Icarus 341 (2020) 113575

Contents lists available at ScienceDirect

Icarus

journal homepage: www.elsevier.com/locate/icarus

ATM: An open-source tool for asteroid thermal modeling and its application to NEOWISE data

Joachim Moeyens [a,*,1], Nathan Myhrvold [b], Željko Ivezić [a]

[a] Department of Astronomy and the DIRAC Institute, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195, USA
[b] Intellectual Ventures, Bellevue, WA 98005, USA

ARTICLE INFO

Keywords:
Asteroids
Near-Earth objects
NEATM
Data reduction techniques
Radiative transfer

ABSTRACT

We publicly release ATM, a Python package designed to model asteroid flux measurements to estimate an asteroid's size, surface temperature distribution, and emissivity. A number of the most popular static asteroid thermal models (NEATM, STM, FRM) are implemented with the reflected solar light contribution and Kirchhoff's law accounted for. Priors for fitted parameters can be easily specified and the solution, including the full multi-dimensional posterior probability density function, is found using Markov Chain Monte Carlo (MCMC). We describe the package's architecture and discuss model and fitting validation. Data files with ~10 million WISE flux measurements for ~150,000 unique asteroids and additional Minor Planet Center data are also included with the package, as well as Python Jupyter Notebooks with examples of how to select subsamples of objects, filter and process data, and use ATM to fit for the desired model parameters. The entirety of the analysis presented here, including all the figures, tables, and catalogs, can be easily reproduced with these publicly released Notebooks. We show that ATM can match the best-fit size estimates for well-observed asteroids published in 2016 by the NEOWISE team (Mainzer et al., 2016) with a sub-percent bias and a scatter of only 6%. Our analysis of various sources of random and systematic size uncertainties shows that for the majority of over 100,000 objects with WISE-based size estimates random uncertainties (precision) are about 10%; systematic uncertainties within the adopted model framework, such as NEATM, and with assumed emissivity for WISE W3 and W4 bands, are likely in the range of 10–20%. Hence, the accuracy of WISE-based asteroid size estimates is approximately in the range of 15–20% for most objects, except for unknown errors due to a possibly over-simplified modeling framework (e.g., spherical asteroid approximation). We compare model families to data in WISE color-color diagrams and derive a simple method to estimate size that only uses WISE W3 band flux; we show that it matches estimates based on all four WISE bands to within 10%. We also study optical data collected by the Sloan Digital Sky Survey (SDSS) and show that correlations of optical colors and WISE-based best-fit model parameters indicate robustness of the latter. Our analysis gives support to the claim by Harris & Drube (2014) that candidate metallic asteroids can be selected using the best-fit temperature parameter and infrared albedo. We investigate a correlation between SDSS colors and optical albedo derived using WISE-based size estimates and show that this correlation can be used to estimate asteroid sizes with optical data alone, with a precision of about 17% relative to WISE-based size estimates. After accounting for systematic errors, the difference in accuracy between infrared and optical color-based size estimates becomes less than a factor of two.

1. Introduction

Asteroid thermal flux modeling aims to estimate an asteroid's size (cross-section-equivalent diameter, hereafter diameter), its surface temperature distribution, and sometimes other physical properties such as emissivity, from measured infrared (IR) fluxes. The largest dataset of infrared flux measurements for asteroids was recently contributed by the WISE survey (Wright et al., 2010) and analyzed by the associated NEOWISE team (Mainzer et al., 2011 and references therein). Flux measurements in four WISE bands provide strong constraints on asteroid sizes and emissivities as a function of wavelength. A series of papers that produced size estimates for about 164,000 asteroids, as well as

* Corresponding author.
E-mail address: moeyensj@uw.edu (J. Moeyens).
[1] LSSTC Data Science Fellow.

https://doi.org/10.1016/j.icarus.2019.113575

constraints on asteroid emissivity properties, was reviewed and summarized by Mainzer et al. (2015).

It appears that the pioneering analysis by the NEOWISE team can be improved in various ways, as argued by Myhrvold (2018a,b). In particular, ignoring reflected sunlight can induce biases in estimated asteroid sizes and lead to underestimated size uncertainties. In addition, best-fit sizes can be biased due to assumptions on the priors for an asteroid's emissivity as a function of wavelength, which varies with the chemical composition of an asteroid's surface. Further biases are introduced by the use of fairly simplistic static thermal models to calculate cross-section-equivalent diameters. The need for improvements in data analysis was recently acknowledged by Wright et al. (2018), but it appears that a number of data analysis issues remain open. In particular, the behavior of systematic and random (statistical) uncertainties for the best-fit parameters remains an active research topic (see for example Masiero et al., 2018a; Mommert et al., 2018; Wright, 2007).

Given that the WISE dataset is by far the largest of its kind and will not soon be surpassed, it is prudent to focus on improved data analysis. We aim to contribute to such improvements by publicly releasing a new Python modeling tool, ATM (Asteroid Thermal Modeling[2]), designed to enable easy fitting of the most common static thermal models to WISE and other asteroid flux measurements. Data analysis software is often a crucial component in delivering scientific results, as vividly exemplified by this case, and thus discussions of scientific reproducibility and transparency can be greatly enhanced by collaborative software development and code sharing. Thanks to rapidly developing tools and technologies, such as Jupyter Notebooks, Python, and GitHub, performing these steps in open-source environment is now easier than ever. By releasing ATM we aim to increase reproducibility – a fundamental tenet of the scientific process.

In Section 2 we describe the mathematical and physical model underlying ATM, its Python implementation, and we discuss model validation using observational and modeled data for asteroid sizes from the literature. The capabilities of ATM, with emphasis on various treatments of Bayesian priors when also fitting emissivity are further illustrated in Section 2 using three well-observed asteroids. In Section 3 we apply ATM to a "gold" sample of ~7000 best-observed asteroids from the NEOWISE dataset; in addition to best-fit diameters, we also obtain best-fit values for the emissivity of each asteroid across WISE bands W1 and W2. The best-fit sizes are compared to the values obtained by the NEOWISE team (the 2016 Planetary Data System version[3] (Mainzer et al., 2016)). In Section 4 we match the "gold" sample to optical data from the SDSS Moving Object Catalog, compute visual albedo p_V, and study its correlation with optical colors. We also discuss how optical colors can be used to estimate p_V and asteroid sizes when adequate infrared data are not available. Our results are summarized in Section 5.

As part of the ATM package release, we also include all data files used in this work including NEOWISE, Minor Planet Center, and SDSS Moving Object Catalog data. We provide Jupyter Notebooks with examples of how to select subsamples of objects, process and filter their flux measurements, and how to use ATM to estimate their diameters and infrared emissivities. All of the analysis presented here, including all the figures and tables, can be easily reproduced using the Notebooks released with the ATM package on GitHub. In particular, https://github.com/moeyensj/atm_notebooks/blob/master/paper1/README.md lists the Notebooks required to reproduce each figure in this paper. In addition to the Python code and Jupyter Notebooks, we also provide Docker containers tagged to coincide with the publication of this paper which allow the user to access and use all the components necessary (operating system, software dependencies, data files, the ATM Python code, and the analysis Notebooks) to reproduce every figure in this paper in a container-based environment.

2. ATM: an open-source tool for asteroid thermal modeling

A detailed discussion of the relevant physics involved in asteroid thermal flux modeling, and a summary of models proposed in the literature, are presented in Myhrvold (2018a) and references therein. Here we only summarize the main results needed to understand how ATM works and what it computes. We also describe its Python implementation and discuss model validation using observational and model data for asteroid sizes from literature.

Static or instantaneous thermal models, like the ones discussed in this work, offer a relatively simple but effective method to estimate an asteroid's diameter. These models assume the asteroid is a non-rotating or simply rotating sphere with a well-defined surface temperature distribution which is often expressed as a function of angular distance from the subsolar point. Static thermal models differ significantly from more robust thermophysical models which seek to model conductivity, take into account spin axis, rotation, and orientation, consider the granularity and composition of the surface, and also how composition changes with depth. However, with added complexity comes an added increase in computational cost and an increased need for high quality and quantity of observations. Static thermal modeling, while more simple, offers a computationally feasible method to get estimates of diameters and other properties but with the bonus that they can be applied at scale. And thus, instantaneous thermal modeling is a natural choice for the WISE dataset. We now summarize the underlying physics that drive these models.

2.1. The asteroid flux model summary

We first introduce $F_\nu(\lambda)$: the specific flux (flux per unit frequency, ν) from an object. The SI units for F_ν are W m^{-2} Hz^{-1} ($= 10^3$ erg cm^{-2} s^{-1}). The specific flux can also be defined per unit wavelength, F_λ, using energy conservation $F_\nu|d\nu| = F_\lambda|d\lambda|$ and $\lambda\nu = c$. The choice of F_ν, as opposed to F_λ, is completely arbitrary. Similarly, the running variable can be either λ or ν, and the choice of λ is more convenient in this context.

The model flux from an asteroid, $F_\nu^{ast}(\lambda)$, corresponding to flux detected by the observer, $F_\nu^{obs}(\lambda)$, is the sum of the emitted thermal flux controlled by the asteroid's surface temperature distribution, and the portion of the incident solar flux reflected by the asteroid,

$$F_\nu^{ast}(\lambda) = F_\nu^{th}(\lambda) + F_\nu^{ref}(\lambda). \quad (1)$$

A given model spectrum $F_\nu^{ast}(\lambda)$, corresponding to observational quantity $F_\nu^{obs}(\lambda)$, is integrated over the bandpass (assumed known hereafter) to obtain observed in-band model fluxes for a given instrument. For example, Wright (2013) has derived simple quadrature formulae[4] that can be used to efficiently and accurately compute in-band fluxes for the four WISE bands from model flux $F_\nu^{ast}(\lambda)$.

Both $F_\nu^{th}(\lambda)$ and $F_\nu^{ref}(\lambda)$ depend on the relative positions of the Sun, the asteroid, and the observer, the asteroid's diameter, D, and its emissivity, $\epsilon(\lambda)$, which controls the balance between absorbed/emitted and reflected incident flux. The observing geometry is fully described by the asteroid-Sun distance, r, the asteroid-observer distance, Δ, and the angle subtended on the asteroid's surface by the lines of sight towards the Sun and the observer, the so-called phase angle, α. Hereafter we assume that asteroid orbital parameters are known and that r, Δ, and α can be easily computed using standard and readily-available tools (e.g., the JPL HORIZONS service[5]; OpenOrb, Granvik et al., 2015). To simplify nomenclature, we do not explicitly list these independent variables, unless

necessary to avoid confusion.

2.1.1. The emitted flux

The observed thermal flux is obtained by integrating the emitted thermal flux per unit area over the visible surface of the asteroid,

$$F_\nu^{th}(\lambda) = \left(\frac{D}{2\,\Delta}\right)^2 \epsilon(\lambda) \int_{-\pi/2}^{\pi/2} \int_{-\pi/2+\alpha}^{\pi/2+\alpha} \pi B_\nu(T(\theta,\phi),\lambda)\cos(\phi-\alpha)\cos^2(\theta)\,d\phi\,d\theta, \tag{2}$$

where θ, the angle of latitude, and ϕ, the angle of longitude, are the integration variables over the asteroid's surface (here we use geographic coordinates with $\theta = 0$ and $\phi = 0$ at the subsolar point, and both ranging from $-\pi$ to π; note that Myhrvold, 2018a used ISO coordinates with $0 < \theta < 2\pi$), and B_ν is the Planck function.

In this context of thermal emission, it is implied that asteroids are approximated as perfect spherical Lambertian emitters which follow Lambert's cosine rule. In reality, both scattered light and thermal emission may exhibit non-Lambertian properties. However, the effect in scattered light (which shows a strong opposition surge – a peaked reflectance for phase angles near zero) is stronger than the effect of non-Lambertian beaming in the case of thermal emission (see e.g. Rozitis and Green, 2011).

The temperature distribution across an asteroid's surface, $T(\theta,\phi)$, is model dependent. ATM implements the three most common instantaneous thermal models: the Standard Thermal Model (STM) (Lebofsky et al., 1986), the Fast Rotating Model (FRM) (Lebofsky and Spencer, 1989), and the Near-Earth Asteroid Thermal Model (NEATM) (Harris, 1998). For a comparison of these models and an analysis of their validity, please see Wright (2007), Mommert et al. (2018), and references therein. Common to all models is a temperature scale set by the so-called subsolar temperature, T_{ss}, which is the highest temperature on the asteroid's surface. Depending on the adopted model, the temperature variation across the surface can range in complexity from being constant to having a strong temperature gradient from the subsolar point (e.g., see Fig. 1 in Mommert et al., 2018). The temperature distribution for the STM and NEATM models is equivalent to

$$T(\theta,\phi) = \begin{cases} T_{ss}(\cos\theta\cos\phi)^{1/4} & \text{if} \quad \theta \le \pi/2, \phi \le \pi/2 \\ 0 & \text{if} \quad \theta > \pi/2, \phi > \pi/2 \end{cases} \tag{3}$$

In addition to differences in surface temperature distribution, the three implemented models may use different integration variables and different methods to account for phase angle. Eq. (2) follows innovations introduced by NEATM, where the observer's direction cosine is directly integrated over the surface of a model asteroid.

To account for effects such as surface roughness, the subsolar temperature is modulated using the "beaming parameter", η. For a more detailed discussion of this parameter, please see Myhrvold (2018a). The η parameter is not constant and can vary with observing geometry, particularly, phase angle (Wright, 2007).

Given a model parameter for the temperature scale, T_1, defined as the subsolar temperature when $r = 1$ au, the subsolar temperature is simply

$$T_{ss} = \left(\frac{1\text{ au}}{r}\right)^{1/2} T_1. \tag{4}$$

Unlike T_{ss}, which depends on the asteroid-Sun distance, T_1 is nearly constant for a given asteroid (not exactly constant because of its slight dependence on η, see Eq. (5) below).

The energy balance, discussed below, connects T_1 with the incident solar flux, an asteroid's physical properties and other model parameters. Myhrvold (2018a) made an important point that when fitting a model to data, all the other parameters are not directly relevant – it is only T_1 that

2.1.2. The energy balance and the meaning of best-fit parameter T_1

Using the energy balance equation that equates the absorbed incident solar flux with the flux emitted by an asteroid, it can be shown (e.g., see Myhrvold, 2018a and references therein) that

$$T_1 = \left(\frac{S(1-A)}{\sigma\eta\epsilon_B}\right)^{1/4}, \tag{5}$$

where S is the solar constant at 1 au ($S = 1360.8$ Wm^{-2}), $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4} is the Stefan-Boltzmann constant, and A and ϵ_B are appropriately wavelength-averaged values of $1 - \epsilon(\lambda)$ and $\epsilon(\lambda)$ over the incident and thermal flux distributions, respectively (for details, please see Section 3 in Myhrvold, 2018a). It is usually assumed that $\epsilon_B = 0.9$, but the exact value is not crucial to modeling once the beaming parameter η is introduced – it is only the $\eta\epsilon_B$ product that can be constrained using best-fit T_1. Another conventional approximation is

$$A \approx A_V = p_V\, q, \tag{6}$$

where A_V is the Bond albedo (limited to the range 0–1), p_V is the geometric albedo in the visible band (can be larger than 1), and q is the empirically-derived phase integral. In the H-G magnitude system introduced by Bowell et al. (1989), $q(G) = 0.29 + 0.684G$, where G is the slope parameter of the phase function. A common assumption, when G is unknown, is $G = 0.15$. Therefore, the best-fit T_1 can be used to estimate p_V, given the value of the $\eta\epsilon_B$ product.

The geometric albedo can also be constrained, when the asteroid's diameter, D, is known, using flux measurements at wavelengths sufficiently short for flux to be dominated by the reflected incident solar flux. For example, in the visual band

$$p_V = \left(\frac{1329\text{ km}}{D}\right)^2 10^{-0.4H}, \tag{7}$$

where H is the asteroid's absolute magnitude in the visual band. When such measurements are available, the two constraints for p_V can be used to estimate the beaming parameter, η, or to simply check the model's internal consistency.

2.1.3. The reflected flux

Via Kirchhoff's law, an asteroid's reflectivity, $\rho(\lambda)$, is

$$\rho(\lambda) = \frac{1-\epsilon(\lambda)}{q(G)} \tag{8}$$

The reflected flux is then equivalent to

$$F_\nu^{ref}(\lambda) = \left(\frac{D}{2\,\Delta}\right)^2 \Psi(\alpha,G)\left[\frac{1-\epsilon(\lambda)}{q(G)}\right] F_\nu^{\odot}(\lambda) \tag{9}$$

where the H-G phase function, $\Psi(\alpha,G)$, and empirically derived phase integral, $q(G)$, are purely geometric quantities that account for phase effects (Bowell et al., 1989).

The incident solar flux, $F_\nu^{\odot}(\lambda)$, at a distance r from the Sun, is given by

$$F_\nu^{\odot}(\lambda) = \left(\frac{R_\odot}{r}\right)^2 \pi\, B_\nu(T_\odot,\lambda), \tag{10}$$

where $T_\odot = 5778$ K and $R_\odot = 0.00465$ au. Therefore, when $\epsilon(\lambda)$ is known, the reflected flux is fully determined for a given observing geometry (note also that a fully known $\epsilon(\lambda)$ implies a given value of p_V). In Appendix D we examine the effect of using a more accurate solar spectrum on certain modeling scenarios presented in the latter sections of this work.

2.1.4. Fitting model fluxes to data

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi(\sigma_i^2 + \Sigma^2)}} \exp\left(\frac{-(F_i^{obs} - F_i^{ast})^2}{2\,(\sigma_i^2 + \Sigma^2)}\right), \quad (11)$$

where σ_i is the flux measurement uncertainty and Σ accounts for variability. Assuming flat Bayesian priors for fitted model parameters, maximizing this likelihood function is equivalent to maximizing the Bayesian posterior probability density function. ATM also supports the Jeffreys' priors (flat distributions in $\log(D)$ and $\log(T_1)$, for details see Chapter 5 in Ivezić et al., 2014).

The above expression assumes that the scatter of flux measurements around predicted model values follows a Gaussian distribution with standard deviation (measurement uncertainty) σ_i. In practice, most asteroids show variability with amplitudes (~1 mag) exceeding typical measurement uncertainties ($\leq$0.2 mag for SNR$\geq$ 5). This variability is mostly due to non-spherical shapes, which are not captured by the model. For this reason, the likelihood expression includes the Σ^2 term. Given that $\Sigma \gg \sigma_i$ for reasonably high SNR, we have $\sigma_i^2 + \Sigma^2 \approx \Sigma^2$, which is presumed constant for a given asteroid. While it does not influence the values of the best-fit parameters, Σ does control their uncertainty. Following Masiero et al., 2018b, we adopt $\Sigma = 0.2$ mag as typical uncertainty due to variability.

When emissivity $\epsilon(\lambda)$ is assumed known, the best-fit model parameters are effectively obtained by maximizing the log-likelihood in magnitude space, as a function of two free parameters, D and T_1 (for a discussion of emissivity fitting, see the next section),

$$ln(L) = \text{const.} - \sum_{i=1}^{N} \left[m_i^{obs} - m_i^{ast}(D, T_1)\right]^2, \quad (12)$$

where m_i^{obs} are observed magnitudes and m_i^{ast} are their model predictions ($m = -2.5\,\log_{10}(F) + \text{const.}$).

2.1.5. The treatment of emissivity, $\epsilon(\lambda)$

The assumed priors for $\epsilon(\lambda)$ play an important role in determining an asteroid's size. Technically, all fitted parameters require their prior probability distributions to be specified (for a discussion of Bayesian priors, see, for example, Chapter 5 in Ivezić et al., 2014). When $\epsilon(\lambda)$, or any other parameter, is assumed to be known, their priors effectively become Dirac δ functions.

When $\epsilon(\lambda)$ is assumed known, in principle even just two infrared flux measurements can constrain an asteroid's diameter D and temperature T_1. In essence, the measured color (i.e. the ratio of the two measured fluxes) constrains T_1 and the overall flux level constrains D. When measurement errors are present, the constraints become more complex, as discussed further below in the model validation section (Section 2.3). If more than two infrared flux measurements are available, if Σ is known from light curve analysis, and if measurement errors are sufficiently small, the model fluxes can be checked for internal consistency using statistical tests such as χ^2.

When $\epsilon(\lambda)$ is treated as a free parameter, care must be taken to avoid the $D\epsilon$ degeneracy when only thermal fluxes are measured: in Eq. (2) parameters D and ϵ appear only as a product $D^2\epsilon$. This degeneracy can be broken when flux measurements at wavelengths where the reflected light component is non-negligible are available. Because in Eq. (9) the reflected flux is proportional to $D^2(1-\epsilon)$ rather than proportional to $D^2\epsilon$, the $D-\epsilon$ degeneracy is broken and both ϵ and D can be estimated. The larger the contribution of the reflected flux to the total flux, the less correlated are the best-fit values of ϵ and D. We return to this point in Section 2.3.

In reality, a rather substantial variation in $\epsilon(\lambda)$ is observed for asteroids (e.g., see Fig. 1 in Myhrvold, 2018a). If more than two flux measurements are available, as is the case for NEOWISE data with four bands, parameters from a judiciously chosen parameterization of $\epsilon(\lambda)$ can also be fit to data. In the case of four bandpasses, up to two such parameters can be well constrained[6] (in addition to D and T_1, for the total of four fitted parameters).

Unknown functions are often specified as power laws with two free parameters: the power-law index and its overall normalization. This common approximation would work poorly given the $\epsilon(\lambda)$ functions observed for asteroids. Instead, motivated by the observed behavior of the $\epsilon(\lambda)$ functions, we consider an ansatz ("a simplified procedure", or "a provisional mathematical assumption") where we fit ϵ_{W1W2}, the value of $\epsilon(\lambda)$ in the WISE bands 1 and 2, and ϵ_{W3W4}, the value of $\epsilon(\lambda)$ in the WISE bands 3 and 4. In other words, we approximate $\epsilon(\lambda)$ as a step function $\epsilon(\lambda) = \epsilon_{W1W2}$ for λ<6 μm, and $\epsilon(\lambda)=\epsilon_{W3W4}$ for λ>6 μm. Because the two shortest WISE bands include substantial contribution of reflected flux (e.g., see Fig. 1 in Myhrvold, 2018a), constraints can be placed on ϵ_{W1W2} and ϵ_{W3W4} for well-observed asteroids. We discuss these constraints in more detail in the next section. We note that asteroids with sizes estimated by other means (e.g., from occultation or radar measurements) represent an invaluable sample for constraining the behavior of $\epsilon(\lambda)$ among the asteroid population.

2.2. Model implementation: ATM

We now briefly describe the ATM code architecture and highlight some aspects of its functionality. ATM is built on two bespoke Python classes: the 'observatory' and 'model' classes. The observatory class is used to describe an observatory's filter throughput curves as a function of wavelength. In the case of WISE, these are the modified quadrature formulae provided by Wright (2013). The model class has placeholder functions for how a thermal model describes the surface temperature distribution on a model asteroid, how the observing geometry is taken into account, and how the model integrates emitted flux over the surface of a model asteroid of unit diameter. The instances of the two classes, called objects, can then be passed as arguments with physical parameters (D, T_{ss}, ϵ, p, G) and observing parameters (α, r, Δ) to a series of functions. These functions calculate predicted in-band fluxes with and/or without reflected sunlight for the given observatory. Additionally, the model class can be passed separately to a different series of functions to generate model spectral energy distributions (SEDs) at a range of desired wavelengths.

As indicated by Myhrvold (2018a), fitting for different desired parameters and integrating the emitted flux can be slow. To make modeling and subsequent fitting more computationally tractable, ATM comes with emitted flux lookup tables generated for each thermal model at each WISE quadrature wavelength and at a range of wavelengths at 1 μm intervals between 1 and 30 μm. The lookup tables are functions of a grid of T_{ss} values, ranging between 100 to 1200 K in steps of 0.5 K, and α values, ranging between 0 and π radians in steps of $\pi/360$. Each table therefore has ~800,000 saved integration evaluations. Flux values of phase angles and subsolar temperatures that do not fall on this grid are interpolated between nearest neighbors using a bivariate spline approximation. Critically, interpolation is not conducted between wavelengths and so lookup tables need to be created for wavelengths not already included in the package.

To conduct MCMC sampling, ATM uses the *pymc3* (Salvatier et al., 2016) package. *pymc3* is a robust Python probabilistic programming and Bayesian inference package and it allows for multi-threaded sampling of

the posterior; the ATM lookup table implementation in combination with *pymc3* allows 4000 samples to be extracted in approximately 40 s on a moderate CPU. Each sampling chain requires a single thread, hence, sampling the posterior with 10 chains would require 10 threads and, because *pymc3* support parallelization, it takes the same amount of time to sample as a single chain. Given a set of asteroid flux measurements for either a single object or multiple objects, with assumptions on priors for fitting parameters, the user can fit for any combination of parameters as defined by Eq. (2). Should the user decide to constrain emissivity as described in Section 2.1.5, this task too is made simple by the use of several keyword arguments in the fitting function that describe how emissivity and albedo should be constrained and calculated. We note that MCMC sampling is an inherently stochastic process and so best-fit values obtained may differ slightly between different runs (well within stated uncertainties) unless an explicit value is set for the random seed (which needs to be done for every chain or walker).

We refer the reader to the GitHub repository for further information on the modeling code. A list of the ATM's primary requirements[7] can be found in the repository. The Notebooks used to validate the code are located here: https://github.com/moeyensj/atm_notebooks/tree/master/paper1/validation/. The Notebooks used to analyze the WISE sample of data (presented in the following sections) are located here: https://github.com/moeyensj/atm_notebooks/tree/master/paper1/analysis/. A Notebook showing how different assumptions regarding emissivity and albedo affect the modeling and retrieval of physical parameters is located here: https://github.com/moeyensj/atm_notebooks/tree/master/paper1/analysis/single_object_90367.ipynb.

2.3. ATM validation

We validate and verify the ATM code in two steps. First, we generate synthetic fluxes from a hypothetical asteroid using given D, T_1, and $\epsilon(\lambda)$, and then use ATM to perform Markov Chain Monte Carlo (MCMC) fitting to recover the input parameters. This step serves as a validation of internal code consistency. Second, physical validation is performed using a few studies from the literature, where asteroid sizes are known independently from infrared flux measurements (e.g., from radar measurements and stellar occultation measurements).

2.3.1. Validation of the internal code consistency

We generate synthetic fluxes in the four WISE bandpasses using an asteroid with $D = 1$ km, $T_1 = 422$ K, $\epsilon(\lambda) = 0.7$, and asteroid-Sun distance of 3 au. The observatory is placed at 2 au from the asteroid. The synthetic fluxes generated using the NEATM model are shown in Fig. 1. The four symbols with error bars correspond to the four in-band fluxes for the WISE observatory. As evident, the flux in the bluest band is dominated by the reflected flux component, while the two reddest bands are dominated by the emitted flux component. When fitting for D and T_1 using "observational" constraints in the four WISE bands, we add Gaussian scatter to each flux with a standard deviation of 20%. This scatter reflects the variability amplitude Σ from Eq. (11).

The synthetic model flux, generated using NEATM, is fit using MCMC and three different thermal models implemented in ATM: NEATM, STM, and FRM. We use 20 chains, each with 3000 samples and 500 burn-in samples. Depending on the prior and the complexity of the posterior, it may take a number of steps or iterations before the MCMC sampler (or similar) approaches the vicinity of the maximum likelihood solution. To avoid biasing statistics with these samples, it is common practice to define a number of first samples to discard, these are known as burn-in samples. 3000 samples with 500 burn-in typically means 2500 effective samples of the posterior per chain, although it should be noted this depends on implementation. The best-fit parameters obtained using the STM and FRM models are very similar, but not identical, to the

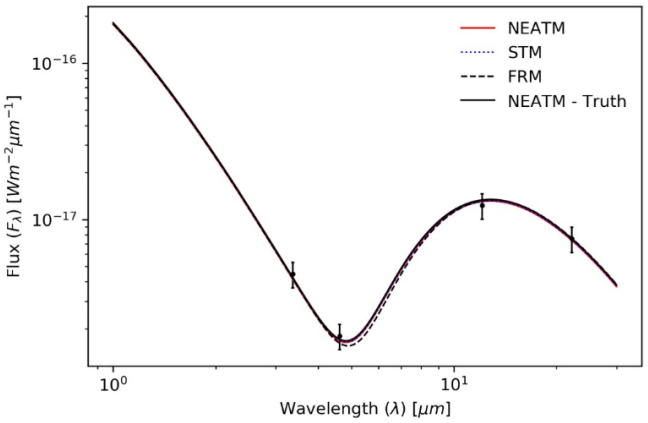

Fig. 1. The solid line shows the synthetic spectral energy distribution (SED) generated using the NEATM model for an asteroid with $D = 1$ km, $T_1 = 422$ K, $\epsilon(\lambda) = 0.7$, asteroid-Sun distance of 3 au, and an observer at 2 au from the asteroid. The four symbols with error bars correspond to the four in-band fluxes for the WISE observatory at the same distance from the asteroid. The error bars reflect the assumed variability amplitude but the implied scatter is not shown in this plot. The three other lines show the best fits obtained with ATM (see text). The two solid lines are indistinguishable. This figure was generated using example_synthetic_changingNumObservations.ipynb.

parameters obtained with the correct NEATM model. The three best-fit SEDs for the fiducial validation case are shown in Fig. 1.

The posterior probability density function (pdf) for the NEATM fitting case is shown in Fig. 2. Note that the use of full multi-dimensional pdf for the analysis of constraints on the model parameters is superior to using the so-called "point estimates" obtained by maximum likelihood methods (e.g., the least squares method). A strong covariance between D

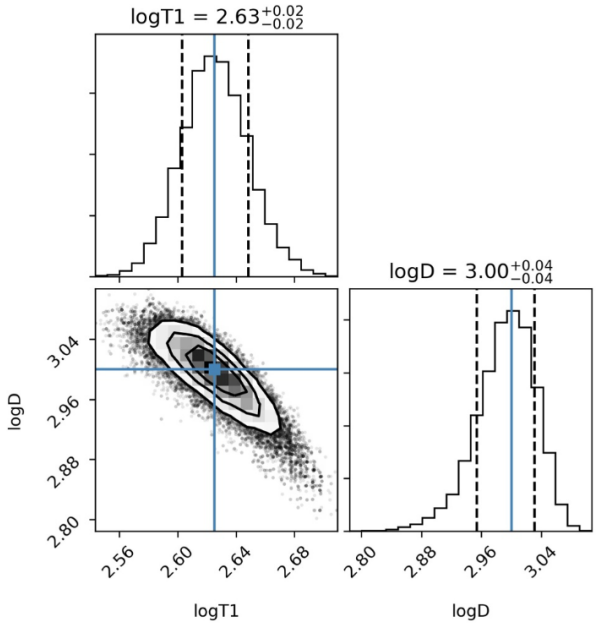

Fig. 2. The posterior probability density function (pdf) for the case of fitting asteroid diameter D (in m) and temperature parameter T_1 (in Kelvin; see Eq. (5)). The equivalent $1-\sigma$ uncertainties for D and T_1, measured as standard deviation of the marginal probability distributions (shown as histograms; the vertical solid lines are true input values and dashed lines mark $1-\sigma$ un-

and T_1 is clearly visible: a smaller D is compensated by higher T_1, with a bias (as well as scatter) in best-fit D twice as large as the corresponding T_1 bias (because the observed flux approximately scales with $D^2T_1^4$). The equivalent $1-\sigma$ uncertainties for D and T_1 are 9.1% and 5.4%, respectively. These uncertainty levels can be approximately explained from first principles: given the flux uncertainty of 20% and 4 observations (1 in each band), the overall flux normalization uncertainty is about 10% ($=20\%/\sqrt{4}$). If T_1 were known, the best-fit uncertainty for D would be about 5%, and if D were known, the best-fit uncertainty for T_1 would be about 2.5%. Given the covariance between best-fit D and T_1, their expected uncertainties are about twice as large (see Fig. 2), in good agreement with the MCMC results.

When the number of epochs N is increased, we find that the fitting uncertainty for the best-fit parameter decreases as $N^{-1/2}$, as expected. The fitting uncertainty increases approximately linearly with the assumed flux uncertainty (Σ=0.2 mag, presumably due to variability, rather than due to measurement uncertainties). For example, with $\Sigma =$ 0.1 mag and 25 observing epochs (each with four WISE fluxes), the fitting uncertainty for the best-fit parameters D and T_1 decreases by about a factor of ten, to 0.8% and 0.5%, respectively. Of course, here we are using the same model for fitting as was used to generate the synthetic flux – in reality, such small uncertainties are *essentially impossible* due to numerous systematic shortcomings of idealized thermal models (and uncertain $\epsilon(\lambda)$ – see below).

These tests assumed that the *correct* values of $\epsilon(\lambda)$ are known a priori. When we set ϵ_{W1W2} to an incorrect value of 0.8, instead of the true input value of 0.7, we found that the best-fit value of D was biased by as much as 24%. Therefore, *even though formal fitting precision can be high, the accuracy of fitted parameters can be significantly worse.*

We investigated next how the precision of the best-fit D and T_1 decreases when emissivity $\epsilon(\lambda)$ is also fit. Following discussion in Section 2.1.5, we fit ϵ_{W1W2}, the value of $\epsilon(\lambda)$ in the WISE bands 1 and 2, and assume that ϵ_{W3W4}, the value of $\epsilon(\lambda)$ in the WISE bands 3 and 4, is known. In other words, we assume a flat prior in the range 0 to 1 for ϵ_{W1W2}, and a Dirac δ function for ϵ_{W3W4}, centered on its true value. The posterior pdf is shown in Fig. 3. The differences compared to the posterior pdf in Fig. 2 are substantial; for example, the equivalent $1-\sigma$ uncertainties for D and T_1 increased by about a factor of three: from 9.1% and 5.4% to 31% and 13%, respectively. In addition to the covariance between D and T_1, both parameters show covariances with ϵ_{W1W2}. A larger ϵ_{W1W2} can be compensated by a lower T_1 or a larger D. For example, a 0.1 uncertainty in ϵ_{W1W2} can induce a ~25% uncertainty in D! Given that typically $\epsilon(\lambda)$ is not known to better than 0.05, it follows that uncertainty of best-fit D is unlikely below ~10% (in agreement with recent analysis by Masiero et al., 2018a; Wright et al., 2018).

Compared to D and T_1, the posterior constraint on ϵ_{W1W2} is much weaker. The standard deviation for the marginal distribution of ϵ_{W1W2} shown in Fig. 3 is 0.23, but note that this marginal pdf is far from Gaussian. The fitting constraints on all three free parameters, and ϵ_{W1W2} in particular, improve as the number of data points increases and the scatter Σ decreases. Fig. 4 shows the posterior pdf when synthetic data include 25 epochs (each with four WISE fluxes) instead of one in Fig. 3, and with flux uncertainty of $\Sigma = 0.1$ mag, instead of 0.2 mag. As evident, the posterior constraints for all three fitted parameters are significantly improved. The standard deviations for the marginal distributions are 2.1% for D, 1.0% for T_1 and 0.011 for ϵ_{W1W2} (recall that in the case of fitting only D and T_1, their uncertainties were equal to 0.8% and 0.5%, respectively, using the same number of observations and the same value of Σ). Therefore, adding ϵ_{W1W2} as a free parameter results in about twice as large statistical uncertainties for the best-fit size D.

2.3.2. Physical validation

several real asteroids, with sizes known independently from infrared flux measurements, to physically validate the code and implemented models.

First, we consider the observations of asteroid 433 Eros by Lebofsky and Rieke (1979). We use best-fit model parameters from Harris (1998) and demonstrate in Fig. 5 that ATM produces model fluxes in agreement with measurements. An additional example, with data and analysis for near-Earth asteroid 1991 EE from Harris et al. (1998), also includes and validates the reflected light component, as shown in Fig. 5. Finally, ATM is also validated against NEOWISE results, as described in detail in Section 3.

2.4. An illustration of ATM fitting capabilities: different emissivity priors

The choice of Bayesian priors when fitting asteroid spectral energy distributions, such as priors for emissivity, can have a significant impact on the best-fit parameters. We illustrate ATM's capabilities for fitting asteroid spectral energy distributions, with emphasis on the treatment of Bayesian priors, using three well-observed asteroids: (25916), (54789), and (90367).

We study how the choices of priors affect the posterior pdf for fitted parameters and we study the resulting biases in point estimates derived from these pdfs (that is, "best-fit parameters"). We consider five different models, where in addition to fitting for asteroid diameter, D, and temperature parameter, T_1, we treat emissivity $\epsilon(\lambda)$ as follows:

- **Model 1:** $\epsilon(\lambda) = \epsilon_0 =$ const. and we fit for ϵ_0 using a flat prior $0 < \epsilon_0 < 1$.
- **Model 2:** We set $\epsilon_{W3} = \epsilon_{W4} = 0.9$ (that is, the prior is a Dirac δ function) and fit for unknown $\epsilon_{W1} = \epsilon_{W2} = \epsilon_{W1W2}$, where subscripts indicate WISE bands. Here, and in models below, we model $\epsilon(\lambda)$ as a step function, with transition wavelengths[8] between bands at $\lambda_{W1W2} = 3.9$ μm, $\lambda_{W2W3} = 6.5$ μm, and $\lambda_{W3W4} = 18.5$ μm. Note that in SED plots discussed below emissivity ϵ_{W1} is extrapolated to wavelengths shorter than λ_{W1W2}, implying a pre-set value for the visual albedo, p_V. The same is true for emissivity in the most redward band, ϵ_{W4}; it is extrapolated to wavelengths beyond the W4 bandpass for plotting purposes.
- **Model 3:** We set $\epsilon_{W3} = 0.70$ and $\epsilon_{W4} = 0.86$, motivated by ensemble analysis presented later in Section 3.3, and fit for ϵ_{W1W2}.
- **Model 4:** We set $\epsilon_{W3} = 0.80$ and $\epsilon_{W4} = 0.98$ and fit for ϵ_{W1W2}. The $\epsilon_{W4}/\epsilon_{W3}$ ratio is the same as in the previous model.
- **Model 5:** We set $\epsilon_{W3} = 0.80$ and $\epsilon_{W4} = 0.98$ and fit for ϵ_{W1} and ϵ_{W2}. Compared to the previous model, here we do not enforce $\epsilon_{W1} = \epsilon_{W2}$.

Fig. 6 shows the posterior pdf for four fitted parameters in the case of model 5 and asteroid (54789). We have also considered a case where emissivity values in all four bands are free fitting parameters, for a total of six fitted parameters, but concluded that it is sufficient to start the discussion here with model 5. As the panels in the bottom row of Fig. 6 show, the pdf for ϵ_{W2} is quite wide and not too dissimilar to its prior, while D, T_1 and ϵ_{W1} have pdfs that are much narrower than their priors, and with a well defined peak. In other words, ϵ_{W2} is much less constrained by the data than the other three parameters. This conclusion is also valid for the other two asteroids considered here.

Following the behavior of emissivity inferred from laboratory spectra for different materials, as illustrated in Fig. 1 from Myhrvold (2018a), we enforce an additional fitting constraint: $\epsilon_{W1} = \epsilon_{W2} = \epsilon_{W1W2}$ and fit for three free parameters (models 2, 3, and 4). Fig. 7 shows the posterior pdf for fitted parameters in the model 4 case. The fitted parameter ϵ_{W1W2} now has a pdf that is much narrower than its priors and with a well defined peak. The same behavior is observed for the other two asteroids. As a result of this analysis, we conclude that for the robust

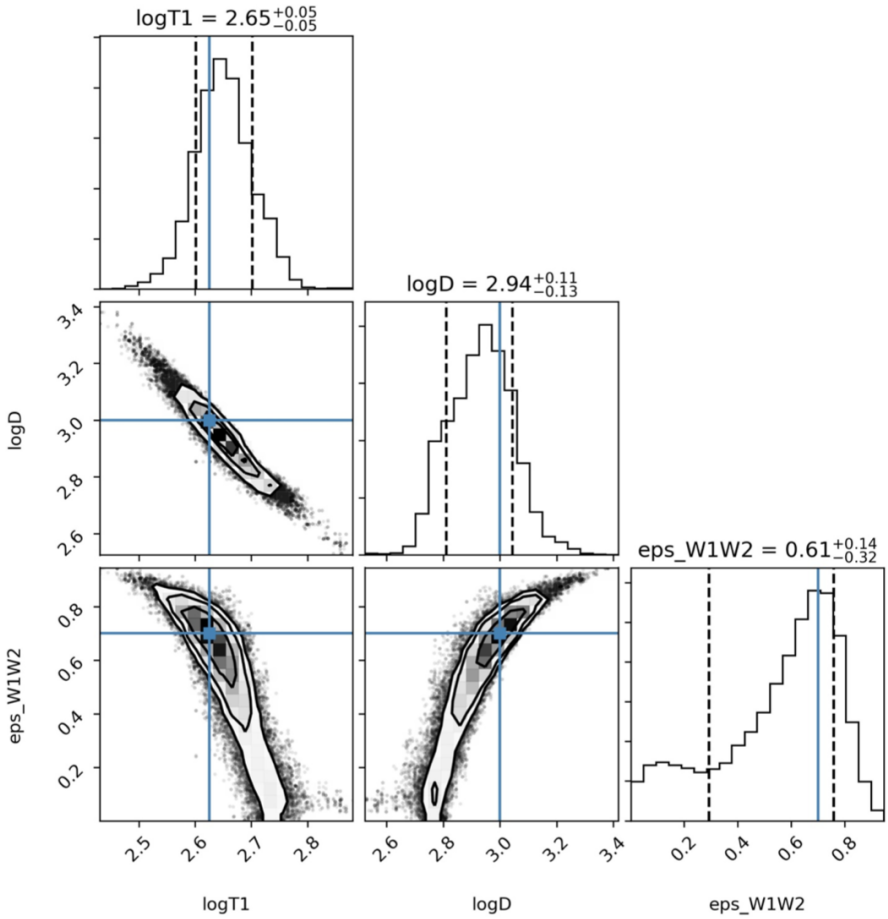

Fig. 3. Analogous to Fig. 2, but here for the case of fitting three free parameters: asteroid diameter D, temperature parameter T_1, and the value of $\epsilon(\lambda)$ in the WISE bands 1 and 2, ϵ_{W1W2}. Although all other simulation and fitting parameters are the same as is in Fig. 2, the differences between the posterior pdfs are substantial. For example, the equivalent $1-\sigma$ uncertainties for D and T_1 increased from 9.1% and 5.4% to 31% and 13%, respectively. In addition to the covariance between D and T_1, both parameters show covariances with ϵ_{W1W2}. Compared to D and T_1, posterior constraints on ϵ_{W1W2} are much weaker (the $1-\sigma$ uncertainty for ϵ_{W1W2} is 34%!). This figure was generated using example_synthetic_changingNumObservations.ipynb.

fitting of these and other less observed asteroids, we need to only fit for ϵ_{W1W2}, and not for ϵ_{W1} and ϵ_{W2} separately.

There is a remaining question of what to choose for the values of ϵ_{W3} and ϵ_{W4} that are not fitting parameters. As Table 1 shows, models 2, 3, and 4 result in values of best-fit D varying by about 10–15 %, depending on the chosen values of ϵ_{W3} and ϵ_{W4}. In addition, adopting $\epsilon(\lambda) = \epsilon_0$ and fitting for ϵ_0 can result in a change of D as large as 16% (compare model 1 and model 2 for asteroid 54789). This approach also leads to a taxonomy-dependent D bias (systematic uncertainty) because the actual bias depends on detailed deviations of $\epsilon(\lambda)$ from the assumed constant value of ϵ_0. We will return to the discussion of optimal choice of ϵ_{W3} and ϵ_{W4} in Section 3.3.

Fig. 8 compares spectral energy distributions for models 1–5 in the case of asteroid (90367). As evident, all models are equally successful in explaining observed fluxes in bands W1 and W2. Models 3, 4 and 5 are more successful than models 1 and 2 in explaining observed fluxes in bands W3 and W4. This improvement is due to the choice $\epsilon_{W4} = 1.22\epsilon_{W3}$, instead of $\epsilon_{W4} = \epsilon_{W3}$ for models 1 and 2, and is discussed in detail in Section 3.3. This model degeneracy implies a bias in best-fit asteroid size in the range 10–20 %, as discussed above.

Note that the prediction for reflected flux at wavelengths below ~2 μm varies by about 30% due to variation of best-fit D and best-fit ϵ_{W1}, which implies an extrapolated value of p_V. Nevertheless, a measurement of flux at optical wavelengths could only break the model degeneracy if there were a strong prior reason to believe that emissivity values at optical wavelengths were somehow fully determined by the value of

3. Application of ATM to WISE data and comparison with NEOWISE analysis

In this section, we apply ATM to a "gold" sample of ~7000 best-observed asteroids by WISE. In addition to the best-fit diameter D and characteristic temperature T_1 for each object, we also obtain best-fit emissivity across bands W1 and W2, ϵ_{W1W2}. We compare ATM best-fit parameters to their values as published by the NEOWISE team, discuss physical implications of the best-fit parameters, and derive an approximate method for estimating D from WISE W3 measurements that doesn't require model fitting and is applicable to the majority of asteroids with WISE measurements.

3.1. Selection of high-quality WISE data and reliable ATM fits

To compare ATM best-fit parameters with their values as published by the NEOWISE team, we select a relatively small subsample (~5% of the full sample) with the highest quality and quantity of WISE data. Observations of asteroids by WISE were obtained using the same criteria described by NEOWISE papers. The list of observations were obtained from the Minor Planet Center database for observatory code C51. The positions and times were then used to search the WISE All-Sky Single Exposure Level 1b Source catalog using a cone search with radius 10 arcsec, and an observation time tolerance of ± 2 s. The distance from the asteroid to the Sun, distance from the asteroid to WISE, and the phase angle were obtained with the JPL HORIZONS service[9], using the asteroid

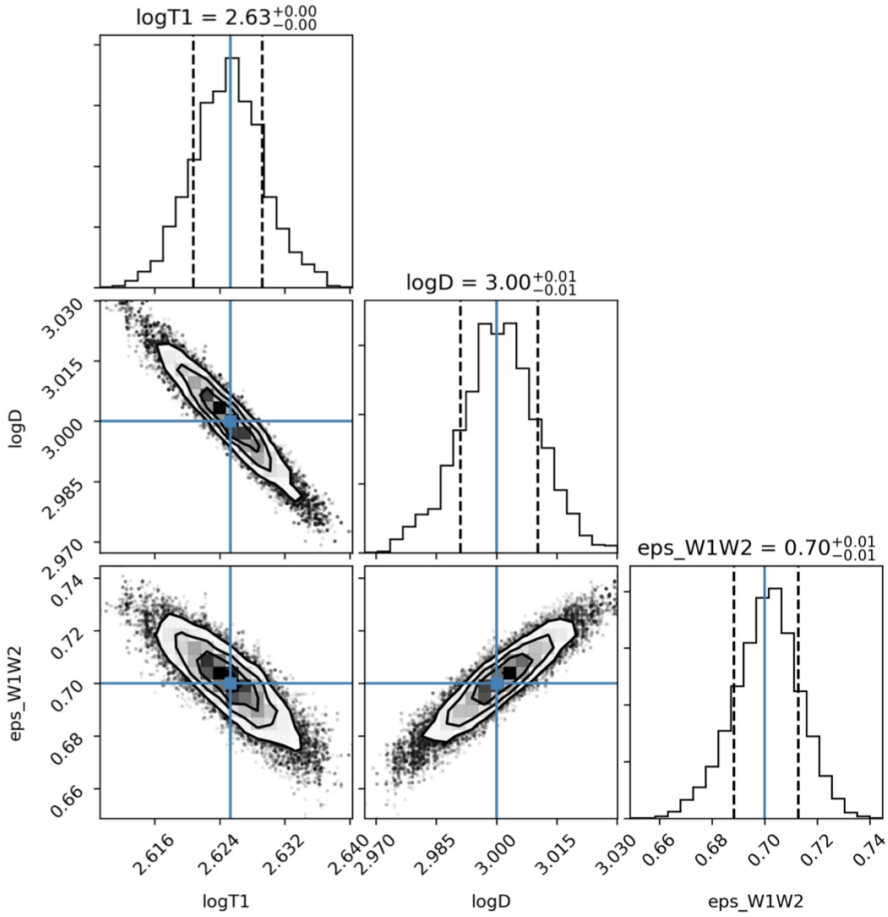

Fig. 4. Analogous to Fig. 3, but here with better observational constraints: 25 epochs (each with four WISE fluxes) instead of one, and with flux uncertainty Σ of 0.1 mag, instead of 0.2 mag. Note the significant improvement in best-fit constraints, as well as persistent covariance between the fitted parameters. This figure was generated using example_synthetic_changingNumObservations.ipynb.

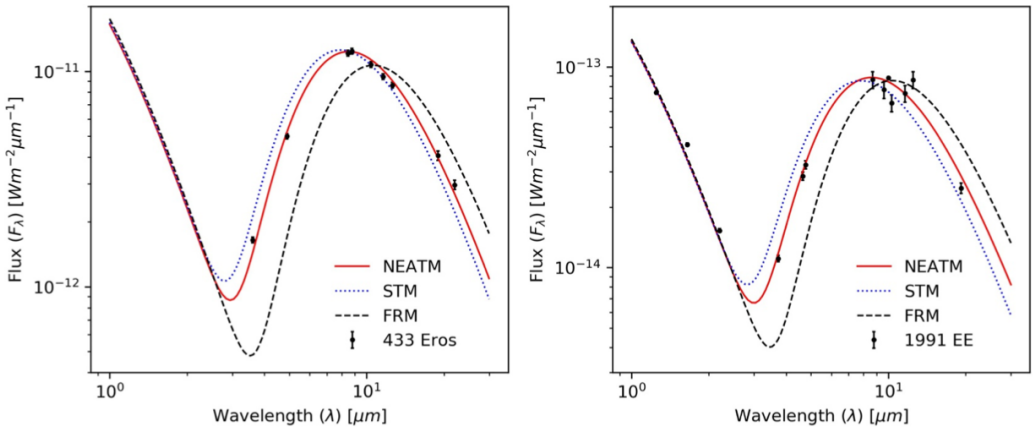

Fig. 5. In the left panel, physical model validation using data for 433 Eros (symbols) from Lebofsky and Rieke (1979) and models evaluated using best-fit parameters from Harris (1998). This panel recreates Fig. 1a from Harris (1998). In the right panel, we plot physical model validation using data for asteroid 1991 EE from Harris et al. (1998). This panel recreates Fig. 3 from Harris et al. (1998) and was generated using example_1991EE&Eros.ipynb.

designation and observation time. To ensure high quality measurements, we require[10]:

- at least 3 observations in each band with photometric signal-to-noise ratio of at least 4, and
- artifact flag = 0 and quality flag = A, B, C, but not U or X.

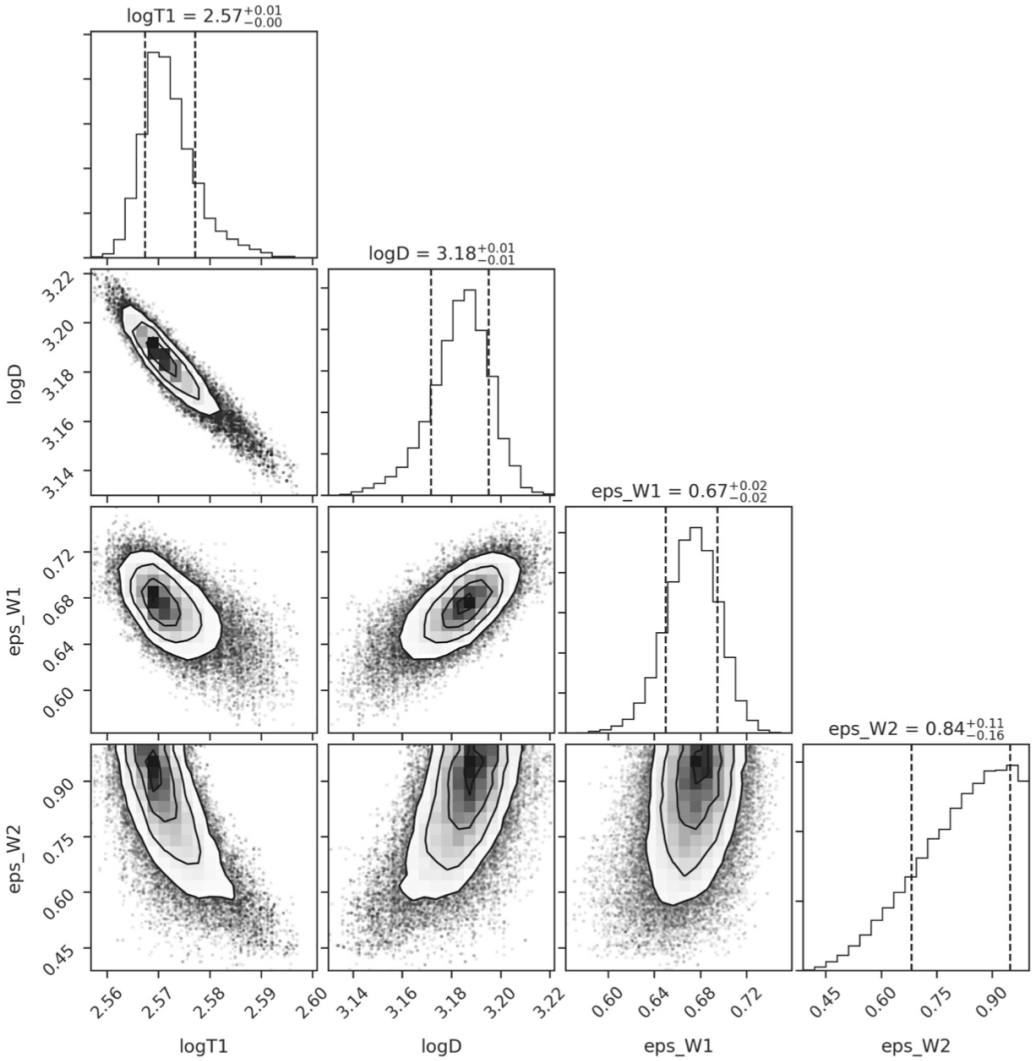

Fig. 6. ATM "corner" plot for asteroid (54789) assuming model 5. Here both ϵ_{W1} and ϵ_{W2} are free parameters; compare to Fig. 7 where an additional constraint $\epsilon_{W1} = \epsilon_{W2}$ is imposed. This figure was generated using single_object_54789.ipynb.

10, while a quality flag of C indicates $2 < \text{SNR} < 3$, while U or X indicate an upper limit on magnitude or a non-detection, respectively. This set of criteria reduces ~10 million observations of ~150,000 asteroids to ~350,000 observations of 9672 asteroids.

For some objects, measurements still include outliers despite the above quality cuts. We clip outliers using a simple automated iterative algorithm. For each asteroid, using only the observations that are within 1 mag of the median magnitude a linear time versus magnitude model is fit for each band. Once the models have been fit, all observations that are more than 1 mag away from the best-fit model in each band are flagged as magnitude outliers. The observations may be flagged regardless of if they were or were not used to build the linear model. This outlier clipping reduced the aforementioned 9672 asteroids down to 7363 (~14% of observations at this stage were cut as magnitude outliers). Finally, 4 objects were removed as they had missing assumptions on slope parameter, G. The final "gold" sample contains ~308,000 observations (~77,000 four-band observations) of 7359 objects.[11]

For each object, we fit for diameter D, characteristic temperature T_1 and emissivity across the W1–W2 wavelength range, ϵ_{W1W2}. We produced four sets of fits that differ in priors for emissivity across the W3–W4 wavelength range. Following NEOWISE analysis, we adopted $\epsilon_{W3} = \epsilon_{W4} = 0.90$ for the first set. Informed by discrepancies between the observed and modeled distributions of objects in the W3–W4 vs. W2–W3 color-color diagram (see below for detailed discussion), for the other three sets we adopted (ϵ_{W3},ϵ_{W4}) = (0.80, 0.98), (0.76, 0.93) and (0.70, 0.86).

The top right panel in Fig. 9 shows the distribution of χ^2 per degree of freedom as a function of the total number of data points in all 4 bands. The computation of χ^2, as well as fitting, assumes an intrinsic scatter due to variability of $\Sigma = 0.15$ mag (see Eq. (11)). The distribution of χ^2, with a mode at 0.88, validates this choice. We note that the mean χ^2 for the entire 7359 sample is 1.77, when cutting outliers with $\chi^2 < 6$ then the mean reduces to 1.11. There is no correlation of the χ^2 distribution with the number of data points; typically, there are ~50 observations per object.

3.2. Comparison to NEOWISE fits

To compare our best-fit parameters to those obtained by the

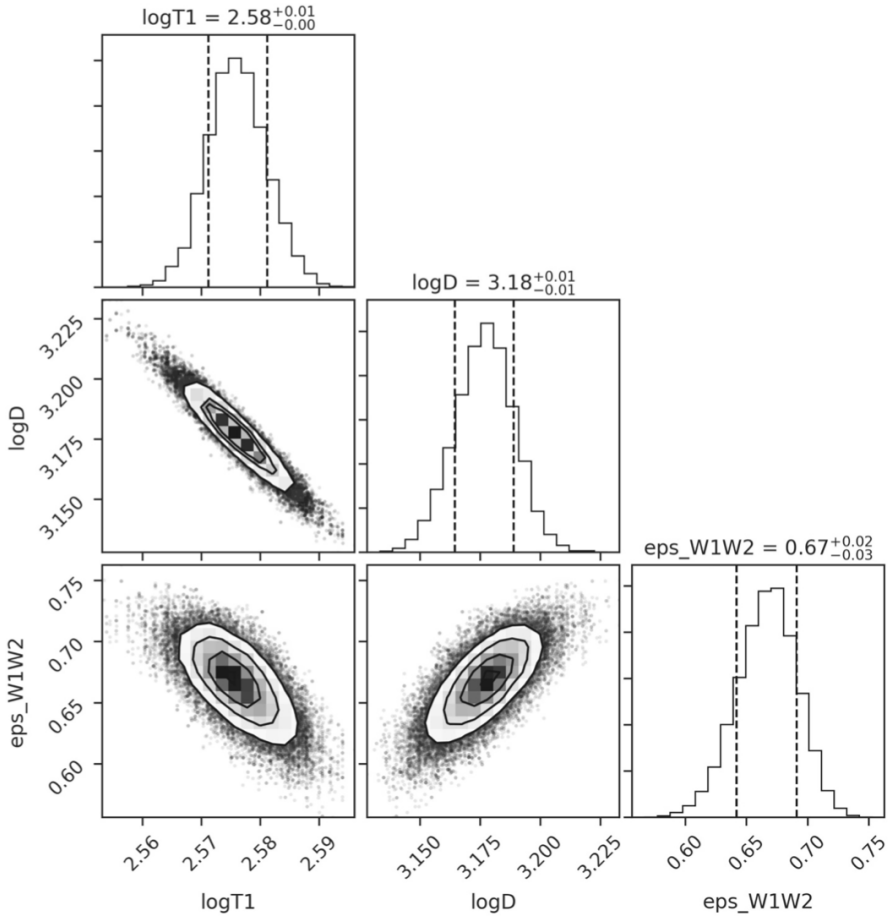

Fig. 7. ATM "corner" plot for asteroid (54789) assuming model 4. This model imposes a constraint $\epsilon_{W1} = \epsilon_{W2}$. Compare to Fig. 6 where both ϵ_{W1} and ϵ_{W2} are free parameters. This figure was generated using single_object_54789.ipynb.

Table 1
The Best-fit NEATM Parameters[a] for Asteroids (25916), (54789) and (90367).

Model	log(D/m)	Δ D [%][b]	log(T_1/K)	p_{W1}[c]	p_{W2}[c]	ϵ_{W1}	ϵ_{W2}	ϵ_{W3}	ϵ_{W4}
Asteroid (25916)									
1	3.690	0.00	2.587	0.322	$=p_{W1}$	0.876	$=\epsilon_{W1}$	$=\epsilon_{W1}$	$=\epsilon_{W1}$
2	3.682	−2.02	2.589	0.328	$=p_{W1}$	0.874	$=\epsilon_{W1}$	(0.90)	(0.90)
3	3.701	2.45	2.586	0.301	$=p_{W1}$	0.885	$=\epsilon_{W1}$	(0.70)	(0.86)
4	3.684	−1.56	2.589	0.321	$=p_{W1}$	0.877	$=\epsilon_{W1}$	(0.80)	(0.98)
5	3.662	−6.36	2.603	0.243	0.927	0.907	0.644	(0.80)	(0.98)
Asteroid (54789)									
1	3.240	0.00	2.558	0.675	$=p_{W1}$	0.741	$=\epsilon_{W1}$	$=\epsilon_{W1}$	$=\epsilon_{W1}$
2	3.166	−15.7	2.575	0.904	$=p_{W1}$	0.653	$=\epsilon_{W1}$	(0.90)	(0.90)
3	3.197	−9.49	2.571	0.795	$=p_{W1}$	0.695	$=\epsilon_{W1}$	(0.70)	(0.86)
4	3.177	−13.6	2.576	0.867	$=p_{W1}$	0.667	$=\epsilon_{W1}$	(0.80)	(0.98)
5	3.184	−12.1	2.571	0.852	0.431	0.673	0.834	(0.80)	(0.98)
Asteroid (90367)									
1	3.230	0.00	2.580	0.075	$=p_{W1}$	0.971	$=\epsilon_{W1}$	$=\epsilon_{W1}$	$=\epsilon_{W1}$
2	3.254	5.74	2.576	0.072	$=p_{W1}$	0.972	$=\epsilon_{W1}$	(0.90)	(0.90)
3	3.278	11.6	2.572	0.067	$=p_{W1}$	0.974	$=\epsilon_{W1}$	(0.70)	(0.86)
4	3.261	7.49	2.575	0.066	$=p_{W1}$	0.975	$=\epsilon_{W1}$	(0.80)	(0.98)
5	3.261	7.36	2.575	0.065	0.118	0.975	0.955	(0.80)	(0.98)

[a] The values in parenthesis represent priors.
[b] Change in diameter is measured relative to Model 1 diameters.
[c] Geometric albedo is calculated using Eq. (8) and is stated for convenience.

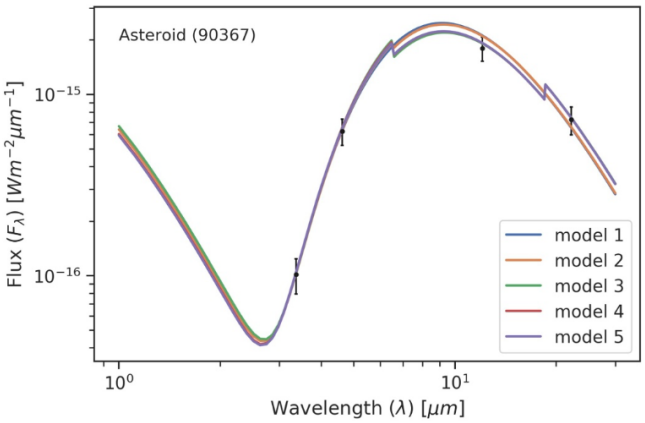

Fig. 8. The symbols show the median fluxes and their errors for WISE data for asteroid (90367). The errors include the contribution of assumed variability amplitude of 0.15 mag. The solid lines show five NEATM models that differ in chosen priors for emissivity (see text). The best-fit parameters for these models are listed in Table 1. The sharp changes in modeled fluxes at $\lambda_{W2W3} = 6.5$ μm and $\lambda_{W3W4} = 18.5$ μm is due to a simplistic approximation of emissivity as a function of wavelength as a step function. This figure was generated using single_object_90367.ipynb.

NEOWISE team, we match objects from our dataset to objects from the 2016 Planetary Data System version of the NEOWISE Diameters and Albedos database[12] (Mainzer et al., 2016), hereafter "NEOWISE values". For this initial comparison, we also require NEOWISE fit code "DVBI" and at least three observations in each of the four WISE bands. A NEOWISE fit code of "DVBI" indicates that diameter (D), visual albedo (V), beaming parameter (B), and infrared albedo (I) were fit for. If a fit code contains a hyphen (-) instead of a letter, the parameter which the hyphen replaced was not fit for and instead had an assumed value. For example, the most numerous fit code with over 87,000 results, "DV– ", fits for diameter and visual albedo while beaming parameter and infrared albedo had assumed values. We limit our selection of fit codes to "DVBI" as this is the most NEATM-like fitting scenario and is the NEOWISE fit most similar to our modeling assumptions. As "DVBI" fits comprise only about 4600 of the total fits published by NEOWISE, crossmatching with this requirement immediately reduced the number of matches. The additional requirement of having at least three observations in each band further reduces number of matches as the vast majority of NEOWISE results did not use observations in all four bands. We also selected only the NEOWISE results for each asteroid that used the most data points as some asteroids had multiple fitting results under the same fit code. These criteria yielded a final comparison sample of 2807 asteroids. The ratio of best-fit ATM and NEOWISE values vs. χ^2 for diameter D and emissivity ϵ_{W1W2} is shown in the bottom panels in Fig. 9. As evident, there is a good agreement between the two sets of best-fit parameters.

A more quantitative comparison is shown in the top two panels in Fig. 10, for the fitting case with $\epsilon_{W3} = \epsilon_{W4} = 0.90$. Both the best-fit diameters and emissivity agree on average to within 0.4%, with a scatter of 5.5% for diameters and 1.1% for emissivity. Plausible reasons for the scatter include different outlier rejection algorithms, different treatments of Kirchhoff's law[13] (for a detailed discussion, see Myhrvold, 2018a), and how ATM accounts for intrinsic variability, although we cannot exclude other causes (e.g., a slight error in the quadrature formula for the W3 band, see Appendix A for details). Whatever the reason, the discrepancies are encouragingly small.

Our ATM results faithfully match the tri-modal distribution of ϵ_{W1W2} emissivity discovered by Masiero et al. (2014). The two bottom panels in Fig. 10 show the resulting distributions of emissivity and corresponding albedo based on ATM results. The parameters of a best-fit 3-component Gaussian mixture model are listed in Table 2. Note that the fractions of three components listed in Table 2 (20%, 30% and 50%) are not corrected for the various sample selection effects.

This tri-modality is related to taxonomic classes; according to Masiero et al. (2014) the low and high p_{W1W2} albedo peaks correspond to the low and high visible albedo groups observed previously (C/D/P and S groups, respectively), while the intermediate albedo peak corresponds to intermediate visible albedo values that are blended with the high albedo objects in visible albedo distributions.

3.3. Understanding ATM best fits

Despite good agreement between ATM and NEOWISE fitting results, there are statistical problems with best-fit models when $\epsilon_{W3} = \epsilon_{W4} = 0.90$: the magnitude residuals for all objects are offset from zero by 0.1–0.2 mag in most bands. These offsets indicate that the thermal models used are not fully capable of explaining WISE data. The distributions of observed objects in WISE color-color diagrams offer an efficient way to study possible causes of model deficiencies. We note that the use of a full sample to constrain priors for individual objects is known as Hierarchical Bayesian modeling (e.g., see Chapter 5 in Ivezić et al., 2014).

For each object, we compute the median observed magnitudes in each WISE band and the median observed colors. The resulting color-color diagrams are shown in Fig. 11. The position along the model sequences in the W3-W4 vs. W2-W3 color-color diagram (see the bottom right panel in Fig. 11) is by and large controlled by the asteroid-Sun distance r, while both r and emissivity ϵ_{W1W2} (or, equivalently, albedo) determine the position in the W2-W3 vs. W1-W2 color-color diagram (bottom left panel). Note the three clearly delineated observed sequences in this diagram. Their finite width is influenced by both the distribution of ϵ_{W1W2} values, and the distribution of observational phase angles (increase in ϵ_{W1W2} and/or phase angle moves model sequences from the bottom left to the top right). Note also that the W2–W3 and W3–W4 colors become redder as r increases, while the W1–W2 color becomes redder for decreasing r (because of the increasing relative contribution of the blue reflected component).

The placement of model tracks in the W3–W4 vs. W2–W3 color-color diagram is controlled by emissivity across the W3 and W4 wavelength range, parameterized as ϵ_{W3} and ϵ_{W4}, while the position along the tracks is controlled by the asteroid's temperature (itself controlled by T_1 and the asteroid-Sun distance). The three model tracks shown by dashed lines correspond to $\epsilon_{W3} = \epsilon_{W4} = 0.90$, and are strongly ruled out by the data: they don't overlap the majority of data points.

A small fraction of data points (about 3%) with very blue W3–W4 colors could be, at least in principle, either measurement outliers, or interesting objects with unusual ϵ_{W3} and ϵ_{W4} values. However, all of them are extremely bright in W3 (W3<1.8) and essentially represent the top 3% brightest sources. Hence, it is likely that their peculiar behavior in the W3–W4 vs. W2–W3 color-color diagram is simply due to overestimated W3 flux by about 0.5 mag. This conclusion is consistent with the recently reported flux corrections for saturation in the W3 band by Wright et al. (2018).

The only way to move model tracks to reach the data distribution is to assume $\epsilon_{W3} > \epsilon_{W4}$. After some experimentation, we adopted $\epsilon_{W3} \sim$

[12] See https://sbn.psi.edu/pds/resource/neowisediam.html.

[13] We have run ATM with a fiducial dataset both ways: with Kirchhoff's law properly implemented and with the NEOWISE ansatz. The distribution of the

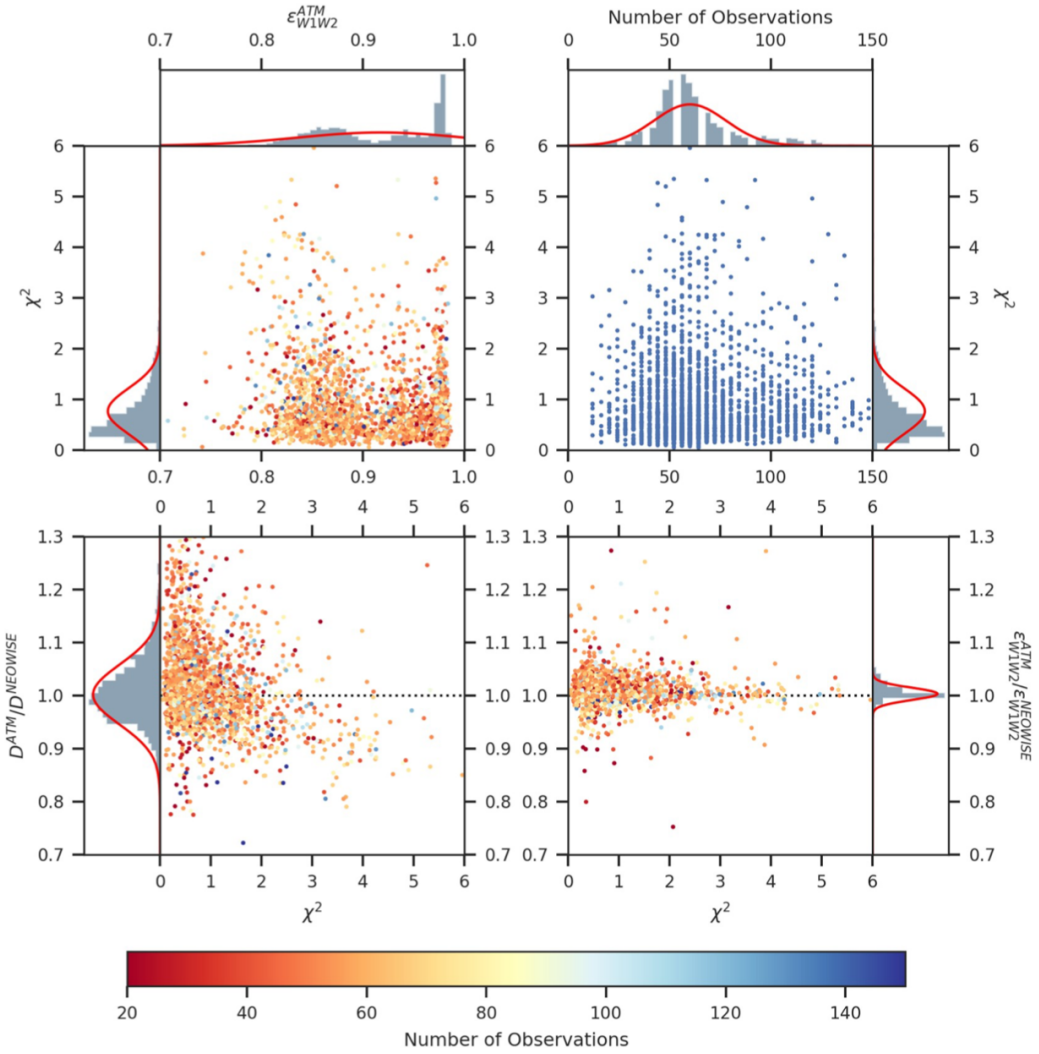

Fig. 9. Analysis of ATM fit quality. The top left panel shows the χ^2 per degree of freedom vs. ϵ_{W1W2} diagram for asteroids that pass initial quality cuts (see Section 3.1), have NEOWISE fit code "DVBI" and have at least 3 observations in each of the four WISE bands. The best-fit χ^2 and emissivity ϵ_{W1W2} were obtained using ATM. The symbols are color-coded using the total number of data points, according to the legend below the panel. The top right panel shows the χ^2 vs. the total number of data points diagram. The bottom panels show the ratios of the best-fit diameter (left) and ϵ_{W1W2} (right) obtained by ATM and the 2016 NEOWISE release values vs. χ^2, color-coded the same way as in the top left panel. This figure was generated using analysis.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

thermal emission component, as discussed earlier).

We investigated three pairs of values: (ϵ_{W3},ϵ_{W4}) = (0.80,0.98), (0.76, 0.93) and (0.70, 0.86). All three pairs produce model tracks in Fig. 11 that are in a much better agreement with data than $\epsilon_{W3} = \epsilon_{W4} = 0.9$. The model tracks could be moved further to completely overlap the data distribution by adopting $\epsilon_{W4} \approx 1.34\ \epsilon_{W3}$ (e.g., $\epsilon_{W3} = 0.73$, $\epsilon_{W4} = 0.98$). We leave detailed investigation of such models to future work.

While in all three cases the agreement between the data and the models in color-color diagrams is essentially unchanged, the values of best-fit diameters change in inverse proportion to the values of ϵ. Relative to the best-fit diameters for (ϵ_{W3},ϵ_{W4}) = (0.80,0.98) case, which on average agree with the NEOWISE values to better than 1%, the sizes for (0.76, 0.93) case are on average 4% larger, and 10% larger for (0.70, 0.86) case. Note that best-fit diameters scale with ϵ faster ($\propto \epsilon^{-0.7}$) than the naively expected $1/\sqrt{\epsilon}$ due to non-linear fitting effects. We chose (ϵ_{W3},ϵ_{W4}) = (0.80,0.98) for the rest of analysis here because of agree-

biases could be larger for the $\epsilon_{W4} \approx 1.34\epsilon_{W3}$ case.

This simple analysis illustrates how data can constrain model parameters, and also justifies our updated choice of priors for ϵ_{W3} and ϵ_{W4}. We now proceed with the analysis of best-fit model parameter distributions, their relationship to data properties, and derive an approximate size estimator that can be applied to the majority of objects in WISE sample.

3.4. An approximate single-band W3-based size estimator

The availability of WISE measurements varies greatly with bandpass and only a small minority of asteroids have data in all four bands. The majority of asteroids have data only in W3 and W4 bands, or in the W3 band alone. In fact, 45% of published NEOWISE fits (70,000+ out of ~160,000) used only W3 band data, while only 9% of published fits

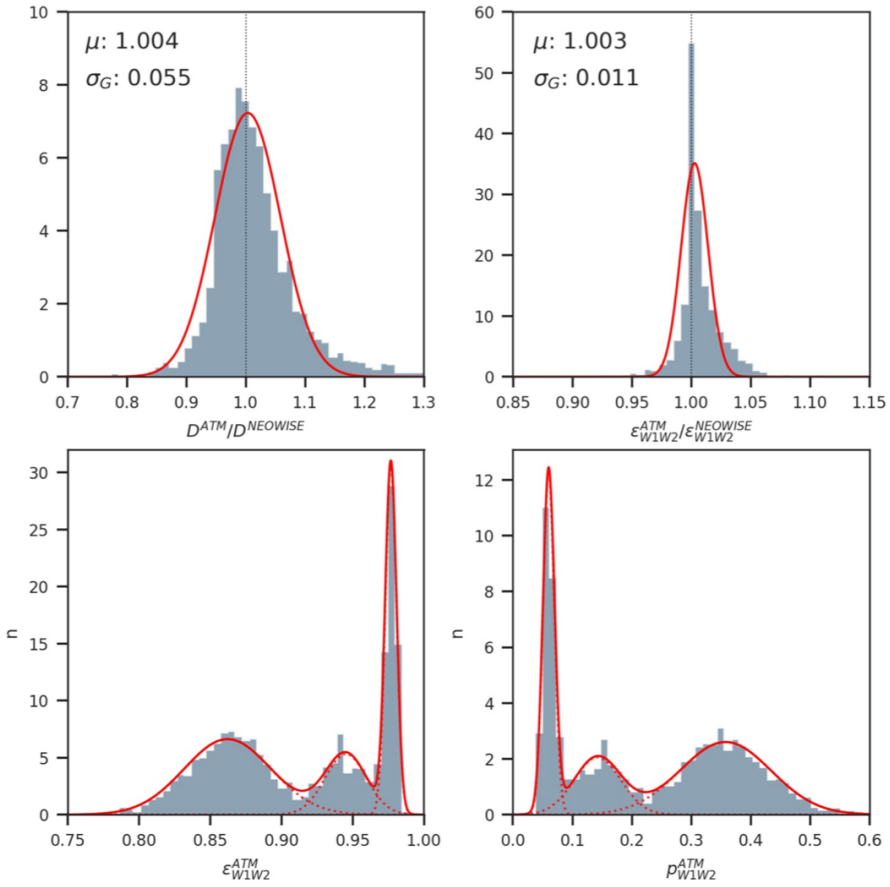

Fig. 10. Comparison of ATM best-fit parameters to results from NEOWISE. The top panels show histograms of the ratios of the best-fit diameter (left) and ϵ_{W1W2} (right) ATM and the 2016 NEOWISE release values, for 2656 asteroids with $\chi^2 < 3$ (ATM) and at least 28 observations. The red lines are single Gaussian fits, with their mean and standard deviation shown in each panel (e.g., ATM matches NEOWISE fits with a bias of 0.4% and a scatter of 5.5%, though note the distribution is leptokurtic). The bottom panels show distributions of emissivity ϵ_{W1W2} (left) and corresponding albedo (right) for the same objects. The solid lines show best-fit 3-component Gaussian mixtures (fit to individual data points rather than to histograms which are shown only for illustration). The best-fit parameters are listed in Table 2. This figure was generated using analysis.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The Best-fit Parameters for Emissivity and Albedo Distributions[a].

Quantity	fraction$_1$	μ_1	σ_1	fraction$_2$	μ_2	σ_2	fraction$_3$	μ_3	σ_3
Emissivity ϵ_{W1W2}	0.299	0.977	0.004	0.195	0.945	0.015	0.506	0.862	0.030
Albedo p_{W1W2}	0.296	0.060	0.010	0.203	0.142	0.039	0.501	0.358	0.077

[a] Best-fit 3-component Gaussian mixture shown in two bottom panels in Fig. 10.

scheme:

1. In the first step, size, temperature parameter and infrared emissivity ϵ_{W1W2} are fit for several thousand asteroids with WISE data in all four bands. The subset of these objects with direct size measurements is used to calibrate and validate NEATM model parameters and priors (not done here but implicitly used since the NEOWISE team validated their best-fit sizes using about 100 such objects, which we reproduce here with a subpercent bias).
2. In the second step, only W3 measurements are used to estimate the object's size, and the method is calibrated and validated using the four-band sample from the first step.

Hence, for studies requiring large samples of objects with size estimates (e.g., tens of thousands), the second step can provide a uniform dataset with well understood random and systematic size uncertainties from the 4-band sample, which is in turn calibrated using objects with direct size measurements.

The top left panel in Fig. 12 shows the distribution of objects in the results further below. As evident, the three sequences displayed by data distribution are related to albedo, and to its tri-modal distribution shown in Fig. 10. As was demonstrated in Fig. 11, the position along each sequence is by and large controlled by the asteroid temperature, which itself is controlled by the best-fit T_1 and asteroid-Sun distance. These relations are illustrated in the top right panel in Fig. 12. The distribution of best-fit T_1 is rather narrow and only weakly correlated with near-IR albedo, as shown in the bottom left panel in Fig. 12.

As expected from Eq. (5), T_1 does not strongly depend on asteroid-Sun distance. Given this independence and a rather narrow distribution of T_1, it is possible to derive an approximate but relatively precise predictor for the asteroid flux, given the observing geometry (asteroid-Sun distance, r, and asteroid-observer distance, Δ). We first define a "pseudo-absolute" magnitude that accounts for the dependence of observed flux on asteroid size ($\propto D^2$) and its distance from the observer ($\propto \Delta^{-2}$),

$$M = W + 5\log\left(\frac{D}{\mathrm{km}}\right) - 5\log\left(\frac{\Delta}{\mathrm{au}}\right) \quad (13)$$

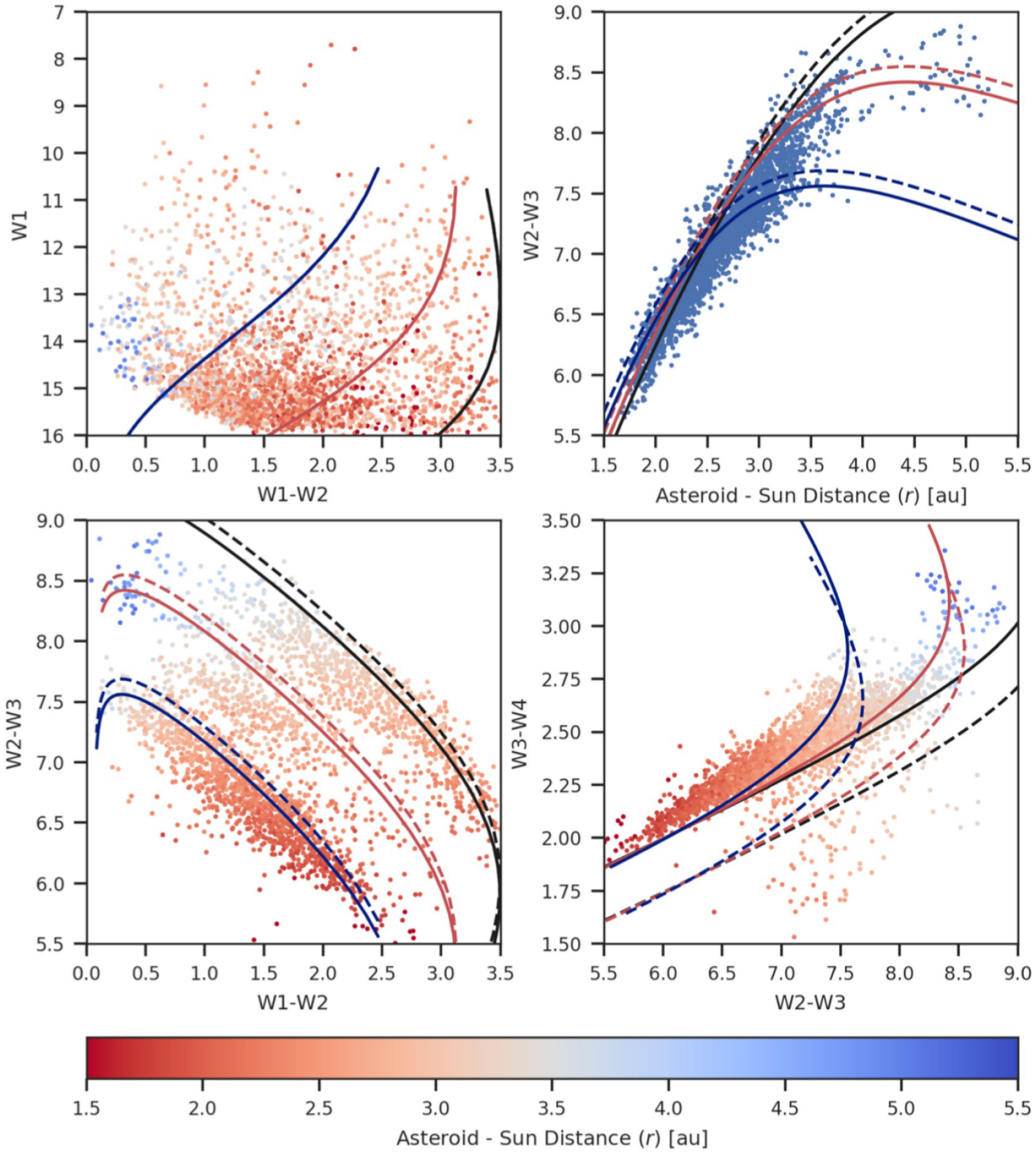

Fig. 11. Comparison of model families with WISE color-color diagrams. The symbols correspond to 2656 asteroids with $\chi^2 < 3$ (ATM) and at least 28 WISE observations (the same sample as those plotted in Fig. 10). Except for the top right panel, symbols are color-coded according to the median asteroid-Sun distance, r. The top right panel shows the variation of W2–W3 color with r. The three solid lines in each panel are model tracks produced with three fixed pairs of characteristic temperature, T_1, and emissivity across the W1 and W2 wavelength range, ϵ_{W1W2} (0.98, 406 K; 0.95, 399 K; 0.86, 391 K; for black, red and blue tracks, respectively). The position along each track is controlled by the asteroid-Sun distance (models are computed for the range 1.5–5.5 AU). Colors do not depend on the object size and distance from the observer; however, the position in the W1 vs. W1–W2 color-magnitude diagram (top left) depends on both of these quantities and model tracks assume a fiducial asteroid diameter of 10 km at a distance of 1 au. Emissivity across the W3 and W4 wavelength range, parameterized as ϵ_{W3} and ϵ_{W4}, controls the placement of model tracks in the W3–W4 vs. W2–W3 color-color diagram. The model tracks shown by solid lines are computed using $\epsilon_{W3} = 0.80$ and $\epsilon_{W4} = 0.98$. Model tracks shown by dashed lines correspond to the usually adopted values $\epsilon_{W3} = \epsilon_{W4} = 0.90$, and are strongly ruled out by the data distribution in the W3–W4 vs. W2–W3 color-color diagram. This figure was generated using analysis.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of $M3$ is rather narrow because both T_1 and ϵ_{W3} distributions are narrow.

A linear fit to $M3$ as a function of asteroid-Sun distance,

$$M3 = 0.863 \left(\frac{r}{\text{au}}\right) + 7.859. \tag{14}$$

is adequate and matches observed values with a scatter of 0.21 mag, as shown in the top left panel in Fig. 13. Using this fit and the definition of $M3$ given by Eq. (13), we derive an approximate estimator for asteroid

$$5 \log\left(\frac{D_{approx}}{\text{km}}\right) = 7.859 + 0.863 \left(\frac{r}{\text{au}}\right) - W3 + 5 \log\left(\frac{\Delta}{\text{au}}\right) \tag{15}$$

As shown in the top right panel in Fig. 13, this single-band estimator matches best-fit diameters based on data in all four bands with a scatter of only about 10%. This scatter is primarily due to scatter in unknown T_1 and ϵ_{W3} around their typical values implicitly assumed in Eq. (15). For the same reason, systematic errors are correlated with albedo, at about

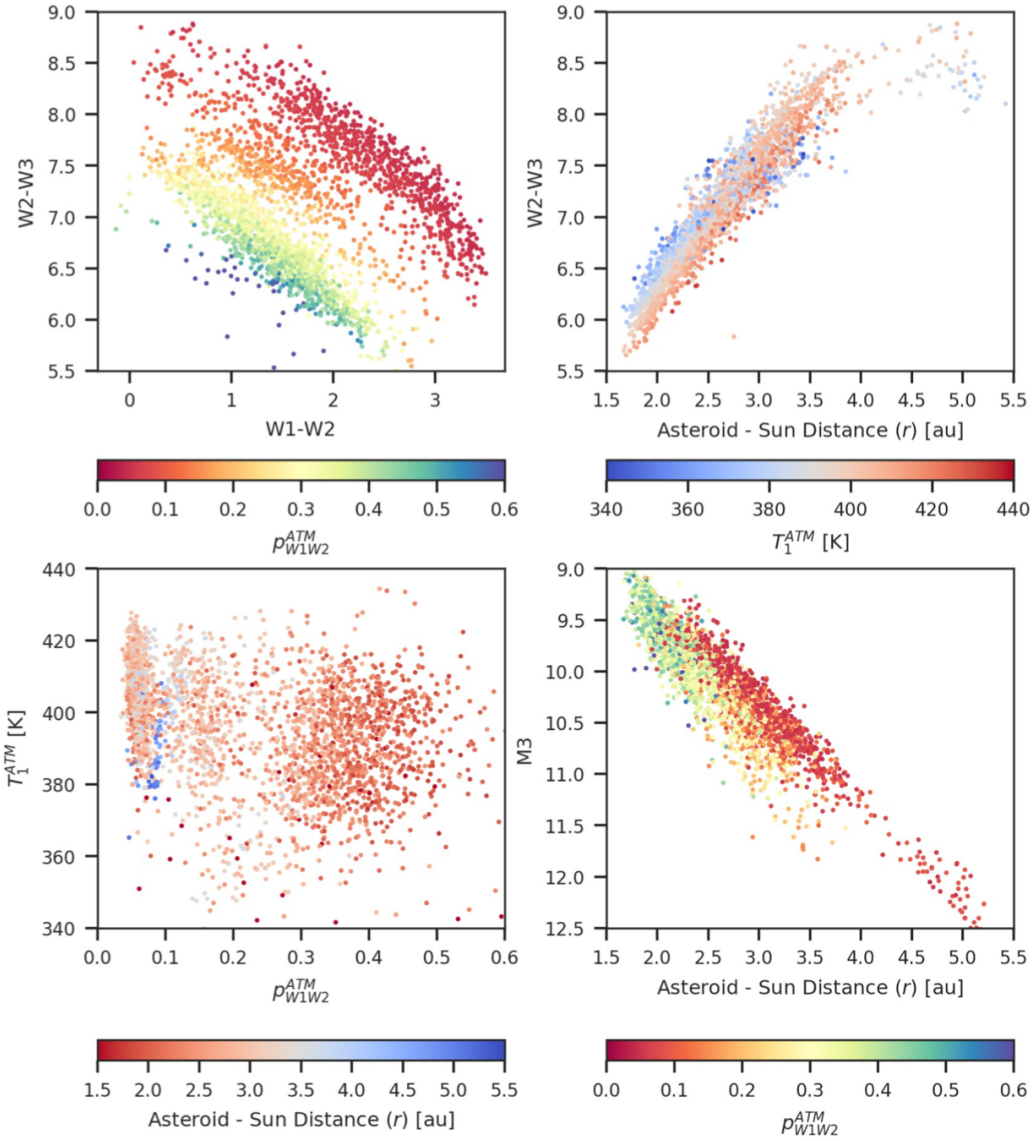

Fig. 12. Analysis of how various observables constrain model parameters. The symbols correspond to 2707 asteroids with $\chi^2 < 3$ (ATM) and at least 4 observations in each band. The top left panel shows the W2–W3 vs. W1–W2 color-color diagram, with symbols color-coded by the best-fit values of ϵ_{W1W2}. The three observed sequences clearly correspond to different taxonomic groups. The top right panel shows that the W2–W3 color is by and large controlled by the asteroid-Sun distance, r, while characteristic temperature, T_1, has only a secondary influence due to its small dynamic range. The bottom left panel shows a weak correlation between albedo (or equivalently ϵ_{W1W2}) and T_1 (the albedo distribution is tri-modal, see the bottom right panel in Fig. 10). The bottom right panel shows a correlation of "absolute" magnitude (corrected for the asteroid size and asteroid-observer distance) with r. The preponderance of low-albedo objects at large r reflects the structure of the asteroid belt. This figure was generated using analysis_W3.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

primarily that it can be applied to a much larger sample of objects than the 4-band sample discussed above. We computed the median W3 magnitude for 128,660 unique objects, that also have NEOWISE size estimates, using about 2.3 million W3 measurements. The formal size uncertainties based on scatter in observed W3 magnitudes and an intrinsic variability of 15% are in the range 2–6 % for objects with W3 < 8 and about 10% at the sample faint limit (W3∼10).

The two size estimators are compared in Fig. 14, as a function of the NEOWISE fit code. They agree on average with a scatter of about 10%, and without appreciable biases for the high data quality "DVBI" subsample (as expected, as this subsample is closely related to the training although not certain, that biases are introduced by the NEOWISE size estimator. As a possible clue, we find that the bias increases with the formal uncertainty in median W3 magnitude.

We note that because of a strong correlation between the median W3 flux and asteroid size (because of the finite dynamic range of observed distances), the biases with respect to the median W3 flux propagate to biases with respect to the object size and thus may introduce biases when estimating size distributions. Furthermore, WISE observed asteroids at a relatively small range of phase angles, and as stated in Section 2.1.1, the beaming parameter η can vary with phase angle. An increase in range of phase angles observed would create larger variations in

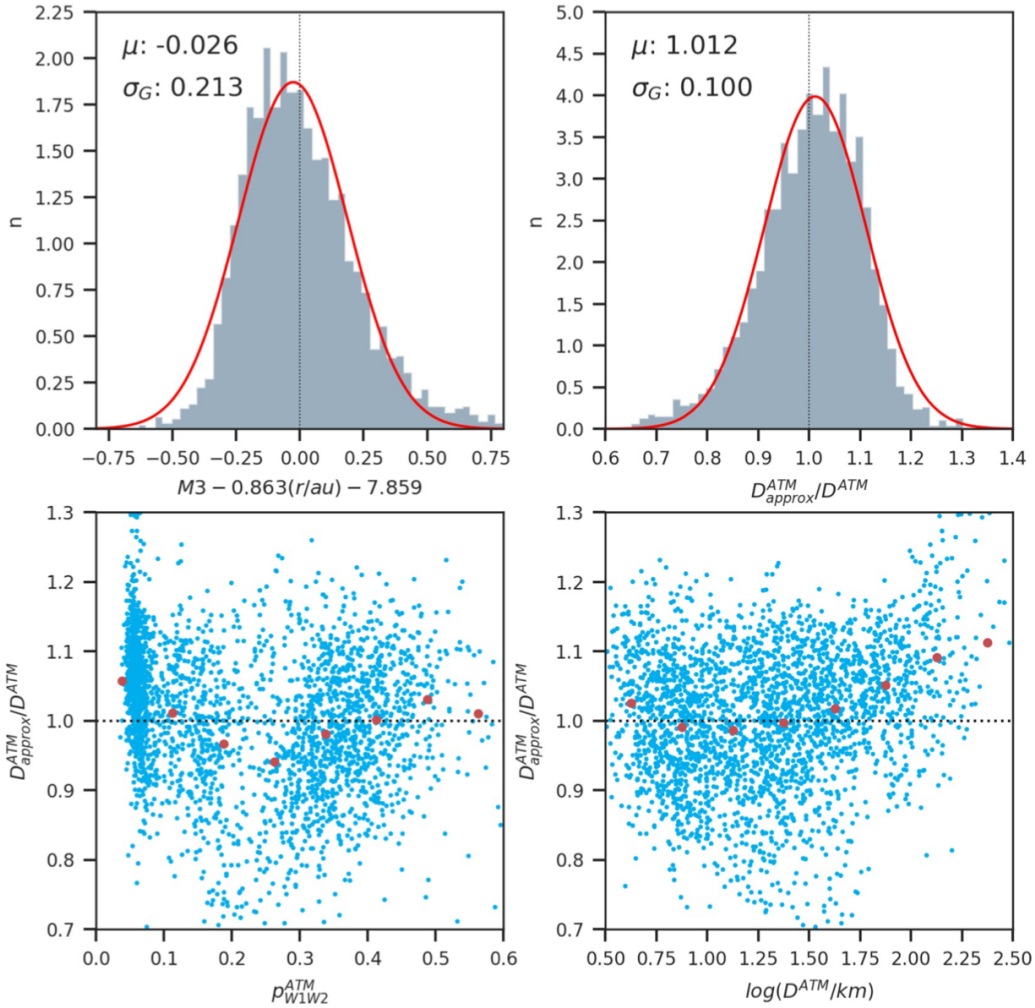

Fig. 13. Analysis of systematic uncertainty for the single-band size estimator. The top left panel shows a histogram of residuals after a straight line is fit to the "absolute" magnitude vs. the asteroid-Sun distance relationship shown in the bottom right panel in Fig. 12. The histogram in the top right panel shows the distribution of the ratio of an approximate asteroid diameter estimate based on W3 measurements alone and the best-fit ATM values based on all four WISE bands. The red lines are single Gaussian fits, with their mean and standard deviation shown in each panel. Note that single-band estimates match 4-band estimates with a scatter of 10% and a bias of 1% (this scatter is not dominated by the formal random uncertainties of single-band diameter estimates). The bottom two panels illustrate systematic uncertainty in this approximate estimate as a function of albedo and diameter. The large symbols are the median values for bins along the horizontal axis. This figure was generated using analysis_W3.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Comparison of WISE-based model parameters and SDSS data

In this section, we match the "gold" sample of 7359 best-observed asteroids from the WISE dataset (see Section 3.1 for selection criteria) to asteroids with optical observations listed in the Sloan Digital Sky Survey Moving Object Catalog (hereafter SDSSMOC; Ivezić et al., 2001, 2002, Jurić et al., 2002, Parker et al., 2008). Following Mainzer et al. (2012), we aim to study correlations between optical and infrared properties, such as colors and albedo. The presence of such correlations provides support that infrared emission models and best-fitting parameters are robust because the two datasets are essentially independent. We quantitatively explore the color vs. albedo correlation and develop an approximate method to estimate asteroid size from optical data alone. In addition, in Appendix B we discuss the selection of candidate M type asteroids using WISE-based best-fit parameters. Their optical color distribution is different than that for the whole sample and this fact

The 4th SDSSMOC release[14] lists astrometric and photometric data for 471,569 moving objects observed by SDSS prior to March 2007. Of those, 220,101 objects are linked to 104,449 unique objects with orbits. A match based on object designation to 7359 objects with WISE-based fits yields 1574 objects.

In the remainder of the analysis here, we use optical absolute magnitude, H, based on SDSS measurements (field 47 in SDSSMOC) because it represents an observationally uniform dataset, and because the values obtained from Minor Planet Center were found to have errors (both biases and random scatter) at the level of a few tenths of magnitude (for detailed discussion, see Section 2.3 in Parker et al., 2008). However, we note that the net offset between the SDSS and MPC values of H, reported by Parker et al. (2008), disappears when using the June 2018 version of the MPC catalog MPCORB.DAT, and the scatter is

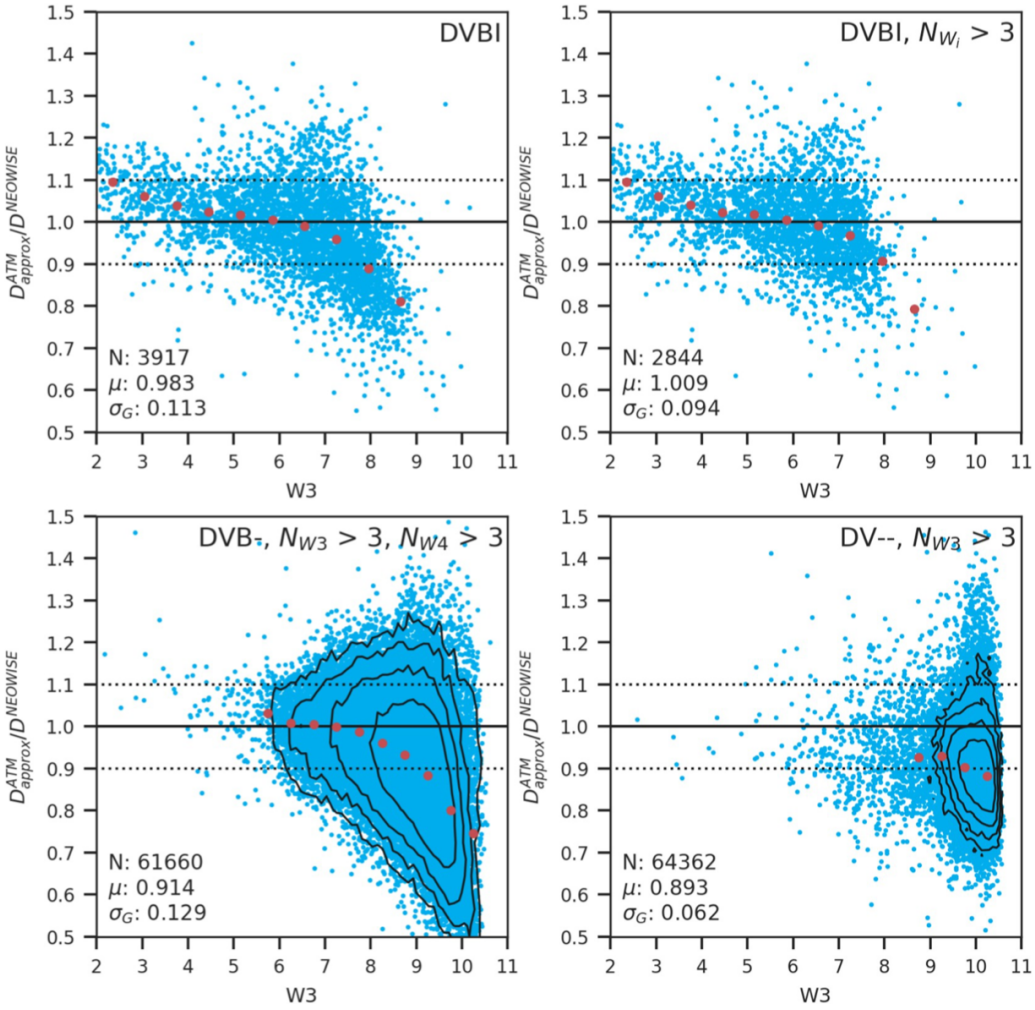

Fig. 14. Analysis of systematic uncertainty for the single-band size estimator compared to different NEOWISE thermal modeling fits. The panels show the ratio of W3-based size estimate and the 2016 NEOWISE estimate for subsamples of objects selected by their NEOWISE fitting code and the number of observations (symbols in top two panels and contours in bottom two panels due to much larger subsamples), as a function of the median observed W3 magnitude. The large symbols show the median values of size ratio in bins of W3. The three numbers in each panel list the subsample size, the median value of size ratio, and its robust (interquartile based) standard deviation. Top panels show "DVBI" sample, without (left) and with (right) a limit on the minimum number of observations in all four WISE bands. Bottom panels show "DVB-" subsample (left) and "DV–" subsample (right), with a limit on the minimum number of observations as shown next to the label. This figure was generated using analysis_W3.ipynb.

reduced to 0.22 mag.

4.1. Estimates of optical albedo from WISE-based best-fit sizes

Using SDSS-based absolute magnitude H and WISE-based best-fit size D, we estimate optical albedo using Eq. (7). This estimate implies that p_V is a free fitting parameter whose prior is decoupled from emissivity and albedo values at IR wavelengths probed by WISE. If some prior data implied a strong emissivity/albedo relationship across the entire probed wavelength range (e.g., when assuming a constant unknown value of emissivity ϵ), then a joint fit would be more appropriate and D would be constrained by both optical and infrared data. Given that our knowledge of the emissivity vs. wavelength curve for individual objects is usually poor, and that observed emissivity values span a much smaller dynamic range than albedo values, it is better to first estimate size at wavelengths where thermal emission dominates total observed flux, and then use that best-fit size to estimate albedo at wavelengths were reflected light

4.2. Optical albedo-color correlation

Following Fig. 5 from Mainzer et al. (2012), the top two panels in Fig. 15 show the SDSS a color vs. $i-z$ color-color diagram, with symbols color coded using the V band albedo from Eq. (17) (left) and IR albedo derived from best-fit ϵ_{W1W2} (right). The a color is the first principal axis for the asteroid distribution in the SDSS $r-i$ vs. $g-r$ color-color diagram, defined as (Ivezić et al., 2001)

$$a = 0.89(g-r) + 0.45(r-i) - 0.57, \tag{16}$$

and enables easy separation of C type asteroids ($a < 0$) from S type asteroids ($a > 0$).

As discussed by Mainzer et al. (2012), the addition of IR albedo improves the definition of taxonomic regions, and presumably of subsets of asteroids with different optical albedo distributions, in the optical $r-i$ vs. $g-r$ color-color diagram. Motivated by the morphology of diagrams shown in the top two panels, in addition to the $a = 0$ separator, we add

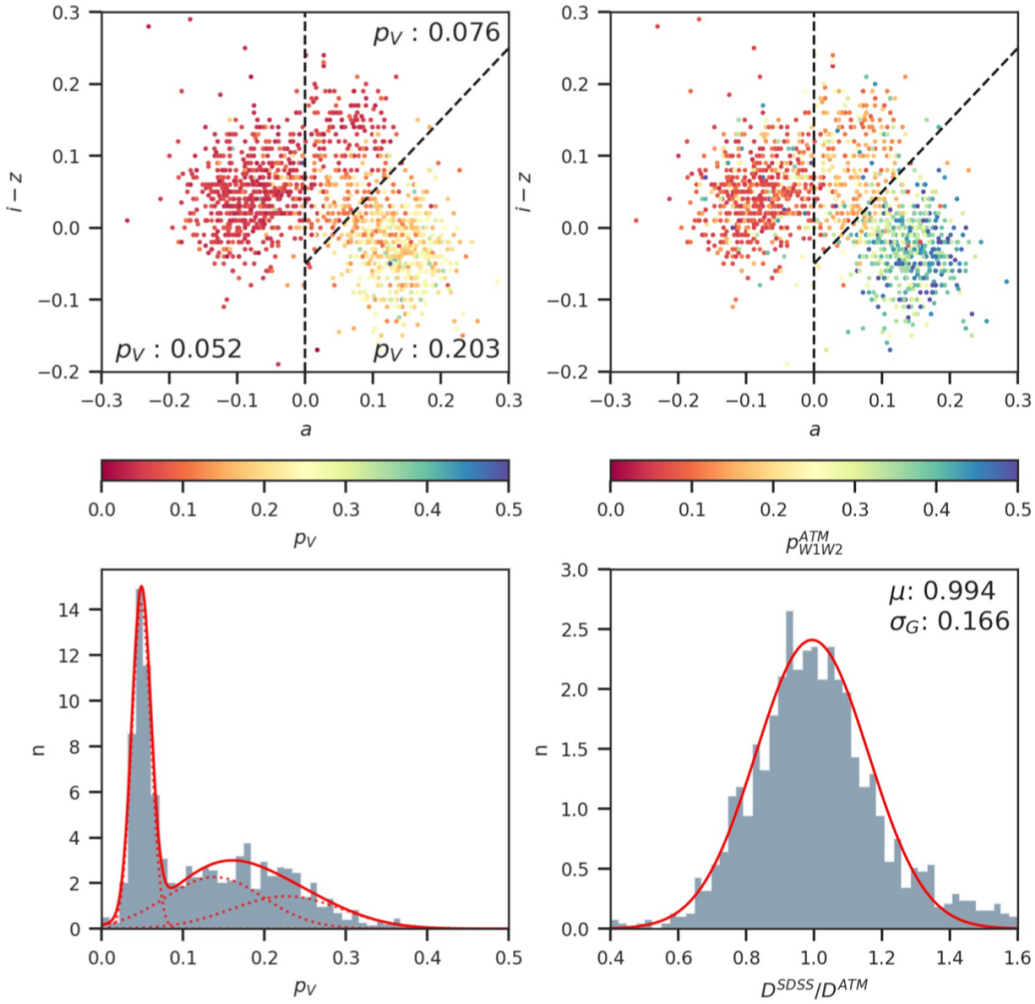

Fig. 15. Analysis of the correlations between WISE-based model parameters and optical colors measured by SDSS. The top panels show the a vs. $i - z$ SDSS color-color diagram, where a color is defined as $a = 0.89 * gr + 0.45 * ri - 0.57$ (Ivezić et al., 2001), for 1574 asteroids from SDSS MOC4 catalog that also have WISE-based diameters and IR albedo estimated with ATM. The symbols are color-coded by WISE-based IR albedo (right) and by V-band albedo (left) obtained using WISE-based diameter and SDSS-based absolute magnitude. The vertical dashed line shows the separation between C and S taxonomic classes from Ivezić et al. (2001), while the solid line is a separator between low-albedo and high-albedo objects derived here. The histogram in the bottom left panel shows the distribution of V-band albedo and the dashed lines show the best-fit 3-component Gaussian mixture (fit to individual data points). The darkest component is centered on $p_V = 0.052$. The histogram in the bottom right panel shows the distribution of the ratio of an approximate asteroid diameter estimate based on SDSS measurements alone and the best-fit ATM values based on all four WISE bands. The SDSS estimate is based on three color-assigned values of p_V, listed in the top left panel. The mean and standard deviation listed in the bottom right panel show that SDSS-based diameters match WISE-based best-fit ATM values with a bias of 0.6% and a scatter of 17%. This figure was generated using analysis_SDSS.ipynb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in Table 3.

4.3. An approximate SDSS-based size estimator

Asteroid size can be estimated from optical data alone by transforming Eq. (7) into

$$D = 1329 \text{ km } \frac{10^{-0.2H}}{\sqrt{p_V}} \tag{17}$$

Table 3
The median and robust standard deviation for the visual albedo, p_V.

Color-selected sample	Median	St.dev.[a]

and adopting a median albedo. Due to the large dynamic range of visual albedo, the scatter of such size estimates around the true values is large, about 50–60 %, and non-Gaussian. However, the separation of asteroids using optical colors into three subclasses, each with a much narrower albedo distribution than for the whole sample, greatly improves such estimates. We use the following simple algorithm to assign p_V:

$$\begin{array}{lllll} p_V = 0.052 & \text{if} & a < 0 & & \\ p_V = 0.076 & \text{if} & a > 0 & \text{and} & i - z > a - 0.05 \\ p_V = 0.203 & \text{if} & a > 0 & \text{and} & i - z < a - 0.05 \end{array} \tag{18}$$

As shown in the bottom right panel in Fig. 15, such SDSS-based size estimates match WISE-based estimates with a scatter of only 17%, and a nearly Gaussian error distribution. For reference, single-band W3-based estimates have an intrinsic precision of about 10%, so the size estimates

be at least partially due to the variability-induced scatter in single-epoch SDSS estimates of H.

5. Discussion and conclusions

Motivated by a desire to enable transparency and reproducibility of results, and to foster collaborative software development, we release ATM, a general tool for interpreting infrared flux measurements of asteroids. With adequate infrared data, this tool can be used to estimate asteroid sizes, constrain asteroid emissivities at infrared wavelengths, and when optical data are available, also estimate visual albedos. The package also includes data files and example Jupyter Notebooks that can help significantly reduce the time to reproduce published results. All the analysis presented here, including all the figures, tables, and catalogs, can be easily reproduced with these Notebooks.

We emphasize that our analysis presented here, and corresponding catalogs with best-fit sizes and other parameters, are far from definitive and can be improved in various ways. For example, an analysis of emissivity in W1 and W2 bands using Hierarchical Bayesian modeling, similar to our analysis of emissivity in W3 and W4 bands (which could also be improved by optimizing the adopted $\epsilon_{W4}/\epsilon_{W3}$ ratio), would likely further decrease systematic uncertainties. As another example, modern machine learning methods could be used to improve our simplistic algorithm for assigning visual albedo using optical colors (Eq. (18)). We leave these improvements for future studies by us and the community.

Nevertheless, our results presented here already yield a number of useful conclusions. We show that ATM can match the best-fit size estimates for best-observed objects published in 2016 by the NEOWISE team with a sub-percent bias and a scatter of only 6%. Plausible reasons for this scatter include different outlier rejection algorithms, different treatments of Kirchhoff's law, and ATM accounting for intrinsic variability, although we cannot exclude other causes. Whatever the reason, the discrepancies are encouragingly small.

Our analysis of various sources of random and systematic size uncertainties show that for the majority of over 100,000 objects with WISE-based size estimates random uncertainties (precision) are about 10% (using W3-based estimates calibrated using high-quality 4-band subsample, see Section 3.4), and systematic uncertainties within the adopted model framework, such as NEATM, are in the range 10–20 %. We estimate that the accuracy of WISE-based asteroid size estimates is in the range 15–20 % for most objects, except for unknown errors due to an inadequate modeling framework (such as spherical asteroid approximation). Of course, there is no implied guarantee of Gaussianity and these statements need to be interpreted with care. This result is consistent with the statement that accuracy is about 15% by Mainzer et al. (2016), but somewhat larger than the claim of "errors better than 10%" in Masiero et al. (2011).

The treatment of priors for emissivity $\epsilon(\lambda)$ has a direct and fundamental impact on biases in best-fit size estimates. We note that given this role of $\epsilon(\lambda)$, which is presumably shared by all members of an asteroid family, resulting systematic errors will be shared by all family members and thus cannot be seen when analyzing the per-family scatter in WISE-based optical albedo (Masiero et al., 2018a). Studies of such scatter are insensitive to systematic errors due to incorrect $\epsilon(\lambda)$ and thus cannot be used to constrain the absolute uncertainty, or accuracy, of flux-based size estimates. Only direct size measurements can enable a full understanding of the accuracy of size estimates based on thermal models.

Our ATM results faithfully recover the tri-modal distribution of ϵ_{W1W2} emissivity related to taxonomic classes discovered by Masiero et al. (2014). Correlations of SDSS colors and WISE-based best-fit model parameters indicate the robustness of the latter, and also give support to the claim that candidate metallic asteroids can be selected using best-fit temperature parameter and IR albedo (Harris and Drube, 2014).

However, it should be noted that the condition on priors $\epsilon_{W1} = \epsilon_{W2} = \epsilon_{W1W2}$, introduced because WISE data do not strongly constrain ϵ_{W2}, may not be optimal. For example, we did not investigate ansatz $\epsilon_{W2} = k\,\epsilon_{W1}$, with k different from unity, as we did for ϵ_{W3} and ϵ_{W4}. We noticed some evidence for $k < 1$ when both ϵ_{W1} and ϵ_{W2} are free fitting parameters (see Section 2.4 and Table 1). Investigation of the optimal value of k using the full high-quality sample and Hierarchical Bayes methodology (as we did for the $\epsilon_{W4}/\epsilon_{W3}$ ratio, see Section 3.3) may shed new light on the behavior of infrared emissivity/albedo (especially if attempted for taxonomic subsamples defined by optical colors).

We utilized a correlation between SDSS optical colors and optical albedo derived using WISE-based size estimates and developed a method to estimate asteroid sizes with optical data alone, with an uncertainty of about 17% relative to WISE-based size estimates. When systematic errors are included, this small difference in accuracy between IR-based and optical size estimates is further diminished. This remarkable result bodes well for future optical asteroid surveys, such as the Large Synoptic Survey Telescope (Ivezić et al., 2019), which might deliver such size estimates for over 5 million asteroids (Jones et al., 2018, and references therein).

Nevertheless, we point out that adequate infrared data are crucial for breaking the degeneracy between emissivity and asteroid size. An infrared survey with appropriately placed (at least) three bandpasses, and sensitivity to match the LSST sample, could provide a major breakthrough in our knowledge of the emissivity distribution for asteroid population. We also note that laboratory measurements of emissivity can greatly contribute to this endeavor by providing more robust priors for $\epsilon(\lambda)$. Last but not least, direct asteroid size measurements are of paramount importance for validating thermal asteroid models and quantitatively estimating their intrinsic biases, and they should be greatly encouraged and supported.

Acknowledgments

The authors thank the two anonymous reviewers for their comments that helped improve the presentation of our results. We would also like to thank Siegfried Eggl, Dino Bektešević, Lynne Jones and Mario Jurić from the University of Washington for their comments and expert advice.

We thank the organizers of an asteroid modeling workshop at the University of Washington (October 3–4, 2016), where this project was conceived: LSST Corporation, NASA NEO Office and the B612 Foundation.

J. Moeyens and Ž. Ivezić acknowledge support from the University of Washington College of Arts and Sciences, Department of Astronomy, and the DIRAC Institute. The DIRAC Institute is supported through generous gifts from the Charles and Lisa Simonyi Fund for Arts and Sciences, and the Washington Research Foundation. J. Moeyens thanks the LSST Corporation Data Science Fellowship Program, his time as a Fellow has benefited this work.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project 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 work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Software: pymc3 (Salvatier et al., 2016), numpy (Oliphant, 2006), scipy (Jones et al., 2001), pandas (McKinney, 2010), astropy (Astropy Collaboration et al., 2013, 2018), matplotlib (Hunter, 2007), corner (Foreman-Mackey, 2016), seaborn (Waskom et al., 2018), astroML (VanderPlas et al., 2012)

Appendix A. A correction to the quadrature formula for the W3 band

Wright (2013) has derived simple quadrature formulae that can be used to compute in-band fluxes for the four WISE bands from model flux $F_\nu^{ast}(\lambda)$. It appears that the provided coefficients for the W3 band can be improved (N. Myhrvold, in prep.). The corrected coefficients are available in the ATM package, in method bandpassLambda that can be found in file atm/obs/wise.py. We validated the new coefficients by comparing the approximate integral obtained using the quadrature formula to the exactly integrated flux for a $T = 100$ K blackbody. The new coefficients match the exact integral to better than 0.2%, while the original coefficients result in a 35% smaller flux. We have verified that the original coefficients for the other three bands match exact integrals to sub-percent accuracy.

Appendix B. Selection of M type asteroids using WISE-based best-fit parameters

The joint analysis of optical and infrared properties discussed in Section 4 is focused on objects with "typical" properties. It shows a good correlation between WISE-based best-fit infrared albedo and optical colors measured by SDSS. Such correlations provide support that infrared emission models and best-fitting parameters are robust because the two datasets are essentially independent. Given this independence, we can also improve our understanding of outliers in each dataset. As discussed below, we select a judicious subsample of outliers using only IR parameters and show that their optical color distribution is different than that for the whole sample. This fact further demonstrates that infrared best-fit model parameters are robust – if instead IR outliers were random measurement or modeling failures, their optical color distribution would not differ from that for the whole sample.

Harris and Drube (2014) argued that the best-fit IR albedo (p_{IR} in the WISE context, p_{W1W2} in the ATM case) and beaming parameter (η) can be used to select metallic asteroids (M taxonomic type). Their main argument is that objects with high radar albedo values, indicative of metallic objects, display a very narrow distribution of IR albedo ($p_{W1W2}\sim0.2$), while a larger fraction of objects with unusually high beaming parameter values are seen in the same albedo range. Therefore, objects with large η and $p_{W1W2}\sim0.2$ are good candidates for metallic asteroids. Since WISE data are available for orders of magnitude more objects than radar observations, and metallic objects are interesting in many ways (for discussion see Harris and Drube, 2014), it is prudent to critically examine this method.

Due to degeneracy discussed in Section 2.1.1, ATM fits only for temperature parameter T_1 and not for η. The relationship between T_1 and η is given by Eq. (5); high η corresponds to low T_1. Therefore, an implication of analysis from Harris and Drube (2014) is that low T_1 objects with $p_{W1W2}\sim0.2$ are good candidates for metallic asteroids. We now examine whether IR data discussed here suggest that such outliers exist, and if so, whether they have distinct optical colors.

The right panel in Fig. 16 shows that the SDSS-WISE sample does contain objects at the low end of T_1 range that have $p_{W1W2}\sim0.2$. By requiring $340 < T_1/K < 370$ and $0.1 < p_{W1W2} < 0.20$, we select 13 objects out of 1574 objects in the SDSS-WISE sample (there are 32 selected objects out of 2479 objects in the high-quality WISE sample), or about 1% of the sample. These candidates for metallic asteroids are listed in Table 4.

Table 4
Candidates for M type (metallic) asteroids[a].

Designation	$g-r$	$r-i$	$logD$	σ_{logD}	$logT_1$	σ_{logT_1}	p_{W1W2}	$\sigma_{p_{W1W2}}$	Tax. T.	Tax. B-D.
(497)	0.48	0.15	4.717	0.024	2.542	0.009	0.141	0.015	M	–
(844)	0.50	0.19	4.715	0.029	2.541	0.011	0.146	0.019	–	X
(1349)	0.51	0.16	4.466	0.040	2.539	0.015	0.175	0.033	–	–
(1546)	0.52	0.19	4.468	0.029	2.546	0.011	0.180	0.022	–	–
(1670)	0.62	0.14	4.365	0.016	2.549	0.006	0.170	0.012	–	–
(1730)	0.58	0.13	4.213	0.025	2.538	0.010	0.166	0.018	–	Xe
(1732)	0.63	0.21	4.397	0.028	2.560	0.011	0.155	0.019	–	–
(1860)	0.52	0.20	4.270	0.012	2.554	0.004	0.130	0.009	–	X
(1977)	0.56	0.15	4.284	0.021	2.552	0.008	0.164	0.016	–	Sq
(2294)	0.61	0.15	4.206	0.017	2.557	0.006	0.158	0.014	–	–
(2573)	0.61	0.17	4.325	0.020	2.564	0.007	0.110	0.011	–	–
(2904)	0.49	0.19	4.209	0.017	2.557	0.006	0.188	0.015	–	–
(4813)	0.56	0.24	4.256	0.027	2.560	0.011	0.147	0.018	–	–

[a] $g-r$ and $r-i$ are SDSS colors. Diameter D is in meters, T_1 in Kelvin. Tholen and Bus-DeMeo (SMASSII) taxonomies drawn from the JPL Small-Body Database Browser (see: https://ssd.jpl.nasa.gov/sbdb.cgi).

As the two-dimensional color scheme[15] in Fig. 16 illustrates, the IR-selected objects have significantly different optical a color distribution than the full sample: the a color mean and standard deviation for selected objects are -0.02 and 0.027, respectively (see the left panel in Fig. 16). This difference demonstrates that these objects are not random outliers in the p_{W1W2} vs. T_1 diagram and provides support to the hypothesis about metallic asteroids advanced by Harris and Drube (2014). A Kolmogorov-Smirnoff test using the optical a colors of both the metallic asteroid candidates and the remainder of the full SDSS-WISE sample yields only a 1.7% probability that both samples are drawn from the same distribution (1.9% if the optical a color selection is limited to $a < 0.4$). Yet, the optical colors of these metallic candidates are not sufficiently unique for an efficient selection using only optical data – for example, a restrictive selection based on a and $i-z$ colors that selects only eight out of 13 WISE-selected candidates (a selection completeness of ~50%) still results in only 10% sample selection purity (that is, there are about 10 times as many other objects in the SDSS subsample selected by the cut).

Out of the 13 candidates, only five have measured taxonomies. One of which has been classified as metal (M), while three are possibly metal-like (X, Xe). There is one potential interloper with classification Sq. Given the simple selection criteria using only modeled albedo and characteristic temperature parameter T_1, to have four out of five classified objects be metal or possibly metal-like shows promise.

Infrared data are required to efficiently select candidates for metallic asteroids. As the right panel in Fig. 16 shows, T_1 selection is more restrictive than p_{W1W2} selection. Therefore, in the context of selecting metallic candidates with a hypothetical two-band survey, having W3 and W4 data would be more useful than W2 and W3 data. On the other hand, for studies requiring p_{W1W2}, the W1 or W2 band would be a more useful addition to the W3 band

than the addition of the W4 band.

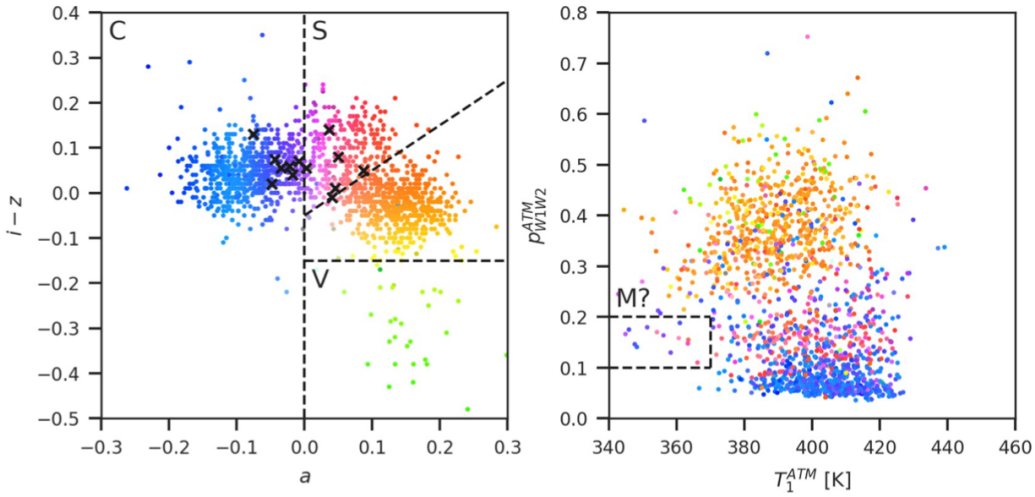

Fig. 16. Illustration of the selection method for candidate M type asteroids. The colored symbols in the left panel show the a vs. $i-z$ SDSS color-color diagram for the same objects as in Fig. 15. The symbols' color code is two-dimensional, according to a and $i-z$ colors (for algorithmic details see Ivezić et al., 2002). The same color coding is used in the right panel to visualize the correlation of optical colors and WISE-based best-fit values of IR albedo (p_{W1W2}) and temperature parameter T_1. It is easy to discern that, for example, objects with $i-z<-0.15$ have high IR albedo, while objects with $a<0$ have predominantly low IR albedo. The dashed lines in the right panel outline selection of 13 candidates for metallic asteroids. Their distribution of optical colors is different from the color distribution for the full sample, as further visualized by showing them as crosses in the left panel. The dashed lines in the left panel outline the distribution of main taxonomic classes, as marked in the panel (C, S and V). This figure was generated using analysis_SDSS.ipynb.

Appendix C. NEOWISE PDS 2016 vs. 2019

Since the processing of our results, the NEOWISE team released a newer version of their diameters and albedos for some 160,000 minor planets. In Table 5, we compare the ratios of diameters from each dataset to best-fit diameters and emissivity in bands W1 and W2 from our "gold" sample. We regenerated the top half of Fig. 10 with diameters and albedos, or equivalently, emissivities from the 2019 dataset and summarize the results in Table 5. Comparing ATM diameters to those published in 2019 still yields sub-percent bias (from +0.4% in the 2016 version to −0.4 % in the 2019 version) with the same scatter (5.5%). The ratio of emissivities improved with even smaller sub-percent bias (from 0.3% to 0.2%) and a slightly smaller scatter (from 11% to 10%).

Table 5
NEOWISE Diameters and Albedos 2016 vs. 2019.

Dataset	$D^{ATM}/D^{NEOWISE}$ [$\mu \pm \sigma_G$]	$\epsilon_{W1W2}^{ATM}/\epsilon_{W1W2}^{NEOWISE}$ [$\mu \pm \sigma_G$]	Number[b]
NEOWISE Diameters and Albedos v1.0 (2016)	1.004 ± 0.055[a]	1.003 ± 0.11[a]	2656
NEOWISE Diameters and Albedos v2.0 (2019)	0.996 ± 0.055	1.002 ± 0.10	3072

[a] These values are the same as those in the top panels of Fig. 10.
[b] We used the same selection criteria for both comparisons: NEOWISE fit code "DVBI", $\chi^2 < 3$ (ATM), and at least 28 WISE observations. The number of crossmatched minor planets in the 2019 comparison sample increased relative to its 2016 counterpart in part due to more fits being published and some designations being updated.

Appendix D. True solar spectrum vs. idealized blackbody

In the ATM modeling framework, the incident solar flux on a model asteroid is described by the Planck function (see Eq. (9)). This assumes that the Sun's spectral energy distribution can be idealized as a blackbody. We now describe the effect of using a more accurate solar spectrum on best-fit parameters for the three asteroids discussed in Section 2.4 and for several outliers with unusually strong reflected components from the "gold sample" described in Section 3.

We took the three asteroids described in Section 2.1.5 and changed their reflected fluxes by the ratio of a more accurate solar spectrum to that of an idealized blackbody spectrum. We used the best-fit diameter, characteristic temperature T_1, and emissivity from the Model 1 case in Table 1. We then calculated the NEATM flux with and without the solar component. This allowed us to isolate the thermal and reflected components. We selected the yearly-averaged solar spectrum for 2010 (when the majority of the observations in our "gold" sample were made) from Coddington et al. (2016)[16]. The ratios of this solar spectrum to that of an idealized blackbody integrated over each WISE band are 0.98, 0.88, 0.85, 0.81, in W1, W2, W3, W4, respectively. We then modified the isolated reflected components using these ratios and added them back to the emitted component to create a set of modified fluxes. Using these modified fluxes, MCMC fitting was re-run to examine the effect on the best-fit parameters for the three asteroids in question. We found that the diameters are at worst within 0.2% of previously calculated diameters.

However, the three asteroids selected do not have strong reflected components compared to their thermal emission components. We decided to find a sample of asteroids with unusually dominant reflected components, ie, outliers. We selected the fits from our "gold" sample with $\chi^2 < 3$ and then

filtered for asteroids where the difference between their median magnitude between W1 and W3 is low number using $(W1 - W3) < 3$. This is a simple way to find those asteroids whose observations suggest strong reflected fluxes. This yielded only 14 asteroids of which we randomly selected ten and then we repeated the modeling process applied to the three asteroids chosen for the earlier case study. The mean difference in diameter between those in our gold sample and those in this experiment is 1.99% with the highest difference being 3.85%. While this is a not-insignificant difference, the NEATM model itself assumes simple blackbody emission and it seems a more consistent approach to carry the same approximation to incident solar radiation by modeling the Sun as a blackbody. Another key aspect is that the approximation of the Sun as a blackbody follows NEOWISE's modeling assumptions and allows for a better comparison our results.

References

Astropy Collaboration, Robitaille, T.P., Tollerud, E.J., et al., 2013. Astronomy & Astrophysics 558, A33. https://doi.org/10.1051/0004-6361/201322068.

Astropy Collaboration, Price-Whelan, A.M., Sipőcz, B.M., et al., 2018. Astronomical Journal 156, 123. https://doi.org/10.3847/1538-3881/aabc4f.

Bowell, E., Hapke, B., Domingue, D., et al., 1989. In: Binzel, R.P., Gehrels, T., Matthews, M.S. (Eds.), Asteroids II. University of Arizona Press, Tucson, AZ, pp. 524–556.

Coddington, O., Lean, J.L., Pilewskie, P., Snow, M., Lindholm, D., 2016. Bulletin of the American Meteorological Society 97, 1265–1282. https://doi.org/10.1175/BAMS-D-14-00265.1.

Foreman-Mackey, D., 2016. J. Open Source Softw. 24. https://doi.org/10.21105/joss.00024.

Granvik, M., Virtanen, J., Oszkiewicz, D., Muinonen, K., 2015. OpenOrb: open-source asteroid orbit computation software. Astrophysics Source Code Library. http://ascl.net/1502.002.

Harris, A.W., 1998. Icarus 131, 291–301. https://doi.org/10.1006/icar.1997.5865.

Harris, A.W., Drube, L., 2014. Astrophysical Journal Letters 785, L4. https://doi.org/10.1088/2041-8205/785/1/L4.

Harris, A.W., Davies, J.K., Green, S.F., 1998. Icarus 135, 441–450. https://doi.org/10.1006/icar.1998.6002.

Hunter, J.D., 2007. Comput. Sci. Eng. 9, 90. https://doi.org/10.1109/MCSE.2007.55.

Ivezić, Ž., Connolly, A.J., VanderPlas, J.T., Gray, A., 2014. Statistics, data mining, and machine learning in astronomy: a practical Python guide for the analysis of survey data. In: Princeton Series in Modern Observational Astronomy. Princeton University Press, Princeton, NJ.

Ivezić, Ž., Tabachnik, S., Rafikov, R., et al., 2001. Astronomical Journal 122, 2749–2784. https://doi.org/10.1086/323452.

Ivezić, Ž., Lupton, R.H., Jurić, M., et al., 2002. Astronomical Journal 124, 2943–2948. https://doi.org/10.1086/344077.

Ivezić, Ž., Kahn, S.M., Tyson, J.A., et al., 2019. ApJ 873, 111. https://doi.org/10.3847/1538-4357/ab042c.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/.

Jones, R.L., Slater, C.T., Moeyens, J., et al., 2018. Icarus 303, 181. https://doi.org/10.1016/j.icarus.2017.11.033.

Jurić, M., Ivezić, Ž., Lupton, R.H., et al., 2002. Astronomical Journal 124, 1776–1787. https://doi.org/10.1086/341950.

Lebofsky, L.A., Rieke, G.H., 1979. Icarus 40, 297–308. https://doi.org/10.1016/0019-1035(79)90074-5.

Lebofsky, L.A., Spencer, J.R., 1989. In: Binzel, R.P., Gehrels, T., Matthews, M.S. (Eds.), Asteroids II. University of Arizona Press, Tucson, AZ, pp. 128–147.

Lebofsky, L.A., Sykes, M.V., Tedesco, E.F., et al., 1986. Icarus 68, 239–251. https://doi.org/10.1016/0019-1035(86)90021-7.

Mainzer, A., Bauer, J., Grav, T., et al., 2011. Astrophysical Journal 731, 53. https://doi.org/10.1088/0004-637X/731/1/53.

Mainzer, A., Masiero, J., Grav, T., et al., 2012. Astrophysical Journal 745, 7. https://doi.org/10.1088/0004-637X/745/1/7.

Mainzer, A., Usui, F., Trilling, D.E., 2015. In: Michel, P., DeMeo, F.E., Bottke, W.F. (Eds.), Asteroids IV. University of Arizona Press, Tucson, AZ, pp. 89–106.

Mainzer, A.K., Bauer, J.M., Cutri, R.M., et al., 2016. NASA Planetary Data System.

Masiero, J.R., Mainzer, A.K., Grav, T., et al., 2011. Astrophysical Journal 741, 68. https://doi.org/10.1088/0004-637X/741/2/68.

Masiero, J.R., Grav, T., Mainzer, A.K., et al., 2014. Astrophysical Journal 791, 121. https://doi.org/10.1088/0004-637X/791/2/121.

Masiero, J.R., Mainzer, A.K., Wright, E.L., 2018a. A Family-based Method of Quantifying NEOWISE Diameter Errors. Astron. J. 156, 62. https://doi.org/10.3847/1538-3881/aacbd4.

Masiero, J.R., Redwing, E., Mainzer, A.K., Bauer, J.M., Cutri, R.M., Grav, T., Kramer, E., Nugent, C.R., Sonnett, S., Wright, E.L., 2018b. Small and Nearby NEOs Observed by NEOWISE During the First Three Years of Survey: Physical Properties. Astron. J. 156, 60. https://doi.org/10.3847/1538-3881/aacce4.

McKinney, W., 2010. In: van der Walt, S., Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference, pp. 51–56.

Mommert, M., Jedicke, R., Trilling, D.E., 2018. Astronomical Journal 155, 74. https://doi.org/10.3847/1538-3881/aaa23b.

Myhrvold, N., 2018a. Icarus 303, 91–113. https://doi.org/10.1016/j.icarus.2017.12.024.

Myhrvold, N., 2018b. Icarus 314, 64–97. https://doi.org/10.1016/j.icarus.2018.05.004.

Oliphant, T.E., 2006. A Guide to NumPy, Vol. 1. Trelgol Publishing USA.

Parker, A., Ivezić, Ž., Jurić, M., et al., 2008. Icarus 198, 138–155. https://doi.org/10.1016/j.icarus.2008.07.002.

Rozitis, B., Green, S.F., 2011. Monthly Notices of the Royal Astronomical Society 415, 2042–2062. https://doi.org/10.1111/j.1365-2966.2011.18718.x.

Salvatier, J., Wiecki, T.V., Fonnesbeck, C., 2016. PeerJ Comput. Sci. 2, e55. https://doi.org/10.7717/peerj-cs.55.

VanderPlas, J., Connolly, A.J., Ivezić, Z., Gray, A., 2012. In: Proceedings of Conference on Intelligent Data Understanding (CIDU), pp. 47–54, 2012., 47-54.

Waskom, M., Botvinnik, O., O'Kane, D., et al., 2018. mwaskom/seaborn: v0.9.0 (July 2018). https://doi.org/10.5281/zenodo.1313201.

Wright, E.L., 2007. https://arxiv.org/abs/astro-ph/0703085 arXiv e-prints, astro.

Wright, E.L., 2013. American Astronomical Society Meeting Abstracts. In: American Astronomical Society Meeting Abstracts #221, Vol. 221, p. 439.05.

Wright, E.L., Eisenhardt, P.R.M., Mainzer, A.K., et al., 2010. Astronomical Journal 140, 1868–1881. https://doi.org/10.1088/0004-6256/140/6/1868.

Wright, E., Mainzer, A., Masiero, J., et al., 2018. https://arxiv.org/abs/1811.01454. ArXiv.