

IMPERIAL

Imperial College of Science, Technology and Medicine

Literature Review

Multi-wavelength validation beyond the optical/NIR (0.3–5 μm), extending to MIR/FIR/submm and incorporating X-ray and radio diagnostics

Project code:	MLBD_2025_17
Name:	Mihir Koka
Supervisor(s):	David Clements & Boris Leistedt
Assessor:	Alan Heavens
Word count:	2731
Date of submission:	23rd January 2026

Abstract

This literature review outlines the motivation, context, and data resources for validating and extending the pop-cosmos galaxy population synthesis framework beyond its current optical and near-infrared training range. Understanding galaxy evolution would require models that produce consistent predictions across multiple observational methods. The pop-cosmos framework provides a population-level generative model of galaxies, combining physical modelling, machine learning, and statistical inference, and has been shown to reproduce optical and near-infrared observations of the COSMOS field. The main aim of this project is to test and extend pop-cosmos by validating its predictions against multi-wavelength observations beyond its current $0.3\text{--}5\ \mu\text{m}$ training range. In particular, the project will assess the consistency between model predictions and far-infrared, submillimeter, X-ray, and radio data that trace star formation and active galactic nucleus activity through distinct physical mechanisms. By comparing predicted and observed star formation rates, and source classifications across these wavelengths, the project hopes to establish whether pop-cosmos provides a robust and physically consistent description of the galaxy population.

1 Introduction

The study of galaxy evolution seeks to understand how galaxies form and change over time, and how their observable properties are connected to the underlying physical processes such as star formation, dust production, and black hole growth. Modern astronomical surveys observe galaxies across a wide range of wavelengths, producing catalogues containing millions to billions of sources. Extracting reliable physical information from such datasets increasingly requires population-level modelling approaches that can combine physical understanding with flexible statistical inference.

The pop-cosmos framework was developed to address this challenge by modelling the galaxy population as a whole rather than treating individual galaxies in isolation. Using a combination of physical modelling, machine-learning techniques, and statistical inference, pop-cosmos has been shown to reproduce observed distributions of galaxy properties within the COSMOS field using optical and near-infrared data in the wavelength range $0.3\text{--}5\ \mu\text{m}$. These results demonstrate that the model can generate consistent galaxy populations that match observational constraints within our current wavelengths.

However, optical and near-infrared observations capture only part of the physical processes involved in galaxy evolution. In many galaxies, a significant fraction of star formation is hidden by dust, which absorbs ultraviolet and optical light and re-emits it at far-infrared and submillimeter wavelengths. As a result, far-infrared observations provide a better measure of star formation. In addition, X-ray and radio observations are sensitive to emission from actively accreting supermassive black holes and to non-thermal processes that are not traced by stellar light alone. These new wavelengths therefore probe physical processes that are independent of the optical and near-infrared data used to train existing pop-cosmos models.

The aim of this project is to test and extend the pop-cosmos framework by validating its predictions against independent multi-wavelength observations beyond the optical and near-infrared. By comparing model predictions with far-infrared, X-ray, and radio data, the project seeks to assess whether pop-cosmos produces consistent galaxy populations across multiple EM wavelengths. Establishing such consistency would strengthen confidence in the underlying assumptions of the model and its applicability to future large-scale surveys, while issues would provide valuable feedback for future model improvement.

2 The pop-cosmos framework and current predictive scope

The pop-cosmos framework is a population-level generative model designed to describe the joint distribution and evolution of galaxy properties across cosmic time. The model is calibrated using 26-band photometry for approximately 141,000 galaxies in the COSMOS field ($r < 25$), covering wavelengths from the ultraviolet to the near-infrared ($0.3\text{--}5\ \mu\text{m}$) Alsing et al. 2024. At its core, pop-cosmos employs a flexible Stellar Population Synthesis (SPS) model to link intrinsic galaxy properties to observed photometry.

The framework predicts a wide range of physical quantities, including stellar mass, redshift, star formation rate (SFR), and stellar metallicity. In addition, it models active galactic nucleus (AGN) activity through parameters describing the fractional AGN contribution to the total bolometric luminosity (f_{AGN}) and the optical depth of the surrounding dust torus (τ_{AGN}) Alsing et al. 2024.

Within its validated wavelength range, pop-cosmos has been shown to reproduce key population-level observables, such as galaxy colours, stellar mass functions, and the star-forming sequence, out to redshifts of $z \sim 4$. However, the model is not intended to extrapolate far below the flux limits of its training data, and its current formulation does not explicitly test emission at longer wavelengths or high-energy regimes. In particular, mid- and far-infrared, sub-millimetre, X-ray, and radio emission—tracing dust-obscured star formation and black hole accretion—are not yet directly validated within the framework. Extending the model to these regimes is therefore necessary to assess whether its physical assumptions remain consistent when confronted with independent tracers of galaxy evolution.

3 Motivation for multi-wavelength validation beyond $0.3\text{--}5\ \mu\text{m}$

Optical and near-infrared observations provide strong constraints on stellar populations, galaxy colours, and redshifts, but they probe only a subset of the physical processes governing galaxy evolution. Emission in the $0.3\text{--}5\ \mu\text{m}$ range is dominated by starlight and is strongly affected by dust attenuation, making key quantities such as star formation rates and active galactic nucleus (AGN) activity indirect and potentially degenerate. As a result, agreement between models and optical/NIR data alone does not guarantee that the underlying physical assumptions remain valid when confronted with independent tracers.

At longer wavelengths, different components of galaxy emission become accessible. Far-infrared and sub-millimetre radiation primarily trace dust-reprocessed starlight, providing a more direct and less extinction-sensitive measure of star formation, particularly in heavily obscured systems. Radio continuum emission offers an almost dust-independent probe of star formation, while also revealing AGN activity through radio jets and excess emission. X-ray observations directly trace accretion onto supermassive black holes and are useful for identifying obscured AGN that could be missed in optical or infrared surveys.

Validating a population-level model such as pop-cosmos across these additional wavelength regimes is therefore essential. Consistent predictions across optical, infrared, radio, and X-ray observations would demonstrate that the model captures the coupled evolution of stellar mass growth, dust physics, and black hole accretion in a physically meaningful way. On the other hand, discrepancies at long wavelengths would indicate limitations in the model assumptions that are not evident from optical/NIR data alone. Multi-wavelength validation would therefore

provides a stronger test of galaxy evolution models beyond their original training area.

4 Inventory of potentially useful multi-wavelength datasets

Table 1: Summary of candidate multi-wavelength datasets relevant for validating *pop-cosmos* predictions beyond the optical and near-infrared.

Catalogue	Regime	Wavelengths	Cross-matching	Reference
COSMOS2025 (COSMOS-Web JWST catalogue)	Near/Mid- IR	0.3–8 μm	Combined JWST/HST pho- tometry with SED fitting	Shuntov et al. (2024)
S-COSMOS MIPS 24 μm catalogue	Mid-IR	24 μm	Cross-matched with COSMOS photometric red- shifts	Sanders et al. (2007)
Herschel/MIPS de- blended COSMOS catalogue	FIR / sub- mm	24–500 μm	Prior-based de- blending using COSMOS2020	Wang et al. (2024)
Super-deblended FIR–(sub)mm COS- MOS catalogue	FIR / sub- mm	100–1200 μm	Prior-based PSF deblending using VLA/MIPS priors	Jin et al. (2018)
VLA COSMOS-XS radio catalogue	Radio	3, 10 GHz	Multi-frequency synthesis source extraction	van der Vlugt et al. (2020)
Chandra COSMOS- Legacy X-ray cata- logue	X-ray	0.5–10 keV	Likelihood-based source detection and COSMOS cross-matching	Civano et al. (2016)

For each of these papers I will briefly cover the dataset observed, The physical processes traced e.g FIR/sub-mm for dust-obscured star formation, What was released ?, It’s relevance to this research project and potential limitations or shortcomings.

4.1 COSMOS2025 (baseline optical/NIR + JWST)

The COSMOS2025 catalogue provides optical to near and mid-infrared observations of galaxies in the COSMOS/COSMOS-Web field, spanning approximately 0.3–8 μm . It combines imaging from JWST/NIRCam and MIRI with complementary space and ground-based data from HST/ACS, Subaru/HSC, and VISTA/UltraVISTA. In this wavelength range, the data primarily trace stellar emission and largely unobscured star formation. The catalogue releases multi-band photometry for over 700,000 galaxies, along with photometric redshifts, derived physical parameters such as stellar mass and star formation rate, rest-frame colours, and morphological measurements. COSMOS2025 therefore serves as the optical–NIR backbone of the COSMOS field, providing positions, redshifts, and host-galaxy properties for cross-matching

with longer-wavelength datasets. A potential limitation for the present project is that the catalogue does not itself include far-infrared, submillimetre, X-ray, or radio measurements so cross matching with other datasets might be required.

4.2 MIPS 24 μm (MIR bridge)

The S COSMOS Spitzer Legacy Survey provides infrared imaging of the COSMOS field using the IRAC and MIPS instruments, covering wavelengths from 3.6 to 160 μm . For the present project, the most relevant component is the MIPS 24 μm mid infrared data, which trace warm dust emission associated with obscured star formation and active galactic nuclei. While the survey also includes far infrared bands at 70 and 160 μm , these longer wavelength regimes are covered in the other papers/catalogues.

In the mid to far infrared, the data primarily trace dust heated by star formation and, in some cases, by AGN activity. The authors release source catalogues with flux measurements in all seven Spitzer bands and describe the survey strategy, background properties, achieved sensitivities, and preliminary number counts. For this project, the S COSMOS survey provides mid infrared measurements that can be cross matched with COSMOS2025 galaxies, enabling infrared based constraints on star formation and obscured activity as part of a multi wavelength validation of pop cosmos predictions.

4.3 Herschel / super-deblended FIR–sub-mm

Wang et al. present a probabilistic deblending approach for far infrared and sub millimetre surveys, applied to the COSMOS field, in which fluxes are extracted at the positions of known galaxies rather than through blind source detection. The method uses Spitzer MIPS 24 μm data as an initial prior and progressively deblends emission in the Herschel PACS and SPIRE bands up to 500 μm , making use of existing COSMOS2020 and radio catalogues. The resulting data trace dust reprocessed emission from star formation and are therefore sensitive to obscured star formation that is not captured by optical or near infrared observations. The authors release deblended source catalogues with associated uncertainties that are cross matched with multi wavelength data by construction. For the present project, these catalogues provide far infrared measurements that can be directly compared with pop cosmos star formation rate predictions as an independent validation of the model at long wavelengths.

4.4 Super-deblended FIR–sub-mm COSMOS catalogue

Jin et al. present a super deblended far infrared to sub millimetre photometric catalogue for the COSMOS field, extending wavelength coverage from the mid infrared to millimetre regimes. The catalogue combines data from Spitzer MIPS, Herschel PACS and SPIRE, and longer wavelength sub millimetre and millimetre observations, with radio and mid infrared measurements used as priors to deblend the low resolution far infrared images. These data trace dust reprocessed emission associated with obscured star formation, which is not directly accessible from optical or near infrared observations.

The authors release photometric and value added catalogues for a large sample of prior galaxies, identifying over eleven thousand sources individually detected in the far infrared or sub millimetre bands. For the present project, this dataset provides long wavelength measurements that enable validation of pop cosmos star formation rate predictions in a regime dominated by

dust emission. A key limitation of the catalogue is the strong source confusion inherent to far infrared and sub millimetre observations, as well as reduced depth in some bands, which leads to increased uncertainties for faint sources.

4.5 VLA COSMOS-XS radio catalogue

The COSMOS XS survey provides ultra deep radio continuum observations of a sub region of the COSMOS field using the Karl G Jansky Very Large Array at 3 and 10 GHz. Radio emission traces synchrotron radiation associated with star formation and active galactic nuclei and is largely unaffected by dust attenuation. The authors release a radio source catalogue detected at 3 GHz with multi wavelength counterparts identified for the majority of sources.

For the present project, this dataset offers dust independent measurements that can be used to test pop cosmos predictions of star formation and AGN activity, and to cross check classifications derived from optical and infrared data. A key limitation is the small sky area covered by the ultra deep observations, which increases sensitivity to cosmic variance, as well as uncertainties introduced by limited spectral information and the use of photometric redshifts for part of the sample.

4.6 Chandra COSMOS-Legacy X-ray catalogue

The Chandra COSMOS Legacy survey provides deep X-ray observations of the COSMOS field, covering 2.2 square degrees with a total exposure of 4.6 Ms. X-ray emission traces accretion onto supermassive black holes and therefore offers a direct diagnostic of active galactic nuclei, including systems that may be heavily obscured at optical or infrared wavelengths. The authors release a point source catalogue containing over four thousand sources, with measured fluxes in the soft, hard, and full X-ray bands, along with hardness ratios used as a proxy for intrinsic obscuration.

For the present project, this dataset enables validation of pop cosmos AGN classifications using independent high energy observations, complementing infrared and radio diagnostics. A key limitation is that hardness ratios provide only an approximate measure of obscuration and can be affected by redshift and spectral complexity, while a small fraction of sources lack secure multi wavelength identifications or redshifts.

5 Model–data comparison strategy

The current section of this review details the methodology for extending the pop-cosmos framework into long-wavelength and high-energy regimes.

What pop-cosmos Predicts

The pop-cosmos framework functions as a comprehensive generative model that predicts the joint distribution of physical characteristics (ϕ) for the galaxy population Alsing et al. 2024. Utilizing a state-of-the-art 16-parameter Stellar Population Synthesis (SPS), the framework predicts key physical properties including stellar mass, stellar metallicity, and redshift. By assuming a flexible non-parametric star-formation history, the model derives accurate Star Formation Rates (SFR). Furthermore, it predicts AGN probability through two specific parameters: the fractional contribution of the AGN to total bolometric luminosity (f_{AGN}) and the optical depth of the AGN dust torus (τ_{AGN}). These physical parameters are then trans-

lated into synthetic photometry from the UV through the infrared via emulators of the Flexible Stellar Population Synthesis (FSPS) code. This allows the model to faithfully predict galaxy colors and magnitudes while accounting for internal degeneracies.

What the Observations Provide

To validate these predictions, we incorporate datasets tracing high-energy and dust-obscured processes. FIR/sub-mm observations from the "super-deblended" COSMOS catalog provides a tracer of dust-obscured star formation, extending detections to very high redshifts Jin et al. 2018. Complementary MIR data from Spitzer/MIPS and IRAC trace warm dust emission associated with both obscured star formation and AGN activity, identifying obscured activity that might be missed in optical surveys Sanders et al. 2007. Deep radio diagnostics from COSMOS-XS trace dust-independent star formation via synchrotron radiation and identify AGN activity through radio-excess signatures Vlucht et al. 2020. Finally, the Chandra COSMOS-Legacy X-ray survey provides direct evidence of SMBH accretion, identifying both obscured and unobscured AGN through high-energy diagnostics. Together, these observations provide a multi-wavelength census of the energy output driven by both stellar and black hole growth Civano et al. 2016.

Comparison Approach

The validation strategy relies on population-level comparisons between mock catalogs and observed data. Rather than fitting individual galaxies, we employ a forward-modeling approach. This approach tests consistency across redshift and mass by validating that the predicted mass functions and redshift distributions ($n(z)$) statistically match the multi-wavelength samples. We specifically cross-check AGN classifications by comparing the model's predicted f_{AGN} against observed X-ray excesses and MIR color-wedge selections. Furthermore, the framework enables testing of known physical relations, such as the FIR–radio correlation we compare the model's predicted long-wavelength fluxes against the observed ratios in radio-selected samples to ensure the dust-radio nexus is correctly captured. This comprehensive comparison accounts for selection effects and photometric noise to verify the model's performance across a broad wavelength range.

Interpretation of Agreement and Disagreement

Agreement would mean that the synthetic populations generated by pop-cosmos reproduce the observed far-infrared, radio, and X-ray distributions of galaxies, without requiring wavelength-dependent re-tuning of the underlying SPS assumptions. It confirms that our assumed relations, such as the star-forming sequence, are consistent across different emission mechanisms. On the other hand, disagreement likely reveals residual systematics in the photometry or fundamental limitations in the SPS parametrization, such as failing to account for dust temperature variations or complex AGN physics. Discrepancies in radio or X-ray spaces would suggest the model requires more careful handling of the data/feedback or non-thermal emission processes Vlucht et al. 2020

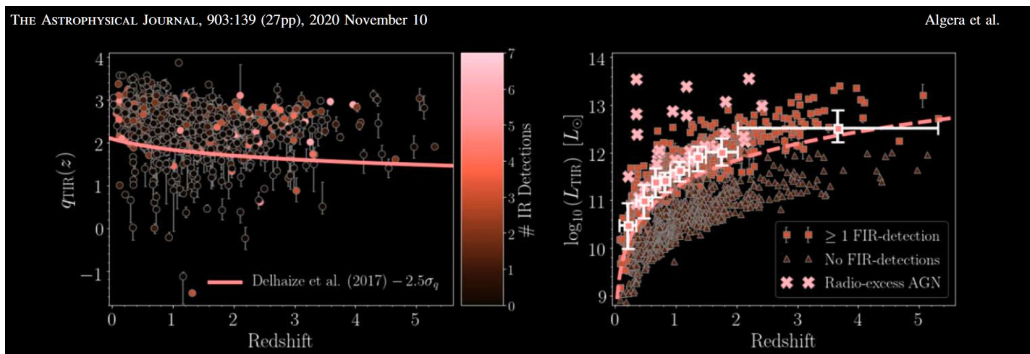


Figure 1: Multiwavelength photometry used for AGN classification.

Overall, we identify a total of 110 radio-excess AGNs via this method. However, this number may be affected by the lack of radio spectral indices for $\sim 80\%$ of the sample.

6 Conclusion

This literature review has outlined the scientific motivation and data landscape for extending the pop-cosmos framework beyond its current optical and near-infrared validation range. While pop-cosmos has been shown to reproduce key galaxy population properties using $0.3\text{--}5\ \mu\text{m}$ photometry, optical and near-infrared data alone cannot fully constrain dust-obscured star formation or black hole accretion. Independent validation using longer-wavelength and high-energy observations is therefore required to test the physical consistency of the model.

The review identified a set of well-established multi-wavelength datasets in the COSMOS field that together provide complementary constraints on galaxy evolution. Far-infrared and sub-millimetre surveys trace dust-reprocessed star formation, radio observations offer dust-independent probes of star formation and AGN activity, and X-ray surveys directly trace accretion onto supermassive black holes. Combined with deep optical and near-infrared catalogues, these datasets enable a comprehensive population-level comparison between model predictions and observations.

By establishing both the motivation and the available observational resources, this review provides a clear foundation for the project work. The next stage will involve generating extended model predictions and performing systematic population-level comparisons across multiple wavelength regimes, with the aim of assessing whether the physical assumptions of pop-cosmos remain valid when confronted with independent tracers of galaxy evolution.

AI Use Declaration

Portions of this document were assisted by AI tools for drafting and editing. Further uses include assistance in understanding some technical details within academic papers and LaTeX formatting.

References

- Alsing, Justin et al. (2024). “pop-cosmos: A Comprehensive Picture of the Galaxy Population from COSMOS Data”. In: *The Astrophysical Journal Supplement Series* 274, p. 12. DOI: 10.3847/1538-4365/ad5c69.
- Civano, F. et al. (2016). “The Chandra Cosmos Legacy Survey: Overview and Point Source Catalog”. In: *The Astrophysical Journal* 819, p. 62. DOI: 10.3847/0004-637X/819/1/62.
- Jin, Shuowen et al. (2018). ““Super-deblended” Dust Emission in Galaxies. II. Far-IR to (Sub)millimeter Photometry and High-redshift Galaxy Candidates in the Full COSMOS Field”. In: *The Astrophysical Journal* 864, p. 56. DOI: 10.3847/1538-4357/aad4af.
- Sanders, D. B. et al. (2007). “S-COSMOS: The Spitzer Legacy Survey of the Hubble Space Telescope ACS 2 deg2 COSMOS Field I: Survey Strategy and First Analysis”. In: *The Astrophysical Journal Supplement Series* 172, pp. 86–98. DOI: 10.1086/517885.

- Shuntov, Marko et al. (2025). “COSMOS2025: The COSMOS-Web galaxy catalog of photometry, morphology, redshifts, and physical parameters from JWST, HST, and ground-based imaging”. In: *Astronomy and Astrophysics* 704, A339. DOI: 10.1051/0004-6361/202555799.
- Thorp, Stephen, Justin Alsing, et al. (2024). “pop-cosmos: Scaleable Inference of Galaxy Properties and Redshifts with a Data-driven Population Model”. In: *The Astrophysical Journal* 975, p. 145. DOI: 10.3847/1538-4357/ad7736.
- Thorp, Stephen, Hiranya Peiris, et al. (2025a). *pop-cosmos: COSMOS-like mock galaxy catalogs*. DOI: 10.5281/zenodo.15622325.
- (2025b). *pop-cosmos: Galaxy property and redshift catalog for COSMOS2020*. DOI: 10.5281/zenodo.17426655.
- Thorp, Stephen, Hiranya V. Peiris, et al. (2025). “pop-cosmos: Insights from generative modeling of a deep, infrared-selected galaxy population”. In: *The Astrophysical Journal* 993.2, p. 240. DOI: 10.3847/1538-4357/ae0936.
- Vlugt, D. van der et al. (2020). “A Multiwavelength Analysis of the Faint Radio Sky (COSMOS-XS): the Nature of the Ultra-faint Radio Population”. In: *The Astrophysical Journal* 903, p. 139. DOI: 10.3847/1538-4357/abb77a.
- Wang, Lingyu et al. (2024). “Probabilistic and progressive deblended far-infrared and sub-millimetre point source catalogues. I. Methodology and first application in the COSMOS field”. In: *Astronomy and Astrophysics* 688, A20. DOI: 10.1051/0004-6361/202349055.