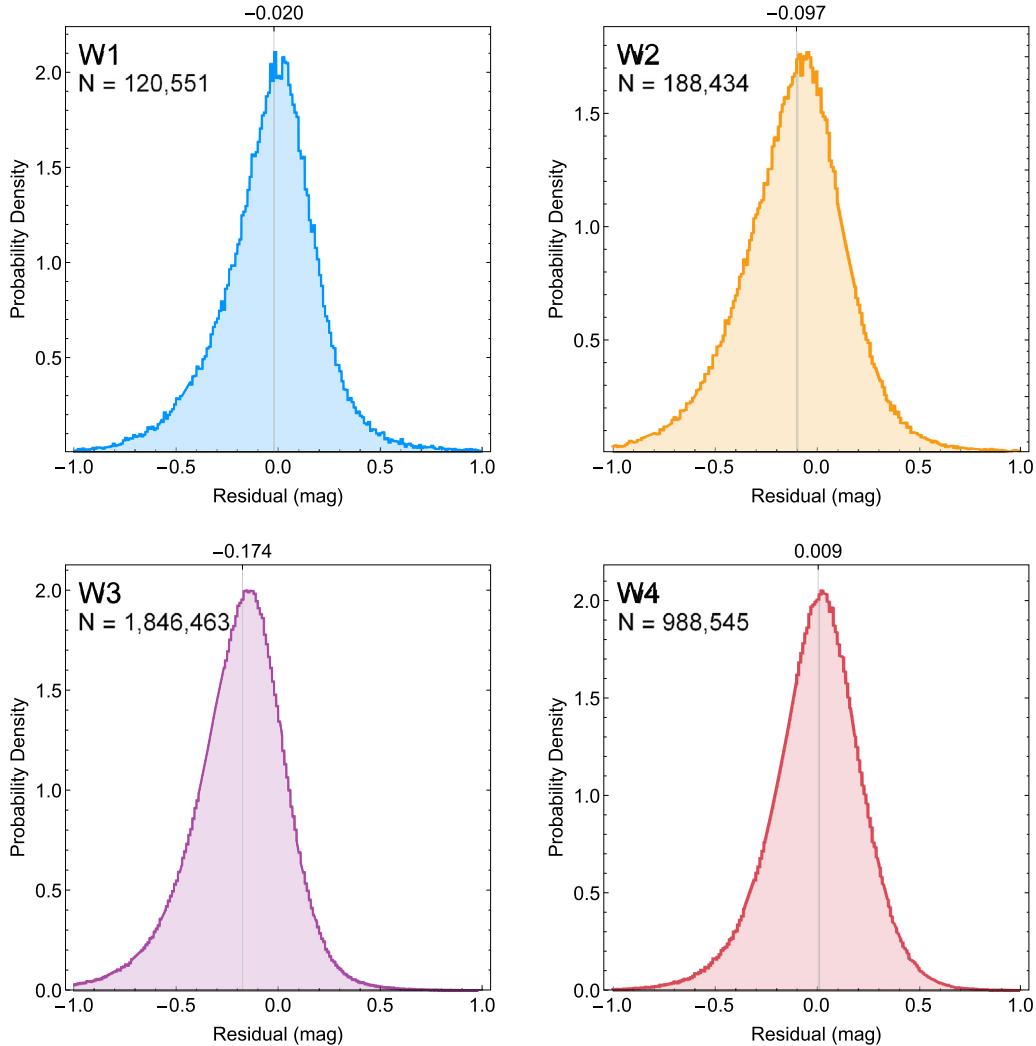


Fig. 5. Residuals in each band from NEOWISE fits from the FC mission (SI Figure S8). The residuals are approximately symmetric, and the median (vertical lines) are near zero, demonstrating that for each band the NEOWISE curves are approximately balanced between being too low and too high. No systematic shift is apparent between the fluxes calculated here and those calculated by NEOWISE papers during model fitting. The median for W3 has the largest deviation from zero among the four bands, possibly due to a linear correction to the W3 band that is mentioned briefly in the NEOWISE papers without supporting details.

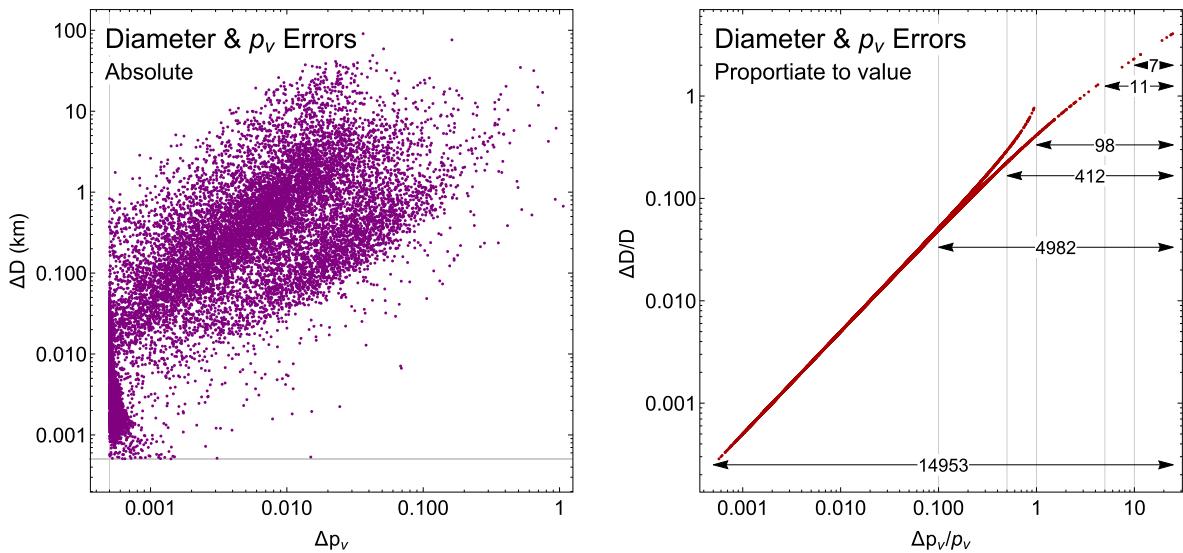


Fig. 6. NEOWISE fits that violate the definition of visible-band, geometric albedo p_v . The geometric albedo of a spherical asteroid is defined by Eq. (1). Left panel: $(\Delta p_v, \Delta D)$ from Eq. (4) for 14,953 NEOWISE fits from the PDS that have one of $(\Delta p_v, \Delta D)$ greater than 0.0005, the value expected from the fact that published NEOWISE fits have p_v and D rounded to 0.001. Right panel: the same points divided by the values $(\Delta p_v/p_v, \Delta D/D)$. Arrows indicate the number of points summed across the range of Δp_v , for all points (bottom arrow), and those for which $\Delta p_v/p_v \geq 0.1, 0.5, 1.0, 5$, and 10 , respectively.

In extreme cases, p_v changes by large factors ranging from 0.03 to 43.66. For the asteroids in Masiero/MB:NIR, using updated MPC H values results in 59% of asteroids changing p_v by 10% or more, with extreme cases altering p_v by factors ranging from 0.07 to 3.41. Alternatively, one could view the effect on p_v as neutral and the effect on D as a change by roughly the square root of these factors.

Any application of p_v values from NEOWISE should take note of these issues. The exponential dependence of Eq. (1) on H makes p_v and D sensitive to any changes. Fortunately, it is straightforward to update the values of p_v and D by using Eq. (1) (Harris and Harris, 1997).

8. Issues with error analysis

The error analysis presented in the NEOWISE papers is based on a series of Monte Carlo trials that were performed while varying the parameters. Mainzer et al., (2011d) explained the process thus:

Error bars on the model magnitudes and subsolar temperatures were determined for each object by running 25 Monte Carlo (MC) trials that varied the objects' H values by 0.3 mag and the *WISE* magnitudes by their error bars using Gaussian probability distributions.

Other NEOWISE papers state that the H - G slope parameter G was also varied. Just 25 Monte Carlo trials are too few to yield an accurate representation of the errors. None of the NEOWISE studies ran more than 100 trials (see SI Table S4), which is too few to use as a basis for reliable error estimates, especially when varying multiple parameters simultaneously.

As explained in Section 5, the assumption used by the NEOWISE Monte Carlo analysis that *WISE* observational errors can be modeled as a normal distribution with standard error given by the *WISE* pipeline estimated σ is contradicted by the empirical statistics of the double detections. The observational errors have thus been systematically underestimated in each band, by differing factors. Not only are the true observational errors larger, but their distribution is not well approximated by a normal distribution.

There are other issues with the Monte Carlo error analysis. All of the NEOWISE result papers omit details of the procedure that are typically reported as a matter of standard practice. For example, the papers fail to specify whether each central reported value is the original fit or is statistically derived from the trial fits by calculating a mean, a median, or some other statistic. Variations in the techniques used could explain some of the inconsistencies noted in this study, as discussed in Section 7.

Reported errors for estimates of D and other model parameters should include all sources of error, including sampling error as well as non-sampling error. The NEOWISE papers unfortunately addressed only non-sampling errors that occur in measurement—i.e., systematic errors caused by fundamental characteristics of *WISE* instruments and observations. Sampling error was ignored, but it is non-negligible because the NEATM, like most simple thermal models, assumes that asteroids are spherical and exhibit uniform surface properties (p_v , p_{IR} , η , ϵ). Masiero et al. (2011) correctly recognized that this idealized form only approximates real asteroids when a large number of observations can be averaged to generate estimates based on its spherical equivalent:

Although we do not expect all, or even most, asteroids to have a spherical shape, our observations covering \sim 36 h smooth out rotation effects and allow us to determine the effective diameter of a spherical body with the same physical properties. Long period rotators ($P \sim$ days) with large amplitudes, e.g., binary asteroids with mass fraction \sim 1, will have poor fits resulting in a moderate mismeasurement of albedo and diameter.

Crucially, however, these factors affect, to some degree, *all* asteroid observations fit to the model—the impact is not restricted to binary asteroids or those having a long rotational period. Variations in albedo or curvature will be observed at different rotational phases; they are the source of the light-curve effect used to infer asteroid shapes. From a statistical point of view, the variations are a source of population-sampling error, which is unavoidable when working with a finite sample of observations of asteroids' rotational phases. These effects can be quite large.

Observational error for NEOWISE is typically much less than 0.25 magnitude (see Fig. 1 for a sample), with a median of \approx 0.1 magnitude. Light-curve amplitude differences can be much larger. In recent studies, the debiased median light-curve amplitude was roughly \approx 0.2 magnitude, and the upper limit is 1.2 (Durech et al., 2015; Masiero et al., 2009). Although these studies are much smaller than NEOWISE, they suggest that population sampling error from the light-curve effect can be twice or more of the observation sampling error (i.e., 2σ or more) for at least half the asteroid population.

It is important to quantify this sampling error as part of the NEOWISE error analysis. That has not been done to date, and as a result the total error in parameter estimates has been systematically underestimated for the majority of asteroids, with significant consequences for the rest of the analysis.

Minimum χ^2 curve fitting (also known as weighted least-squares regression or WLS) has commonly been used for such analyses in astronomy. The least-square error for each observational data point i is weighted by $1/\sigma_i^2$, where σ_i is the estimated error (Carroll and Ruppert, 1988; Ryan, 2008). WLS is appropriate only if the σ_i accurately estimate *all* known sources of error. The approach is known to fail if the error estimates σ_i are small relative to the true error because the weights over-constrain possible least-squares solutions (Carroll and Ruppert, 1988; Feigelson and Babu, 2013; Ivezić et al., 2014; Ryan, 2008; Wall and Jenkins, 2012).

I attempted to apply WLS using the *WISE* observational σ_i and found that in many cases the fits failed due to being over constrained. That result is unsurprising because the σ_i from the *WISE* pipeline are intended to estimate only instrumental measurement errors, not population sampling error, and also because they appear to have been systematically underestimated (see Section 6).

WLS/minimum χ^2 is an inappropriate statistical tool for analyzing error in *WISE* observational data because we lack sufficient *a priori* estimates of the total error, and in particular of the population sampling error. The NEOWISE papers do not describe the details of the fitting and error analysis that went into them. If WLS was employed, ad hoc workarounds—such as the 40% rule and per-epoch fitting to arbitrary epochs—that pruned outliers and discarded data could have increased the likelihood that WLS or minimum χ^2 would generate fits when using the *WISE* observational σ_i . But the quality of those fits would in many cases be so poor that they would be rejected in a conventional statistical analysis. The results of the analysis presented in Section 6 is consistent with this scenario.

9. Comparison with diameters estimated by other methods

Error estimates measure the precision with which a given thermal-modeling statistical analysis internally estimates parameters, but such estimates are constrained by the assumptions in the model. To determine the true accuracy of diameter estimates produced by thermal modeling, one must compare them to estimates made by independent “gold standard” measurements, such as those obtained from radar returns, stellar occultations, or space-craft flybys. None of the NEOWISE papers examined here included a direct comparison of this kind, and only Mainzer/TMC offers a detailed accuracy analysis of any kind. The approach used in that

study was to obtain modeled NEATM fluxes for 50 hypothetical spherical objects whose diameters had been set equal to the ROS diameter of two moons and 48 asteroids (see Section 3.2). The calculated NEATM fluxes were then compared to observed *WISE* fluxes for the same objects. Table 2 in Mainzer/TMC presents the mean and standard deviation of the residuals of the differences between modeled and actual fluxes, with values reported separately for each object and *WISE* band. The authors noted:

The offsets and errors given in Table 2 can be regarded as the minimum systematic errors in magnitude for minor planets observed by *WISE*/NEOWISE. Since diameter is proportional to the square root of the thermal flux (Eq. (1)), the minimum systematic diameter error due to uncertainties in the color correction is proportional to one-half the error in flux. These magnitude errors result in a minimum systematic error of ~5%–10% for diameters derived from *WISE* data; they are of similar magnitude to the diameter uncertainties of most of the underlying radar and spacecraft measurements, which are ~10% (references are given in Table 1).

In summary, the mean difference in flux across the synthetic objects (i.e., a measure of flux accuracy) was used to estimate the diameter accuracy, on the assumption that diameter error scales as the square root of flux error. Unfortunately this approach is fraught with several difficulties.

In addition to diameter, flux also depends on p_{IR} (or $p_{\text{IR1}}, p_{\text{IR2}}$), p_v , and η (or, equivalently, T_{ss}). Those quantities can be determined *only* by a nonlinear curve fit to the observed fluxes in the four bands. If all of the fluxes were displaced uniformly in magnitude, then p_{IR} and T_{ss} (or equivalently p_v and η) would remain unchanged, and the diameter would change in response. Scaling might be reasonably applied in this case, with care.

It is clear from Table 2 of Mainzer/TMC, however, that each band had a *different* mean flux offset and a different standard deviation in flux differences. That is unsurprising: each band is captured by a different sensor, whose distinct hardware errors and photon count statistics contribute to a band-specific noise background and SNR. These differences among sensors result in the differing statistics in double detections explored in Section 5. But displacements by different amounts in each band shift the apparent temperatures and near-IR albedos of asteroids, affecting the estimates of their diameters. The *only* way to determine the impact of the band variations on diameter is to perform the full NEOWISE model fit procedure to these 50 objects.

A second issue with the data in Table 2 of Mainzer/TMC is that, with just 46 to 52 observations used across the four bands to assess 50 objects, the analysis involves only about one *WISE* observation per asteroid, per band. It is unclear how the single observation used was chosen from the dozens to hundreds of four-band observations NEOWISE has collected for these asteroids, but regardless of how it was chosen, it would seem difficult to justify using a *single observation* to provide a reasonable test of *typical* flux error. Mainzer/TMC offers no justification for this approach.

Although Mainzer/TMC labelled its offsets and errors as “minimum systematic error” in diameter estimates, in subsequent NEOWISE papers, all accuracy claims, including those of “better than 10% in diameter” (Maisiero/MB:Pre) are supported only by a reference to Mainzer/TMC. Unfortunately, an (incorrect) estimate of the *lower bound* on systematic accuracy has been widely misinterpreted as a *typical expected value* for *total* accuracy.

There is an urgent need to make a proper estimate of the NEOWISE diameter errors by applying identical model fitting and diameter estimation to all $\approx 164,000$ asteroids in the NEOWISE corpus, and then comparing the resulting diameters to ROS estimates for the small subset of asteroids for which they are available. Only small steps have yet been taken in this direction.

One NEOWISE study (Mainzer et al., 2011d) compared modeled diameters from IRAS (Tedesco et al., 2002) and Ryan and Woodward (2010) to ROS diameters. It also compared NEOWISE results to IRAS and RW. The IRAS to NEOWISE and RW to NEOWISE comparisons are based on NEOWISE results that include results with diameters that exactly match prior ROS diameters. The intermingling of those ROS-matching results with fit results considerably affects the outcome of the comparisons Table 10.

No numerical quality metrics are presented for the IRAS to NEOWISE comparison by Mainzer et al. (2011d). Instead the IRAS versus NEOWISE results are graphed against one another in the manner of Fig. 8 (*top right*) and from the graph non-numerical conclusions are drawn about the IRAS results (Mainzer et al., 2011d):

It can be seen that the values from Tedesco et al. (2004a) tend to be slightly smaller than NEOWISE-derived diameters. The NEOWISE diameters diverge widely from IRAS for the smallest objects; however, these objects are observed by NEOWISE with high signal-to-noise ratio.

The results presented here that are most comparable to those of Mainzer et al. (2011d) can be seen in the top row of Table 11 labeled “ApJ FC Pre.” This includes results for the first round of NEOWISE papers: Maisiero/MB:Pre, Mainzer/NEO:Pre, Grav/JT:Pre, and Grav/Hilda. Note that this row includes the results that have a diameter that exactly matches prior ROS diameters. The Monte Carlo-based ratio analysis here shows a median ratio $D_{\text{NEOWISE}}/D_{\text{IRAS}}$ of 1.024 and a 68.27% CI of -16.3% to $+17.1\%$. This is consistent with the remark that the NEOWISE diameters are slightly smaller (by 2.4%) than those for IRAS.

If we remove the exact matches and include the subsequently published results from ApJ (i.e., Maisiero/MB:NIR and Grav/MB:NIR), the results are quite different: a median ratio of 0.635 and a 68.27% CI of -35.9% to 113.4% (Table 10). The IRAS diameters are systematically 36.6% larger than the NEOWISE results (having a ratio <1). This is due to the smaller objects that “diverge widely” between NEOWISE and IRAS, as remarked in the quoted passage above.

Numbers for the PDS archive fits are considerably better, with a median ratio $D_{\text{NEOWISE}}/D_{\text{IRAS}}$ of 0.935 and a 68.27% CI of -29.1% to 24.0% when -VB- results are removed. The IRAS diameters are still systematically larger than NEOWISE, but by a much smaller factor (6.6%). Evidently the changes between ApJ and PDS (Tables 2 and 3) strongly impact the comparison to IRAS.

With respect to the RW results, the NEOWISE papers conclude that “the diameters computed by Ryan & Woodward (2010) are systematically higher.” This is consistent with the ratios found in (Tables 10–12, where the median ratio $D_{\text{model}}/D_{\text{RW NEATM}}$ (Table 12) for the ApJ FC Pre case is 0.929 and the 68.27% CI spans -14.6% to 13.5% . Whether exact matches are removed or included has little impact on the overall median ratio of RW diameters to those published in ApJ by NEOWISE.

Mainzer et al. (2011d) also make a graphical comparison to ROS diameters for IRAS and RW results (with graphs analogous to Fig. 7 *top left* and *bottom*), concluding that those studies are “systematically biased.” No similar comparison or graph was made between ROS and NEOWISE results. Such a graph would not be meaningful because the relevant NEOWISE results at that time all had diameters that exactly match the very ROS diameters they would be compared to, so the graph would be the line $y = x$.

Because no quantitative estimates relating IRAS, RW, ROS, and NEOWISE were made, there is no objective support for the conclusion of (Mainzer et al., 2011d):

We conclude that the diameters for minor planets derived from NEOWISE are generally in good agreement with those found by IRAS and are likely more free of systematic biases than the diameters provided in either Tedesco et al. (2002) or Ryan &

Table 10

Comparison of NEOWISE diameter estimates from PDS to IRAS diameters. The method of computing the results and confidence intervals is described in SI Section 7. ApJ FC Pre refers to NEOWISE results from Masiero/MB:Pre, Mainzer/NEO:Pre, Grav/JT:Pre, and Grav/Hilda. Other ApJ data sets includes results from later papers on the FC mission (see Table 1). The method of computing the results and confidence intervals is based on a Monte Carlo approach described in SI Section 7.

Model fits	Grouping	Count	Median	$D_{\text{PDS}} / D_{\text{IRAS}}$	
				68.27% CI	CI length
ApJ FC Pre	Overall	1726	1.024	-16.3%	17.1%
	W1 W2 W3 W4	1362	1.016	-12.9%	15.6%
	W1 W2 W3	17	0.862	-14.5%	23.8%
	W1 W3 W4	53	0.939	-18.4%	15.7%
	W2 W3 W4	149	0.961	-18.2%	19.0%
ApJ	W1 W3	2	0.648	-18.4%	16.6%
non-exact	W2 W3	none	none	none	none
	W3 W4	194	0.893	-26.9%	21.1%
	W3	2	0.352	-46.4%	213.6%
	W4	3	0.901	-10.8%	12.6%
	Overall	1782	0.632	-35.7%	116.0%
	W1 W2 W3 W4	1352	1.015	-12.8%	15.6%
	W1 W2 W3	16	0.849	-13.6%	23.3%
	W1 W3 W4	53	0.939	-18.4%	15.7%
ApJ FC Fits	W2 W3 W4	146	0.960	-18.1%	19.0%
with exact matches removed	W1 W3	2	0.645	-18.7%	16.8%
	W2 W3	none	none	none	none
	W3 W4	194	0.894	-26.9%	21.1%
	W3	2	0.358	-46.9%	208.4%
	W4	3	0.900	-10.6%	12.5%
	Overall	1768	0.635	-35.9%	113.4%
	W1 W2 W3 W4	83	0.994	-13.0%	12.0%
	W1 W2 W3	1	1.094	-10.1%	10.1%
ApJ FC fits	W1 W3 W4	none	none	none	none
only exact matches	W2 W3 W4	3	0.969	-19.4%	15.0%
	W3	none	none	none	none
	Overall	87	0.993	-13.4%	12.2%
	W1 W2 W3 W4	1278	0.990	-11.6%	12.8%
	W1 W2 W3	235	0.992	-15.7%	18.8%
	W1 W3 W4	74	0.929	-19.3%	16.6%
	W2 W3 W4	206	0.976	-18.1%	16.9%
PDS fits	W1 W2	1387	0.984	-26.5%	27.0%
with exact matches removed	W1 W3	6	0.818	-23.4%	25.3%
	W2 W3	48	1.012	-16.5%	20.3%
	W3 W4	201	0.901	-26.3%	20.5%
	W1	none	none	none	none
	W2	144	0.903	-37.5%	46.1%
	W3	37	0.955	-33.0%	25.6%
	W4	2	0.900	-11.0%	12.6%
	Overall	3618	0.935	-29.1%	24.0%
	DVBI	237	0.982	-15.1%	14.9%
	DVB2	1206	0.990	-11.5%	12.6%
	DVB-	473	0.947	-20.3%	18.5%
PDS fits	DV-I	824	1.029	-17.9%	17.3%
	DV-2	20	0.862	-17.3%	35.4%
	DV--	1122	0.950	-32.8%	37.4%
	-VB-	none	none	none	none
	Overall	3882	0.951	-26.7%	28.4%
RW	NEATM	116	1.089	-7.3%	8.1%
	STM	116	0.969	-11.3%	10.1%
IRAS	All	2228	1.000	-7.8%	8.4%
AKARI	All	1985	1.020	-11.8%	13.3%
					25.6%

Woodward (2010). Together with Mainzer et al. (2011b), this demonstrates that the NEOWISE data set will produce good quality physical parameters for the $> 157,000$ minor planets it contains.

It is inconsistent to conclude that the IRAS and RW results have systematic biases by comparing them to ROS estimates and to then

neglect to make the same comparison for NEOWISE results. Without quantitative comparisons, “generally good agreement” has no well-defined meaning. Neither the criticisms of IRAS and RW nor visual comparisons of the fits support a conclusion about “good quality” parameters. A direct and quantitative comparison of ROS to NEOWISE is required, but that could not be done because the

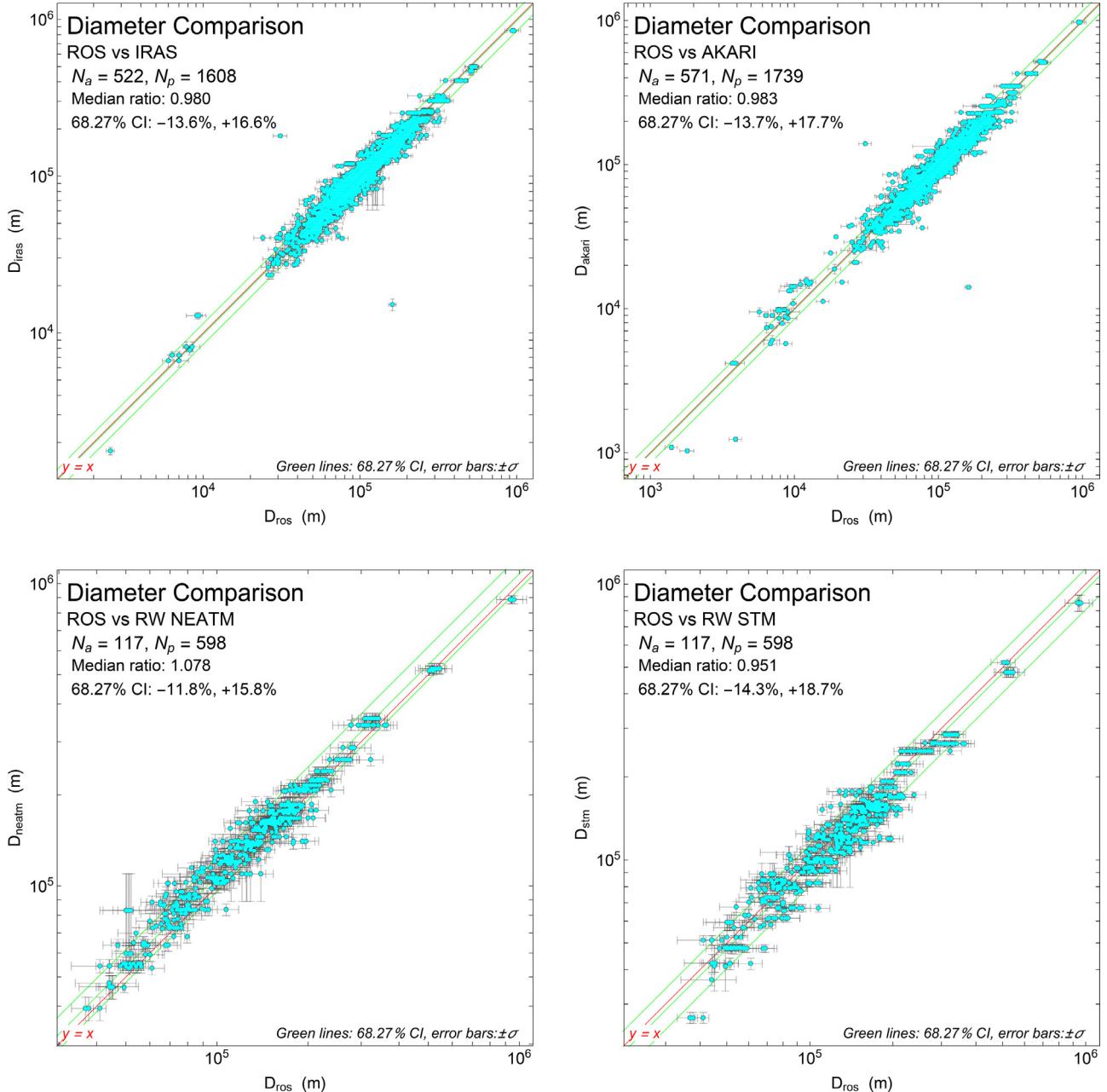


Fig. 7. Comparison of diameter estimates derived from non-NEOWISE thermal modeling to estimates made from radar, occultation, and spacecraft (ROS) observations. D_{ROS} , diameters for asteroids from the ROS literature, with $\pm\sigma$ error estimates. D_{IRAS} , diameters from thermal modeling by IRAS (Tedesco et al., 2002). D_{NEATM} , diameters from Ryan and Woodward (2010). D_{akari} , diameters from AKARI, (Usui et al., 2014, 2012, 2011a, 2011b). N_a , the number of asteroids. N_p , the number of data points (multiple points for some asteroids).

NEOWISE diameters exactly matched the ROS diameters available at that time.

A further complication is that the NEOWISE results come from 10 different models applied to 12 different combination of WISE bands. In order to properly assess the accuracy of the result set as a whole, one must first assess the accuracy for each model code and band combination in Table 4. There is no reason to believe that the accuracy of one model and band combination should imply the same or similar accuracy for different models and different bands. This is particularly true between the four-band models DVBI and DVB2, which comprise a small percentage of the overall results, and the single band results that are the majority. Asteroids for which ROS diameters are available tend to be large or close to Earth (or both), and thus easier to observe by those methods. The

ideal approach would be to fit each ROS asteroid with as many of the model code and band combinations in Table 4 as possible, given the observational data available. That would give direct information on the difference between estimation with multiple bands and with just one band, as well as on the differences among various models.

As an example, there are only 17 results for ROS asteroids that use the W3 band alone (Table 10), which is unfortunate since there are 74,853 W3-only results (45% of total) in the NEOWISE result set (Table 4). Any observation of a ROS asteroid that includes W3 (almost all of them) could have been used to make a W3-only fit, and that would greatly increase the statistical precision at which we could estimate the accuracy of W3-only fits.

Table 11

Comparison of NEOWISE and other diameter estimates to ROS diameters. Model fits marked “exact matches removed” are based on data pruned to remove diameter results that exactly match prior ROS diameters. Fits marked “only exact matches” contain only fits having diameters that exactly match prior ROS diameters. ApJ FC Pre refers to NEOWISE results from Masiero/MB:Pre, Mainzer/NEO:Pre, Grav/JT:Pre and Grav/Hilda. Other ApJ data sets include results from later papers from the FC mission (see Table 1). The method of computing the results and confidence intervals is based on a Monte Carlo approach described in SI Section 7.

Model fits	Grouping	Count	Median	$D_{\text{model}}/D_{\text{ROS}}$	
				68.27% CI	CI length
ApJ FC Pre	Overall	483	1.013	-14.5%	18.8%
	W1 W2 W3 W4	434	0.988	-13.7%	18.6%
	W1 W2 W3	3	1.046	-10.5%	11.9%
	W1 W3 W4	12	1.023	-16.1%	24.2%
	W2 W3 W4	11	0.999	-27.4%	17.6%
ApJ non-exact	W1 W3	none	none	none	none
	W2 W3	none	none	none	none
	W3 W4	19	1.052	-18.3%	18.7%
	W3	12	1.013	-24.6%	28.8%
	W4	none	none	none	none
	Overall	491	1.028	-21.6%	23.6%
ApJ FC Fits with exact matches removed	W1 W2 W3 W4	423	0.985	-13.6%	18.7%
	W1 W2 W3	2	1.055	-10.1%	12.2%
	W1 W3 W4	10	1.018	-15.3%	22.3%
	W2 W3 W4	4	0.934	-87.9%	22.0%
	W1 W3	none	none	none	none
	W2 W3	none	none	none	none
	W3 W4	19	1.052	-18.4%	18.6%
	W3	10	1.029	-27.8%	31.2%
	W4	none	none	none	none
	Overall	468	1.034	-25.8%	25.0%
ApJ FC fits only exact matches	W1 W2 W3 W4	93	0.991	-14.1%	17.7%
	W1 W2 W3	1	1.027	-11.0%	11.8%
	W1 W3 W4	2	1.077	-22.2%	24.1%
	W2 W3 W4	7	1.004	-15.6%	18.5%
	W3	3	1.000	-15.6%	19.0%
	Overall	106	0.994	-14.4%	17.8%
PDS fits with exact matches removed	W1 W2 W3 W4	429	0.963	-13.1%	17.4%
	W1 W2 W3	52	0.980	-18.9%	20.9%
	W1 W3 W4	11	1.015	-14.5%	20.4%
	W2 W3 W4	15	1.045	-33.1%	18.7%
	W1 W2	462	0.972	-28.3%	32.2%
	W1 W3	1	1.233	-15.7%	20.0%
	W2 W3	2	1.123	-23.0%	27.3%
	W3 W4	18	1.058	-17.6%	18.9%
	W1	none	none	none	none
	W2	10	1.095	-38.5%	55.0%
	W3	17	1.048	-24.4%	32.1%
	W4	none	none	none	none
	Overall	1017	1.050	-22.7%	27.3%
PDS fits	DVBI	50	0.995	-15.2%	17.7%
	DVB2	401	0.961	-13.0%	17.2%
	DVB-	49	1.020	-24.0%	19.0%
	DV-I	279	1.013	-19.3%	23.2%
	DV-2	3	1.224	-10.5%	12.5%
	DV--	351	0.950	-32.9%	40.0%
	-VB-	5	1.002	-18.4%	22.2%
	Overall	1138	0.981	-28.3%	30.2%
RW	NEATM	117	1.078	-11.8%	15.8%
	STM	117	0.951	-14.3%	18.7%
IRAS	All	522	0.980	-13.6%	16.6%
AKARI	All	571	0.983	-13.7%	17.7%
					30.9%

This was not the approach taken by the NEOWISE group, and their results cannot now be replicated, so any approach to using ROS asteroids to assess the accuracy is regrettably crude. Nevertheless, I made a rough assessment by using the NEOWISE diameters from the ApJ papers and the PDS archive. ROS diameters were obtained from the literature (see SI Table S5). Although the NEOWISE diameters for many asteroids exactly match the diameter of prior ROS results (Section 4 above), many new ROS results have been published since then, and these diameters are included in the analysis here (Brozović et al., 2017; Greenberg et al., 2016; Hanuš et al., 2013; Lawrence et al., 2015; Lehtinen et al., 2016; Naidu et al., 2015; Shepard et al., 2017; 2015). I also compared NEOWISE diameters to estimates derived from thermal models published by IRAS, RW, and AKARI (see Table 1). Because multiple ROS estimates and NEOWISE model fits often exist for an asteroid, a Monte Carlo procedure was used to determine the ratio of NEOWISE and ROS sources $D_{\text{neo}}/D_{\text{ros}}$ —see SI Section 7 for details.

Table 11 summarizes the results of this analysis for NEOWISE fits. The NEOWISE ApJ papers (with the diameters that exactly match a prior ROS result removed) are presented first. For the largest group of these fits (434 total, all based on four WISE bands), systematic deviation was -3% in the median, and the 68.27% CI was -13.7% to $+18.6\%$, significantly larger than the claimed error of “better than 10%,” despite this being arguably the most favorable case.

Weighting the results by the number of observations in each category, we can extrapolate a very crude estimate of the overall accuracy of the FC data set, with the caveat that very few fits exist for some of the most important examples (just 10 for W3 only), and many other band combinations are not available at all. The result is that NEOWISE results from the FC mission, taken overall, have systematic bias of $\approx +4\%$ (i.e., about 4% too large) and 68.27% CI (i.e., 1σ) random variations from -26.0% to $+24.6\%$, or two to three times less accurate than claimed.

The same approach was applied to the NEOWISE FC results as published, including the diameters that exactly match prior ROS results. Unsurprisingly, those errors are much smaller, though nonzero because multiple ROS diameter estimates have been published for some asteroids.

My analysis of the PDS data set produced similar results, with the caveat that the omissions, revisions, and new results change some of the numbers slightly. The introductory text to the PDS archive (in file dataset.cat) states that the diameter accuracy is $\pm 10\%$ when all parameters are fitted, and $\pm 15\%$ for the results that use an assumed value of a parameter (i.e., DV-I, DV--, and DVB--, discussed in Section 3). The best tested model (with 407 results) in the PDS results is DV2, which has systematic underestimate of diameter by $\approx -6\%$ and 68.27% CI from -15.3% to $+18.3\%$, significantly less accurate than claimed. The best tested model (with 358 results) for models that have assumed parameters is DV--, which has a systematic underestimate of $\approx -7\%$ and 68.27% CI from -33.5% to $+41.4\%$, also significantly less accurate than claimed.

Overall, a crude extrapolation of these results to the entire corpus suggests that the PDS results have a systematic bias of $\approx +4\%$ and 68.27% CI from -23% to $+27.4\%$, when extrapolated using a weighted average on bands used. The systematic bias is $\approx -3\%$ and the 68.27% CI spans -29.4% to $+31.3\%$ when extrapolating on a per-model basis. There are insufficient examples to extrapolate on the basis of both model and band.

I also compared the thermal modeling results from IRAS, RW, and AKARI to ROS estimates (Table 11 and Fig. 7). The results contradict the conclusions of Mainzer et al. (2011d) that NEOWISE is more accurate and freer of systematic bias. IRAS and AKARI both have $\approx -2\%$ systematic biases and 68.27% CI of -13% to $+18\%$. The bias is comparable to the bias found in NEOWISE results, and the

68.27% CI is smaller. The RW NEATM and STM results have larger bias and comparable errors, although the comparison is difficult to make fairly because the RW data set is so much smaller than the IRAS and AKARI data sets.

Cross-comparisons among the NEOWISE data sets and IRAS, RW, and AKARI are shown graphically in Fig. 8 and numerically in Tables 10–13. Broad surveys often assume that only systematic errors are relevant when sampling the overall population of asteroids because random errors average out, however that is true only if the random errors are uncorrelated to diameter or other physical properties. Such a correlation makes the random errors become, in effect, systematic errors. Fig. 8 and SI Figure S10 show this effect clearly. Some model codes (Fig. 8) and some band combinations (SI Figure S10) are associated primarily with either the smaller or larger end of the diameter range covered by NEOWISE. Each of these different combinations has distinct systematic and random errors; it is far from clear whether these will average out.

Usui et al. (2014) also performed a cross comparison of NEOWISE, IRAS, and AKARI results. Note that they used a different methodology and data sets than those used here. For example, they compared the diameter estimates directly, rather than using a Monte Carlo approach that takes into account the uncertainty in the estimates (see SI Section S7).

The NEOWISE dataset that Usui et al. (2014) used is most closely comparable to the ApJ FC Pre fits in the top row of Table 13, which shows a median ratio $D_{\text{model}}/D_{\text{AKARI}}$ of 0.971. The exactly matching cases (row ApJ FC only exact matches to ROS) by themselves have a median ratio of 1.030. These values are in approximate agreement with Usui et al. (2014), who find a mean ratio between AKARI and NEOWISE of 0.982 using a different method. In comparing AKARI to IRAS, their mean ratio of 0.982 is similarly consistent with the median value of 0.980 in Table 13 (row labeled IRAS).

10. Discussion

The initial NEOWISE studies listed in Table 1, as well as subsequent papers by the NEOWISE research group (e.g., Mainzer et al., 2015b) have suggested that the task of understanding the WISE/NEOWISE asteroid data is a solved problem. The NEOWISE results were claimed to be highly accurate, providing diameter estimates to within 10%, as well as less affected by systematic bias than previous thermal modeling studies. The NEOWISE diameters were already at the accuracy limits found from ROS estimates, according to the most recent review (Mainzer et al., 2015b).

These statements strongly suggest that there is little more to be done in analyzing the raw observational data. This is endorsed by the fact that since these studies appeared, most of the research effort by the NEOWISE team has concentrated on applying the NEOWISE results rather than exploring ways to improve the analysis of the observational data. Although some re-analysis of previous data has been done (Masiero/MB:NIR, Grav/JT/Tax), together they cover less than 2% of the asteroids observed. No external researchers have been able to replicate the NEOWISE fit results by reaching the same answers by calculation from the observational data, in large part because the NEOWISE analysis procedure has not been fully described publicly (see SI Section S4), and this has frustrated attempts to provide independent analysis.

The results presented in this study show that substantial work remains to be done in analyzing data from the WISE/NEOWISE mission and interpreting it for asteroid science. Far from being a closed topic, it urgently needs attention. This should be viewed as a great opportunity for the field of asteroid studies: much of the NEOWISE data has yet to be analyzed at all, because it was discarded by the analysis done so far. While this may require developing new techniques, the potential benefit is great.

Table 12

Comparison of NEOWISE diameter estimates from PDS to Ryan and Woodward diameters. The method of computing the results and confidence intervals is described in SI Section 7. ApJ FC Pre refers to NEOWISE results from Masiero/MB:Pre, Mainzer/NEO:Pre, Grav/JT:Pre and Grav/Hilda. Other ApJ data sets includes results from later papers (see Table 1). The method of computing the results and confidence intervals is based on a Monte Carlo approach described in SI Section 7.

Model fits	Grouping	Count	$D_{\text{PDS}} / D_{\text{RW NEATM}}$			$D_{\text{PDS}} / D_{\text{RW STM}}$				
			Median	68.27% CI	CI length	Median	68.27% CI	CI length		
ApJ FC Pre	Overall	86	0.929	-14.6%	13.5%	26.1%	1.045	-15.5%	20.2%	35.7%
ApJ non-exact	W1 W2 W3 W4	83	0.908	-11.9%	12.3%	22.0%	1.028	-13.4%	18.2%	32.5%
	W1 W2 W3	2	1.005	-8.5%	10.0%	18.6%	1.210	-11.2%	9.3%	24.8%
	W1 W3 W4	none	none	none	none	none	none	none	none	none
	W2 W3 W4	2	0.826	-14.6%	28.8%	35.8%	0.891	-11.8%	20.9%	29.2%
	W1 W3	none	none	none	none	none	none	none	none	none
	W2 W3	none	none	none	none	none	none	none	none	none
	W3 W4	none	none	none	none	none	none	none	none	none
	W3	none	none	none	none	none	none	none	none	none
	W4	none	none	none	none	none	none	none	none	none
	Overall	87	0.861	-13.4%	21.9%	30.0%	0.949	-12.5%	19.8%	30.6%
ApJ FC Fits with exact matches removed	W1 W2 W3 W4	75	0.902	-10.0%	10.6%	18.6%	1.032	-11.5%	16.8%	29.2%
	W1 W2 W3	1	0.982	-7.1%	7.4%	14.2%	1.265	-6.8%	6.8%	17.3%
	W1 W3 W4	none	none	none	none	none	none	none	none	none
	W2 W3 W4	none	none	none	none	none	none	none	none	none
	W1 W3	none	none	none	none	none	none	none	none	none
	W2 W3	none	none	none	none	none	none	none	none	none
	W3 W4	none	none	none	none	none	none	none	none	none
	W3	none	none	none	none	none	none	none	none	none
	W4	none	none	none	none	none	none	none	none	none
	Overall	76	0.903	-10.0%	10.6%	18.6%	1.033	-11.5%	16.7%	29.1%
ApJ FC fits only exact matches	W1 W2 W3 W4	59	0.906	-15.5%	13.9%	26.7%	1.024	-16.3%	18.8%	35.9%
	W1 W2 W3	1	1.038	-10.2%	10.5%	21.5%	1.135	-10.3%	10.3%	23.4%
	W1 W3 W4	none	none	none	none	none	none	none	none	none
	W2 W3 W4	2	0.828	-14.7%	28.0%	35.4%	0.890	-11.8%	20.5%	28.7%
	W3	none	none	none	none	none	none	none	none	none
	Overall	62	0.902	-15.4%	14.9%	27.2%	1.015	-15.9%	18.8%	35.3%
PDS fits with exact matches removed	W1 W2 W3 W4	81	0.891	-9.8%	10.3%	17.8%	1.020	-13.3%	16.2%	30.0%
	W1 W2 W3	8	0.933	-11.2%	25.9%	34.6%	1.150	-23.4%	19.2%	49.1%
	W1 W3 W4	none	none	none	none	none	none	none	none	none
	W2 W3 W4	2	0.937	-6.8%	8.1%	14.0%	0.999	-4.2%	4.3%	8.5%
	W1 W2	90	0.894	-24.5%	29.7%	48.5%	1.018	-25.0%	31.1%	57.2%
	W1 W3	none	none	none	none	none	none	none	none	none
	W2 W3	none	none	none	none	none	none	none	none	none
	W3 W4	none	none	none	none	none	none	none	none	none
	W1	none	none	none	none	none	none	none	none	none
	W2	none	none	none	none	none	none	none	none	none
	W3	3	0.976	-16.9%	16.2%	32.3%	1.126	-18.6%	24.8%	48.9%
	W4	none	none	none	none	none	none	none	none	none
	Overall	184	0.965	-16.3%	16.3%	31.4%	1.107	-17.8%	23.3%	45.8%
PDS fits	DVBI	7	0.932	-10.2%	25.8%	33.5%	1.159	-23.6%	14.6%	44.3%
	DVB2	71	0.891	-9.6%	9.6%	17.1%	1.023	-11.5%	15.6%	27.7%
	DVB-	10	0.874	-9.4%	10.4%	17.3%	0.925	-14.8%	15.8%	28.3%
	DV-I	47	0.943	-14.3%	20.7%	33.0%	1.078	-15.7%	26.3%	45.3%
	DV-2	2	1.188	-8.2%	7.0%	18.0%	1.399	-9.2%	11.4%	28.8%
	DV--	72	0.886	-32.2%	35.5%	60.0%	0.999	-32.0%	38.6%	70.5%
	-VB-	none	none	none	none	none	none	none	none	none
	Overall	209	0.882	-21.6%	24.1%	40.4%	0.971	-24.0%	28.1%	51.2%
RW	NEATM	118	1.000	-6.9%	7.4%	14.3%	1.134	-11.6%	13.8%	28.9%
	STM	118	0.882	-12.1%	13.2%	22.3%	1.000	-5.0%	5.2%	10.2%
IRAS	All	116	0.918	-7.5%	7.9%	14.2%	1.032	-9.2%	12.8%	22.6%
AKARI	All	118	0.917	-9.5%	11.1%	18.9%	1.036	-12.3%	15.7%	29.0%

Table 13

Comparison of NEOWISE diameter estimates from PDS to AKARI diameters. The method of computing the results and confidence intervals is described in SI Section 7. ApJ FC Pre refers to NEOWISE results from Masiero/MB:Pre, Mainzer/NEO:Pre, Grav/JT:Pre and Grav/Hilda. Other ApJ FC data sets includes results from later papers on the FC mission (see Table 1). The method of computing the results and confidence intervals is based on a Monte Carlo approach described in SI Section 7.

Model fits	Grouping	Count	Median	$D_{\text{PDS}}/D_{\text{AKARI}}$		CI length
				68.27% CI	CI length	
ApJ FC Pre	Overall	3998	0.971	-15.5%	16.0%	31.5%
	W1 W2 W3 W4	2097	0.985	-12.3%	13.6%	25.4%
	W1 W2 W3	22	0.884	-16.3%	22.6%	34.4%
	W1 W3 W4	149	0.929	-13.5%	14.6%	26.2%
	W2 W3 W4	622	0.942	-15.7%	16.3%	30.2%
ApJ non-exact	W1 W3	1	0.898	-12.0%	13.4%	22.8%
	W2 W3	1	1.159	-7.6%	9.0%	19.3%
	W3 W4	1237	0.916	-16.4%	17.4%	31.0%
	W3	15	0.965	-29.5%	41.5%	68.6%
	W4	7	0.988	-15.5%	28.7%	43.7%
	Overall	4151	0.944	-22.7%	29.1%	49.3%
ApJ FC Fits with exact matches removed	W1 W2 W3 W4	2087	0.984	-12.1%	13.5%	25.2%
	W1 W2 W3	21	0.874	-15.7%	22.8%	33.7%
	W1 W3 W4	148	0.928	-13.5%	14.4%	25.9%
	W2 W3 W4	618	0.941	-15.7%	16.3%	30.1%
	W1 W3	1	0.899	-11.6%	13.4%	22.5%
	W2 W3	1	1.158	-7.6%	8.9%	19.1%
	W3 W4	1237	0.916	-16.4%	17.4%	31.0%
	W3	14	0.936	-28.7%	44.1%	68.2%
	W4	7	0.987	-15.5%	28.4%	43.4%
	Overall	4134	0.929	-22.3%	30.4%	49.1%
ApJ FC fits only exact matches	W1 W2 W3 W4	91	1.008	-13.9%	12.9%	27.1%
	W1 W2 W3	1	1.059	-9.4%	9.5%	20.0%
	W1 W3 W4	1	1.519	-6.8%	6.9%	20.9%
	W2 W3 W4	4	1.029	-29.1%	23.8%	54.5%
	W3	1	1.282	-10.8%	11.2%	28.1%
	Overall	98	1.030	-14.6%	13.4%	28.7%
PDS fits with exact matches removed	W1 W2 W3 W4	1868	0.964	-11.2%	11.2%	21.6%
	W1 W2 W3	386	0.962	-17.6%	19.9%	36.1%
	W1 W3 W4	212	0.932	-11.9%	13.5%	23.7%
	W2 W3 W4	773	0.946	-15.7%	15.7%	29.8%
	W1 W2	2283	0.946	-25.7%	27.4%	50.2%
	W1 W3	10	0.953	-15.6%	17.2%	31.3%
	W2 W3	192	0.956	-18.0%	22.4%	38.6%
	W3 W4	1260	0.917	-16.2%	17.0%	30.5%
	W1	none	none	none	none	none
	W2	761	0.897	-33.6%	43.3%	69.0%
	W3	162	0.986	-22.3%	25.1%	46.7%
	W4	6	1.013	-18.7%	28.3%	47.5%
	Overall	7913	0.951	-20.1%	22.3%	40.4%
PDS fits	DVBI	426	0.955	-12.5%	16.3%	27.6%
	DVB2	1744	0.963	-11.4%	10.9%	21.5%
	DVB-	2164	0.929	-16.1%	16.6%	30.4%
	DV-I	1297	0.994	-17.3%	17.7%	34.8%
	DV-2	22	0.899	-17.5%	34.5%	46.8%
	DV--	2351	0.918	-31.3%	37.0%	62.7%
	-VB-	1	1.287	-14.3%	14.5%	37.0%
	Overall	8005	0.925	-24.1%	27.3%	47.4%
RW	NEATM	118	1.090	-10.0%	10.5%	22.3%
	STM	118	0.966	-13.5%	14.0%	26.6%
IRAS	All	1985	0.980	-11.7%	13.4%	24.6%
AKARI	All	5120	1.000	-5.6%	5.9%	11.5%

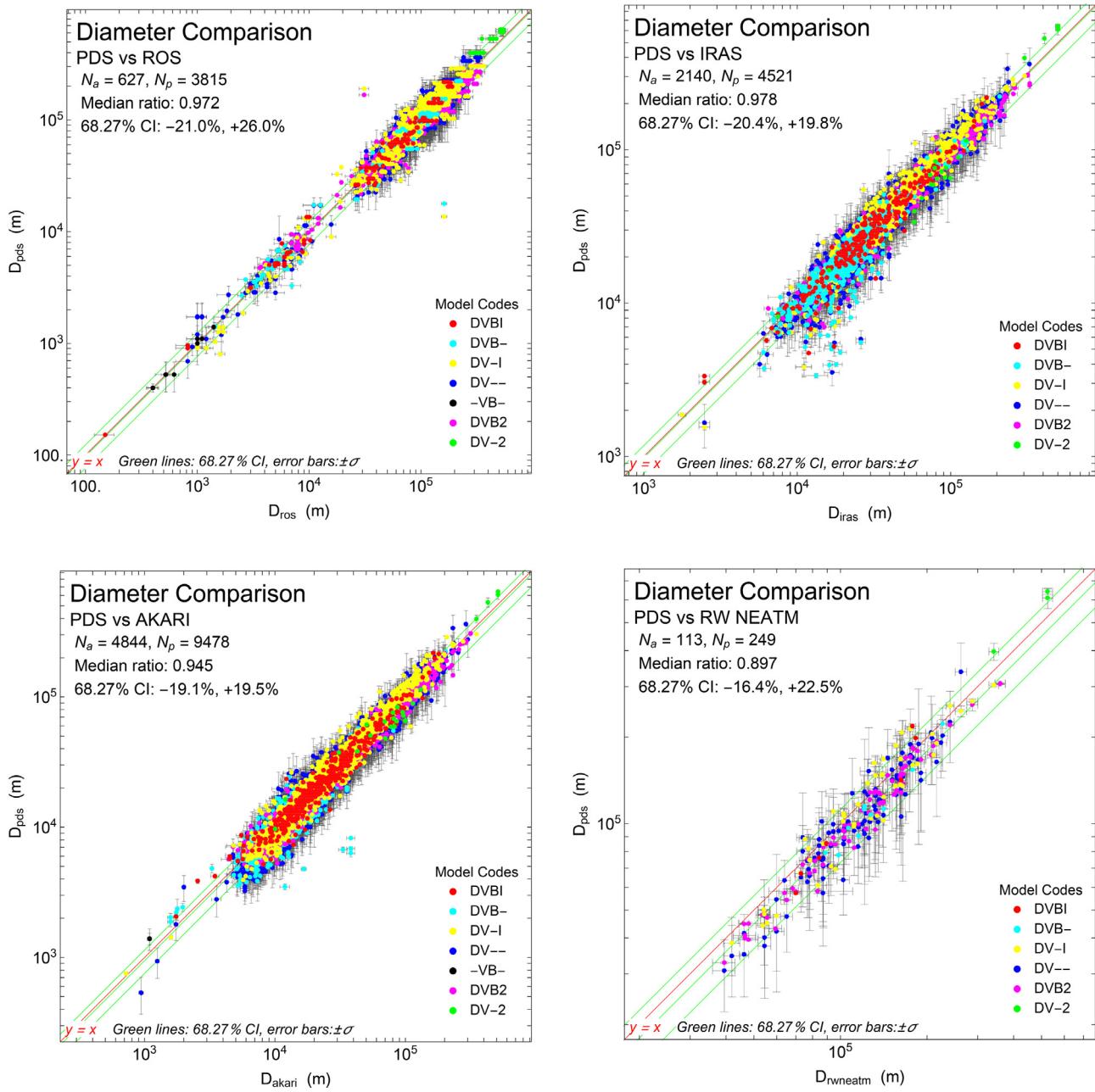


Fig. 8. Comparison of NEOWISE diameter estimates with other sources. D_{ROS} , diameters for asteroids from the ROS literature, with $\pm \sigma$ error estimates. D_{IRAS} , diameters from thermal modeling by IRAS (Tedesco et al., 2002). D_{RWNEATM} , diameters from Ryan and Woodward (2010). D_{PDS} , diameters from the PDS archive of NEOWISE results (Table 1). D_{akari} , diameters from AKARI, (Usui et al., 2014, 2012, 2011a, 2011b). N_a , the number of asteroid fits. N_p , the number of data points (there are multiple points for many asteroids). See SI Figure S10 for a version of these plots colored by bands used rather than model codes.

The data that has already been processed also has tremendous potential for reanalysis. As shown in the sections above, when the analytical methods employed by the NEOWISE group are closely examined, one finds many potential avenues for improvement. It is important to mine the existing data because no current or planned mission will provide comparable four-band IR measurements of as many asteroids.

Some of these areas for improvement are straightforward: avoiding model fits that miss entire bands of data; computing all diameters by thermal modeling rather than mixing in some that exactly match prior ROS sources. Other areas may require more work, such as creating a modeling approach that uses all of the available data, including across mission phases rather than limiting analysis to artificial 3- to 10-day epochs.

An important issue for future modeling of the WISE/NEOWISE data is the treatment of exceptions and special cases. The NEOWISE analysis procedure appears to create exceptional parameter values that break the general rules of assumptions about η or p_{IR} (Tables 5 and 6). Unusually high and low values of p_v or p_{IR} appear to have been clipped (Table 6 and SI Figure S1), and $\approx 14,000$ cases have an incorrect relationship between p_v , D , and H . While there is a legitimate role for outlier rejection if observations are truly faulty, and for special treatment of truly unusual cases, altering the analytical method for special cases runs the risk of hiding cases that could be of scientific interest. Even modeling results that produce non-physical parameter estimates can be useful if they point us to problems with our simple thermal models or their application.

Until the NEOWISE observation data is reanalyzed in an open, transparent, and replicable fashion, the existing NEOWISE results may be the best available for many asteroids. IRAS, RW, and AKARI results in general have comparable or lower systematic bias and errors, but they are available for many fewer asteroids.

The sections above show that it is important for a user of the NEOWISE results to be aware of the disparate sources of the results (47 different combinations of bands and models – [Table 4](#)) and cognizant that the data differs between the original ApJ publications and the more recent PDS archive ([Tables 2](#) and [9](#)). More than half of those results come from fits to a single band of data—and more than 9,000 NEOWISE results were analyzed by relying on W2 data alone ([Table 4](#)). The W2 band does not have dominant thermal emission, so the NEOWISE analysis relies on assumptions about η and p_{IR} , and those assumptions largely determine the outcome.

Observational error is a reality for astronomy, and correctly accounting for it is central to any study based on observations. Building on the work of Hanuš and co-workers ([Hanuš et al., 2015](#)), I have shown here that the double detections provide an empirical check on *WISE* observational errors. The standard errors in each band are much larger than the estimated σ provided by the *WISE* pipeline, and the distribution of errors is not a normal distribution ([Figs. 1](#) and [2](#)). The Monte Carlo error analysis in all of the NEOWISE papers is based on the assumption of normal distribution with *WISE* standard errors, so this finding largely refutes its foundational assumptions.

Perhaps the biggest practical issue to using NEOWISE model fits is that many do not actually fit the data they purport to ([Figs. 3](#) and [4](#)). About a third of the single-band model fits entirely miss the band of data that they are based on. Such results are not “results” based on the data at all—they are entirely a product of the assumptions and analysis that created them and constrained the model from finding a better fit. Of the thermal-model fits based on multiple thermal bands, only 48% to 58% hit data in every band used, based on a very conservative metric ([Fig. 4](#)). A Monte Carlo simulation that takes into account overcounting of hits in my analysis due to insufficient information on the epochs used suggests that the true fraction of multi-band fits which miss one or more bands is likely well above 50%.

However, the problem is not just confined to the cases where the models miss the data. The constraints in the process that prevented the models from intercepting data points in some cases may bias the results from achieving a good fit even in cases where the model does pass through the clouds of data points ([Fig. 3](#)).

I conclude that analysis of the *WISE*/NEOWISE observational data is far from a solved problem. There is much still to do in documenting the results published to date, as well as in applying wider and more robust data analysis. The *WISE*/NEOWISE data is a potential treasure trove of data that could provide insight into asteroids and other attendant issues in our understanding of the solar system. New insights may lurk in the data waiting to be found by novel analyses.

Acknowledgments

This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint mission of the University of California, Los Angeles and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This publication also makes use of data products from NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the Planetary Science Division of the National Aeronautics and Space Administration. Wayt Gibbs, Dhileep Sivam, and Daniel Mc-

Coy assisted in the preparation of the manuscript, and Andrei Modoran assisted with data downloading.

I also wish to thank helpful discussions, correspondence, and data from Victor Alí Lagoa, Francesca DeMeo, José Galache, Bruce Hapke, Alan Harris, Alan Harris (DLR), Željko Ivezić, Seth Koren, Jean-Luc Margot, Tony Pan, Thomas Statler, Edward Wright, and reviews by Alan Harris and an anonymous reviewer.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.icarus.2018.05.004](https://doi.org/10.1016/j.icarus.2018.05.004).

Appendix

This appendix contains methodological and analytical details of points discussed in the text.

Discarded data

The use of epochs. Prior studies that used either the NEATM or the STM treated the results of thermal modeling, including D and η , as intrinsic and fixed properties of the asteroid ([Delbó et al., 2007](#); [Delbó and Harris, 2002](#); [Delbó and Tanga, 2009](#); [Harris, 2005, 1998](#); [Harris and Lagerros, 2002](#); [Kim et al., 2003](#); [Lagerros, 1998](#); [Lebofsky et al., 1986](#); [Mueller et al., 2005](#); [Müller, 2007](#); [Rivkin et al., 2005](#); [Rozitis and Green, 2011](#); [Ryan and Woodward, 2010](#); [Tedesco et al., 2002](#); [Usui et al., 2014, 2011a, 2011b](#); [Wolters and Green, 2009](#)). More recent work has established a physical interpretation for η as a parameter connected to the metal content of the asteroid ([Harris and Drube, 2014](#)). In contrast, the NEOWISE approach generally yields values of D , p_v , and η that vary from epoch to epoch and thus clearly cannot correspond to any fixed physical properties of the asteroid—nor to any physical interpretation because the epochs are determined without reference to the rotational phase of the asteroid, and are instead driven by *WISE* observing cadence.

The only attempt in the NEOWISE papers to justify this departure from the conventional fitting approaches appears in [Mainzer et al., \(2011d\)](#):

However, since 87% of the objects with multi-epoch observations have diameters that agree to within 10%, we conclude that multiple observational epochs do not contribute significantly to the differences observed between IRAS and NEOWISE diameters.

This note fails, however, to answer the obvious question of why a single least-square fit was not performed on all of the observations together—the natural approach one would use if different epochs do provide observational approximations to the asteroid’s true physical parameters. Indeed, fitting aggregated observations is the only way that one can obtain an accurate average across possibly irregular shape and surface properties, as was discussed further in [Section 8](#).

In cases such as observations made at very different phase angles, or observations made at very different distances from the Sun, it may be useful to analyze subsets of asteroid observations differently. Neither of these conditions apply broadly to NEOWISE, however. The observational geometry of *WISE* limits the range of phase angles, and the vast majority of the asteroids involved are in the main belt at distances from the Sun that varied little during the nine-month FC mission.

The separation of thermal modeling into epochs impacts the data available in several ways. The first is that the minimum number of data points in a band is three (after filtering by quality and artifact flags, and applying an SNR or σ threshold). Bands having

Table A1

Summary of NEOWISE observations by band. The sensitivity of the *WISE* sensors and the flux received from asteroids combine to make the number of observations in each band quite different. The column ratio to W3 count is the ratio of counts for a band to the count for W3 in the same mission.

Mission	Band	Count	Ratio to W3 count	Flux magnitude		σ		SNR	
				Min	Max	Min	Max	Min	Max
FC	W1	388,877	0.2	5.394	17.269	0.001	0.543	2	1169.9
	W2	516,716	0.28	4.105	15.817	0.004	0.543	2	261.3
	W3	1,823,223	1.0	-7.071	11.642	0.003	0.543	2	335.3
	W4	1,546,476	0.85	-4.526	8.041	0.001	0.543	2	821.2
3B	W1	27,098	0.38	6.868	17.137	0.006	0.543	2	196.9
	W2	42,497	0.60	6.282	15.802	0.002	0.543	2	538.3
	W3	70,665	1.0	0.147	11.368	0.002	0.543	2	629.1
PC + Re	W1	221,098	na	1.88	17.2	0.007	0.543	2	148.7
	W2	269,836	na	1.789	15.64	0.005	0.543	2	220.2

fewer than three points are discarded. Because the count is done on a per-epoch basis, entire bands of data could be discarded for an asteroid observed during two epochs, if fewer than three data points were collected in a band in each epoch. Those discarded observations would be available for analysis if data points were pooled across epochs. For example, an asteroid seen at two data points in W1 in two different epochs could have four data points in W1 if the epochs were pooled, but no W1 data at all after application of the minimum rule.

The NEOWISE analysis is also restricted because the same bands may not be present in each epoch, due to artifacts, noise, or sensitivity limits. Thus, an asteroid with W1 and W3 data in one epoch but W2 and W4 data in another epoch could be analyzed using data from all four bands if the two epochs are pooled, but analysis would otherwise be limited to only two bands in each of the individual epochs.

The 40% rule. Within each epoch, counts of data points in each band are compared, and all data are discarded from any band that has fewer than 40% of the highest count, even if the number of data points is greater than the minimum number of three. The numbers of data points vary among bands because some observations are too bright or dim relative to the background for the signal to be usable. Noise or background objects may affect each band differently. Alternatively, an artifact may contaminate an observation in one band, but not in others. Note that each band has a separate focal plane array, and a cosmic ray would thus likely affect only one band per observation.

As shown in Table A1, the number of data points for the FC mission was greatest in band W3 due to the favorable combination of asteroid thermal emission, *WISE* sensitivity, and background noise in this band. The count of W1 data points is only 21% of the count for the W3 band; for W2, the figure is 28%. The 40% rule would thus discard the W1 and W2 bands entirely from analysis of the FC mission data if applied at the mission level.

NEOWISE instead applied the 40% rule to individual epochs, asserting that this tactic ameliorates the effects of low-level noise and cosmic rays (Grav et al., 2011b; 2011a; Mainzer et al., 2011d). However, cosmic rays are already handled in the *WISE* pipeline by artifact flags, and low-level noise would be better handled by placing limits on magnitude, observational error, or SNR. Moreover, the overall statistics shown in Table A1 do not seem to support an expectation that normal observations will generate band counts that fall within 40% of each other. On the contrary, observations that pass the 40% rule *must be atypical* of the archive as a whole. Additional issues raised by the 40% rule, along with other details of the data analysis, are discussed in more detail in the Supplemental Information.

Distribution parameters for double detections

The parameters for the Student's t-distributions that best fit near-simultaneous *WISE* observations (see Section 5) are similar for W1 and W2 across all cases of σ_1 , σ_2 , especially the parameter ν_t , which at high values allows the Student's t-distribution to approximate a normal distribution. To test whether dim or noisy observations could explain the effects observed, I filtered the data set to retain observations for which σ_1 , $\sigma_2 \leq 0.25$. Deviations from a normal distribution (as reflected by declining ν_t) grew in magnitude, particularly in bands W3 and W4, and were greater still for observations limited to σ_1 , $\sigma_2 \leq 0.1$ (Table A2). The result that the best (lowest-error) observation points depart most from the normal distribution demonstrates that non-normality of the error is a robust effect.

The Student's t-distribution was originally derived as the sampling distribution for small sample sizes of a normal distribution $\mathcal{N}(\mu_{\text{true}}, \sigma_{\text{true}})$ with unknown true mean and standard error μ_{true} , σ_{true} . If one draws n samples f_i from this distribution, then the statistic t defined by

$$\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{f} - f_i)^2,$$

$$t = \frac{\sqrt{n}}{s} (\bar{f} - \mu_{\text{true}})$$
(A1)

will follow a Student's t-distribution $\mathcal{S}(0, 1, n - 1)$. The key consequence of Eq. (A1) is that means and standard deviations based on small sample sizes will not be normally distributed even if the fundamental distribution being drawn from is normal.

This result forms the basis for many hypothesis tests in classical statistics. An example is that, given two sets of samples of size n , one can use the statistic t and $\mathcal{S}(0, 1, n - 1)$ to assess the probability that both sets are from the same distribution $\mathcal{N}(\mu_{\text{true}}, \sigma_{\text{true}})$.

In the context of the *WISE*-estimated flux and flux uncertainty, the base distribution is probably not exactly a normal distribution, but instead depends on photon counting statistics in the detector arrays. The process of making the estimate is not formally identical to a small sample size. However, the fact that the distribution is well approximated by the Student's-t distribution shows that the analogy has some conceptual value.

Linear correction to the W3 band

The NEOWISE papers reference the *WISE* Explanatory Supplement (WES) (Cutri et al., 2015) as the origin of the linear correction to the W3 band, but no such correction appears in the WES,

Table A2

Student's t-distribution parameters for near-simultaneous observations. The distribution of the statistic \hat{f} of Eq. (3) is well described by a generalized Student's t-distribution with these parameters (see Fig. 2 for plots). Medians and confidence intervals ($\pm 34.1\%$) were calculated by performing 2000 bootstrap resampling trials.

σ	Limit	SNR Limit	Case	Band (# Obs)	μ	σ	ν	Band (# Obs)	μ	σ	ν	
0.532	1.88	Median	W1	17,528	-0.012	0.856	2.865	W2	24,801	0.054	0.911	5.227
				CI Low	-0.020	0.848	2.789		0.047	0.904	5.017	
				CI High	-0.004	0.863	2.949		0.061	0.918	5.450	
		Median	W3	133,216	-0.063	1.385	9.696	W4	102,122	0.072	1.170	25.768
				CI Low	-0.067	1.380	9.406		0.069	1.165	23.750	
				CI High	-0.059	1.389	10.011		0.076	1.174	28.130	
0.25	4	Median	W1	7834	-0.023	0.944	2.650	W2	11,202	0.081	1.013	5.324
				CI Low	-0.036	0.932	2.547		0.070	1.002	4.983	
				CI High	-0.008	0.956	2.761		0.092	1.024	5.684	
		Median	W3	125,318	-0.062	1.331	8.473	W4	59,049	0.095	1.185	14.921
				CI Low	-0.066	1.326	8.243		0.090	1.180	14.066	
				CI High	-0.058	1.335	8.718		0.100	1.191	16.019	
0.1	10	Median	W1	2808	-0.068	1.042	2.762	W2	3999	0.144	1.073	5.604
				CI Low	-0.094	1.020	2.598		0.124	1.052	5.045	
				CI High	-0.042	1.063	2.973		0.164	1.095	6.438	
		Median	W3	61,588	-0.078	1.318	6.625	W4	20,733	0.145	1.189	10.161
				CI Low	-0.084	1.311	6.412		0.135	1.180	9.500	
				CI High	-0.072	1.324	6.852		0.154	1.198	10.928	

Table A3

Normal parameters for near-simultaneous observations. The distribution of the statistic \hat{f} of Eq. (3) is not well described by a normal distribution, but we can use maximum likelihood to find the parameters of the best-fitting normal distributions (see Fig. 2 for plots). Medians and confidence intervals ($\pm 34.1\%$) were calculated by performing 2000 bootstrap resampling trials.

σ	Limit	SNR Limit	Case	Band (# Obs)	μ	σ	Band (# Obs)	μ	σ	
0.532	1.88	Median	W1	17,528	-0.005	1.972	W2	24,801	0.059	1.295
				CI Low	-0.020	1.889		0.050	1.262	
				CI High	0.010	2.058		0.067	1.330	
		Median	W3	133,216	-0.070	1.593	W4	102,122	0.073	1.219
				CI Low	-0.075	1.586		0.069	1.216	
				CI High	-0.066	1.601		0.077	1.223	
0.25	4	Median	W1	7834	-0.016	2.484	W2	11,202	0.084	1.468
				CI Low	-0.043	2.338		0.071	1.411	
				CI High	0.011	2.637		0.097	1.533	
		Median	W3	125,318	-0.070	1.565	W4	59,049	0.095	1.277
				CI Low	-0.075	1.557		0.089	1.272	
				CI High	-0.066	1.574		0.100	1.281	
0.1	10	Median	W1	2808	-0.072	2.732	W2	3999	0.133	1.504
				CI Low	-0.123	2.432		0.109	1.400	
				CI High	-0.018	3.033		0.157	1.620	
		Median	W3	61,588	-0.088	1.648	W4	20,733	0.141	1.334
				CI Low	-0.095	1.634		0.131	1.325	
				CI High	-0.082	1.664		0.150	1.344	

nor was one known to E. Wright (email communication, 2016)—see further discussion in SI Section 4. Other authors have recently described a linear correction to the WISE bands (Lian et al., 2014), but it is very slight.

My calculations did not include a linear correction because none has yet been specified. But to test the sensitivity of my code to this possibility, the analysis was rerun, using the empirical medians from the residuals in Fig. 5 to offset the calculated fluxes in each band. While the number of band misses declined, qualitative results were the same: many NEOWISE fits still missed many bands of data entirely.

References

- Alí-Lagoa, V., de León, J., Licandro, J., Delbó, M., Campins, H., Pinilla-Alonso, N., Kelley, M.S., 2013. Physical properties of B-type asteroids from WISE data. *Astron. Astrophys.* 554, A71. <https://doi.org/10.1051/0004-6361/201220680>.
- Alí-Lagoa, V., Lionni, L., Delbó, M., Gundlach, B., Blum, J., Licandro, J., 2014. Thermo-physical properties of near-Earth asteroid (341843) 2008 EV5 from WISE data. *Astron. Astrophys.* 561, A45. <https://doi.org/10.1051/0004-6361/201322215>.
- Bauer, J.M., Grav, T., Blauvelt, E., Mainzer, A., Masiero, J., Stevenson, R., Kramer, E., Fernández, Y.R., Lisse, C.M., Cutri, R.M., Weissman, P.R., Waszczak, A., Daley, J., Masci, F.J., Walker, R.G., Nugent, C.R., Meech, K.J., Lucas, A., Pearman, G., Wilkins, A., Watkins, J., Kulkarni, S., Wright, E.L., 2013. Centaurs and scattered disk objects in the thermal infrared: analysis of WISE/NEOWISE observations. *Astrophys. J.* 773, 22. <https://doi.org/10.1088/0004-637X/773/1/22>.
- Benner, L.A.M., Ostro, S.J., Magri, C., Nolan, M.C., Howell, E.S., Giorgini, J.D., Jurgens, R.F., Margot, J.L., Taylor, P.A., Busch, M.W., Shepard, M.K., 2008. Near-Earth asteroid surface roughness depends on compositional class. *Icarus* 198, 294–304. <https://doi.org/10.1016/j.icarus.2008.06.010>.
- Bowell, E., Hapke, B., Lumme, K., Harris, A.W., Domingue, D., Peltoniemi, J., 1988. Application of Photometric Models to Asteroids, in: *Asteroids II*. University of Arizona Press, Tucson, AZ, pp. 524–556.
- Brozović, M., Benner, L.A.M., Magri, C., Scheeres, D.J., Busch, M.W., Giorgini, J.D., Nolan, M.C., Jao, J.S., Lee, C.G., Snedeker, L.G., Silva, M.A., Lawrence, K.J., Slade, M.A., Hicks, M.D., Howell, E.S., Taylor, P.A., Sanchez, J.A., Reddy, V., Dykhuus, M., Corre, L.L., 2017. Goldstone radar evidence for short-axis mode non-principal-axis rotation of near-Earth asteroid (214869) 2007 PA8. *Icarus* 286, 314–329. <https://doi.org/10.1016/j.icarus.2016.10.016>.
- Carroll, R.J., Ruppert, D., 1988. *Transformation and Weighting in Regression*, 30th ed. CRC Press, New York.
- Chihara, L.M., Hesterberg, T.C., 2011. *Mathematical Statistics with Resampling and R*, 1st ed. Wiley.
- Cutri, R.M., Wright, E.L., Conrow, T., Bauer, J.M., Benford, D., Brandenburg, H., Daley, J., Eisenhardt, P.R.M., Evans, T., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair, C., Harbut, M., Hoffman, D., Jarrett, T., Kirkpatrick, J.D., Liu, W., Mainzer, A., Marsh, K., Masci, F., McCallon, H., Padgett, D., Ressler, M.E., Royer, D., Skrutskie, M.F., Stanford, S.A., Wyatt, P.L., Tholen, D., Tsai, C.W., Wachter, S., Wheelock, S.L., Yan, L., Alles, R., Beck, R., Grav, T., Masiero, J., McCollum, B., McGehee, P., Wittman, M., 2011. Explanatory Supplement to the WISE Preliminary Data Release Products (WWW Document). Explan. Suppl. to WISE Prelim. Data Release Prod.. URL <http://wise2.ipac.caltech.edu/docs/release/prelim/expsup/>.
- Cutri, R.M., Wright, E.L., Conrow, T., Bauer, J.M., Benford, D.J., Brandenburg, H., Daley, J., Eisenhardt, P.R.M., Evans, T., Fajardo-Acosta, S.B., Fowler, J., Gelino, C., Grillmair, C., Harbut, M., Hoffman, D., Jarrett, T.H., Kirkpatrick, J.D., Leisawitz, D.,

- Liu, W., Mainzer, A., Marsh, K., Masci, F., McCallon, H., Padgett, D., Ressler, M.E., Royer, D., Skrutskie, M.F., Stanford, S.A., Wyatt, P.L., Tholen, D., Tsai, C.-W., Wachter, S., Wheelock, S.L., Yan, L., Alles, R., Beck, R., Grav, T., Masiero, J., McCollum, B., McGehee, P., Papin, M., Wittman, M., Explanatory Supplement to the WISE All-Sky Data Release Products Wise-Field Infrared Survey Explorer (WWW Document) URL.
- Delb  , M., Dell'Oro, A., Harris, A.W., Mottola, S., Mueller, M., 2007. Thermal inertia of near-Earth asteroids and implications for the magnitude of the Yarkovsky effect. *Icarus* 190, 236–249. <https://doi.org/10.1016/j.icarus.2007.03.007>.
- Delb  , M., Harris, A.W., 2002. Physical properties of near-Earth asteroids from thermal infrared observations and thermal modeling. *Meteorit. Planet. Sci.* 37, 1929–1936. <https://doi.org/10.1111/j.1945-5100.2002.tb01174.x>.
- Delb  , M., Tanga, P., 2009. Thermal inertia of main belt asteroids smaller than 100 km from IRAS data. *Planet. Space Sci.* 57, 259–265. <https://doi.org/10.1016/j.pss.2008.06.015>.
-   urech, J., Carry, B., Delb  , M., Kaasalainen, M., Viikinkoski, M., 2015. Asteroid Models from Multiple Data Sources, Asteroids IV. University of Arizona Press, Tucson, AZ.
- Faherty, J.K., Alatalo, K., Anderson, L.D., Assef, R.J., Gagliuffi, D.C.B., Barry, M., Benford, D.J., Bilicki, M., Birmingham, B., Christian, D.J., Cushing, M.C., Eisenhardt, P.R., Elvis, M., Fajardo-Acosta, S.B., Finkbeiner, D.P., Fischer, W.J., Forrest, W.J., Fowler, J., Gardner, J.P., Gelino, C.R., Gorjian, V., Grillmair, C.J., Gromadzki, M., Hall, K.P., Ivezic,   , Izumi, N., Kirkpatrick, J.D., Kovacs, A., Lang, D., Leisawitz, D., Liu, F., Mainzer, A., Malek, K., Marton, G., Masci, F.J., McLean, I.S., Meissner, A., Nikutta, R., Padgett, D.L., Patel, R., Rebull, L.M., Rich, J.A., Ringwald, F.A., Rose, M., Schneider, A.C., Stassun, K.G., Stern, D., Tsai, C.-W., Wang, F., Westrom, M.E., Wright, E.L., Wu, J., Yang, J., 2015. Results from the Wide-field Infrared Survey Explorer (WISE) Future Uses Session at the WISE at 5 Meeting. arXiv:1505.01923.
- Feigelson, E.D., Babu, G.J., 2013. Statistical Methods for Astronomy. In: Planets, Stars and Stellar Systems. Springer, Netherlands, pp. 445–480.
- Grav, T., Mainzer, A., Bauer, J., Masiero, J., Spahr, T., McMillan, R.S., Walker, R., Cutri, R., Wright, E., Eisenhardt, P.R.M., Blauvelt, E., DeBaun, E., Elsbury, D., Gautier, T., Gomillion, S., Hand, E., Wilkins, A., 2011a. WISE/NEOWISE observations of the jovian trojans: preliminary results. *Astrophys. J.* 742, 40. <https://doi.org/10.1088/0004-637X/742/1/40>.
- Grav, T., Mainzer, A., Bauer, J.M., Masiero, J., Spahr, T., McMillan, R.S., Walker, R., Cutri, R., Wright, E.L., Eisenhardt, P.R.M., Blauvelt, E., DeBaun, E., Elsbury, D., Gautier, T., Gomillion, S., Hand, E., Wilkins, A., 2011b. WISE/NEOWISE observations of the Hilda population: preliminary results. *Astrophys. J.* 744, 197. <https://doi.org/10.1088/0004-637X/744/2/197>.
- Grav, T., Mainzer, A., Bauer, J.M., Masiero, J.R., Nugent, C.R., 2012. WISE/NEOWISE observations of the jovian trojan population: taxonomy. *Astrophys. J.* 759, 49. <https://doi.org/10.1088/0004-637X/759/1/49>.
- Grav, T., Spahr, T., Mainzer, A., Bauer, J.M., 2015. In: Simulating Current and Future Optical Ground Based NEO Surveys, 22. IAU General Assembly, p. 57382.
- Greenberg, A.H., Margot, J.-L., Verma, A.K., Taylor, P.A., Naidu, S.P., Brozovic, M., Benner, L.A.M., 2016. Asteroid 1566 Icarus's size, shape, orbit, and Yarkovsky drift from radar observations. *Astron. J.* 153, 108. <https://doi.org/10.3847/1538-3881/153/3/108>.
- Hanu  , J., Delb  , M.,   urech, J., Al  -Lagoa, V., 2015. Thermophysical modeling of asteroids from WISE thermal infrared data – Significance of the shape model and the pole orientation uncertainties. *Icarus* 256, 101–116. <https://doi.org/10.1016/j.icarus.2015.04.014>.
- Hanu  , J., Marchis, F.,   urech, J., 2013. Sizes of main-belt asteroids by combining shape models and Keck adaptive optics observations. *Icarus* 226, 1045–1057. <https://doi.org/10.1016/j.icarus.2013.07.023>.
- Harris, A.W., 2005. The surface properties of small asteroids from thermal-infrared observations. *Proc. Int. Astron. Union* 1, 449–463. <https://doi.org/10.1017/S1743921305006915>.
- Harris, A.W., 1998. A thermal model for near-earth asteroids. *Icarus* 131, 291–301. <https://doi.org/10.1006/icar.1997.5865>.
- Harris, A.W., Drube, L., 2014. How to find metal-rich asteroids. *Astrophys. J.* 785 (L4). <https://doi.org/10.1088/2041-8205/785/1/L4>.
- Harris, A.W., Harris, A.W., 1997. On the revision of radiometric albedos and diameters of asteroids. *Icarus* 126, 450–454. <https://doi.org/10.1006/icar.1996.5664>.
- Harris, A.W., Lagerros, J.S.V., 2002. Asteroids in the Thermal Infrared, in: Asteroids III. University of Arizona Press, Tucson, pp. 205–218.
- Ivezic,   , Connolly, A., VanderPlas, J., Gray, A., 2014. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide For the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy). Princeton University Press, Princeton, N.J.
- Ivezic,   , Smith, J.A., Miknaitis, G., Lin, H., Tucker, D., Lupton, R.H., Gunn, J.E., Knapp, G.R., Strauss, M.A., Sesar, B., Doi, M., Tanaka, M., Fukugita, M., Holtzman, J., Kent, S., Yanny, B., Schlegel, D., Finkbeiner, D., Padmanabhan, N., Rockosi, C.M., Juric, M., Bond, N., Lee, B., Stoughton, C., Jester, S., Harris, H., Hardling, P., Morrison, H., Brinkmann, J., Schneider, D.P., York, D., 2007. Sloan digital sky survey standard star catalog for Stripe 82: the dawn of industrial 1% optical photometry. *Astron. J.* 134, 973–998. <https://doi.org/http://doi.org/10.1086/519976>.
- Kanji, G.K., 2006. 100 Statistical Tests, 3rd ed SAGE Publications Ltd.
- Kim, S., Lee, H.M., Nakagawa, T., Hasegawa, S., 2003. Thermal models and far infrared emission of asteroids. *J. Korean Astron. Soc.* 36, 21–31. <https://doi.org/10.5303/JKAS.2003.36.1.021>.
- Koren, S.C., Wright, E.L., Mainzer, A., 2015. Characterizing asteroids multiply-observed at infrared wavelengths. *Icarus* 258, 82–91. <https://doi.org/10.1016/j.icarus.2015.06.014>.
- Lagerros, J.S.V., 1998. Thermal physics of asteroids IV . Thermal infrared beaming. *Astron. Astrophys.* 332, 1123–1132.
- Lawrence, K.J., Benner, L.A.M., Brozovic, M., Ostro, S.J., Jao, J.S., Giorgini, J.D., Slade, M.A., Jurgens, R.F., 2015. Goldstone radar imaging of near-Earth Asteroid 2003 MS2. *Icarus* 260, 1–6. <https://doi.org/10.1016/j.icarus.2015.06.001>.
- Lebofsky, L.A., Sykes, M.V., Tedesco, E.F., Veverd, G.J., Matson, D.L., Brown, R.H., Gradie, J.C., Feierberg, M.A., Rudy, R.J., 1986. A refined "standard" thermal model for asteroids based on observations of 1 Ceres and 2 Pallas. *Icarus* 68, 239–251. [https://doi.org/10.1016/0019-1035\(86\)90021-7](https://doi.org/10.1016/0019-1035(86)90021-7).
- Lehtinen, K., Bach, U., Muinonen, K., Poutanen, M., Petrov, L., 2016. Asteroid sizing by radiogalaxy occultation at 5 GHz. *Astrophys. J.* 822, L21. <https://doi.org/10.3847/2041-8205/822/2/L21>.
- Lian, J., Zhu, Q., Kong, X., He, J., 2014. Characterizing AGB stars in wide-field infrared survey explorer (WISE) bands. *Astron. Astrophys.* 564, A84. <https://doi.org/10.1051/0004-6361/201322818>.
- Mainzer, A., Bauer, J.M., Cutri, R.M., Grav, T., Kramer, E.A., Masiero, J., Nugent, C.R., Sonnett, S.M., Stevenson, R.A., Wright, E.L., 2016. NEOWISE Diameters and Albedos V1.0. EAR-A-COMPIL-5-NEOWISEDAM-V1.0. NASA Planet. Data Syst. 247.
- Mainzer, A., Bauer, J.M., Cutri, R.M., Grav, T., Masiero, J., Beck, R., Clarkson, P., Conrow, T., Dailey, J., Eisenhardt, P.R.M., Fabinsky, B., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair, C., Heinrichsen, I., Kendall, M., Kirkpatrick, J.D., Liu, F., Masci, F., McCallon, H., Nugent, C.R., Papin, M., Rice, E., Royer, D., Ryan, T., Sevilla, P., Sonnett, S., Stevenson, R., Thompson, D.B., Wheelock, S., Wiemer, D., Wittman, M., Wright, E.L., Yan, L., 2014a. Initial performance of the NEOWISE reactivation mission. *Astrophys. J.* 792, 30. <https://doi.org/10.1088/0004-637X/792/1/30>.
- Mainzer, A., Bauer, J.M., Grav, T., Masiero, J., Cutri, R.M., Dailey, J., Eisenhardt, P.R.M., McMillan, R.S., Wright, E.L., Walker, R., Jedicke, R., Spahr, T., Tholen, D., Alles, R., Beck, R., Brandenburg, H., Conrow, T., Evans, T., Fowler, J., Jarrett, T.H., Marsh, K., Masci, F., McCallon, H., Wheelock, S., Wittman, M., Wyatt, P., DeBaun, E., Elliott, G., Elsbury, D., Gautier, T., Gomillion, S., Leisawitz, D., Maleszewski, C., Micheli, M., Wilkins, A., 2011a. Preliminary results from NEOWISE: an enhancement to the wide-field infrared survey explorer for solar system science. *Astrophys. J.* 731, 53. <https://doi.org/10.1088/0004-637X/731/1/53>.
- Mainzer, A., Bauer, J.M., Grav, T., Masiero, J., Cutri, R.M., Wright, E., Nugent, C.R., Stevenson, R., Clyne, E., Cukrov, G., Masci, F., 2014b. The population of tiny near-earth objects observed by NEOWISE. *Astrophys. J.* 784, 110. <https://doi.org/10.1088/0004-637X/784/2/110>.
- Mainzer, A., Grav, T., Bauer, J.M., Conrow, T., Cutri, R.M., Dailey, J., Fowler, J., Giorgini, J., Jarrett, T.H., Masiero, J., Spahr, T., Statler, T., Wright, E.L., 2015a. Survey simulations of a new near-earth asteroid detection system. *Astron. J.* 149, 172. <https://doi.org/10.1088/0004-6256/149/5/172>.
- Mainzer, A., Grav, T., Bauer, J.M., McMillan, R.S., Cutri, R.M., Walker, R., Wright, E.L., Eisenhardt, P.R.M., Tholen, D.J., Spahr, T., Jedicke, R., Denneau, L., DeBaun, E., Elsbury, D., Gautier, T., Gomillion, S., Hand, E., Mo, W., Watkins, J., Wilkins, A., Bryngelson, G.L., Del Pino Molina, A., Desai, S., Camus, M.G., Hidalgo, S.L., Konstantopoulos, I., Larsen, J.A., Maleszewski, C., Malkan, M.A., Mauduit, J.-C., Mullan, B.L., Olszewski, E.W., Pfarr, J., Saro, A., Scotti, J.V., Wasserleben, L.H., 2011b. Neowise observations of near-earth objects: preliminary results. *Astrophys. J.* 743, 156. <https://doi.org/10.1088/0004-637X/743/2/156>.
- Mainzer, A., Grav, T., Bauer, J.M., McMillan, R.S., Cutri, R.M., Walker, R., Wright, E.L., Eisenhardt, P.R.M., Tholen, D.J., Spahr, T., Jedicke, R., Denneau, L., DeBaun, E., Elsbury, D., Gautier, T., Gomillion, S., Hand, E., Mo, W., Watkins, J., Wilkins, A., Bryngelson, G.L., Del Pino Molina, A., Desai, S., Camus, M.G., Hidalgo, S.L., Konstantopoulos, I., Larsen, J.A., Maleszewski, C., Malkan, M.A., Mauduit, J.-C., Mullan, B.L., Olszewski, E.W., Pfarr, J., Saro, A., Scotti, J.V., Wasserleben, L.H., 2012a. Physical parameters of asteroids estimated from the WISE 3 band data and NEOWISE post-cryogenic survey. *Astrophys. J. Lett.* 760, L12. <https://doi.org/10.1088/2041-8205/760/1/L12>.
- Mainzer, A., Grav, T., Masiero, J., Bauer, J.M., McMillan, R.S., Giorgini, J.D., Spahr, T., Cutri, R.M., Tholen, D.J., Jedicke, R., Walker, R., Wright, E.L., Nugent, C.R., 2012b. Characterizing subpopulations within the near-earth objects with neowise : preliminary results. *Astrophys. J.* 752, 110. <https://doi.org/10.1088/0004-637X/752/2/110>.
- Mainzer, A., Grav, T., Masiero, J., Bauer, J.M., Wright, E.L., Cutri, R.M., McMillan, R.S., Cohen, M., Ressler, M., Eisenhardt, P.R.M., 2011c. Thermal model calibration for minor planets observed with wide-field infrared survey explorer/Neowise. *Astrophys. J.* 736, 100. <https://doi.org/10.1088/0004-637X/736/2/100>.
- Mainzer, A., Grav, T., Masiero, J., Bauer, J.M., Wright, E.L., Cutri, R.M., Walker, R., McMillan, R.S., 2011d. Thermal model calibration for minor planets observed with WISE/NEOWISE: comparison with infrared astronomical satellite. *Astrophys. J. Lett.* 737, L9. <https://doi.org/10.1088/2041-8205/737/1/L9>.
- Mainzer, A., Grav, T., Masiero, J., Hand, E., Bauer, J.M., Tholen, D., McMillan, R.S., Spahr, T., Cutri, R.M., Wright, E.L., Watkins, J., Mo, W., Maleszewski, C., 2011e. NEOWISE studies of spectrophotometrically classified asteroids: preliminary results. *Astrophys. J.* 741, 90. <https://doi.org/10.1088/0004-637X/741/2/90>.
- Mainzer, A., Masiero, J., Grav, T., Bauer, J.M., Tholen, D.J., McMillan, R.S., Wright, E.L., Spahr, T., Cutri, R.M., Walker, R., Mo, W., Watkins, J., Hand, E., Maleszewski, C., 2012c. NEOWISE studies of asteroids with Sloan photometry: preliminary results. *Astrophys. J.* 745, 7. <https://doi.org/10.1088/0004-637X/745/1/7>.
- Mainzer, A., Usui, F., Trilling, D.E., 2015b. Space-Based Thermal Infrared Studies of Asteroids, in: Asteroids IV. University of Arizona Press, Tucson, AZ, p. 89.
- Masiero, J., Carruba, V., Mainzer, A., Bauer, J.M., Nugent, C., 2015a. The Euphrosyne family's contribution to the low albedo near-Earth asteroids. *Astrophys. J.* 809, 179. <https://doi.org/10.1088/0004-637X/809/2/179>.
- Masiero, J., DeMeo, F.E., Kasuga, T., Parker, A.H., 2015b. Asteroid Family Physical Properties, in: Asteroids IV. University of Arizona Press, Tucson, AZ.

- Masiero, J., Grav, T., Mainzer, A., Nugent, C.R., Bauer, J.M., Stevenson, R., Sonnett, 2014. Main belt asteroids with WISE/NEOWISE: near-Infrared Albedos. *Astrophys. J.* 791, 121. <https://doi.org/10.1088/0004-637X/791/2/121>.
- Masiero, J., Jedicke, R., Durech, J., Gwyn, S., Denneau, L., Larsen, J., 2009. The Thousand Asteroid Light Curve Survey. *Icarus* 204, 145–171. <https://doi.org/10.1016/j.icarus.2009.06.012>.
- Masiero, J., Mainzer, A., Bauer, J.M., Grav, T., Nugent, C.R., Stevenson, R., 2013. Asteroid family identification using the Hierarchical Clustering Method and WISE/NEOWISE physical properties. *Astrophys. J.* 770, 7. <https://doi.org/10.1088/0004-637X/770/1/7>.
- Masiero, J., Mainzer, A., Grav, T., Bauer, J.M., Cutri, R.M., Dailey, J., Eisenhardt, P.R.M., McMillan, R.S., Spahr, T.B., Skrutskie, M.F., Tholen, D., Walker, R.C., Wright, E.L., DeBaun, E., Elsbury, D., Gautier IV, T., Gomillion, S., Wilkins, A., 2011. Main belt asteroids with WISE/NEOWISE. I. Preliminary albedos and diameters. *Astrophys. J.* 741, 68. <https://doi.org/10.1088/0004-637X/741/2/68>.
- Masiero, J., Mainzer, A., Grav, T., Bauer, J.M., Cutri, R.M., Nugent, C., Cabrera, M.S., 2012. Preliminary analysis of WISE/NEOWISE 3-band cryogenic and post-cryogenic observations of main belt asteroids. *Astrophys. J.* 759, L8. <https://doi.org/10.1088/2041-8205/759/1/L8>.
- Mueller, M., Harris, A.W., Delbo, M., 2005. Thermal inertia of small asteroids from detailed thermophysical modeling (WWW Document). Asteroids Comets Meteors. URL <http://www.on.br/acm2005/presentation/P15.14.pdf>.
- Müller, M., 2007. Surface Properties of Asteroids from Mid-Infrared Observations and Thermophysical Modeling. Freie Universität Berlin.
- Myhrvold, N., 2018. Asteroid thermal modeling in the presence of reflected sunlight. *Icarus* 303, 91–113. <https://doi.org/10.1016/j.icarus.2017.12.024>.
- Myhrvold, N., 2016. Comparing NEO search telescopes. *Publ. Astron. Soc. Pacific* 128. <https://doi.org/10.1088/1538-3873/128/962/045004>.
- Naidu, S.P., Margot, J.L., Taylor, P.A., Nolan, M.C., Busch, M.W., Benner, L.A.M., Brozovic, M., Giorgini, J.D., Jao, J.S., Magri, C., 2015. Radar imaging and characterization of the binary near-earth asteroid (185851) 2000 Dp107. *Astron. J.* 150, 54. <https://doi.org/10.1088/0004-6256/150/2/54>.
- NASA/IPAC Infrared Science Archive (WWW Document) URL.
- Nugent, C.R., Mainzer, A., Masiero, J., Bauer, J., Cutri, R.M., Grav, T., Kramer, E., Sonnett, S., Stevenson, R., Wright, E.L., 2015. NEOWISE reactivation mission year one: preliminary asteroid diameters and albedos. *Astrophys. J.* 814, 117. <https://doi.org/10.1088/0004-637X/814/2/117>.
- Nugent, C.R., Mainzer, A., Masiero, J., Grav, T., Bauer, J.M., 2012. The Yarkovsky Drift's influence on NEAs: trends and predictions with NEOWISE measurements. *Astron. J.* 144, 75. <https://doi.org/10.1088/0004-6256/144/3/75>.
- Nugent, C.R., Mainzer, A., Masiero, J., Wright, E.L., Bauer, J., Grav, T., Kramer, E.A., Sonnett, S., 2016. Observed asteroid surface area in the thermal infrared. *Astron. J.* 153, 90. <https://doi.org/10.3847/1538-3881/153/2/90>.
- Pravec, P., Harris, A.W., Kušnírák, P., Galád, A., Hornoch, K., 2012. Absolute magnitudes of asteroids and a revision of asteroid albedo estimates from WISE thermal observations. *Icarus* 221, 365–387. <https://doi.org/10.1016/j.icarus.2012.07.026>.
- Rivkin, A.S., Binzel, R.P., Bus, S.J., 2005. Constraining near-Earth object albedos using near-infrared spectroscopy. *Icarus* 175, 175–180. <https://doi.org/10.1016/j.icarus.2004.11.005>.
- Rozitis, B., Green, S.F., 2011. Directional characteristics of thermal-infrared beaming from atmosphereless planetary surfaces - a new thermophysical model. *Mon. Not. R. Astron. Soc.* 415, 2042–2062. <https://doi.org/10.1111/j.1365-2966.2011.18718.x>.
- Ryan, E.L., Woodward, C.E., 2010. Rectified asteroid albedos and diameters from IRAS and MSX Photometry Catalogs. *Astron. J.* 140, 933. <https://doi.org/10.1088/0004-6256/140/4/933>.
- Ryan, T.P., 2008. *Modern Regression Methods*. John Wiley & Sons.
- Shepard, M.K., Richardson, J., Taylor, P.A., Rodriguez-Ford, L.A., Conrad, A., de Pater, I., Adamkovics, M., de Kleer, K., Males, J.R., Morzinski, K.M., Close, L.M., Kaasalainen, M., Viikinkoski, M., Timerson, B., Reddy, V., Magri, C., Nolan, M.C., Howell, E.S., Benner, L.A.M., Giorgini, J.D., Warner, B.D., Harris, A.W., 2017. Radar observations and shape model of asteroid 16 Psyche. *Icarus* 281, 388–403. <https://doi.org/10.1016/j.icarus.2016.08.011>.
- Shepard, M.K., Taylor, P.A., Nolan, M.C., Howell, E.S., Springmann, A., Giorgini, J.D., Warner, B.D., Harris, A.W., Stephens, R., Merline, W.J., Rivkin, A., Benner, L.A.M., Coley, D., Clark, B.E., Ockert-Bell, M., Magri, C., 2015. A radar survey of M- and X-class asteroids. III. Insights into their composition, hydration state, & structure. *Icarus* 245, 38–55. <https://doi.org/10.1016/j.icarus.2014.09.016>.
- Shevchenko, V.G., Tedesco, E.F., 2006. Asteroid albedos deduced from stellar occultations. *Icarus* 184, 211–220. <https://doi.org/10.1016/j.icarus.2006.04.006>.
- Sonnett, S., Mainzer, A., Grav, T., Masiero, J., Bauer, J.M., 2015. Binary candidates in the Jovian Trojan and Hilda populations from NEOWISE lightcurves. *Astrophys. J.* 799, 191. <https://doi.org/10.1088/0004-637X/799/2/191>.
- Spencer, J.R., Lebofsky, L.A., Sykes, M.V., 1989. Systematic biases in radiometric diameter determinations. *Icarus* 78, 337–354. [https://doi.org/10.1016/0019-1035\(89\)90182-6](https://doi.org/10.1016/0019-1035(89)90182-6).
- Tedesco, E.F., Noah, P.V., Noah, M., Price, S.D., 2002. The supplemental IRAS minor planet survey. *Astron. J.* 123, 1056–1085. <https://doi.org/10.1086/338320>.
- Using Data Tags in Journals [WWW Document] URL.
- Usui, F., Hasegawa, S., Ishiguro, M., Muller, T.G., Ootsubo, T., 2014. A Comparative Study of Infrared Asteroid surveys: IRAS, AKARI, and WISE. Publications of the Astronomical Society of Japan <https://doi.org/10.1093/pasj/psu037>.
- Usui, F., Kasuga, T., Hasegawa, S., Ishiguro, M., Kuroda, D., Müller, T.G., Ootsubo, T., Matsuhara, H., 2012. Albedo properties of main belt asteroids based on the all-sky survey of the infrared astronomical satellite Akari. *Astrophys. J.* 762, 56. <https://doi.org/10.1088/0004-637X/762/1/56>.
- Usui, F., Kuroda, D., Muller, T.G., Hasegawa, S., Ishiguro, M., Ootsubo, T., Ishihara, D., Kataza, H., Takita, S., Oyabu, S., Ueno, M., Matsuhara, H., Onaka, T., 2011a. In: Asteroid Catalog Using AKARI: AKARI / IRC Mid-Infrared Asteroid Survey, 63. Publications of the Astronomical Society of Japan, pp. 1117–1138. <https://doi.org/10.1093/pasj/63.5.1117>.
- Usui, F., Kuroda, D., Müller, T.G., Hasegawa, S., Ishiguro, M., Ootsubo, T., Ishihara, D., Kataza, H., Takita, S., Oyabu, S., Ueno, M., Matsuhara, H., Onaka, T., 2011b. In: Asteroid Catalog Using AKARI: AKARI/IRC Mid-infrared Asteroid Survey, 63. Publications of the Astronomical Society of Japan, pp. 1117–1138. <https://doi.org/10.1093/pasj/63.5.1117>.
- Veeder, G.J., Hanner, M.S., Matson, D.L., Tedesco, E.F., Lebofsky, L.A., Tokunaga, A.T., 1989. Radiometry of near-earth asteroids. *Astron. J.* 97, 1211–1219. <https://doi.org/10.1086/115064>.
- Veres, P., Jedicke, R., Fitzsimmons, A., Denneau, L., Granvik, M., Bolin, B., Chastel, S., Wainscoat, R.J., Burgett, W.S., Chambers, K.C., Flewelling, H., Kaiser, N., Magnier, E.A., Morgan, J.S., Price, P.A., Tonry, J.L., Waters, C., 2015. Absolute magnitudes and slope parameters for 250,000 asteroids observed by Pan-STARRS PS1 - Preliminary results. *Icarus* 261, 34–47. <https://doi.org/10.1016/j.icarus.2015.08.007>.
- Wall, J.V., Jenkins, C.R., 2012. *Practical Statistics for Astronomers (Cambridge Observing Handbooks for Research Astronomers)*, 2nd ed. Cambridge University Press.
- Wolters, S.D., Green, S.F., 2009. Investigation of systematic bias in radiometric diameter determination of near-Earth asteroids: The night emission simulated thermal model (NESTM). *Mon. Not. R. Astron. Soc.* 400, 204–218. <https://doi.org/10.1111/j.1365-2966.2009.14996.x>.
- Wright, E.L., Eisenhardt, P.R.M., Mainzer, A., Ressler, M.E., Cutri, R.M., Jarrett, T., Kirkpatrick, J.D., Padgett, D., McMillan, R.S., Skrutskie, M., Stanford, S.A., Cohen, M., Walker, R.G., Mather, J.C., Leisawitz, D., Gautier, T.N., McLean, I., Benford, D.J., Lonsdale, C.J., Blain, A., Mendez, B., Irace, W.R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A.L., Larsen, M., Cardon, J.G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L., Tsai, C.-W., 2010. The wide-field infrared survey explorer (WISE): mission description and initial on-orbit performance. *Astron. J.* 140, 1868. <https://doi.org/10.1088/0004-6256/140/6/1868>.
- Wright, E.L., Mainzer, A., Masiero, J., Grav, T., Bauer, J., 2016. The Albedo distribution of near earth asteroids. *Astron. J.* 152, 79.