MNRAS **000**, 1–16 (2021) Preprint 8 July 2021 Compiled using MNRAS LaTeX style file v3.0

Modelling type 1 quasar colours in the era of Rubin and Euclid

Matthew J. Temple,[1,2]★ Paul C. Hewett[2] and Manda Banerji[3,2,4]
[1] *Núcleo de Astronomía, Universidad Diego Portales, Av. Ejército Libertador 441, Santiago 8320000, Chile*
[2] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3] *School of Physics & Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[4] *Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We construct a parametric SED model which is able to reproduce the average observed SDSS-UKIDSS-*WISE* quasar colours to within one tenth of a magnitude across a wide range of redshift ($0 < z < 5$) and luminosity ($-22 > M_i > -29$). This model is shown to provide accurate predictions for the colours of known quasars which are less luminous than those used to calibrate the model parameters, and also those at higher redshifts $z > 5$. Using a single parameter, the model encapsulates an up-to-date understanding of the intra-population variance in the rest-frame ultraviolet and optical emission lines of luminous quasars. At fixed redshift, there are systematic changes in the average quasar colours with apparent i-band magnitude, which we find to be well explained by the contribution from the host galaxy and our parametrization of the emission-line properties. By including redshift as an additional free parameter, the model could be used to provide photometric redshifts for individual objects. For the population as a whole we find that the average emission line and host galaxy contributions can be well described by simple functions of luminosity which account for the observed changes in the average quasar colours across $18.1 < i_{\rm AB} < 21.5$. We use these trends to provide predictions for quasar colours at the luminosities and redshifts which will be probed by the Rubin Observatory LSST and ESA-*Euclid* wide survey. The model code is applicable to a wide range of upcoming photometric and spectroscopic surveys, and is made publicly available.

Key words: quasars: general

1 INTRODUCTION

1.1 Wide-field extragalactic surveys in the 2020s

Over the past few decades, larger and deeper surveys of the sky have each built a more complete census of the quasars and active galactic nuclei (AGN) which populate our universe, from the 2dF QSO Redshift Survey (Boyle et al. 2000; Croom et al. 2001, 2004) through iterations of the Sloan Digital Sky Survey (SDSS; York et al. 2000; Richards et al. 2002; Schneider et al. 2010; Pâris et al. 2017; Lyke et al. 2020).

Upcoming photometric surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) and ESA's *Euclid* mission (Laureijs et al. 2011; Euclid Collaboration et al. 2019) will probe new regimes; both pushing fainter in luminosity and also providing unprecedented amounts of data on variability in the time domain. In order to identify as many quasars as possible within these rich upcoming data-sets, we first need to understand what we expect new AGN to look like in those surveys. LSST and *Euclid* will probe deeper than ever before and in so doing will detect tens of millions of new AGN. These apparently fainter AGN will be located in hitherto unpopulated regions of parameter space: both bright objects at higher redshifts and intrinsically fainter objects at high and low redshifts, whether that be due to lower mass black holes or lower accretion rates. For the lower luminosity AGN the contribution from starlight in the host galaxy to the total observed flux will be more significant.

At the same time, spectroscopic instruments such as VLT-MOONS (Maiolino et al. 2020) and 4MOST (Merloni et al. 2019) will conduct surveys of the extragalactic sky. Efficiently selecting AGN for these spectroscopic surveys is crucial in order to allow the community to address the most fundamental questions in studies of galaxy and supermassive black hole (SMBH) evolution, and to maximise the scientific return from the upcoming generation of wide-field photometric surveys.

1.2 The need for parametric SED models

The upcoming generation of wide-field surveys will probe a large dynamic range in apparent magnitude, corresponding to a broader sub-domain of the AGN luminosity function (at any given redshift) than has been sampled to date. Even at fixed redshift the population will exhibit a range of observed quasar properties. Quantifying the intra-population properties with the minimum number of parameters will be a prerequisite for many statistical investigations. Current model spectral energy distributions (SEDs) for AGN based on a single fixed template or composite spectrum will not prove adequate. Here we aim to model the SEDs of luminous, type 1 AGN (quasars) in a way that encapsulates the systematic changes in observed optical

★ E-mail: Matthew.Temple@mail.udp.cl

and near-infrared properties as a function of luminosity, as well as redshift. Notably, we chose parameters which are motivated by known astrophysics, such as the systematic changes in the emission line properties and the relative contribution of the host galaxy. The model allows parameters to be specified independently to explore the intra-population variance at fixed redshift and luminosity.

Since the structure of SMBH accretion discs is not fully understood in a quantitative sense, the majority of existing SED templates are purely empirical. The empirical mean quasar template of Elvis et al. (1994) has seen extensive utilisation in many investigations. Templates based on more recent observations, such as those of Richards et al. (2006), have further parametrized the intra-population dispersion of quasar SEDs but still focus on extended frequency coverage at relatively low resolution. Such templates will continue to see extensive use but they have only limited utility in the context of investigations using precision optical and near-infrared photometry (as will be provided by LSST and *Euclid*) to study the AGN and quasar populations over large ranges of redshift and luminosity. The majority of quasars possess prominent emission features in the rest-frame ultraviolet and optical and a significant subset also show strong absorption features. While optical and near-infrared photometric passbands used for large surveys are normally relatively broad, the majority deliberately possess sharp short- and long-wavelength transmission cutoffs. As a consequence, making accurate (few per cent) predictions for the colours of quasars as a function of redshift requires model SEDs with a resolution of at least $R \approx 500$.

A different approach, which allows investigation across the extended rest-frame wavelength range probed by a single observed-frame photometric passband over a significant redshift range, is the use of templates based on composite quasar spectra. One such template which is highly cited is the composite presented by Vanden Berk et al. (2001). However, such composite spectra are generally derived from samples with a large range in luminosity as a function of rest-frame wavelength, which means that different parts of the template are appropriate for different regions of luminosity-redshift space. In the case of the Vanden Berk composite, there is a significant contribution to the SED at wavelengths longer than ≈4000 Å from host galaxy light associated with the low-redshift AGN used to generate the rest-frame optical composite spectrum. The galaxy contribution can make up ≈40 per cent of the composite SED at long wavelengths. At the same time, the rest-frame ultraviolet region of the composite is derived from higher-redshift quasars which are significantly more luminous. It is therefore desirable to develop model SEDs which allow the luminosity and redshift to change as independent input parameters.

1.3 Previous parametric models

1.3.1 Early quasar SED models

The use of parametric models to help develop selection algorithms and calculate the completeness of quasar samples began with searches for (then) high-redshift quasars of $z \approx 4$ (Warren et al. 1991). At such redshifts, quasars were intrinsically rare and candidate selection was made challenging due to the presence of much more common contaminant populations that possessed similar photometric properties. Comparable models (Fan 1999; Fan et al. 2001) were used to calculate the completeness of the initial SDSS quasar sample (Richards et al. 2002) and extend the high-redshift searches beyond redshift six. All these models focused on predicting the observed-frame colours for wavelengths within the extended 'optical' wavelength interval, ≈3500-10 000 Å, where relatively deep broad-band photometric observations were possible. A similar model was used in an early investigation of photometric redshifts for quasars by Hatziminaoglou et al. (2000).

The availability of wide-field near-infrared surveys, including the Two-Micron All Sky Survey (2MASS; Skrutskie et al. 2006), UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007) and, more recently, *WISE* (Wright et al. 2010) extended the accessible wavelength range for relatively faint quasars, first to 2.5 μm and with *WISE* to 4.5 μm. Parametric models were extended to rest-frame wavelengths of several microns, including contributions to the quasar SED from hot dust (Maddox et al. 2008). Such models provide predictions of the optical through near-infrared colours of an accuracy sufficient to allow the estimation of photometric redshifts. Of particular note was the parametrization of the quasar 'continuum' longward of ≈10 000 Å. Observations of individual quasars (Glikman et al. 2006) and near-infrared photometry of luminous quasars (Maddox & Hewett 2006; Maddox et al. 2008) resulted in the identification of a component that can be reproduced by emission from a blackbody with temperature ≈1200 K (see Temple et al. 2021a, for a recent investigation).

1.3.2 Current state of the art

The two parametric models that have seen considerable use in the prosecution of high-redshift quasar searches are those developed by Hewett (Maddox et al. 2008, 2012) and McGreer and Fan (Fan 1999; McGreer et al. 2013). The former model was a key element in the probabilistic scheme developed by Mortlock et al. (2012) that led to the discovery of the first quasar to be observed with a redshift exceeding seven (Mortlock et al. 2011), and also formed the basis for the SED-fitting scheme used in the selection of $z > 6$ quasars from the Dark Energy Survey and VISTA Hemisphere Survey by Reed et al. (2017, 2019). Comparison of the predictions of the two models for 'mean' high luminosity quasars at significant redshifts ($z \gtrsim 2$) up to redshift $z \approx 8$ show very good agreement (Hewett and McGreer, private communication). The model parametrizations are though quite different. Taking the emission lines for example; the Hewett model uses a template quasar spectrum derived from the Francis et al. (1991) composite spectrum, extended longward of the Hα line, with just two scaling parameters to set the strength of the emission line spectrum and (separately) the Hα line, whereas the McGreer model includes the facility to specify the strengths and velocity-widths of 60 individual emission lines.

1.4 Improvements presented in this work

1.4.1 Emission line properties

Understanding, at least from an observational perspective, of the range of ultraviolet and optical SEDs of quasars has improved significantly over the last 20 years. Of particular relevance are the systematic changes in the equivalent width and kinematics of the strongest emission lines (e.g. Croom et al. 2002; Sulentic et al. 2002, 2007; Richards et al. 2011; Jensen et al. 2016; Rankine et al. 2020; Temple et al. 2020, 2021b). A significant difference in the model presented here is the ability to incorporate the systematic emission-line diversity using just a small number of parameters.

The emission line properties of quasars are fundamentally linked to the ionising SED (Kruczek et al. 2011; Krawczyk et al. 2013, 2015). It has been proposed by Baldwin (1977) that the equivalent width of strong emission lines in quasar spectra is (anti-)correlated with the intrinsic luminosity of the source. However, this correlation

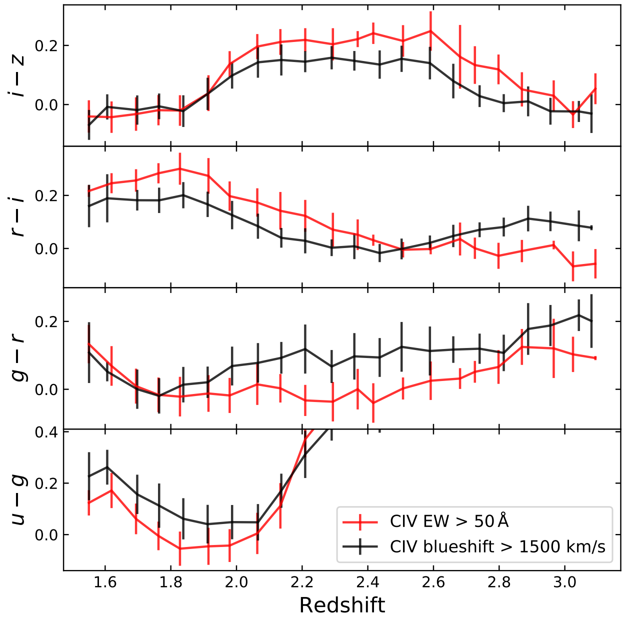

Figure 1. Median SDSS *ugriz* colours as a function of redshift. In red: objects with high equivalent width C IV λ1549 emission also have stronger Ly α emission, leading to bluer *u-g* and *g-r* colours when compared to objects with weaker, highly blueshifted line emission (in black). At redshifts $z > 2.3$, the Lyman break moves through u. Both samples have $18.6 < i_{AB} < 19.1$, and exclude BAL quasars. C IV properties are taken from Rankine et al. (2020). Error bars show the standard deviation within each bin. Even when matching in luminosity, the photometric colours can change by ~0.1 mag for quasars with different emission-line properties.

has also been ascribed as spurious and due to selection effects, with contradictory results in the literature for different samples (see section 3.1 of Sulentic et al. 2000, for a review). In reality, the emission line properties are expected to be driven by the accretion rate (i.e. Eddington fraction) of the AGN, with a secondary dependence on the black hole mass (Bachev et al. 2004; Giustini & Proga 2019). As one looks to higher redshifts, the intrinsic luminosity of a source at any given apparent magnitude increases, and so for any flux-limited sample of AGN, the average Eddington rate will increase as we move to higher redshifts. Such sources are known to generally display weaker emission lines, with the high-ionisation lines in particular showing evidence for outflows through asymmetric and blueshifted emission (see fig. 5 of Temple et al. 2021b). In Fig. 1, we show how these differences can lead to changes in colour of 0.1 mag or more, even when matching in luminosity.

1.4.2 Host galaxy contribution

A second major improvement relates to the determination of host galaxy contributions to the 'quasar' SED and the ability to isolate the pure quasar SED. The Hewett-model was developed using the luminous, $i_{AB} \leq 19.1$, SDSS DR7 quasars to determine the parameters but it is now possible to utilise the much larger sample of quasars included in SDSS DR16. Importantly, the DR16 quasar sample probes a significantly greater dynamic range in luminosity at fixed redshift. Mid-infrared photometry from *WISE*, which provides much improved constraints on the rest-frame $\approx 1\,\mu$m SED, where the relative contribution to the quasar+host SED is maximal, is now also available. As a result, it is possible to determine the relationship between quasar and host galaxy as a function of luminosity far more effectively.

1.5 Structure of this paper

In this work, we present a parametric model for the rest-frame 912 Å to 3 μm region of the quasar SED, which corresponds to the observed-frame optical and near-infrared colours. The quasar SED model builds on the model used by Hewett et al. (2006) and Maddox et al. (2008, 2012), incorporating the improvements outlined above, and is made publicly available for the first time as a PYTHON code.[1] Three examples of our model SED are shown in Fig. 2.

The model which we herein present is capable of reproducing the observed-frame optical and near-infrared colours of the majority of quasars, to a high degree of accuracy, over redshifts $0.2 \lesssim z \lesssim 7.5$. Given the SED properties, the model could be used to simulate the appearance of the quasar population in photometric surveys using either probability density functions for key parameters (e.g. Mortlock et al. 2012) or a monte-carlo sampling scheme (e.g. McGreer et al. 2013). The relatively small number of parameters needed to reproduce the properties of individual quasars means that the parameter distribution(s) for observed quasar populations can be determined though fits of the model to photometry of individual quasars. Where redshifts are not available, model fits for individual objects, with redshift as an additional free parameter, could also be used to generate photometric redshifts.

The structure of the model is described in Section 2. By cross-matching the latest SDSS quasar catalogue to UKIDSS and *WISE*, we obtain a large sample of quasars with photometric data covering the *ugrizYJHKW12* bands, which we use to calibrate our model parameters in Section 3. In Section 4, we show that our model provides good predictions for populations which are significantly fainter than the sample which has been used to calibrate the model, and also for populations at much higher redshift. Our predictions for the average quasar colours which will be seen in LSST and *Euclid* are given as a function of redshift and flux in Section 5, and we compare our model to previous work in Section 6.

All SDSS magnitudes and colours in this work are reported on the AB system (Oke & Gunn 1983). We report near-infrared UKIDSS and VIKING and mid-infrared *WISE* magnitudes and colours on the Vega system, the native magnitude system for those surveys, where Vega is assumed to have zero magnitude in all passbands. Predicted LSST and *Euclid* colours are reported on the AB system. We assume a flat ΛCDM cosmology with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, and $H_0 = 71\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All emission lines are identified with their wavelengths in vacuum in units of Ångströms.

2 THE PARAMETRIC SED MODEL

We begin by constructing a parametric SED model to describe quasar emission in the 912 Å to 3 μm rest-frame wavelength range. In Section 3 the model parameters are then calibrated using the median observed *ugrizYJHKW12* colours of the SDSS DR16 quasar population.

[1] https://github.com/MJTemple/qsogen/

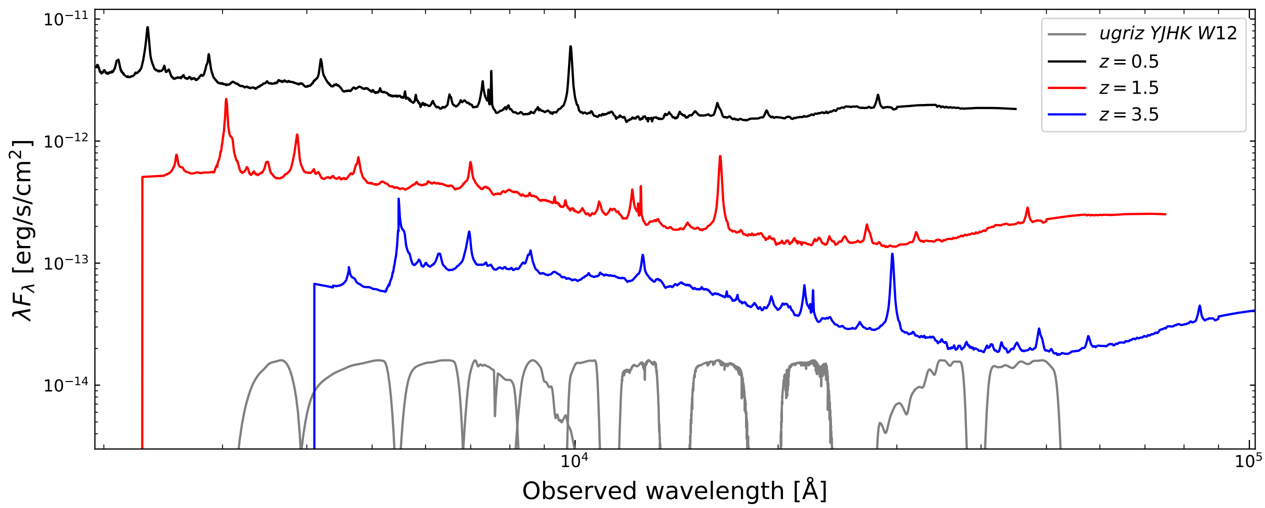

Figure 2. Examples of the model SED in the observed-frame, for three different redshifts. SDSS *ugriz*, UKIDSS *YJHK* and *WISE W12* filter response curves are shown for comparison. Higher-redshift quasars tend to be more luminous and have less contamination from starlight in their host galaxies, resulting in a sharper minimum at ≃1 μm in the rest-frame. The effect of Lyman suppression arising from incomplete transmission through the inter-galactic medium can be seen bluewards of Ly α in the $z = 3.5$ model.

2.1 Ultraviolet–optical continuum

The 900 Å to 1 μm region of a type-1 quasar SED is dominated by a blue continuum, corresponding to the low-frequency tail of the direct emission from the accretion disc. We characterise this continuum using a continuous broken power-law, with both the slopes, $\alpha_{1,2}$, and position of the break, $\lambda_{\rm break}$, able to vary.

$$f_\nu \propto \begin{cases} \nu^{\alpha_1} & \nu > c/\lambda_{\rm break} \\ \nu^{\alpha_2} & \nu < c/\lambda_{\rm break} \end{cases} \tag{1}$$

equivalently

$$f_\lambda \propto \begin{cases} \lambda^{-\alpha_1-2} & \lambda < \lambda_{\rm break} \\ \lambda^{-\alpha_2-2} & \lambda > \lambda_{\rm break} \end{cases} \tag{2}$$

At wavelengths $\lambda \leq 1200$ Å, the power-law slope is modified to become $\alpha_3 = \alpha_1 - 1$ to account for the sharp change in the continuum observed in quasar spectra (Green et al. 1980).

2.2 Hot dust

Near-infrared emission at wavelengths $1 < \lambda < 3\,\mu$m is dominated by emission from hot dust, which we characterise using a single blackbody, as described by Temple et al. (2021a, see Appendix A). The temperature, $T_{\rm BB}$, and normalisation relative to the power-law continuum at 2 μm, `bbnorm`, of this blackbody are free to vary.

$$f_\lambda = \texttt{bbnorm} \times B_\lambda(T_{\rm BB}) \times \left(\frac{k_2 \lambda^{\alpha_2}}{B_\lambda(T_{\rm BB})} \right)_{\lambda = 2\mu m} \tag{3}$$

where $B_\lambda(T)$ is the Planck function:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \tag{4}$$

2.3 Balmer continuum

We use the prescription from Grandi (1982) to describe the broad ~3000 Å 'bump' due to blended emission from high-order Balmer lines, optically thin Balmer continuum, and two-photon emission from neutral Hydrogen:

$$\begin{aligned} f_\lambda &= f_0 \times B_\lambda(T_{\rm BC})(1 - e^{-\tau_\lambda}); \qquad \lambda < \lambda_{\rm BE} \\ \tau_\lambda &= \left(\frac{\lambda}{\lambda_{\rm BE}} \right)^3 \\ f_0 &= \texttt{bcnorm} \times \left(\frac{k_2 \lambda^{\alpha_2}}{B_\lambda(T_{\rm BC})(1 - e^{-\tau_\lambda})} \right)_{\lambda = 3000Å} \end{aligned} \tag{5}$$

where $T_{\rm BC} = 15\,000$ K is the electron temperature of the Balmer continuum, $\lambda_{\rm BE} = 3646$ Å is the wavelength of the Balmer edge, and τ_λ is the optical depth of the Balmer continuum, fixed such that the continuum is optically thick at the edge, i.e. $\tau_{\lambda_{\rm BE}} = 1$. This function is then broadened via convolution with a Gaussian, with a full width at half maximum of 5000 km s^{-1}, to approximate the kinematics of the Balmer-emitting gas in broad line AGN. The normalisation relative to the power-law continuum at 3000 Å, `bcnorm`, is the only parameter free to vary.

2.4 Emission lines

To reproduce the median quasar colours across a wide range of redshifts and intrinsic luminosities, we construct two emission line templates which span the range of observed quasar SEDs (Fig. 3).

We start by defining samples of quasars with spectra from SDSS DR7. DR7 spectra are used in preference to BOSS/eBOSS spectra due to the far more accurate spectrophotometry achieved for the DR7 spectra (Lee et al. 2013; Margala et al. 2016), meaning the intrinsic large-scale shape of each spectrum is retained. The DR7 spectra employed are those without evidence of broad absorption lines, with signal-to-noise ratio (S/N) per pixel ≥ 8 and only a small number of bad pixels. Two samples of quasars at redshifts $z \simeq 2.2$ and $z \simeq 1.2$ provide the information to define the emission line templates in the rest-frame ultraviolet. Specifically, the samples are chosen to display extreme morphologies in the complex of emission lines at ≃1908 Å, which is known to correlate with the blueshift of the C IV emission line (Richards et al. 2011; Rankine et al. 2020; Temple et al. 2020).

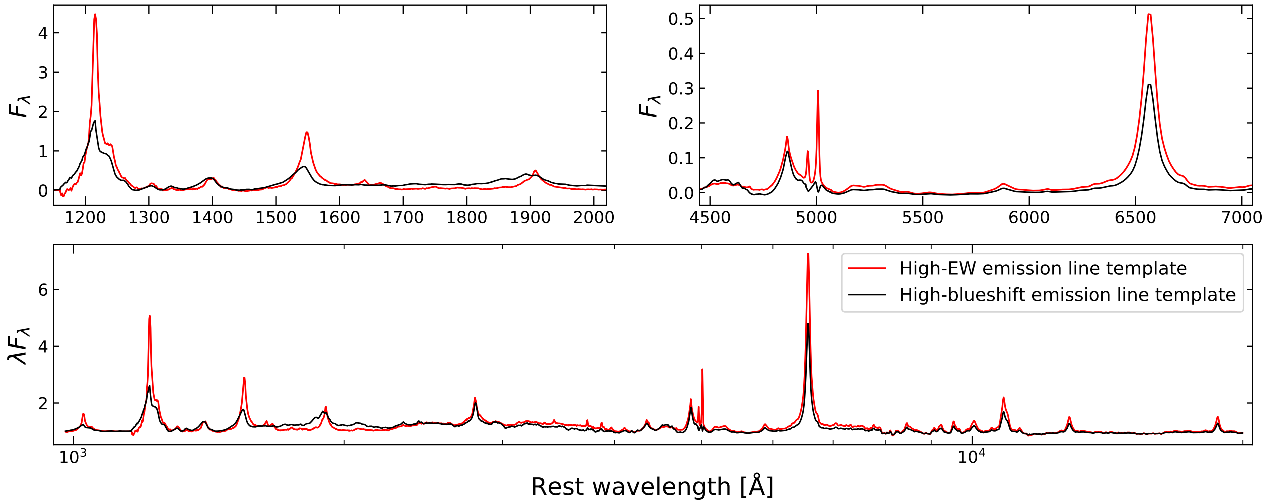

Figure 3. The high blueshift and high equivalent width emission line templates implemented in the quasar SED model. Quasars with higher luminosities and larger Eddington fractions are known to be more likely to display emission lines which are weaker and display larger shifts to the blue in high-ionization ultraviolet transitions (Temple et al. 2021b). Such objects are known to show weaker optical line emission in both narrow [O III] and the broad Balmer series (Fig. 4).

The effect of the Ly α forest on quasar spectra at redshift $z \approx 2.2$ is small, but, nevertheless, the 1050-1215 Å region in each composite is corrected for incomplete transmission through the inter-galactic medium using the prescription of Becker et al. (2013). The composite is then extrapolated down to 970 Å assuming the Ly β line is one-sixth the strength of Ly α. High S/N composite spectra covering 970-3000 and 1850-4000 Å respectively were then generated for each λ1908 morphology, giving an emission line template covering 970-4000 Å.

A third sample of 1666 quasars with redshifts $0.73 \le z \le 0.80$ was defined and a mean-field independent component analysis (MFICA) (Højen-Sørensen et al. 2002; Allen et al. 2013), covering the rest-frame wavelength interval $2200 \le \lambda \le 5100$ Å, performed. MFICA reconstructions that reproduced the 2200 – 4000 Å spectra of the 'extreme' composites generated from the $z \simeq 1.2$ quasar sample then allowed the definition of extreme emission-line spectra redward to the Hβ and [O III] lines. Confidence in the reliability of the resulting emission-line spectra, covering $970 \le \lambda \le 5100$ Å, comes from the reproduction of the empirical correlations between the equivalent widths and blueshifts of the C IV and [O III] emission lines found by Coatman et al. (2019), as shown in Fig. 4.

Finally, we include the H α and Paschen lines using the template of Glikman et al. (2006). These lines are scaled such that the strength of the H α emission line matches the known correlation between the equivalent widths of C IV and H α, using the catalogue of Coatman et al. (2017, 2019). This gives a set of emission-line templates, all covering 970 Å to 2 μm, which span the known diversity of UV-optical emission-line properties. These are shown in Fig. 3.

Within the SED model, the equivalent width of each line in the template is preserved as the continuum changes, subject to four scaling factors: `scal_em` re-scales the total emission line template, `scal_ha` re-scales the Hα and near-infrared region to account for any differences in the continuum subtraction prior to combining the templates, `scal_lya` re-scales the region shortwards of 1350 Å and `scal_nlr` rescales the narrow optical emission lines. We found that preserving the emission line equivalent widths resulted in a better representation of the colours, but the model code also has the ability to instead preserve the relative emission line fluxes.

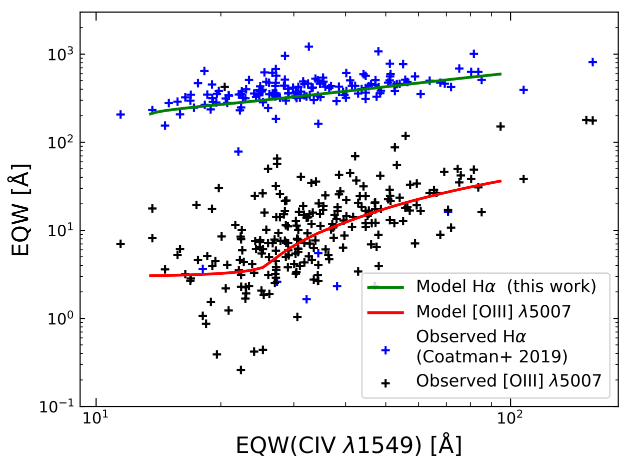

Figure 4. The equivalent widths of the rest-frame ultraviolet C IV λ1549 emission line (Rankine et al. 2020), and the rest-frame optical [O III] λ5007 and H α λ6565 lines (Coatman et al. 2017, 2019). Crosses show measurements in individual objects; the strengths of both broad and narrow optical lines display positive correlations with the strength of C IV emission. The solid lines show how our SED model reproduces these trends by using a single parameter to describe the emission-line properties across the ultraviolet-optical wavelength range.

To implement the Baldwin effect (Section 1.4.1), we introduce a parameter `emline_type` which allows for interpolation between the two extrema described above. The value of this parameter at any given redshift is controlled by the average absolute magnitude M_i at that redshift together with the multiplicative factor `beslope`:

$$\texttt{emline\_type} = \texttt{beslope} \times (M_i(z) + 27) \quad (6)$$

where the zeropoint is chosen to match the average absolute magnitude $M_i = -27$ for the median $\lambda 1908$ complex at $z \simeq 2$.

2.5 Host galaxy flux

We include the facility to incorporate emission from host galaxies. For the luminous quasars we consider, the primary effect of starlight from the host galaxy is to add flux at and around the 1 μm minimum in the total quasar SED. For optically unobscured, blue quasars such as those from SDSS, the strength of emission from any young stellar population is very hard to quantify as the quasar power-law continuum will be much brighter than the host galaxy at shorter wavelengths. It is therefore hard to quantify any change in shape of the host galaxy SED across the luminosity and redshift range we consider. We use an S0 template from the SWIRE library (Polletta et al. 2007; Rowan-Robinson et al. 2008), having found that this choice of template provides the best fit to the SDSS-UKIDSS-unWISE colours of known quasars in Section 3. The galaxy contribution to the median quasar plus host SED will be a combination of a significant fraction of quasars with hosts dominated by old stellar populations and other objects with hosts that have experienced different star-formation histories. The selection of the S0 host is thus not surprising.

Modelling the spectra and colours of individual quasars and the population of fainter AGNs will necessarily involve incorporating different host galaxy SEDs. Our model code is capable of utilising any specified galaxy SED and a specific application is the estimation of photometric redshifts for quasars and AGN utilising a range of host-galaxy SEDs (Section 6). Here, however, we focus on reproducing the median colours of the SDSS quasar population and employ a single host-galaxy SED.

The strength of host galaxy emission is incorporated into the model with two parameters: `fragal`, the fraction of total flux in the wavelength region 4000-5000 Å due to the galaxy for an object at a reference luminosity L_0, and `gplind`, the power-law index controlling how the luminosity of the galaxy component changes as a function of the quasar luminosity:

$$\frac{L_{\text{galaxy}}}{L_0} = \frac{\texttt{fragal}}{1-\texttt{fragal}}\left(\frac{L_{\text{quasar}}}{L_0}\right)^{\texttt{gplind}} \tag{7}$$

For values of `gplind` which lie between 0 and 1, the host galaxy component gets brighter as the quasar gets brighter (e.g. as one moves to higher redshifts within a flux-limited sample), but the fractional contribution of the galaxy to the total flux decreases.

The luminosities used are the absolute i-band magnitudes at redshift 2, derived using the K-correction from Richards et al. (2006) and as reported in the Lyke et al. (2020) catalogue. For each flux cut that is used, the median M_i is calculated in each redshift bin, giving an empirical redshift-luminosity relation for each sample which is then assumed when calculating the galaxy normalisation at any given redshift (see Fig. 11). The reference luminosity L_0 is taken to be $M_i = -23$. This is close to the average luminosity of SDSS DR16 quasars with $18.6 < i_{\text{AB}} < 19.1$ at $z = 0.35$.

2.6 Dust reddening

Attenuation due to dust at the redshift of the quasar using an empirically derived extinction curve can be incorporated in the model. The user can specify the form of the extinction curve by providing an ascii file specifying the value of $E(\lambda - V)/E(B - V)$ as a function of wavelength λ. Our own specification of the extinction curve appropriate for quasar samples (Fig. 5) is very similar to that presented by Gallerani et al. (2010). The curve rises more steeply with decreasing wavelength than the Milky Way curve, has no 2200 Å feature and is similar to the extinction curve of the Small Magellanic Cloud (SMC) for wavelengths $\gtrsim$ 1700 Å. Shortward of 1700 Å, however, the extinction increases significantly less rapidly than for the SMC or, indeed, the LMC and Milky Way curves. In other words the form of the curve shortward of ≈1700 Å is greyer than observed in the Local Group.

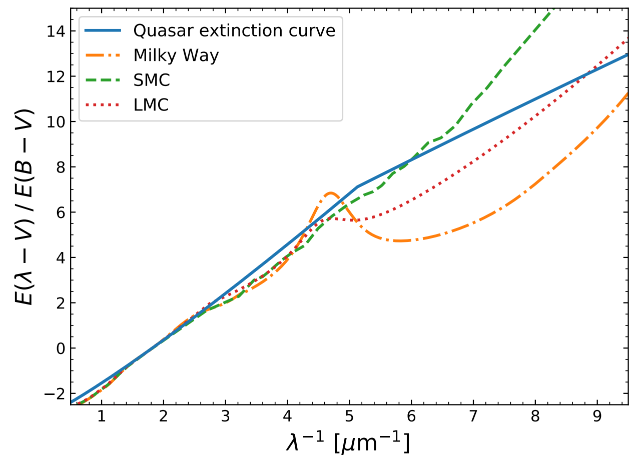

Figure 5. The quasar extinction curve included in the model, compared to commonly-used curves from our own Galaxy and the Small and Large Magellanic Clouds. The quasar extinction curve has no feature at 2200 Å ($\lambda^{-1} = 4.5\,\mu$m^{-1}), and rises less steeply than the SMC extinction curve at wavelengths shorter than 1700 Å ($\lambda^{-1} > 6\,\mu$m^{-1}).

The extinction curve was calculated using i- and K-band photometry of SDSS DR7 quasars with $2.0 < z < 3.0$ to calculate the rest-frame optical (3000-6500 Å) reddening, $E(B - V)$. Composite SDSS quasar spectra for samples with different optical reddenings covering the rest-frame wavelengths 1100-2500 Å then enabled the ultraviolet extinction curve to be defined. While not explicitly published previously the extinction curve has been used extensively in our earlier work (e.g. Maddox et al. 2008; Allen et al. 2011; Maddox et al. 2012). Adopting an extinction curve as steep as local examples at $\lambda < 1700$ Å would be inconsistent with numerous of our previous results. Figure 22 of Allen et al. (2011) provides an example, where the extinction curve provides an excellent match to the reddening of broad absorption line quasars over the redshift range $1.7 < z < 4.0$.

The quasar flux, excluding the host galaxy component, can be reddened using this quasar extinction curve with the $E(B - V)$ as a free parameter. The colour distribution of SDSS quasars is largely Gaussian, with only a small tail due to reddening (Richards et al. 2003; Hopkins et al. 2004). We apply an iterative sigma-clipping procedure to exclude this tail when calculating the median colour in any given redshift bin. The observed colours we report are therefore assumed to be not significantly affected by extinction, and the $E(B - V)$ is held fixed at zero when calibrating the other model parameters.

2.7 Lyman-absorption suppression

We have the ability to include a Lyman-limit system (LLS), which has the effect of setting the model flux to zero at all wavelengths below some cut-off. The default cut-off wavelength is $\lambda_{\text{LLS}} = 912$ Å.

At redshifts $z > 1.4$, flux at $\lambda < 1216$ Å is suppressed to account for the incomplete transmission through the inter-galactic medium of flux shortwards of Lyα. The effect of this suppression on the average colours of a sample of quasars is relatively well-understood, however we note that individual objects can display very different amounts of Lyman transmission depending on their individual sight-lines. We use the prescription of Becker et al. (2013), which has been calibrated using the results of Faucher-Giguère et al. (2008):

$$\tau_{\mathrm{eff,\,Ly}\alpha}(z) = 0.751\left(\frac{1+z}{1+3.5}\right)^{2.90} - 0.132 \tag{8}$$

The suppression due to Lyβ and Lyγ is included, assuming the theoretical ratio of oscillator strengths (e.g. Keating et al. 2020):

$$\begin{aligned}\tau_{\mathrm{eff,\,Ly}\beta}(z) &= 0.160 \times \tau_{\mathrm{eff,\,Ly}\alpha}(z)\\ \tau_{\mathrm{eff,\,Ly}\gamma}(z) &= 0.056 \times \tau_{\mathrm{eff,\,Ly}\alpha}(z)\end{aligned} \tag{9}$$

Finally, the SED is redshifted into the observed-frame and multiplied by the passband filter responses to obtain synthetic photometry.

3 CALIBRATING THE MODEL PARAMETERS

In the previous section, we described the construction of a parametric quasar SED model. Before using this model to provide predictions for future surveys, in this section we first calibrate the model parameters with the average colours of known quasar populations which have been binned in redshift and flux.

3.1 Data

We make use of the SDSS sixteenth data release (DR16) quasar catalogue, the selection of which is summarised in Lyke et al. (2020). As well as being the final data release from SDSS-IV, the DR16 catalogue constitutes the largest sample of spectroscopically confirmed quasars currently available. We use the 'primary' redshifts from the DR16 catalogue, although we take care to exclude outliers when computing the median colour in each redshift bin (Section 3.2), and our results would be unchanged if we instead used the 'PCA' redshift.

Starting with the 750 414 objects in the DR16 'quasar-only catalogue' (Lyke et al. 2020), we exclude 82 objects which have been identified by Flesch (2021) as being non-quasar contaminants. The number of quasars in the catalogue with redshifts $z > 5$ is very low, and their pipeline redshifts are highly uncertain, so we also exclude these objects from the catalogue. Finally, we exclude objects with Galactic extinction $E(B-V) > 0.3$. This yields a total of 748 620 quasars with redshifts $0 < z < 5$.

3.1.1 SDSS photometry

We use *ugriz* point spread function magnitudes from the SDSS quasar catalogue. SDSS reports magnitudes on the 'asinh' system (Lupton et al. 1999). We only use objects with $i_{\mathrm{AB}} < 20.6$, where the majority of SDSS quasars have $z_{\mathrm{AB}} < 20.3$ and their asinh magnitudes and logarithmic Pogson magnitudes differ by less than one per cent in flux. We also exclude u at redshifts $z > 2.3$ and g at $z > 3.0$ where quasars have little flux as they 'drop-out' of these bands. The asinh magnitude is therefore well-approximated by the Pogson magnitude and SDSS zeropoints are converted to the AB standard using

$$\begin{aligned}u_{\mathrm{AB}} &= u_{\mathrm{SDSS}} - 0.04\\ z_{\mathrm{AB}} &= z_{\mathrm{SDSS}} + 0.02\end{aligned} \tag{10}$$

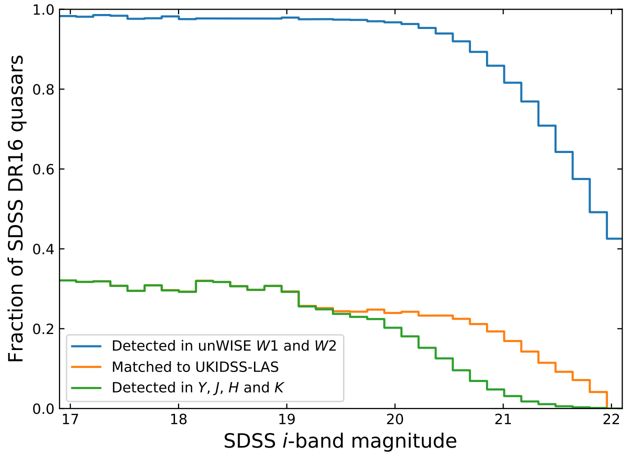

Figure 6. The completeness of the SDSS DR16 quasar catalogue in UKIDSS-LAS DR11plus and unWISE. unWISE covers the whole sky, and provides *W1* and *W2* data for 98 per cent of DR16 quasars with $i < 20.1$. UKIDSS-LAS covers around three tenths of the SDSS footprint. Within that area, almost all DR16 quasars with $i_{\mathrm{AB}} < 19.1$ are detected in all four *YJHK* bands, but the number of matches begins to drop above $i \simeq 19.5$.

The SDSS *gri* zeropoints are believed to match the AB system.[2]

3.1.2 WISE photometry

We cross-match the SDSS DR16 quasar catalogue to the deepest source of mid-infrared *WISE* (Wright et al. 2010) data which is currently available: the unWISE catalogue presented by Schlafly et al. (2019), which makes use of the Meisner et al. (2019) coadds. unWISE incorporates additional data from the reactivation of the satellite as NEOWISE, giving ≈0.7 magnitudes deeper coverage in *W1* and *W2* compared to AllWISE. The Meisner et al. (2019) coadds are also deeper than the Lang (2014) images which were force-photometered to produce the *WISE* measurements reported in the Lyke et al. (2020) catalogue. The unWISE catalogue is matched to SDSS using a 3.0 arcsec matching radius, keeping only sources with unique matches within that radius. unWISE models the *W1* and *W2* pixel data separately, and to reduce the number of contaminants we keep only sources where there is a detection in both *W1* and *W2*. We find 650 551 unWISE sources match to quasars from SDSS DR16. As shown in Fig. 6, this includes more than 98 per cent of all quasars with $i_{\mathrm{AB}} < 20.1$.

3.1.3 UKIDSS photometry

We cross-match the SDSS DR16 quasar catalogue to the eleventh data release of the UKIDSS Large Area Survey (UKIDSS-LAS DR11plus; Lawrence et al. 2007). UKIDSS-LAS covers almost three tenths of the SDSS footprint in the near-infrared *YJHK* bands (Hewett et al. 2006; Hodgkin et al. 2009). Compared to previous releases of UKIDSS data, DR11plus includes a slight increase in the fraction of the SDSS footprint which is covered, and a significant improvement to the *J*-band depth through the co-addition of a second epoch of imaging. We use 'apermag3' values, which are the default

[2] `https://www.sdss.org/dr16/algorithms/fluxcal#SDSStoAB`

Table 1. Passband attenuations $A_\lambda/E(B-V)$ for *ugrizYJHK* filters adopting the $R_V = 3.1$ Galactic reddening law from Fitzpatrick & Massa (2009) and a $z = 2.0$ quasar SED.

Filter	$A_\lambda/E(B-V)$
SDSS u	4.82
SDSS g	3.80
SDSS r	2.58
SDSS i	1.92
SDSS z	1.42
UKIDSS Y	1.12
UKIDSS J	0.80
UKIDSS H	0.52
UKIDSS K	0.33

point source 2.0 arcsec diameter aperture corrected magnitudes. We match 152 946 quasars from SDSS DR16 to unique sources within 0.85 arcsec in the UKIDSS-LAS DR11plus catalogue. This number of matches is slightly larger than the number of force-photometered detections given by Lyke et al. (2020), and we believe the improvement is due to our use of more recent data from UKIDSS. Of the 650 551 unWISE-DR16Q matches, 139 396 quasars are also matched to UKIDSS-LAS DR11plus. However, the limiting magnitude of UKIDSS is brighter than that of SDSS, and so the completeness of *YJHK* matching drops significantly above $i_{\rm AB} \approx 19.1$.

3.1.4 Galactic extinction

We correct the *ugrizYJHK* photometric bands for Galactic extinction using the dust maps of Schlafly et al. (2010) and Schlafly & Finkbeiner (2011). Such extinctions are typically of the order $E(B-V) < 0.1$, as most SDSS quasars lie outside the Galactic plane.

Commonly quoted passband attenuations are correct only when the source SED is similar to that of an elliptical galaxy at low redshift. A type 1 quasar SED is bluer than the SED of a typical star or galaxy, leading to subtle differences in the conversion from $E(B-V)$ to the attenuation in each observed passband. We derive our own passband attenuations using a $z = 2.0$ quasar source SED, assuming $R_V = 3.1$ and the Galactic extinction curve of Fitzpatrick & Massa (2009). The conversions we use are given in Table 1. Using source quasar SEDs in the range $0 \leq z \leq 5$ alters these values by no more than three per cent. The attenuation due to dust in the *WISE* bands is negligible for the extinctions we consider and no correction for Galactic extinction is applied to these data.

3.2 Sample definition and binning

As we expect the average quasar colour at any given redshift to change as a function of luminosity, we calibrate our model parameters using the colours of quasars which have been binned in apparent magnitude (i.e., flux) as well as in redshift. The numbers of objects contributing to each bin are given in Table 2.

First, we take all objects with SDSS-UKIDSS-unWISE *ugrizYJHKW12* photometry in the flux range $18.6 < i_{\rm AB} < 19.1$. The completeness of SDSS quasars within the UKIDSS-LAS drops significantly above $i_{\rm AB} = 19.1$, and the number of objects with $i_{\rm AB} < 18.6$ is relatively low, and so we then take all objects with SDSS-unWISE *ugrizW12* photometry in the flux ranges $17.6 < i_{\rm AB} < 18.1$ and $19.6 < i_{\rm AB} < 20.1$ (i.e. one magnitude brighter and one magnitude fainter than the UKIDSS-matched flux bin).

Table 2. The number of quasars remaining at each stage of the cross-matching of catalogues described in Section 3.1, and in construction of the three flux bins used to calibrate our model parameters in Section 3.2.

	No. of quasars
SDSS DR16 quasar catalogue, $0 < z < 5$	748 620
Matched to unWISE (*ugrizW12*)	650 551
Brighter flux bin:	
$17.6 < i_{\rm AB} < 18.1$	9903
107 redshift bins	$\simeq$93 per bin
Fainter flux bin:	
$19.6 < i_{\rm AB} < 20.1$	103 290
107 redshift bins	$\simeq$965 per bin
Matched to UKIDSS-LAS (*ugrizYJHKW12*)	139 396
Fiducial flux bin:	
$18.6 < i_{\rm AB} < 19.1$	15 203
214 redshift bins	$\simeq$71 per bin

Within each flux-redshift bin, the distribution of each colour is roughly Gaussian, with a tail to redder colours which we ascribe to dust at the quasar redshift. This tail is found to be stronger in the fainter flux bins, which is consistent with the fact that reddening is a side-effect of wavelength-dependant extinction, i.e. the flux in reddened objects is attenuated and so they also appear fainter. We therefore apply an iterative sigma-clipping procedure to exclude this tail, and any other outliers, when calculating the median colour in any given flux-redshift bin. We clip objects more than 2 sigma away from the sample median, and use the standard deviation of the resulting clipped sample as a measure of the dispersion around the median.

3.3 Fitting routine

For every colour in each of the flux-redshift bins described above, we compute the median $M_{z,f}$ and dispersion $\sigma_{z,f}$ of the sigma-clipped colour distribution. We also compute the median redshift z and median absolute magnitude M_i in each bin. For any given set of SED model parameters, the SED model is evaluated these values of z and M_i, giving a set of synthetic colours $S_{z,f}$. The loglikelihood is defined in the usual way as

$$\ell = -\frac{1}{2}\sum_{z,f}\left(\frac{(M_{z,f} - S_{z,f})^2}{\sigma^2_{z,f}} + \log(\sigma^2_{z,f})\right) \tag{11}$$

where we exclude the colours which, for a given redshift, probe rest wavelengths $\lambda < 912$ Å or $\lambda > 3\,\mu$m. In practice this means we exclude $W1 - W2$ at $z < 0.7$, $K - W1$ at $z < 0.3$, $g - r$ at $z > 3.0$ and $u - g$ at $z > 2.3$. Using flat priors on all parameters, samples are drawn from the posterior distribution using an affine-invariant ensemble Markov Chain Monte Carlo method (Goodman & Weare 2010). We use 200 walkers with 750 steps each, which are started near the least-squares solution to Eq. 11. We remove the first 250 steps from each walker, having verified that this removes the burn-in and the chains are subsequently well-mixed.

3.4 Results

In Fig. 7 we show the median sigma-clipped *ugrizYJHKW12* colours in our $18.6 < i_{\rm AB} < 19.1$ flux bins. The number density of objects is such that there are more redshift bins in the range $1 < z < 2$ than there are in $3 < z < 5$. Significant deviations in colour are seen as a function of redshift, due to strong emission lines such as Ly α and

Table 3. Free parameters in our quasar SED model (Section 2) and their values after calibrating to SDSS, UKIDSS and unWISE data in Section 3.

Description	Parameter	Value
blue power-law slope α_1	`plslp1`	-0.349
red power-law slope α_2	`plslp2`	+0.593
power-law break wavelength $\lambda_{\rm break}$ [Å]	`plbrk1`	3880
blackbody normalisation	`bbnorm`	3.96
blackbody temperature $T_{\rm BB}$ [K]	`tbb`	1240
overall emission line scaling	`scal_em`	-0.994
Baldwin Effect 'slope'	`beslope`	0.183
galaxy fraction at $M_i = -23$	`fragal`	0.244
galaxy luminosity power-law index	`gplind`	0.684

H α moving in and out of the different photometric filters. Error bars show the dispersion within every eighth bin.

In Fig. 8 we show the SDSS-unWISE *ugrizW12* colours for our brighter and fainter flux bins. Changes in the average quasar colours (at fixed redshift) are expected with varying luminosity, due to a varying fraction of host galaxy flux contribution. This is expected to be most significant around the 1 μm minimum in the quasar SED, where the optical continuum is falling but the hot dust emission is yet to start rising. At redshifts $1 < z < 2$, this minimum is moving through the unWISE *W1* passband, and in Fig. 8 we can see that there are noticeable changes in the *zW1W2* colours between our bright and faint flux bins. At rest-frame optical wavelengths, the flux due to starlight is generally going to be redder than the quasar continuum, so a larger host galaxy contribution has the effect of reddening the *ugriz* colours at lower redshifts $z < 1$. Finally, the equivalent width of the ultraviolet emission lines is expected to increase as the luminosity of the quasars decreases, the so-called Baldwin effect, leading to slightly larger-amplitude wiggles in the *ugriz* colour-redshift tracks at higher redshifts $z > 1$.

In Figs. 7 and 8, we also show predicted colours from our calibrated model. The maximum likelihood value for each model parameter is given in Table 3, and the full posterior distribution is given in Appendix B. The normalisation of the Balmer Continuum is found to be consistent with zero, which is to be expected given that the emission-line templates we use have not had any Balmer Continuum subtracted. The model is seen to be very good at reproducing the average quasar properties across the vast majority of the colour-redshift-flux space.

We have successfully constructed a model which reproduces the median SDSS-UKIDSS-unWISE colours of a sample of quasars at $17.6 < i_{\rm AB} < 20.1$. However, future surveys such as LSST will probe significantly fainter than this flux limit, so we now need to verify that our model produces accurate predictions at fainter apparent magnitudes.

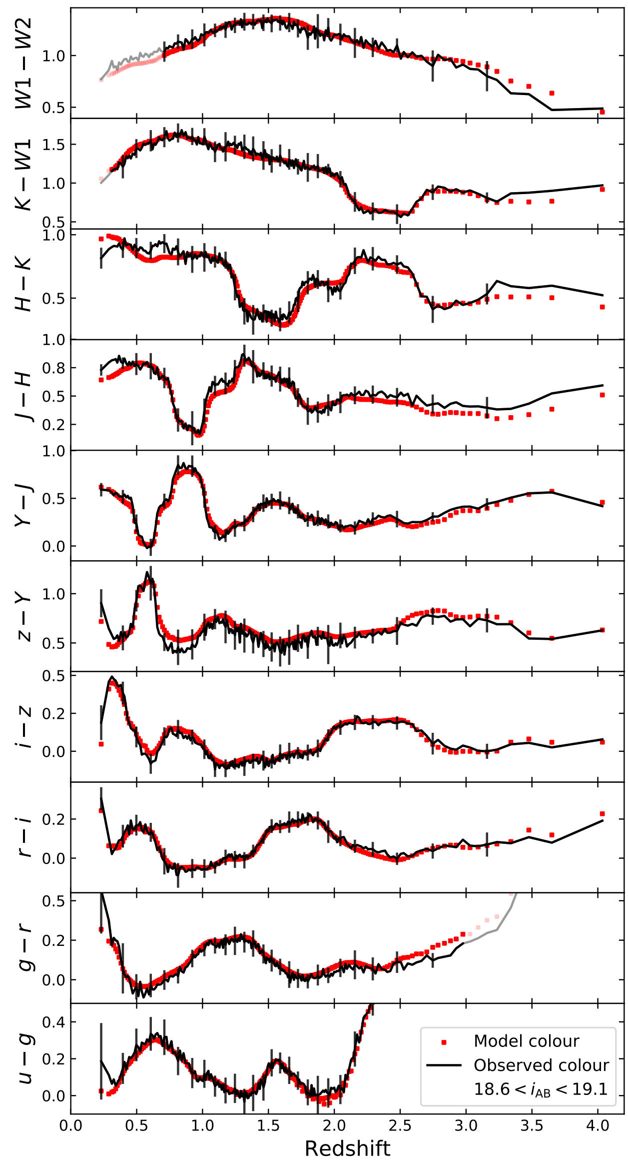

Figure 7. Black: median observed SDSS-UKIDSS-unWISE quasar colours in each redshift bin for our sample of $18.6 < i_{\rm AB} < 19.1$ quasars. Error bars show the dispersion around the median in every eighth bin. Red: the calibrated model. Rest wavelengths $\lambda < 912$ Å or $\lambda > 3\,\mu$m, where the model is not fit to the data, are grayed out.

4 COMPARISON WITH OTHER OBSERVATIONS

We have constructed a model which is capable of reproducing the colours of relatively bright ($i_{\rm AB} < 20.1$) quasars with redshifts $0 < z < 5$ in SDSS, UKIDSS and unWISE. Our model includes two components which adjust the resulting colours as a function of luminosity: both the normalisation of the galaxy component and type of emission-line template depend on the quasar luminosity.

Before providing predictions for upcoming surveys in Section 5, here we verify that our model matches observations of known quasar populations in regions of redshift-luminosity space which have not been used to inform the construction of the model. We compare our predicted model colours to observations of two populations of quasars which have not been used to constrain our model parameters in any way: the near-infrared colours of quasars which are 2.4 magnitudes fainter than the UKIDSS-matched sample used in Section 3, and known $z > 5$ quasars.

4.1 Fainter quasars

We take VISTA *ZYJH* data (González-Fernández et al. 2018) from the fifth data release of the ESO public survey VIKING (Edge et al. 2013). VIKING DR5 is cross-matched to the SDSS DR16 quasar catalogue using a 0.85 arcsec matching radius, yielding a total of 17 047 objects.

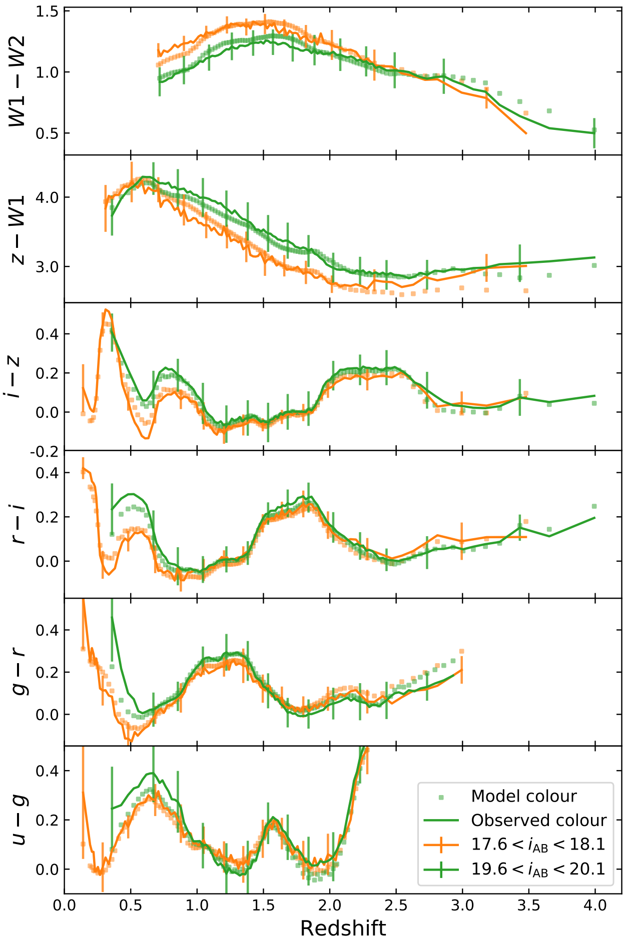

Figure 8. As Fig. 7, but showing the SDSS-unWISE quasar colours in $17.6 < i_{\rm AB} < 18.1$ (orange) and $19.6 < i_{\rm AB} < 20.1$ (green), together with the calibrated model for each apparent magnitude range (squares). The colours of fainter quasars have a larger contribution, due to starlight from the host galaxy, to the quasar SED, which accounts for the variation in *W1* and *W2* and the low-redshift changes in *ugriz*. Smaller differences in *ugriz* at redshifts $z > 1$ are due to changes in the equivalent width of the emission lines, which we model as a Baldwin effect using the templates in Fig. 3.

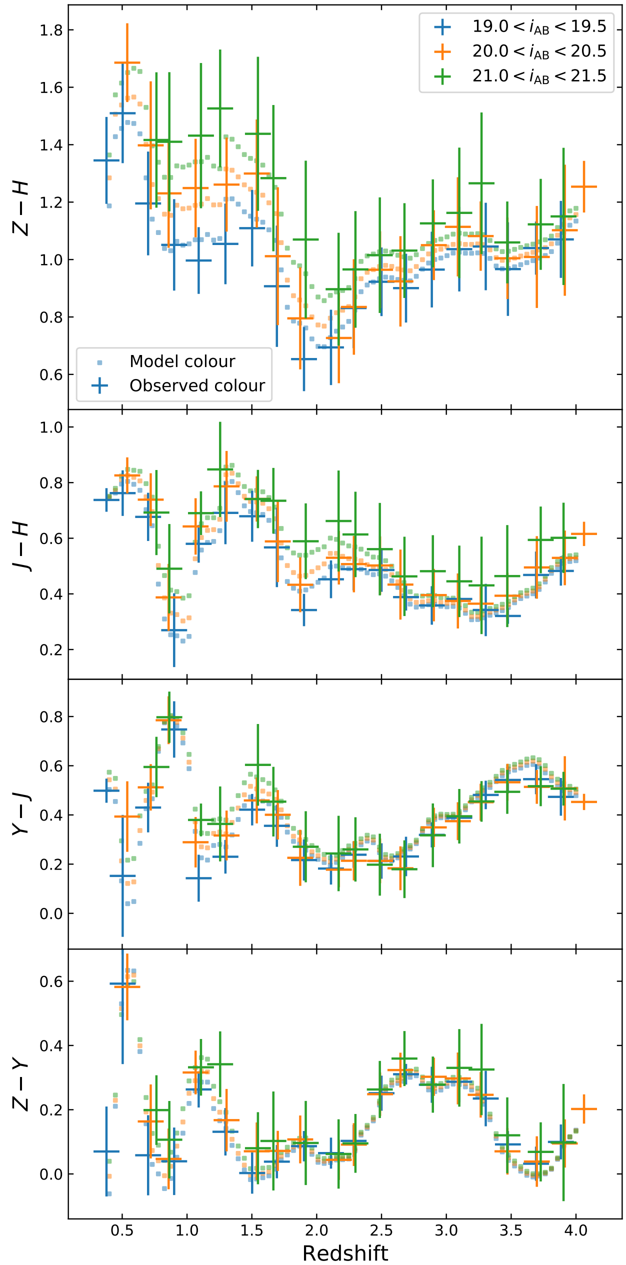

Figure 9. VIKING *ZYJH* colours in three flux bins, probing 2.4 magnitudes fainter than UKIDSS data used to constrain our model. Error bars show the observed average colours and standard deviation within each redshift bin, and squares show the predicted model colours. Our model predictions reproduce the changes in colour which are observed as a function of luminosity, giving us confidence in our ability to extrapolate to LSST-like depths.

In Fig. 9, we show the median sigma-clipped VIKING colours in redshift bins of width $\Delta z = 0.2$. Only bins containing 10 or more objects are shown. The observed colours are seen to change significantly as a function of apparent magnitude, with the *Z-H* colours at $i_{\rm AB} \simeq 21$ up to 0.4 mag redder than those at $i_{\rm AB} \simeq 19$. Our model is seen to reproduce the observed median colours to within ≃0.1 mag, significantly less than the observed intra-sample dispersion, across all redshifts and across all three flux bins.

4.2 High redshift quasars

In Fig. 10, we show the observed colours of known $z > 5$ quasars using force-photometered data from Ross & Cross (2020). We remove all photometric measurements which have uncertainties greater than 0.3 mag. The median uncertainties in each band for the remaining data are $[\sigma(Z), \sigma(Y), \sigma(J), \sigma(H), \sigma(K), \sigma(W1), \sigma(W2), \sigma(W3)] = [0.05, 0.07, 0.08, 0.10, 0.09, 0.04, 0.07, 0.23]$ respectively. The errors on the derived colours are therefore asymmetric: for example the uncertainty of *W1* is significantly less than that of *K* and so the *K-W1* colours preferentially scatter to larger values. We also show our predicted model for three different apparent magnitudes, chosen to be representative of the luminosity range of the $1 < z < 5$ SDSS quasar population. The large changes in predicted colours seen in *K-W1* and *W1-W2* reflect the variation in host galaxy contribution in our model with changing luminosity, while the changes in *Z-Y* and *Y-J* at these redshifts reflect changes in the average emission-line

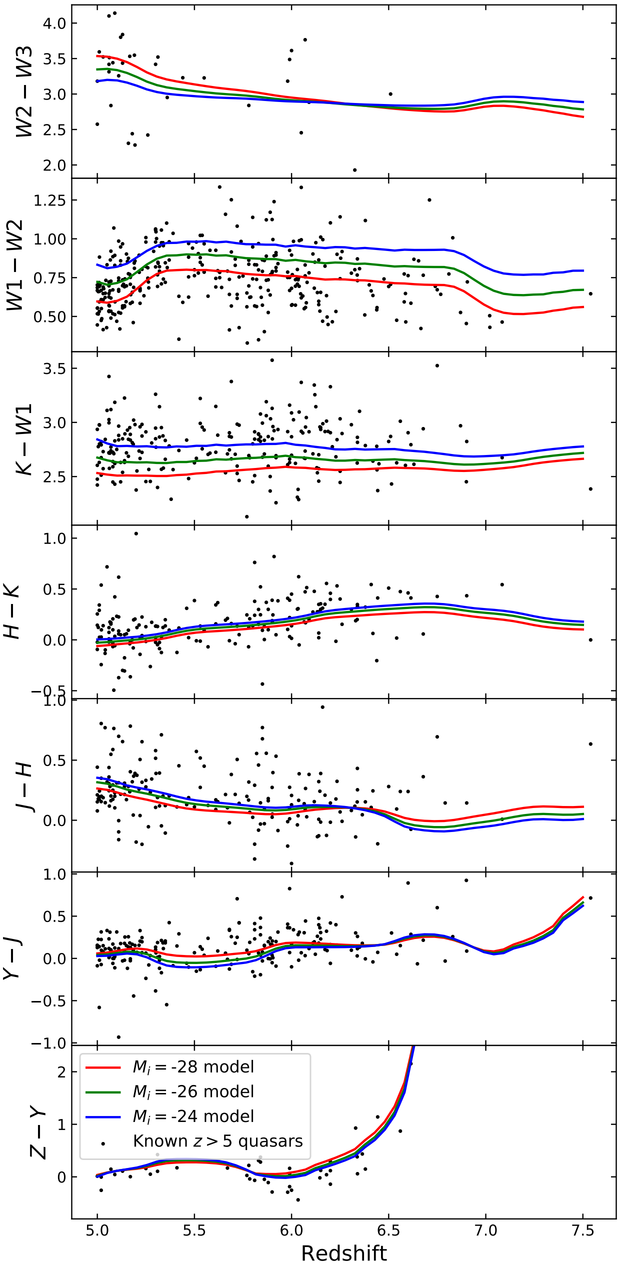

Figure 10. Black points: the observed colours of individual $z > 5$ quasars from Ross & Cross (2020). Any colours for a quasar involving a magnitude uncertainty >0.3 mag are excluded. Note that the errors on these colours are highly asymmetric: for example *W1* is significantly deeper than K and so the *K-W1* colours preferentially scatter to larger values. Coloured lines: our predicted model colours for a range of intrinsic luminosities, corresponding to a range of host galaxy contributions, assuming $E(B-V) = 0$.

properties. Many of the quasars in the Ross & Cross (2020) compilation have errors in the colours >0.2 mag contributing to the extended spread of colours at fixed redshift. Overall, the model predictions are consistent with the observed colours within the uncertainties on the data.

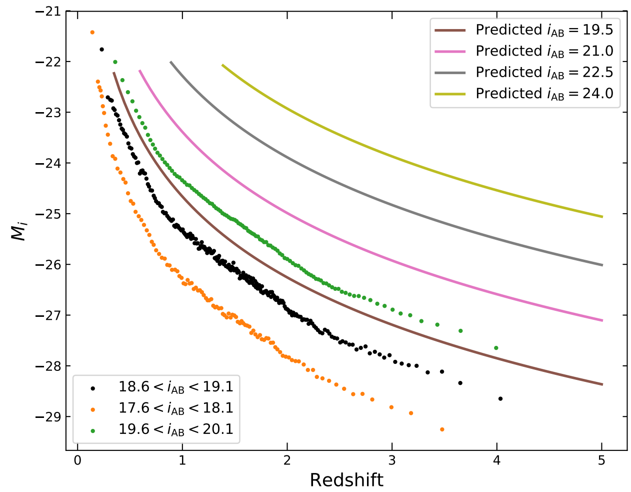

Figure 11. Coloured points: the median M_i in each flux-redshift bin presented in Fig. 7 and Fig. 8. Solid lines: predicted luminosity-redshift relations (Eq. 12) assumed when predicting colours for different flux-limited samples at LSST depths (Fig. 12). We limit our predictions to redshifts and apparent magnitudes where M_i is predicted to be less than −22.0, which has the effect of limiting our predictions at the faintest flux limits to higher redshifts.

5 PREDICTED QUASAR COLOURS FOR LSST AND EUCLID

In order to compute model colours at fainter flux limits, such as those expected to be reached by LSST, it is first necessary to predict the redshift-luminosity relation used to compute the galaxy normalisation (Eq. 7). In Fig. 11, we show the median observed M_i as a function of redshift for each i-band–limited sample. These are found to be approximated by a function of the form

$$M_i = -\left[0.25\left(\frac{i_{AB}}{20}\right) + 5.05\right]\log_{10}(z) - \left[17.4\left(\frac{20}{i_{AB}}\right) + 6.82\right]. \quad (12)$$

The faintest object from SDSS which we have used in calibrating our model is at approximately $M_i = -22.0$, and we chose not to extrapolate any fainter than this luminosity. In effect this limits our model predictions to redshifts $z > 1.5$ for $i_{AB} = 24.0$ and to redshifts $z > 1.0$ for $i_{AB} = 22.5$.

Using our predicted $z - M_i$ tracks, we can make predictions for the average colours of quasars in LSST and *Euclid* down to $i_{AB} = 24$, which is the single-epoch limiting magnitude for LSST. These are shown in Fig. 12, and the predicted flux contribution from the host galaxy to the median colours is shown in Fig. 13. The strong dependence of the fractional galaxy contribution on apparent magnitude leads to significant changes in the predicted *Euclid YJH* colours at redshifts $1 < z < 3$, with fainter objects having a stronger galaxy contribution and hence redder near-infrared colours. Here we have assumed that the average host galaxy SED is the same for SDSS quasars as it will be for the fainter *Euclid* population. There are many reasons why this might not be true, and so it is important not to take our predicted quasar colours as inferring anything about the physics of fainter AGN, but instead merely providing testable predictions for how quasar selection techniques may need to adapt to identify objects in new regimes of luminosity-redshift space.

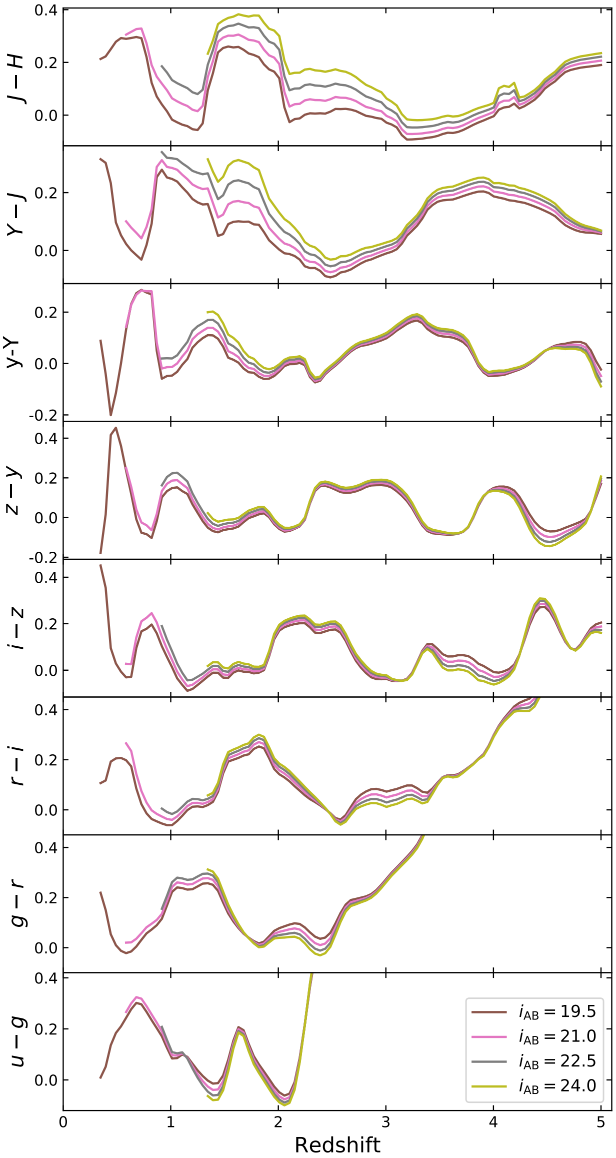

Figure 12. Predicted average colours (assuming $E(B-V) = 0$) as a function of redshift for LSST *ugrizy* and *Euclid YJH*. Colours are presented on the AB system, and shown for different apparent magnitudes. The host galaxy contribution is predicted to have a noticeable effect on the *YJH* colours, even at $z > 2$, as we move towards the LSST single-epoch limiting magnitude of $i_{\rm AB} = 24.0$.

5.1 Very-high redshift quasars in Euclid

In Fig. 14, we show the *Euclid YJH* colour-colour tracks for our quasar model SED with redshifts $z > 5$. Square symbols are plotted at redshift intervals of $\Delta z = 0.1$. Varying the emission-line properties of our model between the extrema seen in lower-redshift quasars produces changes in the *Euclid* colours of up to a few tenths of a magnitude. Quasars with weak, blueshifted lines (as expected for the most luminous, highly accreting sources), and modest dust reddening ($E(B-V) = 0.1$) are predicted to have *J-H* $\simeq 0.5$ at $7 < z < 8$, overlapping the colour tracks of elliptical galaxies of redshift $z \simeq 1$ (cf. Euclid Collaboration et al. 2019, fig. 1). The more common elliptical galaxies may impact on the efficient selection of some

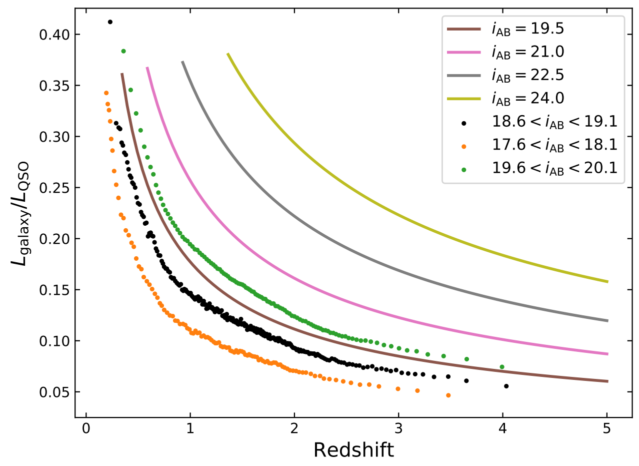

Figure 13. Ratio of integrated 4000-5000 Å luminosities for the galaxy and quasar components of our SED model (Eq. 7). Points: the galaxy fraction in our model for the three observed flux regimes. Solid lines: predicted models from extrapolating those trends to LSST-like depths.

$z > 7$ quasars, reducing the completeness of the census of black hole growth in the early universe.

6 DISCUSSION

A significant improvement over the SED model developed by Hewett is the parametrization of the host-galaxy contribution to the SED as a function of quasar luminosity. In Fig. 15, we show the SED model over a range of redshift. Each model has a luminosity chosen at each redshift to match a flux-limited sample similar to SDSS DR7. At wavelengths longer then 4000 Å, the host galaxy contribution increases as one moves to lower redshifts (with fainter AGN), leading to significant changes in the continuum slope in the rest-frame optical region.

In Fig. 15, we also show the composite of quasar spectra presented by Vanden Berk et al. (2001). The Vanden Berk composite made use of observed-frame optical spectra from the SDSS and, as such, the average redshift of objects changes systematically as a function of rest wavelength: pixels at $\lambda < 1000$ Å in the composite are derived from objects with $4 < z < 5$ while pixels at $\lambda > 5000$ Å are informed mostly by objects with $z < 0.2$. Reassuringly, and without any fine-tuning, our default model is seen to agree very well with the Vanden Berk composite, when we compare any given wavelength region of the composite with the model for the appropriate redshift. We have also verified that our model agrees with the X-Shooter composite presented by Selsing et al. (2016), which made use of much more luminous ($M_i \approx -29$) quasars at redshifts $1.0 < z < 2.1$.

Our blue power-law slope of $\alpha_\nu = -0.349$ corresponds to $\alpha_\lambda = -1.651$, which is consistent with most composite spectra in the literature: e.g. Selsing et al. (2016) find $\alpha_\lambda = -1.70$, Vanden Berk et al. (2001) find $\alpha_\lambda = -1.56$ bluewards of 5000 Å and Francis et al. (1991) find $\alpha_\lambda = -1.68$. Overall there is close agreement between the ultraviolet continuum and the composites of Francis et al. (1991); Vanden Berk et al. (2001); Telfer et al. (2002); Glikman et al. (2006); Lusso et al. (2015). Given the composites span a wide range of luminosities and redshifts, there is no evidence for significant evolution in the shape of the quasar continuum in the rest-frame ultraviolet at these luminosities.

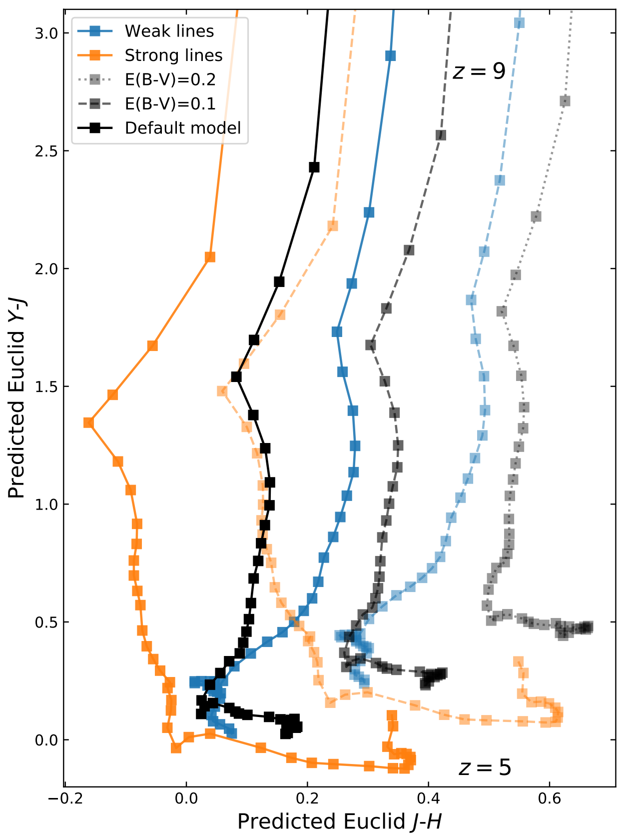

Figure 14. Predicted *Euclid YJH* colour-colour tracks for $5 < z < 9$ quasars (cf. Euclid Collaboration et al. 2019, fig. 1). Colours are presented on the AB system, with each square representing a redshift interval $\Delta z = 0.1$. Different colours show the effect of different emission line templates, and the dashed and dotted lines show the effects of increasing dust reddening. Luminous, high $L/L_{\rm Edd}$ quasars are expected to have relatively weak emission lines, and so have broad-band colours somewhere between the black and blue tracks. The colours of high-redshift quasars in the discovery space opened by *Euclid* could span several tenths of magnitude in *J-H*, which should be taken into account when searching for $z > 7$ quasars.

A feature of our SED model, which distinguishes it from some in the literature, is the flexibility afforded by a broken power-law describing the quasar continuum. The second power-law slope describes the shape of the continuum redward of ≈4000 Å. The value of this red slope is inherently degenerate with host galaxy emission, when the photometric information extends only to the K-band. The behaviour of the longer wavelength *WISE W1-W2* colours as a function of quasar luminosity, however, means it is possible to break the degeneracy between the slope of the red power-law continuum and the contribution from host galaxy. Specifically, the significant range in the *W1-W2* colours as a function of quasar luminosity (at fixed redshift) evident up to $z \approx 2$ are well-reproduced by the changing galaxy contribution as a function of quasar luminosity.

As a result, the dependence of the host-galaxy luminosity on the quasar luminosity, $L_{\rm galaxy} \propto L_{\rm QSO}^{0.684}$, incorporated in the SED model, is more consistent with constraints from other investigations. Richards et al. (2006) find (their section 5.2) a $L_{\rm galaxy} - L_{\rm QSO}$ relation (their eq. 1), which is very similar to our parametrization. Their power-law index, derived from Vanden Berk et al. (2006), is 0.87, compared to our 0.684. However, there is also an $L/L_{\rm Edd}$ term in their eq. 1 which, given brighter quasars will have on average slightly higher accretion rates, will pull their index down closer to ours.

In this work we have assumed a particular functional form (a power-law in luminosity) for the average fractional host galaxy contribution. While this is found to describe the observed data perfectly adequately, we note that there is no *a priori* reason to believe that this is the exact form which the relationship takes. Working under this assumption, the *WISE* data we use is of sufficiently high-quality to infer that the fractional host-galaxy contribution to the rest-frame optical and 1 μm flux of luminous quasars is perhaps higher than previously assumed, around 10 to 20 per cent at $z = 1$ and 5 to 10 per cent at $z = 3$ for SDSS-like flux limits. We note that the relation found by Richards et al. (2006) represents a *minimal* galaxy contribution and that they themselves admit that it could be much greater.

While there remains a slight degeneracy between the strength of the galaxy emission and the other model parameters (Fig. B1), it then follows that the value of the second power-law slope, describing the shape of the quasar continuum redwards of 3880 Å, is significantly steeper than our blue slope, and the overall slope of other quasar templates: $\alpha_\nu = +0.593$ or $\alpha_\lambda = -2.593$.

The situation for the ultraviolet continuum is very different as the quasar dominates for all luminosities we consider. In the 1215-4000 Å region, the only significant change with luminosity, and hence redshift, is due to the rest-frame ultraviolet emission line properties; for higher luminosity objects the high-ionization lines become blue-asymmetric and emission-line equivalent widths are smaller (Figs. 1 and 3). Above redshift $z \approx 2$, the model-flux changes significantly at wavelengths shorter than 1215 Å as a result of absorption due to the inter-galactic medium.

7 SUMMARY

The primary rationale for the information and parametric model presented in this paper is to provide empirically-derived SEDs, covering the rest-frame ultraviolet through near-infrared, for luminous type 1 AGN. These SEDs, combined with host galaxy SEDs, can be used to simulate quasar observed-frame optical and infrared colours for various purposes, such as (a) investigating the SEDs of 'mean' or typical quasars, (b) generating predictions which incorporate the intra-population properties of quasars/AGN, and (c) estimating the properties of individual objects, including photometric redshifts.

As a function of apparent i-band magnitude, there are systematic changes in the average quasar colours at all redshifts, which are well-explained by changes in the strength of the emission lines and the contribution of the host galaxy. Our model is capable of reproducing, to within ~0.1 magnitudes, the optical and infrared colours of tens of thousands of quasars across a wide range of redshift ($0 < z < 5$) and luminosity ($-22 > M_i > -29$). The variation in the *average* emission line and host galaxy contributions can be well described by simple functions of luminosity which account for the observed changes in colour across $18.1 < i_{\rm AB} < 21.5$. To instead model the properties of individual quasars, these parameters can be left free to vary.

Building on work from spectroscopic studies, the model encapsulates our current understanding of the intra-population variance in the ultraviolet and optical emission line properties of luminous quasars. This is achieved through just one parameter, which is similar to the "C IV distance" defined by Richards et al. (2021), and accounts for the observed correlations reported by Coatman et al. (2019) and Temple et al. (2021b).

The relative flux contribution from stars in the host galaxy to the

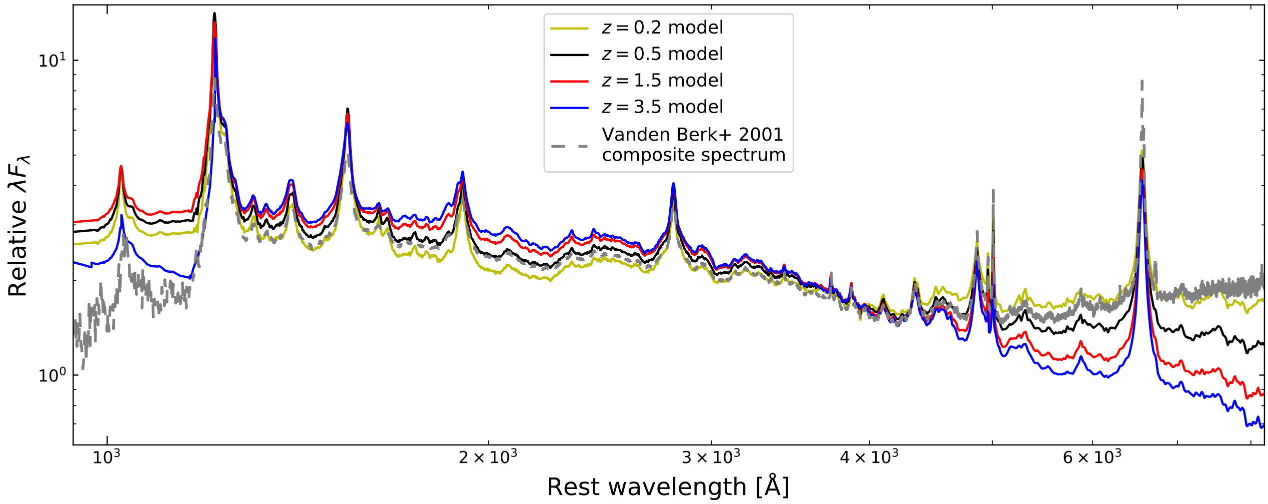

Figure 15. Comparison of our model quasar SED at different redshifts with the composite spectrum from Vanden Berk et al. (2001). The Vanden Berk composite is reassuringly similar to our models across the rest-frame ultraviolet, and is seen to agree very well with our $z = 0.2$ model at rest wavelengths $\lambda > 5000$ Å, consistent with the redshifts of the objects contributing to the composite at those wavelengths. At higher redshifts the average quasar is brighter and has a smaller fractional host galaxy contribution to the rest-frame optical. Similarly the composite agrees well with our $z = 3.5$ model at 1200 Å, as the inter-galactic medium begins to suppress the flux shortwards of Ly α. At ≈1000 Å the composite consists of objects with $z > 4$ where the Lyman suppression is even stronger.

total quasar SED will always be larger in the rest-optical than in the rest-ultraviolet. Assuming that the quasar optical-continuum slope does not change across the population, and that the average fractional galaxy contribution has a power law dependence on the total luminosity, we find that bright, $z > 2$ quasars from SDSS still have significant (greater than 5 per cent) contributions to their rest-frame optical flux by stellar emission from their host galaxies. The next decade will bring several new, advanced large-scale surveys of the extragalactic sky. By assuming that the trends observed in the SDSS population continue to fainter luminosities, we make predictions for the average LSST and *Euclid* colours of soon-to-be-identified populations of new AGN.

ACKNOWLEDGEMENTS

MJT acknowledges support from CONICYT (Fondo ALMA 31190036), and thanks the IoA, Cambridge, for the award of visitor status while this work was completed. PCH and MJT acknowledge funding from the Science and Technology Facilities Council via the Institute of Astronomy, Cambridge, Consolidated Grant and a Ph.D. studentship respectively. MB acknowledges funding from the Royal Society via a University Research Fellowship. We thank Adam Marshall, Matt Auger and Richard McMahon for testing the model code. MJT acknowledges a stimulating examination *viva voce* with Chris Done and Richard McMahon.

The VISTA data flow system, the WFCAM science archive and the VISTA science archive are described in Irwin et al. (2004), Hambly et al. (2008) and Cross et al. (2012). This work made use of ASTROPY (Astropy Collaboration et al. 2013, 2018), CORNER.PY (Foreman-Mackey 2016), EMCEE (Foreman-Mackey et al. 2013), MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020) and Q3C (Koposov & Bartunov 2006). This work also made use of the Whole Sky Database (wsdb) created by Sergey Koposov and maintained at the Institute of Astronomy, Cambridge by Sergey Koposov, Vasily Belokurov and Wyn Evans with financial support from STFC and the European Research Council. This research has made use of the SVO Filter Profile Service (http://svo2.cab.inta-csic.es/theory/fps/) supported from the Spanish MINECO through grant AYA2017-84089.

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, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. *WISE* and NEOWISE are funded by the National Aeronautics and Space Administration.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

DATA AVAILABILITY

The data underlying this article were accessed from the Sloan Digital Sky Survey,[3] the WFCAM science archive,[4] the VISTA science archive,[5] the unWISE catalogue,[6] and the VHzQ GitHub repo.[7] The quasar SED model code described in the text is available on GitHub.[8]

REFERENCES

Allen J. T., Hewett P. C., Maddox N., Richards G. T., Belokurov V., 2011, MNRAS, 410, 860

Allen J. T., Hewett P. C., Richardson C. T., Ferland G. J., Baldwin J. A., 2013, MNRAS, 430, 3510

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

Bachev R., Marziani P., Sulentic J. W., Zamanov R., Calvani M., Dultzin-Hacyan D., 2004, ApJ, 617, 171

Baldwin J. A., 1977, ApJ, 214, 679

Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067

Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C., 2000, MNRAS, 317, 1014

Coatman L., Hewett P. C., Banerji M., Richards G. T., Hennawi J. F., Prochaska J. X., 2017, MNRAS, 465, 2120

Coatman L., Hewett P. C., Banerji M., Richards G. T., Hennawi J. F., Prochaska J. X., 2019, MNRAS, 486, 5335

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Loaring N. S., Miller L., Lewis I. J., 2001, MNRAS, 322, L29

Croom S. M., et al., 2002, MNRAS, 337, 275

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397

Cross N. J. G., et al., 2012, A&A, 548, A119

Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, The Messenger, 154, 32

Elvis M., et al., 1994, ApJS, 95, 1

Euclid Collaboration et al., 2019, A&A, 631, A85

Fan X., 1999, AJ, 117, 2528

Fan X., et al., 2001, AJ, 121, 31

Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831

Fitzpatrick E. L., Massa D., 2009, ApJ, 699, 1209

Flesch E. W., 2021, MNRAS, 504, 621

Foreman-Mackey D., 2016, Journal of Open Source Software, 1, 24

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., Weymann R. J., Morris S. L., 1991, ApJ, 373, 465

Gallerani S., et al., 2010, A&A, 523, A85

Giustini M., Proga D., 2019, A&A, 630, A94

Glikman E., Helfand D. J., White R. L., 2006, ApJ, 640, 579

González-Fernández C., et al., 2018, MNRAS, 474, 5459

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Grandi S. A., 1982, ApJ, 255, 25

[3] `https://www.sdss.org/dr16/`
[4] `http://wsa.roe.ac.uk`
[5] `http://horus.roe.ac.uk/vsa/`
[6] `https://catalog.unwise.me`
[7] `https://github.com/d80b2t/VHzQ/`
[8] `https://github.com/MJTemple/qsogen/`

Green R. F., Pier J. R., Schmidt M., Estabrook F. B., Lane A. L., Wahlquist H. D., 1980, ApJ, 239, 483

Hambly N. C., et al., 2008, MNRAS, 384, 637

Harris C. R., et al., 2020, Nature, 585, 357

Hatziminaoglou E., Mathez G., Pelló R., 2000, A&A, 359, 9

Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454

Hodgkin S. T., Irwin M. J., Hewett P. C., Warren S. J., 2009, MNRAS, 394, 675

Højen-Sørensen P. A., Winther O., Hansen L. K., 2002, Neural Computation, 14, 889

Hopkins P. F., et al., 2004, AJ, 128, 1112

Hunter J. D., 2007, Computing In Science & Engineering, 9, 90

Irwin M. J., et al., 2004, in Quinn P. J., Bridger A., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies. pp 411–422, doi:10.1117/12.551449

Ivezić Ž., et al., 2019, ApJ, 873, 111

Jensen T. W., et al., 2016, ApJ, 833, 199

Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 497, 906

Koposov S., Bartunov O., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, Astronomical Society of the Pacific Conference Series Vol. 351, Astronomical Data Analysis Software and Systems XV. p. 735

Krawczyk C. M., Richards G. T., Mehta S. S., Vogeley M. S., Gallagher S. C., Leighly K. M., Ross N. P., Schneider D. P., 2013, ApJS, 206, 4

Krawczyk C. M., Richards G. T., Gallagher S. C., Leighly K. M., Hewett P. C., Ross N. P., Hall P. B., 2015, AJ, 149, 203

Kruczek N. E., et al., 2011, AJ, 142, 130

Lang D., 2014, AJ, 147, 108

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Lawrence A., et al., 2007, MNRAS, 379, 1599

Lee K.-G., et al., 2013, AJ, 145, 69

Lupton R. H., Gunn J. E., Szalay A. S., 1999, AJ, 118, 1406

Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204

Lyke B. W., et al., 2020, ApJS, 250, 8

Maddox N., Hewett P. C., 2006, MNRAS, 367, 717

Maddox N., Hewett P. C., Warren S. J., Croom S. M., 2008, MNRAS, 386, 1605

Maddox N., Hewett P. C., Péroux C., Nestor D. B., Wisotzki L., 2012, MNRAS, 424, 2876

Maiolino R., et al., 2020, The Messenger, 180, 24

Margala D., Kirkby D., Dawson K., Bailey S., Blanton M., Schneider D. P., 2016, ApJ, 831, 157

McGreer I. D., et al., 2013, ApJ, 768, 105

Meisner A. M., Lang D., Schlafly E. F., Schlegel D. J., 2019, PASP, 131, 124504

Merloni A., et al., 2019, The Messenger, 175, 42

Mortlock D. J., et al., 2011, Nature, 474, 616

Mortlock D. J., Patel M., Warren S. J., Hewett P. C., Venemans B. P., McMahon R. G., Simpson C., 2012, MNRAS, 419, 390

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Pâris I., et al., 2017, A&A, 597, A79

Polletta M., et al., 2007, ApJ, 663, 81

Rankine A. L., Hewett P. C., Banerji M., Richards G. T., 2020, MNRAS, 492, 4553

Reed S. L., et al., 2017, MNRAS, 468, 4702

Reed S. L., et al., 2019, MNRAS, 487, 1874

Richards G. T., et al., 2002, AJ, 123, 2945

Richards G. T., et al., 2003, AJ, 126, 1131

Richards G. T., et al., 2006, ApJS, 166, 470

Richards G. T., et al., 2011, AJ, 141, 167

Richards G. T., McCaffrey T. V., Kimball A., Rankine A. L., Matthews J. H., Hewett P. C., Rivera A. B., 2021, arXiv e-prints, p. arXiv:2106.07783

Ross N. P., Cross N. J. G., 2020, MNRAS, 494, 789

Rowan-Robinson M., et al., 2008, MNRAS, 386, 697

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schlafly E. F., Finkbeiner D. P., Schlegel D. J., Jurić M., Ivezić Ž., Gibson R. R., Knapp G. R., Weaver B. A., 2010, ApJ, 725, 1175
Schlafly E. F., Meisner A. M., Green G. M., 2019, ApJS, 240, 30
Schneider D. P., et al., 2010, AJ, 139, 2360
Selsing J., Fynbo J. P. U., Christensen L., Krogager J. K., 2016, A&A, 585, A87
Skrutskie M. F., et al., 2006, AJ, 131, 1163
Sulentic J. W., Marziani P., Dultzin-Hacyan D., 2000, ARA&A, 38, 521
Sulentic J. W., Marziani P., Zamanov R., Bachev R., Calvani M., Dultzin-Hacyan D., 2002, ApJ, 566, L71
Sulentic J. W., Bachev R., Marziani P., Negrete C. A., Dultzin D., 2007, ApJ, 666, 757
Telfer R. C., Zheng W., Kriss G. A., Davidsen A. F., 2002, ApJ, 565, 773
Temple M. J., Ferland G. J., Rankine A. L., Hewett P. C., Badnell N. R., Ballance C. P., Del Zanna G., Dufresne R. P., 2020, MNRAS, 496, 2565
Temple M. J., Banerji M., Hewett P. C., Rankine A. L., Richards G. T., 2021a, MNRAS, 501, 3061
Temple M. J., Ferland G. J., Rankine A. L., Chatzikos M., Hewett P. C., 2021b, MNRAS, 505, 3247
Vanden Berk D. E., et al., 2001, AJ, 122, 549
Vanden Berk D. E., et al., 2006, AJ, 131, 84
Virtanen P., et al., 2020, Nature Methods, 17, 261
Warren S. J., Hewett P. C., Irwin M. J., Osmer P. S., 1991, ApJS, 76, 1
Wright E. L., et al., 2010, AJ, 140, 1868
York D. G., et al., 2000, AJ, 120, 1579

APPENDIX A: NOTE ON TEMPLE ET AL. (2021A)

In Temple et al. (2021a, hereafter T21), we made use of an earlier version of our model SED to investigate the variation in sublimation-temperature dust in the quasar population. The model employed in T21 made use of a fixed emission-line template, which had been derived from Francis et al. (1991). However, the changes in colour due to the variation in emission-line properties are minimal at the wavelengths > 1 μm where we sought to constrain the hot dust emission, and so this has no effect on our results. In particular, the correlation found between the strength of hot dust emission and the blueshift of the C IV emission line is not driven by a lack of emission-line variation in the SED model.

Second, the model used in T21 had been calibrated to a bright subset of the fourteenth data release of SDSS, as opposed to the multi–flux-binned DR16 used in this work. This meant that the red power-law slope α_2 was -0.16 instead of the value of +0.59 found in this work (Recall also that we assumed zero host galaxy contribution for the relatively bright objects in T21). This change in slope is reflected in the fact that the median blackbody temperature in T21 was 1280 K, as opposed to the value of 1240 K found to be the preferred parameter in this work. While the power-law slope has a direct impact on the quantitative results of T21 (in particular, the blackbody normalisation `bbnorm` is re-scaled by a constant factor of around 1.5 to account for the change of blackbody shape), the qualitative results and correlations in T21 remain unchanged. Further details on the robustness of the results of T21 to different model assumptions can be found in the appendices to that work.

APPENDIX B: PARAMETER DEGENERACIES

The posterior distribution of the likelihood surface of model parameters is shown in Fig. B1. Degeneracies are to be expected between certain parameters: for example the strength of luminosity-evolution of the galaxy contribution `gplind`, the strength of hot dust emission `bbnorm` and the second power-law slope α_2 can all effect the overall shape of the rest-frame near-infrared SED.

This paper has been typeset from a TEX/LATEX file prepared by the author.

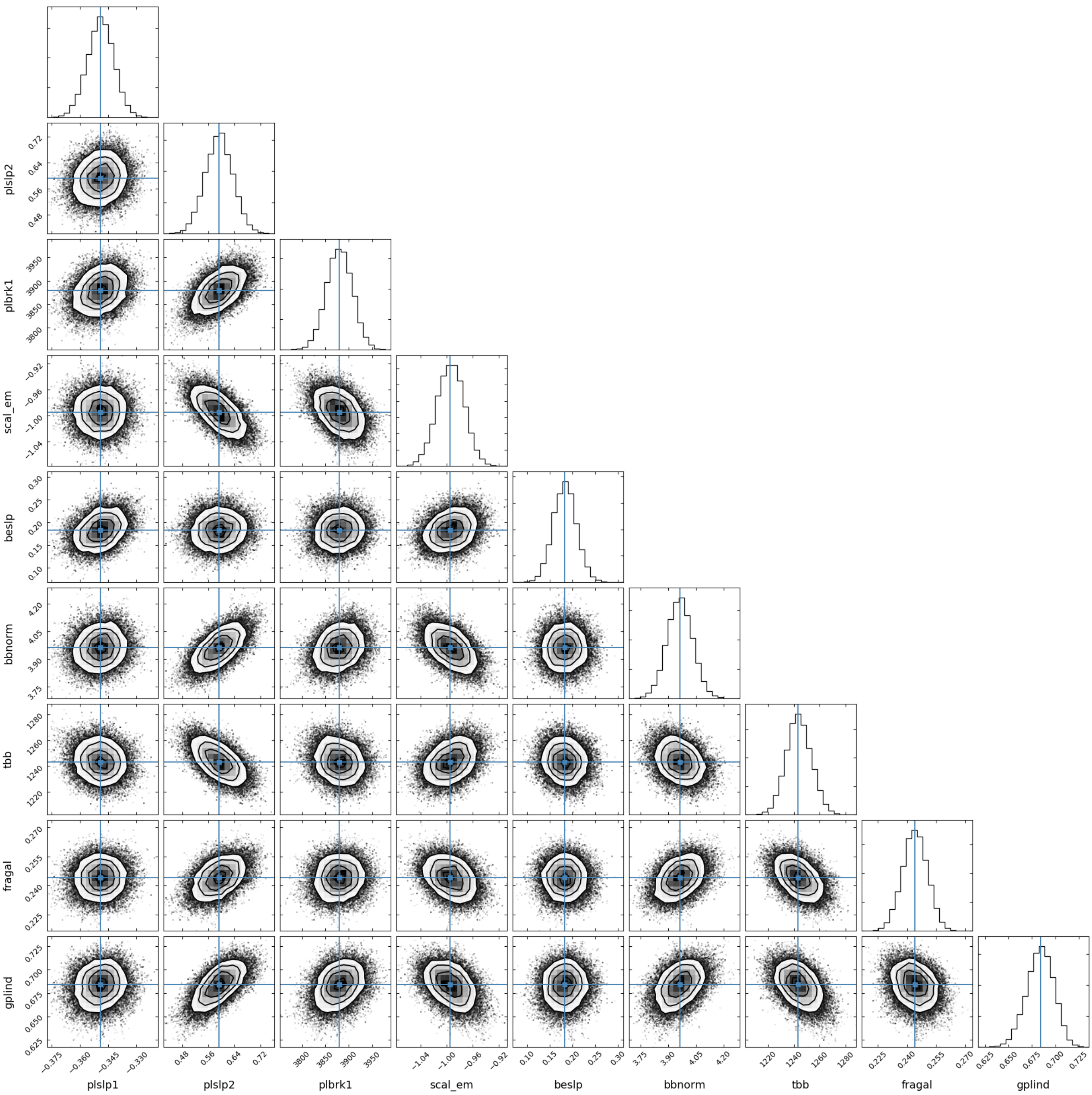

Figure B1. Samples from the posterior distribution of the likelihood surface of SED model parameters defined in equation 11. The posterior is seen to be unimodal in all projections. Some degeneracies can be observed between the strength of luminosity-evolution of the galaxy contribution `gplind`, the strength of hot dust emission `bbnorm` and the second power-law slope `plslp2`, but otherwise all parameters are well-determined. The maximum-likelihood solution is shown in blue.