universe

MDPI

Review

Integrated Galactic Archaeology: An Inverse-Problem Framework for Galaxy Evolution

Tsutomu T. Takeuchi [1,2,*], Karin T. Sakuragi [1], Ryusei R. Kano [1,3] and Sena A. Matsui [1]

[1] Division of Particle and Astrophysical Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan; sakuragi.karin.s4@s.mail.nagoya-u.ac.jp (K.T.S.); kano.ryusei.z5@s.mail.nagoya-u.ac.jp (R.R.K.); sena.ma@nagoya-u.jp (S.A.M.)
[2] The Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa 190-8562, Japan
[3] Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
* Correspondence: tsutomu.takeuchi.ttt@gmail.com

Abstract

Integrated-light spectral energy distribution modeling is widely used to infer the star-formation and assembly histories of galaxies that cannot be resolved into individual stars. However, existing approaches are often discussed primarily in terms of particular fitting codes, star-formation-history parameterizations, or inference algorithms. In this Review, we formulate the recovery of galaxy evolution histories from integrated spectral energy distributions as a unified inverse problem. We separate the physical spectral-generation operator from the observational operator and examine the resulting information loss through non-identifiability, singular-value structure, null directions, effective resolution, regularization, and model discrepancy. We then classify parametric and nonparametric star-formation histories, PCA, MOPED, VESPA, non-negative matrix factorization, deep learning, and simulation-based inference within a common five-component framework consisting of the representation space, forward operator, physical or statistical constraints, inference method, and uncertainty assessment. On this basis, we introduce information-driven adaptive representation as a general design principle in which the complexity of the recovered history is matched to the information supported by the observations. Finally, we extend the framework from star-formation histories to coupled galaxy-evolution states involving chemical enrichment, dust evolution, interstellar-medium conditions, and radiative transfer, and outline a three-layer research program linking controlled mock experiments, inverse-problem theory, and physical forward modeling.

Keywords: galaxy evolution; spectral energy distributions; stellar populations; star formation histories; inverse problems; dust evolution; radiative transfer; statistical inference

Check for updates

Academic Editor: Paolo Salucci

Received: 26 July 2026
Revised: 24 August 2026
Accepted: 25 August 2026
Published: 27 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction: From Integrated Light to Galaxy Histories

Understanding how galaxies assembled their stars, enriched their interstellar media, accumulated and expelled gas, formed dust, and changed their internal structures is a central objective of modern astrophysics. Observations across cosmic time provide direct views of galaxies at different evolutionary stages, but each observation is normally a snapshot of a different system. Reconstructing the history of an individual galaxy therefore requires archaeological approaches that infer past evolution from information retained in its present-day stars, gas, dust, dynamics, and radiation [1–3].

In the Milky Way and the nearest galaxies, individual stars can be resolved and characterized through their ages, chemical abundances, kinematics, and spatial distributions. This resolved-star approach provides exceptionally detailed fossil evidence of galaxy assembly and underlies the classical program of galactic archaeology [1–3]. Its reach is nevertheless limited by distance, crowding, sensitivity, and angular resolution. For the vast majority of galaxies, the available information is instead the light integrated over many stellar populations and, frequently, over a large fraction or the entirety of the galaxy.

Integrated spectra and spectral energy distributions (SEDs) therefore provide the principal observational bridge between nearby stellar archaeology and the evolution of galaxy populations over cosmological time. Their interpretation draws on several decades of development in stellar population synthesis, full-spectrum fitting, broadband and narrowband SED modeling, nebular emission, dust attenuation and re-emission, and statistical inference [4–8]. These methods have enabled estimates of stellar mass, star formation rate, mean stellar age, chemical composition, dust content, and broad star formation histories for galaxies whose constituent stars cannot be studied separately [6–9].

The scientific opportunity is substantial. Integrated-light analysis can be applied over large cosmological volumes and across broad ranges of redshift, environment, mass, morphology, and evolutionary state. It can connect the fossil record retained by low-redshift galaxies to direct observations of galaxy populations at earlier cosmic epochs [8,10] and can combine ultraviolet, optical, infrared, and millimeter information within a common physical description. At the same time, integration fundamentally limits what can be reconstructed. The data do not contain a separately labeled contribution from every stellar generation or physical process. They encode a superposition in which many components have already been mixed before the inference begins.

1.1. Integrated-Light Archaeology and Its Scientific Reach

Building on the established concept of Galactic archaeology and on the literature devoted to recovering stellar populations from unresolved galaxy light, we use the term Integrated Galactic Archaeology to denote a unified framework for inferring galaxy-formation and evolution histories from spatially unresolved spectra and SEDs [1,3,6–8]. The framework extends the archaeological objective of reconstructing past evolution from present-day observables to galaxies for which individual stars cannot be resolved, while treating the information loss inherent in integrated-light measurements as a central part of the problem. The term is not intended to replace the established fields of stellar population synthesis, full-spectrum fitting, or panchromatic SED modeling, but to emphasize their common archaeological objective within an explicit inverse-problem framework.

The term "integrated" has three related meanings in this framework. First, the observed radiation is spatially integrated over unresolved stellar populations and interstellar environments within an aperture or across a galaxy. Second, a galaxy SED is a temporally integrated response to past star formation and to the associated chemical, gaseous, and dust evolution. Third, the framework integrates physical forward modeling, observational response, inverse-problem theory, statistical inference, and uncertainty assessment within one methodological description.

The temporal integration embodied in an SED is the source of both the power and the limitations of the method. Long-lived stars retain information about early star formation, whereas short-lived massive stars, ultraviolet emission, and nebular radiation provide sensitivity to recent activity over tracer-dependent timescales [11]. Metal absorption features, continuum breaks, emission lines, and infrared dust radiation respond to partially different components and timescales [6–8]. Combining them can constrain evolution over intervals

far longer than any direct monitoring program [6–8], but the information is encoded through overlapping responses rather than through an unambiguous temporal sequence.

The observed galaxy SED is shaped by more than the star formation history alone. At the physical level, it also reflects the metallicities and abundance patterns of the stellar populations, the demographics and evolution of binary systems, the state of ionized gas, the amount and composition of dust, the interstellar radiation field, active galactic nuclei and other nonstellar components where present, and the spatial arrangement of stars, gas, and dust. Relating these physical properties to the observed SED requires theoretical ingredients such as an assumed initial mass function, stellar evolutionary tracks, stellar spectral libraries, binary-evolution prescriptions, and models of nebular emission, dust processing, and radiative transfer [6–8]. These components are linked by galaxy evolution: star formation alters the stellar mass and radiation field, stellar evolution returns gas and metals to the interstellar medium, gas flows regulate subsequent star formation and enrichment, and dust both responds to and modifies the radiation from which these histories are inferred.

Integrated spectra should consequently be understood as compressed records of coupled galaxy evolution. Increasing wavelength coverage, spectral resolution, or signal-to-noise ratio can reveal additional information, but it does not generally transform the mapping from history to observation into a one-to-one correspondence. Distinct evolutionary histories can remain observationally similar after their radiation has been temporally superposed, spatially mixed, processed by gas and dust, and projected into a finite set of measured quantities. The central methodological problem is therefore not only to find a history that reproduces the observed SED, but to determine which features of that history are genuinely distinguishable in the available data.

1.2. Galaxy-History Reconstruction as an Inverse Problem

Recovering a galaxy history from integrated light is not a direct conversion from measured fluxes to past physical states. It is an inverse problem: the observations are produced only after the underlying physical histories have been encoded through stellar evolution, chemical enrichment, nebular processing, dust attenuation and emission, radiative transfer, cosmological propagation, and instrumental measurement. Reversing this sequence requires more than a flexible fitting function or a powerful optimization algorithm.

Throughout this Review, we use the term *latent* only for a quantity that is not directly observed but is included in a generative or inferential model and constrained through its observable consequences. This statistical status does not by itself determine physical interpretation: an underlying galaxy history is a physical target, a basis coefficient is a model coordinate, and a learned representation is an internal construction whose coordinates need not correspond one-to-one to physical quantities.

Different histories can generate similar observations for several physical reasons. Stellar populations with neighboring ages often have closely related spectral signatures, particularly after their light has been combined over broad age intervals [7,12,13]. Changes in age can partially compensate for changes in metallicity, while attenuation can mimic changes in age or recent star formation over restricted wavelength ranges [7,12,14,15]. Bursts, periods of quiescence, and smooth variations in star formation can also become difficult to distinguish after temporal integration and projection into a limited number of spectral pixels or photometric bands [13,15–17]. Luminosity weighting introduces an additional asymmetry: a comparatively small mass fraction of young stars can dominate the ultraviolet or optical light and obscure the contribution of an older, more massive stellar population [14,15].

A formally acceptable fit does not remove these ambiguities. An incorrect history can reproduce an observed SED through compensating changes in metallicity, attenuation, nebular emission, redshift, or other nuisance quantities. Conversely, a physically plausible history can be recovered poorly when its representation is too restrictive, when the adopted temporal basis is poorly matched to the information in the data, or when the physical forward model omits relevant processes. Comparisons among contemporary SED-fitting implementations show that common observations can yield materially different estimates, particularly for star formation rate and dust attenuation, when the adopted SFH, nebular, dust, and active-galactic-nucleus prescriptions differ [9]. Agreement between a predicted and observed SED is therefore necessary, but it is not sufficient to establish that the inferred history is accurate or uniquely supported.

At the level of mappings between physical and observational spaces, we write the complete data-generating process schematically as

$$\boldsymbol{y} = \mathcal{O}_{\boldsymbol{\omega}}\left[\mathcal{F}_{\mathrm{phys}}\left(\boldsymbol{h};\boldsymbol{\theta}_{\mathrm{phys}}\right)\right] + \delta_{\mathrm{mod}} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{h}$ denotes the coupled galaxy-evolution histories relevant to the adopted model. The physical operator $\mathcal{F}_{\mathrm{phys}}$ maps those histories to emergent rest-frame radiation under assumptions $\boldsymbol{\theta}_{\mathrm{phys}}$. The observational operator $\mathcal{O}_{\boldsymbol{\omega}}$ maps that radiation into the actual data space under observing configuration $\boldsymbol{\omega}$, including cosmological propagation, wavelength coverage, instrumental response, finite spectral resolution, aperture, masking, and calibration. The term $\boldsymbol{\epsilon}$ represents stochastic measurement error, whereas δ_{mod} represents the difference between the adopted description and the physical and observational processes that generated the data.

Unless otherwise stated, the operator equations in this Review specify the structure and composition of mappings rather than complete computational formulas. Each operator maps inputs defined on physical or statistical spaces, with their associated variables and units, to outputs in a generally different space. For example, an SFH with units of formed stellar mass per unit time is mapped through stellar population synthesis to a luminosity density and then through the observational operator to fluxes or counts in measured channels. The normalization, integration measures, unit conversions, discretization, and internal numerical implementation of each mapping are supplied by the particular physical or observational model. An equality such as Equation (1) is understood in the final data space, so that $\boldsymbol{y}$, δ_{mod}, and $\boldsymbol{\epsilon}$ have the same channel-by-channel units, whereas intermediate inputs and outputs need not share units or coordinate variables. The Jacobians, singular values, and conditioning measures introduced later become quantitative only after the input and output coordinates, coefficient normalization, physical units, covariance model, and reference state have been fixed; they characterize a specified dimensional realization of these mappings rather than invariant numerical properties of the symbolic operators alone.

Separating the physical and observational operators is scientifically important. The historical information available from one intrinsic galaxy changes when the galaxy is placed at a different redshift, observed through different filters, measured over a different wavelength interval, or studied at different spectral resolution and precision. Broadband integration can remove local spectral structure that remains accessible to spectroscopy, while missing wavelength intervals can eliminate responses needed to distinguish particular physical alternatives. At the population level, detection, targeting, measurement success, and quality cuts additionally determine which galaxies enter the analyzed sample.

Equation (1) also separates measurement precision from model adequacy. Increasing the signal-to-noise ratio reduces $\boldsymbol{\epsilon}$ but does not necessarily reduce δ_{mod}. At sufficiently high data quality, limitations in stellar evolutionary tracks, spectral libraries, binary

evolution, abundance responses, nebular modeling, dust geometry, radiative transfer, or instrumental calibration can dominate the residual structure and bias the inferred history [7,18,19]. A reconstruction can therefore be statistically precise while remaining physically inaccurate.

The inverse problem is further conditioned by the representation assigned to the unknown history. A low-dimensional parametric SFH can provide stable estimates when its functional family adequately approximates the relevant evolution, but it can produce persistent bias when the true history lies outside that family. A flexible nonparametric representation reduces this particular restriction, yet its additional coefficients are not automatically constrained by the observations. Weakly informed directions can instead be determined by bin placement, basis normalization, non-negativity, regularization, or the prior distribution [17].

The appropriate inferential target is consequently not an arbitrarily detailed pointwise history. Integrated data constrain combinations of historical variations that produce distinguishable changes in the observations. Some combinations are strongly observable; others are weak and remain sensitive to nuisance freedom, prior structure, or model discrepancy; still others are effectively invisible under the available observing configuration. The number of fitted coefficients therefore need not equal the number of independently measured historical quantities [20,21].

A reconstructed SFH should accordingly be interpreted as a data-, model-, and estimator-dependent projection or averaging of the underlying history. Its effective resolution can vary with lookback time, galaxy state, dust content, chemical composition, redshift, wavelength coverage, and data quality. Fine structure in the plotted representation need not correspond to equally fine temporal information. The relevant questions are which historical combinations are supported, over what resolution they are measured, and which assumptions determine the remaining structure.

This inverse-problem formulation does not diminish the scientific value of integrated-light analysis. It provides a framework for distinguishing robust historical information from details selected by one representation or physical model. Replacing the question "Which detailed history gives the best fit?" with "Which evolutionary combinations are distinguishable under the adopted physical and observational description?" moves integrated-light archaeology from formal reconstruction toward calibrated physical inference.

1.3. Integrated Galactic Archaeology: Scope and Organization of This Review

Integrated Galactic Archaeology is presented here as a conceptual and methodological framework rather than as a single fitting technique. Its central premise is that integrated galaxy spectra contain compressed information about coupled evolutionary histories and that the interpretation of this information requires the physical model, observational response, historical representation, inferential structure, and uncertainty assessment to be considered together. Concretely, an integrated-light reconstruction first chooses what physical history is to be described, for example the star formation rate in a set of lookback-time intervals. Stellar-population, nebular, and dust calculations then predict the emitted SED, while distance, redshift, aperture, bandpass or spectral response, calibration, and noise connect that SED to the measured fluxes or counts. Physical restrictions and statistical assumptions determine which histories are allowed or favored, an optimizer or sampler uses the data to estimate them, and validation tests which derived quantities can be reported reliably. The five-component framework introduced below assigns these familiar modeling steps to distinct parts of the inverse problem.

Throughout this Review, reconstruction methods are compared through five components: the representation of the galaxy history or state, the physical and observational

forward model, the physical constraints or statistical structure added to the problem, the inference or compression procedure, and the characterization and validation of uncertainty. These components are mutually dependent. Changing the historical representation changes the discretized response; changing the physical model changes the observable degeneracies; changing the prior changes the completion of weak directions; and changing the observing configuration changes which combinations can be distinguished. Section 5 formalizes this comparison and applies it to representative method families.

A particular emphasis of this Review is the distinction between nominal model freedom and data-supported historical complexity. Adding time bins, basis coefficients, or additional modeled variables does not by itself increase historical resolution. When their observable responses are nearly indistinguishable, the additional freedom is controlled mainly by regularization, priors, physical boundaries, or model discrepancy. This motivates information-driven adaptive representation: the principle that the dimension and localization of the reported history should be matched to the information supported by the combined physical and observational response. We formulate this as a design principle rather than introducing a particular adaptive algorithm or claiming a completed implementation.

The framework also extends beyond SFH as the only unknown history. Stellar radiation depends jointly on star formation, chemical enrichment, abundance patterns, nebular conditions, dust evolution, the interstellar radiation field, and radiative-transfer geometry. Allowing these components to vary together exposes coupled observable and null directions that are hidden when they are fixed separately. The objective is not to reconstruct every component history point by point, but to identify physically admissible and observationally supported combinations of galaxy evolution.

Reliable development of this framework requires interaction among controlled experiments, inverse-problem analysis, and physical forward modeling. Controlled experiments establish where reconstruction and uncertainty fail; inverse-problem diagnostics explain those outcomes in terms of observable modes, resolution, regularization, and discrepancy; and improved forward models determine whether the inferred states correspond to credible physical evolution. The detailed procedures and feedback among these roles are developed in Section 8. Figure 1 summarizes the conceptual structure of Integrated Galactic Archaeology developed throughout this Review.

The remainder of this Review proceeds from the data-generating process to its inversion. Section 2 develops the physical and observational forward problem. Section 3 defines observability, degeneracy, and effective historical resolution. Section 4 examines representations, physical constraints, regularization, and priors. Section 5 compares existing approaches within the common five-component taxonomy. Section 6 develops the adaptive representation principle. Section 7 extends the inferential target to coupled galaxy-evolution states. Section 8 organizes controlled validation, inverse-problem diagnosis, and forward-model development. Finally, Section 9 considers the implications for multiwavelength observations, survey design, and calibrated historical inference.

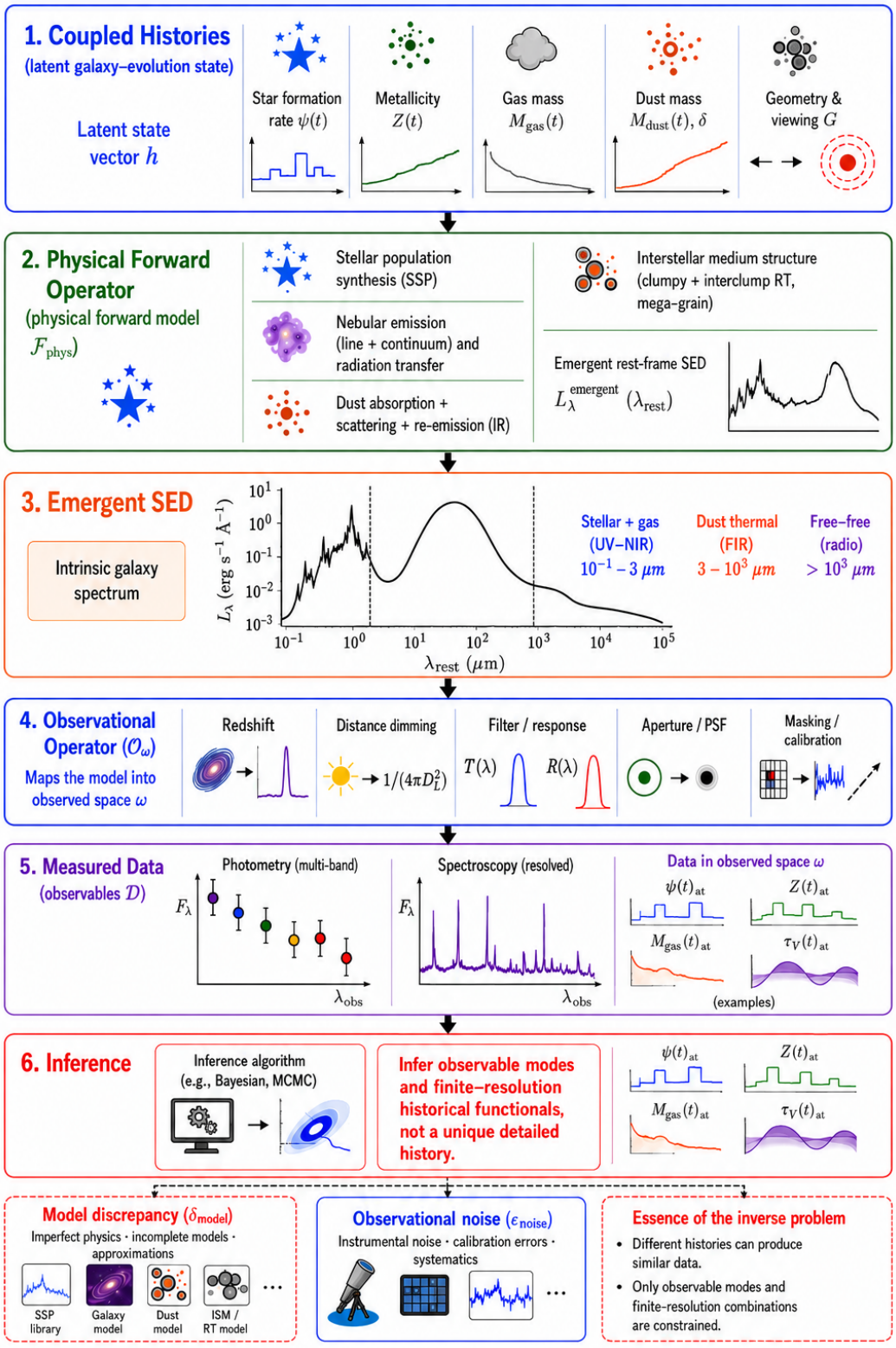


Figure 1. Conceptual overview of integrated galactic archaeology. Coupled histories of star formation, chemical enrichment, gas, dust, and geometry are mapped by the physical forward operator into an emergent rest-frame SED and then by the observational operator into photometric or spectroscopic data. Observational noise and model discrepancy enter as distinct limitations of this mapping. The inverse problem does not generally identify a unique detailed history, but constrains observable modes and finite-resolution historical functionals under the adopted physical, observational, and inferential model. The displayed spectral components and wavelength ranges are schematic.

https://doi.org/10.3390/universe12090259

2. The Physical Forward Problem

The inverse interpretation of integrated galaxy light begins with a forward question: given a galaxy's formation and evolution histories, what radiation should be produced, and how should that radiation appear in a particular observation? Every conclusion about observability, temporal resolution, regularization, or uncertainty depends on the structure and adequacy of this forward description. The forward problem therefore defines not only the predicted SED, but also the physical meaning of the unknown quantities and the pathways through which they affect the data.

For clarity, we separate the complete mapping introduced in Equation (1) into two stages. The physical forward operator generates emergent rest-frame radiation from the galaxy's stellar, chemical, gaseous, dust, and geometrical histories. The observational operator then maps that radiation into the measured data through cosmological propagation, finite wavelength coverage and resolution, instrumental response, aperture and point-spread-function effects, masking, and calibration. Population analyses additionally require an explicit selection model. This separation is important because a physical signature can be weakened by observation, while an apparent residual can originate in either the physical or observational description.

The physical model is itself hierarchical. Stellar population synthesis converts star formation and enrichment into stellar radiation; ionizing photons are reprocessed into nebular line and continuum emission; dust absorbs, scatters, and re-emits radiation; and geometry couples these components through radiative transfer. The aim of this section is not to reproduce the detailed derivation of every implementation, but to identify the physical and observational components that must be specified before the inverse problem can be interpreted.

2.1. Galaxy-Evolution States and Stellar Population Synthesis

A galaxy observed at time $t_{\rm obs}$ is the outcome of a continuous and coupled evolutionary trajectory. A schematic representation of this coupled physical history may be written as

$$\boldsymbol{h}(t) = \left[\psi(t), Z_\star(t), \boldsymbol{a}_\star(t), M_{\rm g}(t), \boldsymbol{Z}_{\rm ISM}(t), \boldsymbol{q}_{\rm ISM}(t), \boldsymbol{d}(t), \boldsymbol{g}(t)\right], \qquad 0 \le t \le t_{\rm obs}, \tag{2}$$

where $\psi(t)$ is the SFH, $Z_\star(t)$ is the total metallicity of the stars formed at time t, and $\boldsymbol{a}_\star(t)$ specifies their elemental abundance pattern, namely how the total metal content is distributed among elements relative to the reference mixture adopted by the stellar-population model. The dimension and entries of $\boldsymbol{a}_\star(t)$ are model-dependent: they can include, for example, $[\alpha/\mathrm{Fe}]$ or selected element-by-element abundance ratios, while a scaled-solar SSP library fixes $\boldsymbol{a}_\star(t)$ rather than treating it as an inferred history. $M_g(t)$ is the gas content, and $\boldsymbol{Z}_{\rm ISM}(t)$ is the total elemental-abundance vector of the star-forming interstellar material before its partition between gas and dust. The gas-phase abundance vector $\boldsymbol{Z}_{\rm gas}(t)$ is derived after depletion into the dust state and enters the nebular-emission calculation. The remaining ISM state $\boldsymbol{q}_{\rm ISM}(t)$ can include density, ionization state, pressure, and radiation-field intensity; $\boldsymbol{d}(t)$ describes the dust mass, composition, grain-size distribution, depletion, and heating state; and $\boldsymbol{g}(t)$ describes the geometry relevant to attenuation and radiative transfer.

Throughout this Review, $\boldsymbol{h}(t)$ denotes the complete coupled galaxy-evolution state, whereas $\psi(t)$ is one component of that state. Under the adopted mixing and depletion model, the stellar birth composition specified by $Z_\star(t)$ and $\boldsymbol{a}_\star(t)$ is inherited from the star-forming material represented by $\boldsymbol{Z}_{\rm ISM}(t)$, while $\boldsymbol{Z}_{\rm gas}(t)$ denotes the portion remaining in the gas phase after depletion into dust. These quantities should not be identified with one another without specifying the chemical and dust model that relates them.

https://doi.org/10.3390/universe12090259

Equation (2) is a conceptual state description, not a claim that every listed function can be independently recovered from one integrated SED. A quantity can be fitted explicitly, represented by a few parameters, derived from an evolution model, or marginalized as nuisance structure. The decomposition therefore determines both the flexibility of the forward model and the physically admissible trajectories it permits. Star formation, enrichment, gas flows, dust processing, and the radiation field are coupled by mass, metal, and energy exchange; treating them as independent functions can admit histories that are mathematically flexible but physically inconsistent.

Stellar population synthesis supplies the fundamental map from the SFH to stellar radiation. Let $\ell_\lambda^{\mathrm{SSP}}(\tau, Z, \boldsymbol{a}; \boldsymbol{\theta}_{\mathrm{SSP}})$ denote the rest-frame luminosity density per unit initial stellar mass of a simple stellar population of age τ, total metallicity Z, and abundance pattern $\boldsymbol{a}$. Here, $\boldsymbol{a}$ denotes the SSP abundance-pattern coordinates corresponding to the model-dependent components specified above. The ingredients $\boldsymbol{\theta}_{\mathrm{SSP}}$ include the initial mass function, evolutionary tracks or isochrones, stellar libraries, mass loss, remnants, short-lived phases, and, where modeled, binary evolution [5,22]. The intrinsic stellar luminosity at the observation time is

$$L_\lambda^\star(t_{\mathrm{obs}}) = \int_0^{t_{\mathrm{obs}}} \psi(t)\, \ell_\lambda^{\mathrm{SSP}}(t_{\mathrm{obs}} - t, Z_\star(t), \boldsymbol{a}_\star(t); \boldsymbol{\theta}_{\mathrm{SSP}})\, \mathrm{d}t. \tag{3}$$

For fixed composition histories and SSP ingredients, this is a linear integral operator acting on ψ. It is a strict time-translation convolution only under additional stationarity assumptions; evolving composition makes the kernel depend explicitly on formation time.

The same integral displays the first fundamental compression of historical information. Each formation epoch contributes over a broad wavelength range, and neighboring SSP ages can produce similar spectra. A localized SFH perturbation therefore generates a distributed response whose strength depends on age, composition, wavelength coverage, and spectral resolution. If enrichment or SSP ingredients are varied jointly with the SFH, the complete mapping becomes nonlinear and age, metallicity, abundance, and stellar-model effects can partially compensate one another [18]. Uncertainty in the SSP kernel is consequently a direct source of forward-model discrepancy rather than a purely numerical error.

2.2. Chemical Enrichment and Nebular Reprocessing

The stellar composition histories in Equation (3) are the outcome of gas consumption, stellar mass return, nucleosynthetic yields, mixing, inflow, and outflow. A physical forward model must therefore specify whether enrichment is fitted freely, inherited from a gas history, or generated by a chemical-evolution model [6,7]. For element X, a schematic total-ISM budget is

$$\frac{\mathrm{d}M_X^{\mathrm{ISM}}}{\mathrm{d}t} = -Z_{\mathrm{ISM},X}(t)\,\psi(t) + E_X\left[\boldsymbol{h}; \boldsymbol{\theta}_{\mathrm{yield}}\right](t) + Z_{\mathrm{in},X}(t)\,\dot{M}_{\mathrm{in}}(t) - Z_{\mathrm{out},X}(t)\,\dot{M}_{\mathrm{out}}(t), \tag{4}$$

where $M_X^{\mathrm{ISM}} = M_{\mathrm{g}} Z_{\mathrm{ISM},X}$, E_X is the stellar return rate, and $\dot{M}_{\mathrm{in}}$ and $\dot{M}_{\mathrm{out}}$ denote gas inflow and outflow. The equation is deliberately schematic: delayed recycling, multiphase mixing, differential metal loading, and spatial transport can require a more detailed state. Grain accretion and destruction redistribute an element between gas and dust and are therefore internal transfers in the total elemental budget, while they directly change $\boldsymbol{Z}_{\mathrm{gas}}$ and the dust state.

Chemical composition affects the stellar SED through both stellar evolution and atmospheric opacity. Total metallicity compresses a detailed abundance pattern, and different elements influence different spectral features. A chemically consistent model can exclude

combinations of SFH and composition that violate its enrichment assumptions, but the resulting separation is then partly model-assisted. Conversely, allowing stellar and gas-phase compositions to vary independently exposes additional degeneracies that the observations may not resolve.

Massive stars also produce photons capable of ionizing hydrogen, helium, and heavier species. Their production rate can be summarized by $Q_{\rm H} = \int_{\nu_{\rm H}}^{\infty} L_\nu^\star/(h_{\rm P}\nu)\,{\rm d}\nu$, while the fraction available to ionize the modeled gas depends on ionizing escape and absorption by dust. The absorbed photons are reprocessed into recombination and collisionally excited lines together with free–free, free–bound, and two-photon continuum emission. We write the nebular component schematically as

$$L_\lambda^{\rm neb} = \mathcal{F}_{\rm neb}\left[L_\nu^\star(\nu \geq \nu_{\rm H}), \boldsymbol{Z}_{\rm gas}(t_{\rm obs}), \boldsymbol{q}_{\rm ISM}(t_{\rm obs}), \boldsymbol{g}_{\rm neb} \,\middle|\, \boldsymbol{\theta}_{\rm neb}\right]. \tag{5}$$

Here, $\mathcal{F}_{\rm neb}$ is a dimensional mapping from an ionizing stellar luminosity density per unit frequency, together with gas-state and geometrical inputs, to an emergent nebular luminosity density per unit wavelength. The use of $L\nu^\star$ as an input and $L_\lambda^{\rm neb}$ as an output therefore does not identify quantities with the same units; the spectral conversion, integration, normalization, and radiative calculation are part of $\mathcal{F}_{\rm neb}$. The vector $\boldsymbol{g}_{\rm neb}$ represents the relevant geometry, while $\boldsymbol{\theta}_{\rm neb}$ includes atomic data, ionizing-source models, stopping conditions, and prescriptions for diffuse gas, shocks, or other excitation sources [7,23].

Nebular emission often provides strong sensitivity to recent star formation, but the conversion between line luminosity and recent SFR is conditional on the initial mass function, metallicity, binary evolution, ionizing-photon escape, dust absorption, and gas conditions. In quiescent systems, evolved stars, shocks, or an active nucleus can contribute to the ionizing field. Strong lines and nebular continuum can also alter broadband fluxes; omitting them can force stellar age, mass, metallicity, or attenuation to absorb the missing signal, while an excessively flexible nebular component can absorb stellar information.

2.3. Dust Evolution, Attenuation, and Infrared Re-Emission

Let $L_\lambda^{\rm src} = L_\lambda^\star + L_\lambda^{\rm neb} + L_\lambda^{\rm other}$ denote the radiation generated before dust processing, where the final term can include an active nucleus, shocks, or other sources when required. A reduced attenuation description can be written as $L_\lambda^{\rm att} = L_\lambda^{\rm src}\exp[-\tau_\lambda^{\rm eff}]$, but the effective optical depth is not a grain-extinction law alone. It also contains source–dust geometry, scattering, orientation, and aperture dependence. In particular, young stars and their nebular regions can remain embedded in dense birth clouds while older populations experience mainly diffuse attenuation, creating an intrinsically age-dependent response [24–26].

The energy absorbed from the source spectrum powers infrared dust emission. In a closed and internally consistent energy-balance description,

$$L_{\rm abs} \equiv \int_0^\infty \left[L_\lambda^{\rm src} - L_\lambda^{\rm esc,src}\right]{\rm d}\lambda, \qquad \int_0^\infty L_\lambda^{\rm dust}\,{\rm d}\lambda \simeq L_{\rm abs}. \tag{6}$$

This relation constrains the integrated energy, not the unique dust temperature distribution, grain composition, heating population, or geometry. Similar infrared luminosities can arise from different dust masses, radiation fields, and spatial configurations [7,27]. Radiative-transfer studies of edge-on spiral galaxies have further revealed a dust energy-balance problem in which models that reproduce optical attenuation can underpredict far-infrared and submillimeter emission. Using mock observations of hydrodynamical galaxy simulations, Saftly et al. showed that smooth axisymmetric fits can miss dust concealed by complex, asymmetric, and inhomogeneous star–dust structure, thereby contributing to this discrepancy [28].

A representative chemically consistent construction builds on the dust-evolution framework of Asano et al. [29] and its SED and radiative-transfer developments by Nishida et al. [30] and Kano et al. [31]. At the level required here, the coupled evolution can be summarized as

$$\left[M_{\rm g}(t), \boldsymbol{Z}_{\rm ISM}(t), \boldsymbol{Z}_{\rm gas}(t), \boldsymbol{d}(t)\right] = \mathcal{E}_{\rm chem+dust}\left[\psi, \dot{M}_{\rm in}, \dot{M}_{\rm out}, \boldsymbol{q}_{\rm ISM} \,\middle|\, \boldsymbol{\theta}_{\rm chem+dust}\right](t), \quad (7)$$

where stellar metal and dust injection, supernova destruction, accretion of gas-phase metals, shattering, and coagulation jointly determine the gas–dust partition and grain-size distribution. The resulting dust state supplies the opacity and emissivity to the radiative-transfer calculation rather than entering as an independent, time-invariant SED template.

A central qualitative result is the nonlinear onset of rapid ISM grain growth after the metallicity reaches a critical regime in which accretion becomes competitive with stellar dust production [29]. The simultaneous evolution of the grain-size distribution changes the attenuation curve and the mid- and far-infrared spectral shape [30]. Compact, optically thick star-forming clumps can further enhance ultraviolet trapping, grain processing, and warm multi-temperature infrared emission [31]. These examples illustrate how chemical state, dust physics, and unresolved geometry can enter one forward model without implying that each internal variable is independently constrained by the SED.

The coupled chemical–dust and radiative-transfer structure of this model family is summarized schematically in Figure 2. Figure 2 should be read as a structural summary rather than as a unique computational sequence: the intermediate metal, dust, grain-size, and geometrical states are generated jointly by the adopted evolution model and need not be independently identifiable from the integrated SED.

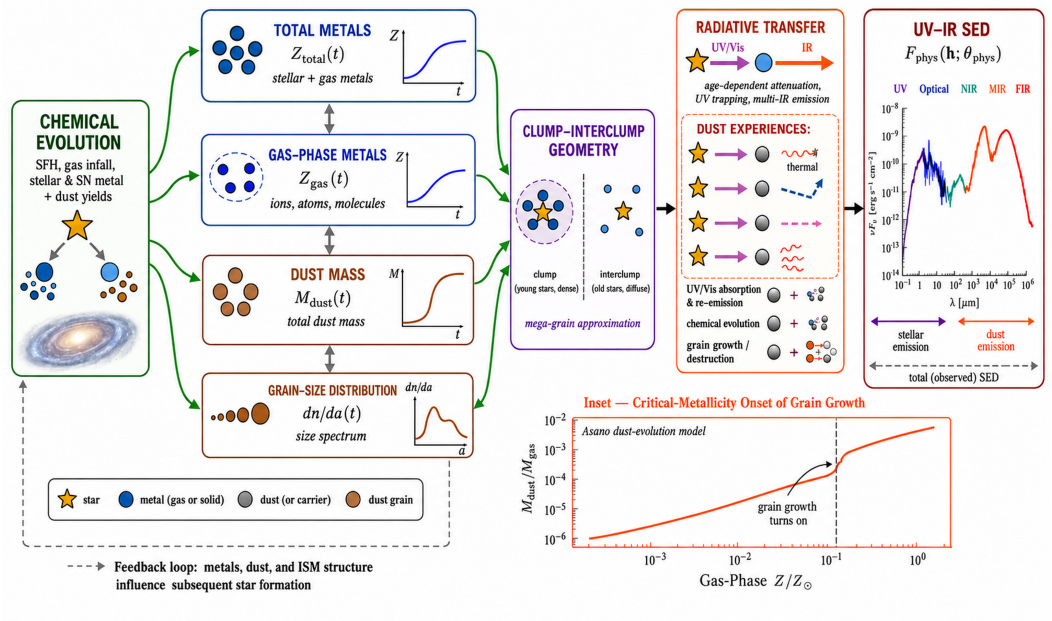


Figure 2. Schematic coupling of chemical evolution, dust evolution, subresolution geometry, and radiative transfer in a chemically consistent galaxy-SED model. Chemical evolution jointly determines the total and gas-phase metal states, dust mass, and grain-size distribution, which are coupled by grain growth, shattering, coagulation, and supernova destruction. Efficient interstellar grain growth produces a rapid increase in dust content after a critical metallicity is reached, as illustrated in the inset. The evolved dust state enters a clump–interclump radiative-transfer model that produces age-dependent attenuation, ultraviolet trapping, infrared re-emission, and the emergent UV–IR SED. All curves and internal tracks are schematic.

Energy conservation still holds for the complete physical system, but an energy-balance constraint can fail operationally when the modeled and observed quantities do not refer to the same physical region and the same set of emitting, absorbing, and re-emitting components. This can occur when the assembled data use inconsistent apertures, radiation escapes anisotropically, foreground attenuation or an active nucleus is omitted, or some dust heating originates outside the modeled region. In such cases, the absorbed luminosity inferred within the adopted model need not match the infrared luminosity included in the assembled data, even though energy is conserved by the complete physical system. Energy balance is therefore a useful physical constraint only when the modeled region, source components, dust components, geometry, and observational apertures are mutually consistent; it is not an automatic property of every assembled multiwavelength dataset.

2.4. Geometry, Radiative Transfer, and the Composite Physical Operator

Attenuation and dust emission are two aspects of radiative transfer through a structured medium. Rather than reproducing a particular transfer solver, we denote the mapping by

$$L_\lambda^{\mathrm{em}}(\hat{\boldsymbol{n}}, t_{\mathrm{obs}}) = \mathcal{T}_{\mathrm{RT}}\left[L^\star, L^{\mathrm{neb}}, L^{\mathrm{other}}; \boldsymbol{d}(t_{\mathrm{obs}}), \boldsymbol{g}(t_{\mathrm{obs}}) \,\middle|\, \boldsymbol{\theta}_{\mathrm{RT}}\right]_{\lambda,\hat{\boldsymbol{n}}}, \tag{8}$$

where L_λ^{em} is the emergent rest-frame luminosity in the selected direction. Spatial integration does not remove geometry because young stars, old stars, ionized gas, diffuse dust, and compact obscured regions can undergo different transfers before their radiation is combined. Inclination, aperture, and scattering can therefore change the integrated SED even for the same underlying stellar populations.

An influential line of galaxy-scale radiative-transfer modeling made these spatial distributions explicit rather than absorbing them into a single attenuation law. The models developed by Popescu and collaborators coupled disk and bulge geometries, diffuse dust, star-forming regions, grain heating, and ultraviolet-to-submillimeter reradiation, first in detailed applications such as NGC 891 and spatially resolved ultraviolet–far-infrared analyses of M101, and subsequently in a general library of dust and polycyclic-aromatic-hydrocarbon emission SEDs for disk galaxies [32–34]. These studies established a physically connected treatment in which the attenuation of stellar radiation and the resulting infrared emission depend jointly on the relative distributions of old stars, young stars, diffuse dust, and localized star-forming regions.

Three-dimensional Monte Carlo frameworks, particularly SKIRT, generalized this program to arbitrary geometries, clumpy media, self-consistent non-LTE dust emission, and spatially resolved synthetic observables [35–37]. The redesigned SKIRT 9 framework prepared the code for kinematics, gas and line transfer, and polarization, while subsequent developments implemented self-consistent non-LTE line transfer and produced multiwavelength synthetic-imaging atlases from cosmological hydrodynamical simulations [38–41]. In this class of models, morphology, inclination, clumpiness, and subresolution structure are components of the physical forward problem rather than corrections applied after an integrated SED has been fitted.

Observational programs such as DustPedia provide a complementary basis for this development by assembling aperture-matched multiwavelength photometry and imagery together with morphological information for nearby galaxies [42,43]. Such data occupy an intermediate regime between resolved-star archaeology and single-aperture SED fitting: individual stars remain unresolved, but spatially resolved surface-brightness structure can constrain the distributions of stellar populations, dust, and heating sources.

Within the operator notation used here, spatially resolved or semi-resolved analyses can be represented by enlarging the data vector $\boldsymbol{y}$ to include image pixels or spatial summaries and by making point-spread-function convolution, registration, aperture weighting, and masking explicit in $\mathcal{O}_\omega$. These additional observables can reduce particular geometric degeneracies, but they do not generally render every component of $\boldsymbol{d}$ and $\boldsymbol{g}$ identifiable. A richer geometrical description can also introduce additional nuisance directions, computational approximations, and sources of model discrepancy.

In the chemically consistent dust-model family discussed above, spatial inhomogeneity is represented through a reduced clump–interclump geometry and the mega-grain approximation. Dusty clumps are replaced by effective absorbers and scatterers whose optical properties depend on internal optical depth and the evolved grain-size distribution. Preferential embedding of young stars produces age-dependent attenuation, while dense compact clumps can remain optically thick and alter both ultraviolet escape and infrared reradiation [30,31]. This is not a fully spatially resolved chemo-radiative-transfer calculation, but it demonstrates how chemical evolution, grain physics, and subresolution geometry can be coupled within one predictive operator.

The fully three-dimensional and reduced geometrical approaches therefore provide complementary levels of physical description. The former can exploit resolved morphology and detailed spatial information, whereas the latter can retain selected effects of clumpiness, age-dependent attenuation, and dust evolution within a lower-dimensional evolutionary calculation. Their adequacy should be judged by the observables and historical functionals under study, together with the corresponding validation results, rather than by geometrical complexity alone.

The physical stages can now be assembled into the operator introduced in Equation (1):

$$\begin{aligned} L_\lambda^{\mathrm{em}} &= \mathcal{F}_{\mathrm{phys}}\left[\boldsymbol{h};\boldsymbol{\theta}_{\mathrm{phys}}\right]_\lambda \\ &\equiv \mathcal{T}_{\mathrm{RT}}\left[\mathcal{F}_\star[\psi \mid Z_\star, \boldsymbol{a}_\star, \boldsymbol{\theta}_{\mathrm{SSP}}], \mathcal{F}_{\mathrm{neb}}\left[L^\star, \boldsymbol{Z}_{\mathrm{gas}}, \boldsymbol{q}_{\mathrm{ISM}}, \boldsymbol{g}_{\mathrm{neb}} \mid \boldsymbol{\theta}_{\mathrm{neb}}\right], L^{\mathrm{other}}; \boldsymbol{d}, \boldsymbol{g} \,\middle|\, \boldsymbol{\theta}_{\mathrm{RT}}\right]_\lambda . \end{aligned} \tag{9}$$

The decomposition is structural rather than a mandatory computational order. Some implementations solve chemical, stellar, nebular, and dust evolution jointly, whereas others replace selected stages by tabulated or empirical relations. When Equation (7) is adopted, $\boldsymbol{Z}_{\mathrm{ISM}}$, $\boldsymbol{Z}_{\mathrm{gas}}$, and $\boldsymbol{d}$ are generated jointly rather than supplied as independent SED components.

The mapping is linear in the SFH only conditionally on fixed stellar composition, nebular state, dust state, and geometry. Once these quantities are varied jointly, the SSP kernel, ionizing spectrum, gas response, dust heating, opacity, and transfer solution all change with the represented galaxy state, making the complete physical operator nonlinear. Increasing physical complexity can remove inconsistent states, but it can also introduce new compensation directions and additional model uncertainty. Stellar-only, stellar-plus-nebular, energy-balanced, chemically coupled, and geometry-dependent models should therefore be regarded as a hierarchy whose appropriate level depends on the target observables and validation results.

2.5. The Observational Operator

The output of $\mathcal{F}_{\mathrm{phys}}$ is an emergent rest-frame SED, not the measured data vector. For an unresolved isotropic-equivalent luminosity, a schematic cosmological propagation relation is

$$f_\lambda^{\mathrm{sky}}(\lambda_{\mathrm{obs}}) = \frac{T_{\mathrm{los}}(\lambda_{\mathrm{obs}}; z, \boldsymbol{\eta}_{\mathrm{los}})}{4\pi D_{\mathrm{L}}^2(z)(1+z)} L_\lambda^{\mathrm{em}}\left(\frac{\lambda_{\mathrm{obs}}}{1+z}\right), \tag{10}$$

where D_{L} is the luminosity distance and T_{los} represents foreground or intergalactic transmission. Lensing, anisotropic emission, surface-brightness structure, and related effects can be included in a more general propagation operator.

The complete observational mapping is written as

$$\begin{aligned}\boldsymbol{\mu} &= \mathcal{O}_{\boldsymbol{\omega}}[L_{\lambda}^{\mathrm{em}}] \\ &\equiv \mathcal{M}_{\mathrm{samp}} \circ \mathcal{K}_{\mathrm{cal}} \circ \mathcal{R}_{\mathrm{inst}} \circ \mathcal{A}_{\mathrm{ap,PSF}} \circ \mathcal{P}_{z}[L_{\lambda}^{\mathrm{em}}],\end{aligned} \tag{11}$$

where $\mathcal{P}_z$ denotes cosmological and line-of-sight propagation, $\mathcal{A}_{\mathrm{ap,PSF}}$ denotes aperture and point-spread-function weighting, $\mathcal{R}_{\mathrm{inst}}$ denotes spectral, photometric, and detector response, $\mathcal{K}_{\mathrm{cal}}$ denotes calibration transformations, and $\mathcal{M}_{\mathrm{samp}}$ denotes sampling, masking, and assembly of the final data vector. The configuration $\boldsymbol{\omega}$ contains the parameters and metadata required by these stages.

For an idealized linear instrumental channel, define the aperture-weighted incident flux $f_{\lambda}^{\mathrm{ap}} = \mathcal{A}_{\mathrm{ap,PSF}} \circ \mathcal{P}_{z}[L_{\lambda}^{\mathrm{em}}]$. The corresponding pre-calibration prediction is

$$\mu_i^{\mathrm{inst}} = \int R_i(\lambda) f_{\lambda}^{\mathrm{ap}}(\lambda)\, \mathrm{d}\lambda, \tag{12}$$

where $R_i(\lambda)$ includes the relevant bandpass, line-spread function, detector convention, and pixel integration. Calibration, resampling, masking, and covariance must be applied at the stage where they arise rather than represented automatically by independent errors on the final channels.

Aperture effects are especially important for joint multiwavelength analysis because stellar populations, nebular emission, and dust emission can have different spatial distributions. Measurements from different facilities need not sample the same physical mixture because their apertures, point-spread functions, and spatial weighting can combine stellar, nebular, and dust-emitting regions differently. This spatial-mixture effect is one component of the observational operator. Other components act at distinct stages: redshift and line-of-sight propagation map rest-frame radiation into observed wavelength and flux coordinates; instrumental response and resolution mix or integrate spectral information; and background treatment and calibration transform the recorded signal into the reported data vector. These quantities can be fixed inputs or inferred nuisance parameters; when uncertain, they modify the mapping from emergent galaxy radiation to measured data and therefore belong inside the observational model.

For population inference, the per-object operator is accompanied by a selection model. Detection, targeting, deblending, redshift success, quality cuts, and follow-up can depend on noisy measured quantities. They do not alter the SED of one galaxy, but they change the probability that a given underlying galaxy state enters the analyzed sample and must therefore be included when reconstructed histories are interpreted statistically.

2.6. Measurement Noise and Model Discrepancy

With the physical and observational operators specified, the mean prediction in the actual data space is

$$\boldsymbol{\mu}\left(\boldsymbol{h}, \boldsymbol{\theta}_{\mathrm{phys}}, \boldsymbol{\omega}\right) = \mathcal{O}_{\boldsymbol{\omega}}\left[\mathcal{F}_{\mathrm{phys}}\left(\boldsymbol{h}; \boldsymbol{\theta}_{\mathrm{phys}}\right)\right]. \tag{13}$$

The observed data differ from this prediction due to stochastic measurement error and the inadequacy of the adopted description. A useful conceptual decomposition is

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{\mu} + \boldsymbol{\delta}_{\mathrm{phys}} + \boldsymbol{\delta}_{\mathrm{obs}} + \boldsymbol{\delta}_{\mathrm{num}} + \boldsymbol{\epsilon}, \\ \boldsymbol{\delta}_{\mathrm{mod}} &\equiv \boldsymbol{\delta}_{\mathrm{phys}} + \boldsymbol{\delta}_{\mathrm{obs}} + \boldsymbol{\delta}_{\mathrm{num}},\end{aligned} \tag{14}$$

where the three discrepancy terms represent incomplete physical ingredients, inadequacy in the observational operator, and numerical or emulator error, respectively. This partition need not be identifiable from one dataset; its purpose is to prevent all coherent residual structure from being attributed automatically to the galaxy history.

In the additive decomposition above, $\boldsymbol{\epsilon}$ denotes stochastic measurement error around the modeled mean for numerical measurements and can be heteroscedastic, correlated, or non-Gaussian. Censoring and non-detections are not additional realizations of $\boldsymbol{\epsilon}$: for such data, the additive equation applies to the numerical measurement before censoring, while the reported datum is an event or limit generated by the sampling and assembly stage $\mathcal{M}_{\mathrm{samp}}$. The corresponding likelihood integrates over the unreported measurement values consistent with that event or limit. The broader observational data-generation process therefore comprises the observational operator and its nuisance parameters, the stochastic error model, and any censoring or selection mechanism, rather than being absorbed into $\boldsymbol{\epsilon}$ alone. An uncertain quantity is not automatically a discrepancy: a calibration offset, redshift, line-spread width, or attenuation parameter that is explicitly represented and inferred belongs to the forward model and nuisance state. Discrepancy denotes remaining structure outside the represented state and modeled nuisance parameters.

At high signal-to-noise ratio, coherent errors in stellar libraries, abundance responses, nebular calculations, dust physics, radiative transfer, calibration, or emulation can dominate the uncertainty budget [7,18,19]. In this model-limited regime, increasing the signal-to-noise ratio does not necessarily drive the inference toward the true history; it can concentrate the likelihood more tightly around an incorrect model manifold, increasing the formal significance of a discrepancy-induced bias even as the statistical interval narrows. A flexible SFH can absorb part of that discrepancy and leave a small residual while remaining physically biased.

This section has specified the route from a galaxy-evolution history to measured data. It has not determined which parts of that route can be inverted reliably. Section 3 begins with Equation (13) and asks which perturbations of the represented state survive temporal integration, physical coupling, radiative transfer, observational projection, measurement uncertainty, and credible model discrepancy.

3. Observability, Degeneracy, and Effective Historical Resolution

The physical forward problem developed in Section 2 maps a galaxy's evolutionary histories into an emergent SED and then into a finite set of measured quantities. The inverse problem attempts to infer those histories from the resulting data. Its difficulty does not arise only from observational noise, numerical optimization, or the dimension of the parameter space. It arises more fundamentally because the forward mapping can be non-injective and poorly conditioned: distinct histories can generate indistinguishable observations, and small changes in the observations can correspond to large changes in some reconstructed directions. For example, consider two trial SFHs that move the same amount of star formation between adjacent old lookback-time bins. If stellar evolution, dust attenuation, and instrumental projection make the resulting spectra differ by less than the combined observational and model uncertainty, the data cannot determine which bin received that star formation. By contrast, moving recent star formation between intervals may change ultraviolet fluxes or nebular lines enough to be distinguished. The terms introduced below describe this practical contrast: identifiability asks whether distinct histories can be distinguished at all; stability asks whether a distinguishable response is strong enough to recover reliably; and effective historical resolution specifies the temporal combinations or averages that the data support.

These two limitations should be distinguished. Non-identifiability occurs when different admissible states produce the same, or observationally equivalent, predictions. Instability occurs when the predictions are mathematically different but separated by an amplitude too small to distinguish reliably. The first defines an exact or effective equivalence class of histories; the second makes inversion sensitive to observational noise, model discrepancy, regularization, and prior structure.

Neither limitation is removed by adopting a more powerful optimizer, sampler, or machine-learning architecture. A flexible inference method can represent a larger family of histories consistent with the data, but it cannot recover a direction whose observable response has been eliminated by temporal integration, physical degeneracy, observational projection, or forward-model uncertainty. The relevant scientific question is therefore not whether a fitted history can be produced, but which combinations of historical variations remain distinguishable under the complete physical and observational model.

This section provides the formal description of that distinction. Section 3.1 first introduces a minimal vocabulary for readers who are less familiar with inverse problems and summarizes how the forward mapping creates observational equivalence. Section 3.2 defines local observable modes, Fisher information, null directions, and effective historical resolution. Section 3.3 finally explains why these quantities depend on galaxy state, observing configuration, parameterization, and model credibility. The detailed physical processes responsible for the responses belong to Section 2; the roles of representation, regularization, and priors belong to Section 4; and their empirical validation belongs to Section 8.

3.1. A Minimal Inverse-Problem Primer

A small set of inverse-problem concepts is sufficient to organize the formal analysis that follows. The *forward problem* predicts the data generated by a specified galaxy-evolution state, whereas the *inverse problem* asks which properties of that state can be inferred from the measured data. The latter is not generally an algebraic reversal of the former. Temporal superposition, physical coupling, observational projection, and finite uncertainty can erase or strongly attenuate distinctions among underlying physical states. Consequently, the natural output of an inverse analysis is often an equivalence class, an observable combination, or a finite-resolution functional rather than a unique pointwise history.

Non-identifiability means that distinct admissible states produce the same, or observationally equivalent, predictions. In that case no inference algorithm can select the correct state from the data alone. *Instability* is a related but distinct limitation: two states can produce mathematically different predictions, yet the difference can be so small that noise, calibration uncertainty, or model discrepancy is strongly amplified when the mapping is inverted. Non-identifiability therefore concerns the absence of a distinguishable response, whereas instability concerns the weakness of a response that is formally present.

An *observable mode* is a combination of perturbations in the represented history, physical state, or nuisance parameters that produces a sufficiently distinctive response in the uncertainty-weighted data space. Such a mode need not coincide with one time bin or one physical parameter. It can combine neighboring formation epochs, stellar composition, attenuation, nebular conditions, geometry, and calibration terms. A *null direction* produces no response under the specified local forward model, while an *effectively null direction* produces a response too small to support the intended scientific claim. The observable content of the data is therefore more naturally described by these combinations than by the nominal number of fitted coefficients.

Effective resolution describes how the complete estimator responds to localized structure in the underlying history. It is not the same as the spacing of the input grid. A narrow averaging response supports localized inference, whereas a broad, displaced, or multilobed

response indicates that the reported value mixes activity over a wider range of epochs. In this sense, nominal resolution belongs to the representation, while effective resolution belongs to the combined forward model, observational operator, representation, constraints, and estimator.

Regularization and *prior structure* stabilize or weight the weakly observable directions. They are essential components of a usable inverse analysis, but they do not create observational information that is absent from the data. A smooth reconstruction, a sparse solution, or a narrow posterior can therefore reflect the chosen completion of weak directions rather than a uniquely measured history. Their role is examined explicitly in Section 4.

Finally, observational noise and model discrepancy have different interpretations. Noise describes stochastic variation under the adopted measurement model. Model discrepancy describes coherent inadequacy in the physical, observational, or numerical forward description. A discrepancy aligned with an observable response can be absorbed into a biased history while leaving a small residual, whereas an orthogonal component remains visible as lack of fit. Observability must therefore be judged relative not only to measurement precision but also to the credibility of the model used to interpret the data.

All of these concepts are conditional. They depend on the reference galaxy state, observing configuration, historical representation, coefficient scaling, covariance or likelihood, nuisance freedom, and scientific target. The following subsections formalize these ideas through observational equivalence, local response modes, Fisher information, resolution operators, and discrepancy projections.

This compression is generally stronger at old stellar ages. The spectra of old populations often evolve more slowly with age than those of young populations, while their luminosity per unit formed stellar mass also declines as short-lived stars disappear. A narrow interval of recent star formation can consequently produce strong ultraviolet, blue-continuum, or nebular signatures, whereas an equally narrow interval in the distant past may alter only subtle absorption features or continuum structure. Historical resolution is therefore not expected to be uniform in lookback time.

A related line of work treats the unobserved SFH as a latent stochastic process in the statistical sense defined above and characterizes its variability across temporal scales through the power spectral density (PSD), the autocovariance function, or an equivalent Gaussian-process description [44–47]. In such formulations, the amplitude, slope, and breaks of the PSD describe the intrinsic temporal correlations of star formation, while physically motivated regulator models connect characteristic breaks to gas-inflow variability, the equilibration time of the galactic gas reservoir, and the life cycle of molecular clouds [44,45]. Comparisons among hydrodynamical, semi-analytic, and empirical galaxy-formation models further show that the predicted distribution of SFH variability across timescales is itself model-dependent, particularly on short timescales and as a function of stellar mass [46]. Physically motivated Gaussian-process SFHs translate this temporal covariance into testable distributions of spectral and photometric observables [47].

This intrinsic star-formation stochasticity is related to, but distinct from, the effective historical resolution defined in this Review. Stochasticity is a property of the underlying SFH or of a population distribution over histories, whereas effective resolution is a property of the physical forward operator, observational operator, representation, constraints, and estimator. The complete forward mapping filters and mixes the temporal modes of the stochastic SFH, so the absence of a recovered short-timescale fluctuation does not establish that the true SFH was smooth, and the presence of intrinsic burstiness does not imply that individual bursts are identifiable from one integrated SED.

The concept of observational equivalence can be expressed without requiring the predicted spectra to be exactly identical.

For a stochastic-process representation, the assumed PSD or autocovariance kernel specifies which temporal fluctuations are probable and how activity at different epochs is correlated. A physically motivated kernel can encode characteristic timescales associated with gas inflow, gas-regulator equilibrium, molecular-cloud evolution, or other processes, while remaining a structural component of the model rather than information automatically measured from the SED [44,45,47]. In the five-component notation introduced in Section 5, such a covariance model contributes jointly to the representation $\mathcal{V}_q$ and added structure $\mathcal{C}_q$; whether its hyperparameters are observable depends on the forward model $\mathcal{F}_q$, the observing configuration, data quality, nuisance freedom, and uncertainty assessment. Population-level distributions of indicators with different temporal responses can constrain aspects of the stochastic process even when the realization of each rapid fluctuation in an individual galaxy is not recoverable [44,47].

The term "nonparametric" does not mean that the reconstruction is free of parameterization or assumptions. The number and placement of time bins, the normalization of the basis functions, the support of the coefficients, and the correlations imposed between epochs all define the allowed histories.

Physical variability timescales and inferential resolution should not be equated when designing an adaptive representation. A candidate component can correspond to a physically plausible stochastic timescale yet remain observationally unresolved after stellar-population integration, attenuation, and instrumental projection; conversely, population-level distributions can contain information about variability power without identifying the timing of individual fluctuations [44,47]. Adaptive refinement should therefore be guided by the transmitted response and achieved averaging kernel, not by the nominal timestep of a simulation or by a prior correlation scale alone. Where a stochastic SFH model is used, a transparent analysis should distinguish the intrinsic variability spectrum specified by the model from the temporal transfer and effective resolution imposed by the complete inverse procedure.

The selected bins should still be interpreted through their effective resolution. A component that remains in an adaptive partition is not automatically localized to its nominal boundaries. Its estimator can average the true history over a broader or asymmetric interval.

The mock suite should also span the amplitude and correlation structure of SFH variability rather than only a small set of visually distinct analytic histories. PSD-based comparisons show substantial differences among galaxy-formation models, while recent simulation studies demonstrate that stellar and active-galactic-nucleus feedback can alter the shape and physical interpretation of predicted SFHs [46,48]. The temporal cadence of the generating calculation must be fine enough to represent the shortest variability scale that the validation is intended to test. For example, recent semi-analytic work at cosmic dawn uses merger trees with typical snapshot spacing of approximately 5–10, Myr to follow rapid assembly and diverse histories containing bursts and mini-quenching episodes, and emphasizes the sensitivity of synthetic photometry to very young stellar populations [49]. If short-timescale structure is removed by coarse simulation outputs or SFH binning before the physical and observational operators are applied, a nominal recovery experiment cannot determine whether the inference procedure would have detected that structure. Matched- and mismatched-model tests should therefore vary the intrinsic SFH stochasticity, forward-model cadence, SFH representation, and inference prior separately whenever these distinctions are scientifically relevant.

Comparisons should be made through scientific quantities that remain meaningful across representations. Examples include stellar mass formed over broad intervals, cumulative mass-assembly fractions, recent mean SFRs, mass-weighted ages, broad transition epochs, metallicity summaries, and predictions for held-out observables.

3.2. Observable Modes and Effective Historical Resolution

Let $\boldsymbol{\vartheta}$ denote the coefficients and physical or nuisance parameters varied in a specified local inverse problem, and let $\boldsymbol{\mu}(\boldsymbol{\vartheta})$ denote the corresponding prediction in the actual data space. Around a reference state $\boldsymbol{\vartheta}_0$, the forward response can be linearized as

$$\Delta \boldsymbol{y} \simeq \boldsymbol{J}_0 \Delta \boldsymbol{\vartheta} + \boldsymbol{\delta}_{\mathrm{mod}} + \boldsymbol{\epsilon}, \qquad \boldsymbol{J}_0 \equiv \left. \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^{\mathsf{T}}} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}, \tag{15}$$

where $\Delta \boldsymbol{\vartheta} = \boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0$, $\Delta \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\vartheta}_0)$, $\boldsymbol{\delta}_{\mathrm{mod}}$ denotes model discrepancy, and $\boldsymbol{\epsilon}$ denotes stochastic measurement error. Each column of $\boldsymbol{J}_0$ is the data-space signature produced by perturbing one coordinate of the represented state while holding the remaining coordinates fixed. At this stage, $\boldsymbol{\mu}(\boldsymbol{\vartheta})$ denotes a fully specified prediction in a chosen data-coordinate system, and the components of $\boldsymbol{\vartheta}$ have fixed definitions, normalizations, and physical units. The Jacobian therefore differentiates a particular dimensional implementation of the forward mapping rather than the earlier schematic operator symbol in isolation.

Equation (15) does not assume that the complete SED model is globally linear. It describes the tangent response around one reference state. Two local parameter directions are difficult to separate when their columns, or linear combinations of columns, are nearly parallel. A direction is weak when its response amplitude is small relative to the uncertainty scale used in the inference.

The relevant geometry is set by the covariance of the adopted data-space model rather than by an unweighted Euclidean distance between spectra. For a positive-definite covariance $\boldsymbol{C}$, define

$$\Delta \widetilde{\boldsymbol{y}} = \boldsymbol{C}^{-1/2} \Delta \boldsymbol{y}, \qquad \widetilde{\boldsymbol{J}}_0 = \boldsymbol{C}^{-1/2} \boldsymbol{J}_0, \qquad \widetilde{\boldsymbol{\delta}}_{\mathrm{mod}} = \boldsymbol{C}^{-1/2} \boldsymbol{\delta}_{\mathrm{mod}}, \qquad \widetilde{\boldsymbol{\epsilon}} = \boldsymbol{C}^{-1/2} \boldsymbol{\epsilon}. \tag{16}$$

When $\boldsymbol{C}$ is the covariance of the complete stochastic residual in the adopted data model, that residual has unit covariance in the whitened space. The norm of a predicted perturbation in this space measures its amplitude relative to the uncertainties and correlations used to judge distinguishability.

The singular value decomposition of the whitened Jacobian is

$$\widetilde{\boldsymbol{J}}_0 = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathsf{T}}, \qquad \boldsymbol{\Sigma}_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \qquad \sigma_1 \geq \cdots \geq \sigma_r > 0, \tag{17}$$

where $r = \mathrm{rank}(\widetilde{\boldsymbol{J}}_0)$. The right singular vectors $\boldsymbol{v}_k$ define orthogonal combinations of perturbations in the adopted parameter or history coordinates. The left singular vectors $\boldsymbol{u}_k$ define the corresponding spectral or photometric response patterns in the whitened data space. They satisfy

$$\widetilde{\boldsymbol{J}}_0 \boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k. \tag{18}$$

Within a specified coefficient normalization, a large singular value indicates a direction whose perturbation generates a strong noise-weighted response. A small singular value indicates a direction whose transmitted signature is weak after physical mixing, observational projection, and covariance weighting. The corresponding direction need not be associated with one time bin. It can combine several epochs, chemical variables, dust parameters, calibration terms, or other nuisance quantities.

The relationship to local Fisher information is direct. For a Gaussian data model with parameter-independent covariance, the data contribution to the Fisher information is

$$\boldsymbol{F}_{\mathrm{data},0} = \boldsymbol{J}_0^{\mathsf{T}} \boldsymbol{C}^{-1} \boldsymbol{J}_0 = \widetilde{\boldsymbol{J}}_0^{\mathsf{T}} \widetilde{\boldsymbol{J}}_0 = \boldsymbol{V}_r \boldsymbol{\Sigma}_r^2 \boldsymbol{V}_r^{\mathsf{T}}. \tag{19}$$

The nonzero Fisher eigenvalues are therefore the squared singular values, and the associated eigenvectors are the right singular vectors. This equivalence is local and conditional. It does not establish global identifiability, represent disconnected solution families, or describe posterior boundaries and non-Gaussian structure.

Projecting Equation (15) onto the singular-vector bases gives

$$\begin{aligned} &\beta_k = \sigma_k \alpha_k + d_k + n_k, \\ &\alpha_k \equiv \boldsymbol{v}_k^{\mathsf{T}} \Delta\boldsymbol{\vartheta}, \qquad \beta_k \equiv \boldsymbol{u}_k^{\mathsf{T}} \Delta\widetilde{\boldsymbol{y}}, \qquad d_k \equiv \boldsymbol{u}_k^{\mathsf{T}} \widetilde{\boldsymbol{\delta}}_{\mathrm{mod}}, \qquad n_k \equiv \boldsymbol{u}_k^{\mathsf{T}} \widetilde{\boldsymbol{\epsilon}}. \end{aligned} \tag{20}$$

Each nonzero singular mode thus behaves locally as a one-dimensional inverse problem. The observed modal coefficient β_k contains the transmitted historical signal, together with the projections of model discrepancy and stochastic noise.

A formal unregularized inversion gives

$$\widehat{\alpha}_k^{\mathrm{unreg}} = \frac{\beta_k}{\sigma_k} = \alpha_k + \frac{d_k + n_k}{\sigma_k}, \qquad \mathrm{Var}\left(\widehat{\alpha}_k^{\mathrm{unreg}}\right) = \frac{1}{\sigma_k^2}. \tag{21}$$

The variance expression assumes correct whitening, unit-variance modal noise, and a fixed model discrepancy. Equation (21) displays the instability explicitly: both random error and discrepancy aligned with a weak mode are amplified by the inverse singular value.

Consistent with the operator convention stated after Equation (1), the numerical singular values are not invariant under arbitrary changes in parameter units. Rescaling one coefficient rescales a column of the Jacobian and changes the singular spectrum. A modal analysis must therefore specify the historical basis, coefficient normalization, and physical scale represented by a unit perturbation. Raw singular values from different SFH bases or coefficient definitions should not be compared unless those perturbations have been placed on a common physical footing.

Within such a specified geometry, three operational regimes can be distinguished. An exact null direction satisfies $\widetilde{\boldsymbol{J}}_0 \boldsymbol{v} = \boldsymbol{0}$ and produces no first-order change in the data. An effectively unobservable direction has a nonzero response but one too small relative to the adopted uncertainty and scientific accuracy. A weakly observable direction produces measurable information, but its amplitude remains sensitive to nuisance freedom, regularization, priors, or plausible changes in the forward model. An observable direction produces a sufficiently strong and credible response for the intended scientific claim. The boundaries between these regimes are conditional and should not be identified with one universal singular-value threshold. Figure 3 illustrates how representative SFH modes are transmitted into the observed spectral space with progressively different strengths.

This distinction also shows why the nominal number of parameters is not the number of measured historical degrees of freedom. A model can contain many SFH coefficients while only a smaller number of their combinations have independently distinguishable responses. The remaining coefficients can still acquire finite or narrow conditional uncertainties from non-negativity, parameter boundaries, regularization, or priors. Such precision belongs to the complete constrained model and should not automatically be interpreted as information supplied by the data alone.

Singular values describe the transmission of local modes by the forward operator. Historical resolution answers a different question: how does the chosen estimator map the true represented state into its expected reconstruction? For a locally linear estimator with inverse response $\boldsymbol{G}_0$,

$$\begin{aligned} \widehat{\Delta\boldsymbol{\vartheta}} &\simeq \boldsymbol{G}_0 \Delta\widetilde{\boldsymbol{y}}, \\ \mathbb{E}_{\boldsymbol{\epsilon}}\left[\widehat{\Delta\boldsymbol{\vartheta}} \,\middle|\, \Delta\boldsymbol{\vartheta}, \boldsymbol{\delta}_{\mathrm{mod}} = \boldsymbol{0}\right] &\simeq \boldsymbol{R}_0 \Delta\boldsymbol{\vartheta}, \qquad \boldsymbol{R}_0 \equiv \boldsymbol{G}_0 \widetilde{\boldsymbol{J}}_0. \end{aligned} \tag{22}$$

The matrix $\boldsymbol{R}_0$ is the local resolution operator. Its rows describe how one reported coefficient averages, mixes, or projects the underlying represented state. A model-discrepancy contribution adds the local bias $\boldsymbol{G}_0\tilde{\delta}_{\mathrm{mod}}$.

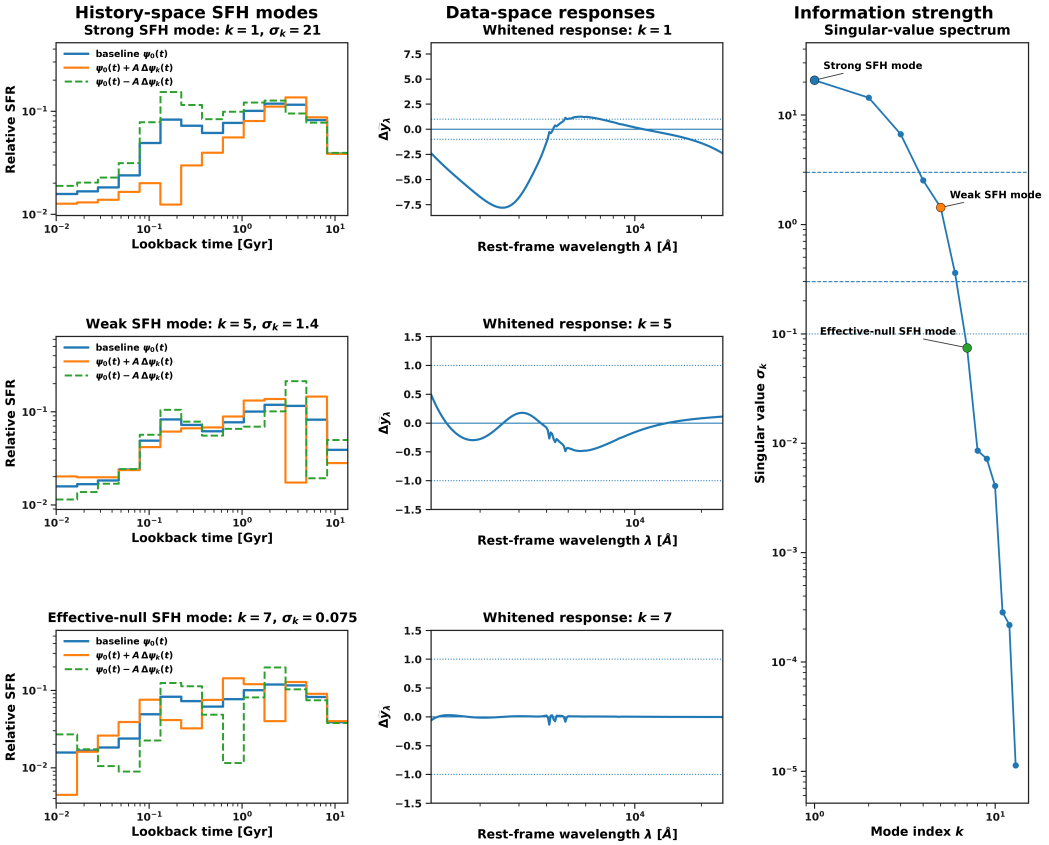


Figure 3. Illustrative singular-mode structure of an SFH inverse problem constructed with a reduced SSP library. The left panels show perturbations of a baseline SFH along representative right singular vectors, the middle panels show the corresponding whitened spectral responses, and the right panel shows the singular-value spectrum. Strong, weak, and effectively null modes illustrate how nominal SFH freedom is compressed into a smaller set of information-supported historical combinations.

A full-rank, unbiased inversion would have $\boldsymbol{R}_0 = \boldsymbol{I}$. A generalized inverse of a rank-deficient problem instead recovers only a projection onto the non-null subspace. Mode truncation, regularization, physical constraints, and prior structure modify this response further. Resolution is consequently a property of the forward problem and the estimator together, not of the observations or temporal grid alone.

For a continuous SFH, the analogous response can be written as

$$\mathbb{E}_{\epsilon}\left[\Delta\widehat{\psi}(t) \,\middle|\, \Delta\psi, \delta_{\mathrm{mod}} = \mathbf{0}\right] \simeq \int \mathcal{R}_{\psi}(t, t')\Delta\psi(t')\,\mathrm{d}t', \tag{23}$$

where $\mathcal{R}_{\psi}(t, t')$ is the SFH resolution or averaging kernel. A localized kernel centered near $t = t'$ indicates localized temporal sensitivity. A broad, asymmetric, displaced, or multilobed kernel indicates that the reported quantity at time t mixes activity over a wider or different range of epochs.

A resolution kernel need not be positive, normalized, or single-peaked. Its width alone is therefore not always a sufficient summary. The complete kernel, or scientifically relevant integrals of it, can be needed to determine which historical functional is actually measured. This perspective follows the resolving-power formulation of inverse theory, in which stable

reconstruction is understood as measurement of a localized average rather than automatic recovery of a pointwise function [50].

The nominal width of an SFH bin specifies the resolution of the representation. It does not specify the resolution of the inference. A fine grid can have broad and strongly overlapping resolution kernels, so adjacent plotted values need not be independent temporal measurements. A coarser representation can instead yield a stable estimate of the mass formed over an interval that the observations genuinely distinguish. Figure 4 illustrates the distinction between nominal temporal binning and the effective historical resolution produced by the complete forward and inference procedure.

Estimator covariance and resolution should likewise not be conflated. Covariance describes stochastic variation in the reported estimate under the adopted model. The resolution operator describes systematic averaging or projection of the underlying state even in the absence of noise. A reconstruction can have a small conditional variance and poor resolution, or broad variance and comparatively localized response.

Regularization and priors modify both quantities. Suppressing small-singular-value modes reduces noise amplification but also alters the expected response to the true history. Their detailed role is treated in Section 4. The SVD analysis of integrated stellar spectra by Ocvirk et al. [20] and the adaptive temporal decomposition in VESPA [21] provide practical precedents for the principle that recoverable temporal complexity is limited by the complete response and data quality rather than by the analyst's input grid.

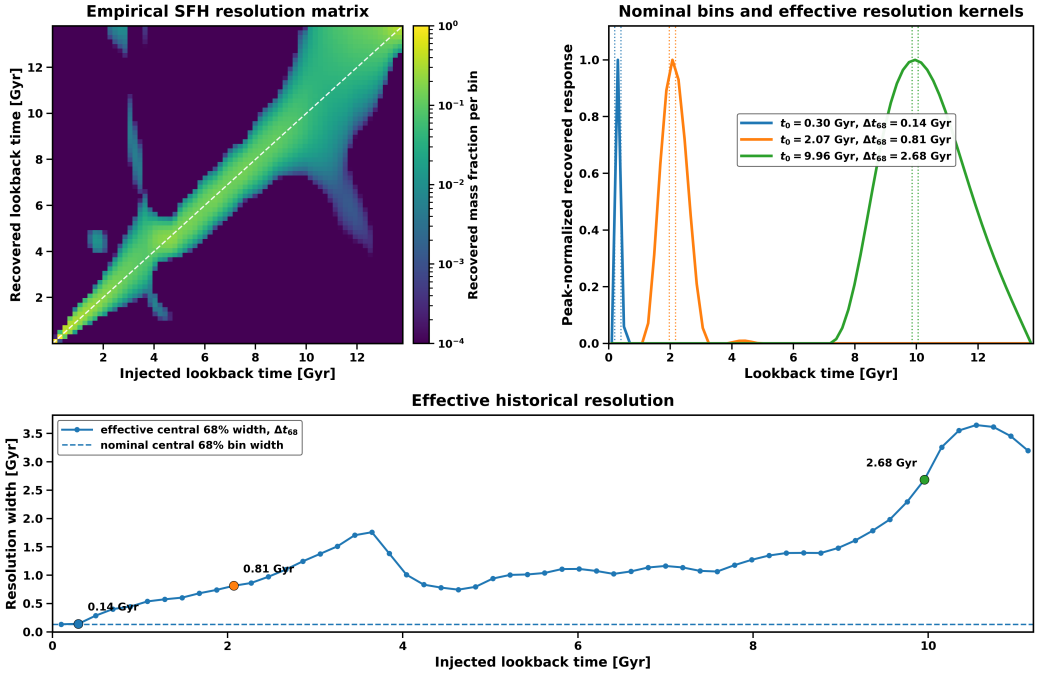


Figure 4. Illustrative injection-and-recovery analysis of nominal and effective SFH resolution. A localized formed-mass component is injected into each uniformly spaced lookback-time bin and recovered using a reduced SSP library, an illustrative UV–NIR observational operator, and a non-negative regularized inversion. The upper-left panel shows the resulting resolution matrix, the upper-right panel shows representative averaging kernels, and the lower panel shows their central 68% widths as a function of lookback time; the dashed line marks the corresponding width of one nominal bin. The broadening of the kernels toward older epochs illustrates that fine input binning does not imply equally fine historical resolution.

3.3. State-Dependent and Model-Limited Observability

The singular modes and resolution structures defined above are properties of a specified local inverse problem. They are not universal properties of a wavelength interval, fitting code, or nominal SFH representation. Schematically,

$$\{\sigma_k, \boldsymbol{u}_k, \boldsymbol{v}_k\} = \mathcal{S}_k\Big(\boldsymbol{\vartheta}_0, \boldsymbol{\omega}, \mathcal{V}, \mathcal{F}_{\mathrm{phys}}, \mathcal{O}_{\boldsymbol{\omega}}, \boldsymbol{C}\Big), \tag{24}$$

where $\boldsymbol{\vartheta}_0$ is the reference galaxy state, $\boldsymbol{\omega}$ is the observing configuration, $\mathcal{V}$ is the representation, $\mathcal{F}_{\mathrm{phys}}$ is the physical forward model, $\mathcal{O}_{\boldsymbol{\omega}}$ is the observational operator, and $\boldsymbol{C}$ is the covariance or metric used to judge distinguishability.

Galaxy state changes the derivatives of the SED with respect to its historical variables. Young and old stellar components need not have the same visibility in quiescent and actively star-forming systems. Dust can weaken direct ultraviolet and optical information while introducing infrared responses whose interpretation depends on the heating and emission model. A fixed observed filter set samples different rest-frame features as redshift changes. These effects alter the orientation and strength of the observable modes rather than merely rescaling one universal information content.

The observing configuration has the same structural role. Changing the spectral resolution, wavelength coverage, aperture, masking, or calibration covariance can remove, add, or rotate response directions. Historical resolution derived for one galaxy class, redshift interval, or instrument should therefore not be transferred automatically to another configuration.

Nonlinearity introduces a further limitation. A direction that is weak near one state can become more strongly observable elsewhere, and the local mode geometry can vary across one posterior region. Moreover, two disconnected solution families can each possess a well-conditioned local Jacobian while remaining globally indistinguishable from one another. SVD and Fisher information describe local response and should not be presented as proofs of global identifiability.

Model discrepancy imposes an additional information limit. Uncertainty in stellar evolution, spectral libraries, abundance responses, nebular emission, dust physics, radiative transfer, calibration, or omitted emission components can create coherent differences between the adopted forward model and real galaxies [7,18,19]. Reducing observational noise does not remove these differences. At sufficiently high signal-to-noise ratio, the reconstruction can become limited by the credibility of the forward model rather than by measurement precision.

When discrepancy can defensibly be represented as an independent, mean-zero stochastic contribution with covariance $\boldsymbol{C}_{\mathrm{mod}}$, one may define

$$\boldsymbol{C}_{\mathrm{eff}} = \boldsymbol{C}_{\mathrm{obs}} + \boldsymbol{C}_{\mathrm{mod}}. \tag{25}$$

Using $\boldsymbol{C} = \boldsymbol{C}_{\mathrm{eff}}$ in the whitening transformation reduces the apparent information in data-space directions affected by known model uncertainty. A correlated discrepancy covariance is generally more appropriate than uniform diagonal inflation when model errors have coherent spectral structure.

Equation (25) is valid only under its assumptions. A discrepancy with a nonzero mean is not removed by covariance inflation. A discrepancy can also depend on age, metallicity, attenuation, galaxy state, or observing configuration. In such cases, increasing the covariance can broaden conditional uncertainty without correcting the direction of the induced bias.

The effect of a deterministic local discrepancy depends on its orientation relative to the response range. Let

$$\begin{aligned} \boldsymbol{P}_J &\equiv \boldsymbol{U}_r \boldsymbol{U}_r^{\mathsf{T}}, \\ \widetilde{\boldsymbol{\delta}}_{\mathrm{mod}} &= \underbrace{\boldsymbol{P}_J \widetilde{\boldsymbol{\delta}}_{\mathrm{mod}}}_{\widetilde{\boldsymbol{\delta}}_{\parallel}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{P}_J\right) \widetilde{\boldsymbol{\delta}}_{\mathrm{mod}}}_{\widetilde{\boldsymbol{\delta}}_{\perp}}, \end{aligned} \tag{26}$$

where $\boldsymbol{P}_J$ projects onto the local range of the whitened Jacobian. The parallel component $\widetilde{\boldsymbol{\delta}}_{\parallel}$ can be reproduced locally by changes in the represented state and can therefore be absorbed into a biased history. The orthogonal component $\widetilde{\boldsymbol{\delta}}_{\perp}$ cannot be removed by a first-order parameter adjustment and remains as lack of fit.

This decomposition explains why a small residual is not sufficient evidence for a correct history. A consequential model error can lie largely within the response range, be absorbed by changes in SFH, metallicity, dust, or nuisance parameters, and leave little residual structure. Conversely, a conspicuous orthogonal residual need not produce a large bias in the inferred historical directions.

Discrepancy modeling is itself subject to confounding. An excessively flexible discrepancy model can absorb real physical information, whereas an overly restrictive one can force missing physics into the inferred history. This is a general difficulty in the calibration of imperfect forward models [51]. Discrepancy assumptions should therefore be informed by physical reasoning, external calibration, repeated observations, or alternative forward models rather than selected solely to improve goodness of fit.

Observability must consequently be judged relative to both measurement precision and model credibility. A direction with a large singular value under $\boldsymbol{C}_{\mathrm{obs}}$ can remain scientifically unreliable when its response is comparable to plausible forward-model discrepancy. A weaker direction can remain usable when its response is supported by well-modeled observables that are not aligned with the principal discrepancy.

The formal objects defined in this section do not determine by themselves which reconstruction should be reported. Representation, physical constraints, regularization, and priors select or weight solutions within the observationally allowed family, as developed in Section 4. The coupled interpretation of the singular directions is developed in Section 7. Matched and mismatched mocks, repeated-sampling calibration, posterior predictive assessment, held-out prediction, and empirical resolution tests are developed in Section 8.

The central conclusion is that an integrated SED constrains a limited set of state- and model-dependent historical combinations. The nominal parameters used to describe a history need not correspond to independently observed quantities, and the temporal grid used to display a reconstruction need not represent its effective resolution. A scientifically meaningful result should therefore identify the observable modes or stable functionals supported by the data, the averaging response through which they are measured, and the model assumptions on which that interpretation depends.

4. Representations, Regularization, and Priors

The inverse problem is specified not only by the physical forward model and the observations, but also by the mathematical representation assigned to the unknown galaxy history. A continuous star formation, enrichment, or dust history must be restricted to a computationally manageable model set before it can be inferred from finite data. This restriction determines which histories can be expressed, which structures are excluded in advance, and how fitted coefficients are translated into physical quantities. For example, choosing six SFH bins means that the fit estimates six bin amplitudes rather than

an unrestricted continuous function. Non-negativity excludes negative star formation rates; a smoothness penalty or continuity prior favors neighboring amplitudes that are similar; and the likelihood determines how strongly the measured SED can override those preferences. A feature in the reconstructed history can therefore arise because the data require it, because the binning makes it expressible, or because the added structure selects it among nearly equivalent fits.

Representation is therefore part of the scientific model rather than a neutral numerical choice. An exponentially declining SFH, a delayed-τ model, a collection of fixed time bins, a spline expansion, a stochastic process, and a learned representation define different admissible histories even when they contain similar numbers of parameters. They encode different assumptions about smoothness, burstiness, temporal localization, and correlations between epochs.

Regularization and prior distributions provide additional structure within the selected representation. Their influence is greatest along the weakly observable directions identified in Section 3, where several histories produce nearly indistinguishable observations. Physical constraints further restrict the admissible solutions through non-negativity, conservation relations, energy balance, or evolutionary consistency.

The reconstructed history is consequently determined jointly by the representation, physical admissibility, regularization or prior structure, and observable information. A feature in a posterior median or regularized solution may be required by the data, inherited from the chosen functional family, selected by a prior correlation scale, or imposed by a physical boundary. This section isolates these roles before Section 5 compares their realization in existing methods and Section 8 develops the corresponding validation program.

4.1. Historical Representations and Admissible Model Spaces

A parametric SFH restricts the unknown history to a finite-dimensional family indexed by parameters $\boldsymbol{\theta}_{\mathrm{SFH}}$:

$$\mathcal{V}_{\mathrm{par}} = \{\psi(\,\cdot\,;\boldsymbol{\theta}_{\mathrm{SFH}}) \,|\, \boldsymbol{\theta}_{\mathrm{SFH}} \in \Theta_{\mathrm{SFH}}\}. \tag{27}$$

Exponentially declining, delayed-τ, lognormal, double-power-law, and burst-plus-continuous histories are familiar examples. Their low dimensionality excludes many rapidly varying histories before the observations are analyzed and can therefore produce stable estimates when the selected family adequately represents the relevant evolution.

The same restriction creates representation bias when the true history lies outside the adopted family. A delayed-τ model, for example, cannot reproduce an arbitrary sequence of quiescence, rejuvenation, bursts, and renewed decline. The fitted model can nevertheless reproduce the observed SED through compensating changes in stellar age, metallicity, attenuation, or stellar mass. A narrow posterior can then reflect the rigidity of the model family rather than uniquely informative data.

Parametric choices also induce prior structure on derived quantities. A prior that appears simple in a timescale, formation time, or power-law slope can generate a highly nonuniform prior on mass-weighted age, recent stellar-mass fraction, quenching time, or specific star formation rate [52]. The physical consequences of the parameterization should therefore be examined in the space of reported quantities rather than inferred only from the algebraic form of its parameters.

A nonparametric SFH uses a more flexible finite-dimensional expansion. In a linear representation,

$$\psi_{\mathcal{V}}(t) = \sum_{j=1}^{N_{\mathrm{SFH}}} a_j \phi_j(t), \qquad \mathcal{V}_{N_{\mathrm{SFH}}} = \mathrm{span}\{\phi_1, \ldots, \phi_{N_{\mathrm{SFH}}}\}, \tag{28}$$

where $\phi_j(t)$ are basis functions and a_j are their coefficients. Piecewise-constant bins, piecewise-linear functions, splines, localized kernels, wavelets, and simulation-derived components are representative choices. Gaussian-process models and nonlinear decoders define more general model sets that need not be linear spans, but they serve the same purpose of restricting an infinite-dimensional history to a manageable family.

The term "nonparametric" does not mean that the reconstruction is free of parameterization or assumptions. The number and placement of time bins, the normalization of the basis functions, the support of the coefficients, and the correlations imposed between epochs all define the allowed histories. A model with ten fixed age bins remains a ten-dimensional parameterized model even though it is not restricted to one analytic SFH form.

Basis normalization determines the physical meaning of the coefficients. For a piecewise-constant basis $\phi_j(t) = \mathbb{1}_{I_j}(t)$, the coefficient a_j represents an average SFR over the interval I_j, and the corresponding formed stellar mass is $a_j \Delta t_j$. If instead $\phi_j(t) = \mathbb{1}_{I_j}(t)/\Delta t_j$, then a_j directly represents the formed stellar mass. A coefficient can alternatively represent a mass fraction, surviving stellar mass, luminosity fraction, or a model coordinate without a direct physical interpretation. These definitions are related but not interchangeable.

Normalization also affects non-negativity, prior widths, smoothness penalties, covariance matrices, and singular-value analyses. The same algebraic prior on average SFRs and formed masses describes different physical behavior when the time bins have unequal widths. Likewise, a regularization parameter cannot be compared meaningfully across bases whose coefficients have different units or scales.

Nominal and effective resolution. Nominal resolution belongs to the representation; effective resolution belongs to the forward model and estimator. Temporal binning introduces an additional resolution assumption. Linear-time and logarithmic-time partitions allocate representational freedom differently, particularly at old ages. Such choices can be motivated by the approximate decline of age sensitivity with lookback time, but their boundaries remain model specifications rather than measurements. Nominal bin width should therefore be distinguished from the effective historical resolution developed in Section 3.2.

A flexible representation reduces the rigid functional mismatch of a low-dimensional parametric family, but its additional coefficients are not necessarily constrained independently. Nonparametric SFH studies show that broad features can be recovered more flexibly than under simple analytic histories, while detailed temporal structure remains sensitive to the prior imposed on changes between neighboring epochs [17]. Regularized full-spectrum methods express the same principle by permitting a detailed age distribution while suppressing combinations that cannot be stably separated in the spectrum [20].

Physical constraints select an admissible subset of the chosen representation. Schematically,

$$\mathcal{A}_{\mathrm{hist}} = \left\{ (\psi, \boldsymbol{\eta}) \in \mathcal{V} \times \mathcal{V}_\eta \;\middle|\; \psi(t) \geq 0, \quad \mathcal{E}_{\mathrm{phys}}[\psi, \boldsymbol{\eta}] = \mathbf{0}, \quad \mathcal{G}_{\mathrm{phys}}[\psi, \boldsymbol{\eta}] \succeq \mathbf{0} \right\}, \tag{29}$$

where $\boldsymbol{\eta}$ represents other physical histories and parameters, $\mathcal{E}_{\mathrm{phys}}$ denotes equality relations, and $\mathcal{G}_{\mathrm{phys}}$ denotes inequality or boundary constraints. An SFH-only model may impose non-negativity and a cosmic-age limit, principally. A coupled model can additionally

enforce mass and metal conservation, chemically consistent enrichment, energy balance, or restrictions derived from gas and dust evolution.

These constraints can eliminate physically impossible solutions, but they do not create observational information. Non-negativity suppresses oscillatory solutions with canceling positive and negative components, yet several distinct non-negative SFHs can remain observationally indistinguishable. Energy balance relates absorbed and re-emitted radiation, but it does not necessarily determine which stellar populations heated the dust or which geometry produced the attenuation.

Representation adequacy in data space is also distinct from accuracy in history space. A restricted model can fail to reproduce the noise-free observables because the true galaxy lies outside its physical or historical model set. Conversely, a flexible model can reproduce the observations while assigning the wrong history because several distinct represented histories map to similar SEDs. Goodness of fit alone therefore cannot determine whether the historical representation is scientifically adequate.

The appropriate representation depends on the target quantity. A coarse basis may be sufficient for estimating the stellar mass formed before and after a broad transition, whereas localized recent components may be justified when ultraviolet or nebular diagnostics provide corresponding sensitivity. No fixed parametric family or temporal grid is uniformly optimal across galaxy state, redshift, wavelength coverage, and data quality. This dependence motivates the adaptive design principle developed in Section 6.

4.2. Regularization and Priors as Additional Structure

Even after a representation and its physically admissible subset have been specified, the data can permit a broad or unstable family of solutions. Regularization and prior distributions provide additional structure for selecting or weighting those solutions. Their scientific role is not to recover information absent from the observations, but to complete the inverse problem in directions that are only weakly constrained.

A general regularized estimator can be written as

$$\widehat{\boldsymbol{\vartheta}}_{\lambda} \in \arg\min_{\boldsymbol{\vartheta}\in\mathcal{A}} \left\{ \Phi_{\mathrm{data}}[\boldsymbol{y}, \boldsymbol{\mu}(\boldsymbol{\vartheta})] + \lambda_{\mathrm{reg}} \mathcal{R}(\boldsymbol{\vartheta}) \right\}, \tag{30}$$

where Φ_{data} measures disagreement between the observations and the forward prediction, $\mathcal{R}$ is a regularization functional, λ_{reg} controls its relative strength, and $\mathcal{A}$ is the admissible parameter set. For Gaussian observational errors, Φ_{data} is commonly a covariance-weighted squared residual, but the same structure applies to more general likelihoods.

Quadratic shrinkage, including zero-order Tikhonov regularization, favors coefficients close to a reference history [53,54]. First- and second-difference penalties are standard discrete derivative regularizers that favor temporal continuity or reduced curvature [53,54]. Their application to SFH reconstruction specializes standard inverse-problem constructions to the chosen historical coordinates rather than introducing new regularizer classes. Sparsity penalties favor histories represented by a small number of active basis components [55], while total-variation penalties favor piecewise-constant behavior separated by a limited number of transitions [56]. These choices encode different structural priors or assumptions about the recovered SFH. A smoothness or sparsity penalty represents a physical prior only when the favored behavior is independently justified by galaxy-evolution physics; otherwise, it is a mathematical regularization choice. The penalties should therefore not be treated as interchangeable numerical stabilizers or as direct statements of required physical behavior.

Their interpretation depends on the representation. Smoothness in linear time differs from smoothness in logarithmic lookback time. A penalty on formed stellar masses

differs from the same algebraic penalty on average SFRs when bin widths vary. lcSparsity in localized bins favors a small number of active epochs, whereas sparsity in a wavelet or learned basis favors a small number of multiscale or learned-basis components. Regularity is therefore always defined relative to a basis, coordinate, normalization, and boundary convention.

For a locally linear and whitened inverse problem, regularization can often be interpreted as filtering the singular modes introduced in Section 3.2. A generic filtered estimator takes the form

$$\widehat{\Delta\boldsymbol{\vartheta}}_{\mathrm{filt}} = \sum_{k=1}^{r} f_k \frac{\beta_k}{\sigma_k} \boldsymbol{v}_k, \qquad 0 \le f_k \le 1, \tag{31}$$

where f_k is the filter factor applied to mode k. The unregularized generalized inverse has $f_k = 1$ for every nonzero singular value. Standard zero-order Tikhonov regularization continuously suppresses modes with small σ_k, while truncated singular-value inversion removes modes below a selected cutoff [54].

Equation (31) displays the essential bias–variance trade-off. Suppressing a weak mode reduces the amplification of observational noise and model discrepancy, but it also attenuates the corresponding component of the true history. A regularized reconstruction is therefore a filtered estimate with a particular resolution and bias structure, not an exact reconstruction made merely more numerically stable.

For derivative penalties or other general quadratic regularizers, the filtered directions need not coincide with the ordinary right singular vectors of the forward response. Their geometry is determined jointly by the forward and regularization operators and can be described through generalized singular values or an equivalent resolution analysis. Nonquadratic penalties and active physical constraints produce nonlinear estimators for which fixed global filter factors are generally unavailable.

Bayesian inference introduces additional structure through a prior distribution. For a fully specified model $\mathfrak{M}$,

$$p(\boldsymbol{\vartheta} \mid \boldsymbol{y}, \mathfrak{M}) \propto p(\boldsymbol{y} \mid \boldsymbol{\vartheta}, \mathfrak{M}) p(\boldsymbol{\vartheta} \mid \mathfrak{M}). \tag{32}$$

The likelihood contains the observational and discrepancy assumptions, while the prior assigns probability over the admissible histories before the current data are analyzed. Bayesian inverse problems use the resulting posterior to describe uncertainty over the model-defined solution family rather than selecting only one regularized point estimate [57].

Under appropriate likelihood and prior forms, the minimizer of Equation (30) has a maximum-a posteriori interpretation. A quadratic penalty corresponds locally to a Gaussian prior, while an ℓ_1 penalty corresponds to a Laplace-type prior in the coordinates on which it is imposed. This correspondence does not make regularized point estimation and posterior inference equivalent: a posterior can be broad, asymmetric, correlated, or multimodal even when its mode resembles a regularized solution.

The prior is itself representation-dependent. Uniform distributions are not preserved under nonlinear changes in variables, and priors on SFRs, logarithmic SFRs, formed masses, mass fractions, or adjacent-bin ratios imply different temporal behavior. Prior assumptions should therefore be described through the physical histories and derived quantities they generate rather than only through the distributions assigned to internal parameters.

For nonparametric SFHs, the prior commonly controls the amplitude and correlation scale of changes between neighboring epochs. Continuity priors favor gradual evolution, while burst-permitting priors assign greater probability to rapid changes. When the

likelihood provides only broad temporal information, these structures can determine the apparent texture of the reconstructed SFH [17].

Population information can also enter through hierarchical priors, simulations, or learned training distributions. Such information can improve inference when the reference population represents the observed galaxies, but it can suppress rare histories or bias weak observations when the population model is incomplete. Selection effects must be included when a prior or population distribution is inferred from a survey sample.

The strength of regularization or the scale of a prior determines how strongly this added structure acts. Weak regularization can retain unstable modes; strong regularization can suppress genuine transitions and force the estimate toward the preferred model. Predictive validation and generalized cross-validation select or assess the scale through held-out or approximately out-of-sample predictive performance [58]. Hierarchical inference treats the scale as a hyperparameter with its own probabilistic model, whereas marginal-likelihood or empirical-Bayes methods estimate it by integrating over the reconstruction parameters [59,60]. Discrepancy principles compare the achieved residual with the level expected from the observational noise model, and sensitivity analysis examines how scientifically relevant conclusions change across a defensible range of scales [53,54]. These procedures answer different questions, so no single criterion should be expected to be uniformly preferable across representations, noise and discrepancy models, forward models, and scientific targets. The selected scale and the criterion used to choose it should therefore be reported together.

Regularization and priors also cannot repair an inadequate forward model. A smoothness assumption can suppress noise-driven oscillations, but it cannot distinguish a real SFH feature from a systematic spectral mismatch aligned with an age-sensitive response. A carefully sampled posterior can propagate uncertainty conditional on the adopted stellar, nebular, and dust models while remaining biased if those models omit relevant physics. Assumption sensitivity and forward-model discrepancy must therefore be evaluated separately.

4.3. Interpreting and Reporting Assumption Dependence

The combined effects of representation, physical constraints, regularization, and priors determine how the inverse problem is completed in directions that the observations do not fully constrain. The resulting history should therefore be interpreted conditionally on the complete model specification rather than as a direct measurement of every fitted coefficient.

Three limiting regimes provide a useful language for this interpretation.

Data-dominated. A reported feature is *data-dominated* when it is associated with a strong observable response and remains stable under reasonable changes in representation, prior structure, nuisance parameters, and forward-model ingredients.

Assumption-sensitive. A reported feature is *assumption-sensitive* when the observations constrain a broad family of possibilities but its detailed amplitude, localization, or temporal texture changes with regularization, prior correlation, or parameterization.

Model-dominated. A reported feature is *model-dominated* when its apparent constraint arises primarily from the adopted stellar, chemical, nebular, dust, radiative-transfer, calibration, or discrepancy model.

These regimes are not mutually exclusive classifications of an entire reconstruction. A broad episode of early star formation may be data-supported, while its subdivision among adjacent ancient bins is assumption-sensitive. Obscured recent activity may be required by ultraviolet and infrared observations, while its distribution among very young components remains dependent on nebular and dust-heating assumptions.

The classification should therefore be applied to specific scientific quantities and levels of temporal resolution.

Posterior contraction relative to the prior is informative but does not by itself establish data dominance. A posterior can narrow because positivity, a parameter boundary, or a rigid physical model truncates a weakly constrained direction. Conversely, a scientifically meaningful quantity can retain broad uncertainty because the observations genuinely support multiple possibilities. The origin of a constraint must be assessed through observability, effective resolution, prior sensitivity, and forward-model sensitivity together.

Comparisons across representations should focus on common physical functionals rather than on internal coefficients. Relevant quantities include stellar mass formed over a broad lookback-time interval, recent mean SFR, cumulative mass-assembly fractions, mass-weighted age, formation-time quantiles, or broad quenching epochs. Individual coefficients from different bases generally do not describe the same estimand.

Effective resolution is essential to this comparison. When the averaging kernel associated with a reported coefficient extends over several neighboring intervals, the scientifically supported result concerns the integrated activity over that broader support rather than the pointwise value in one nominal bin. A fine plotted grid should not be interpreted as fine temporal measurement without the corresponding resolution analysis.

Parameter covariance carries similar information. Adjacent SFH components can be strongly correlated because the data constrain their sums or contrasts more directly than their individual values. Marginal intervals alone can obscure these relationships. Posterior samples, covariance information, cumulative histories, and uncertainties on integrated functionals provide a more faithful description than independent-looking error bars on each time bin.

Prior and representation sensitivity should be separated from physical-model sensitivity. The former can be examined through alternative binnings, functional families, correlation scales, or penalties. The latter requires changes in stellar evolutionary tracks, spectral libraries, abundance treatments, nebular prescriptions, attenuation laws, dust-emission models, radiative-transfer assumptions, or discrepancy descriptions. Agreement across several inference algorithms that share the same forward model does not establish robustness to physical misspecification.

The detailed procedures for testing these dependencies are developed in Section 8. In particular, Layer I distinguishes matched- and mismatched-model mock experiments, repeated-sampling calibration, posterior predictive assessment, held-out prediction, and validation in which every data-dependent representation or hyperparameter choice is repeated within each realization. Layer II then interprets the resulting failures through observable modes, resolution operators, prior curvature, regularization bias, and the projection of model discrepancy into the inferred state.

This division of roles is important. The present section identifies what must be interpreted and reported: the chosen representation, its normalization, the admissible set, the regularization or prior, the effective resolution, and the distinction between data-supported and assumption-selected structure. Section 8 addresses how those claims should be tested empirically and structurally.

A transparent analysis should therefore specify the historical representation and coefficient definitions, the physical constraints, the regularization or prior and its implied temporal behavior, the forward-model ingredients, the observational covariance, and the treatment of model discrepancy. It should identify which alternative assumptions were examined and which scientific conclusions remained stable.

The purpose is not to require all reasonable methods to return the same detailed history. Several physically plausible reconstructions may remain observationally indistinguishable,

and a calibrated analysis should preserve that ambiguity. The objective is to separate stable scientific information from fine structure selected by one particular representation or inferential convention.

Representations, regularization, and priors are thus constitutive components of the inverse problem. They determine which histories can be expressed, how weakly observable directions are completed, and which estimand is ultimately reported. Section 5 now uses this perspective to compare existing reconstruction methods according to the roles played by their representation, forward model, added structure, inference procedure, and uncertainty analysis.

5. A Unified Taxonomy of Existing Methods

Methods for reconstructing stellar populations and galaxy histories are often grouped by their most visible computational feature: parametric versus nonparametric SFHs, photometric versus full-spectrum fitting, optimization versus Bayesian sampling, matrix factorization versus physical modeling, or conventional inference versus machine learning. These labels are useful, but they mix choices made at different stages of the inverse problem. Two methods can use different algorithms while sharing the same historical representation and physical assumptions, whereas two analyses performed with the same software can imply different scientific estimands because they adopt different bases, priors, stellar libraries, or uncertainty models.

A more informative comparison asks which part of the inferential structure a method changes. PCA, MOPED, and VESPA all reduce dimensionality, but PCA represents variance in an ensemble, MOPED compresses the observations for a specified parameterization, and VESPA adapts the historical representation. Non-negative matrix factorization (NMF) learns a constrained basis, a neural point estimator learns a loss-dependent decision rule, and simulation-based inference (SBI) approximates a simulator-defined conditional distribution. None of these operations, by itself, specifies the complete physical model or guarantees calibrated historical inference.

This section therefore classifies representative methods as combinations of reusable components. The aim is not to rank software packages or to reproduce complete algorithmic derivations. We retain the defining equations needed to identify what each method represents, compresses, or returns. Observability and Fisher information are developed in Section 3, representation, regularization, and priors in Section 4, and validation in Section 8.

5.1. A Five-Component Description of Reconstruction Methods

Consider a fit in which the SFH is described by lookback-time bins. The bin edges and coefficient normalization specify what is reconstructed; the stellar-population, nebular, dust, redshift, and instrumental calculations specify how a proposed history predicts the data; non-negativity, smoothness, and priors specify which histories are allowed or favored; the optimizer or sampler specifies how the data are used to estimate them; and posterior checks, mock recovery, or coverage tests specify how uncertainty is assessed. These remain distinct choices even when they are packaged in one code. The notation below assigns them, respectively, to $\mathcal{V}_q$, $\mathcal{F}_q$, $\mathcal{C}_q$, $\mathcal{I}_q$, and $\mathcal{U}_q$ so that methods can be compared by the choices they make rather than only by software name.

Let q denote a fully specified reconstruction method or implementation. We characterize it by

$$\mathfrak{M}_q = (\mathcal{V}_q, \mathcal{F}_q, \mathcal{C}_q, \mathcal{I}_q, \mathcal{U}_q), \qquad q \in \mathcal{Q}, \tag{33}$$

where $\mathcal{Q}$ is the collection of method configurations under comparison. The five components describe the representation, forward model, added structure, inference mechanism, and uncertainty assessment, respectively.

The representation $\mathcal{V}_q$ specifies what is reconstructed and in which coordinates. It can be a low-dimensional analytic SFH family, a fixed or adaptive time grid, SSP weights, a spline or wavelet basis, non-negative learned components, or a nonlinear learned manifold. The definition and normalization of these coordinates determine whether a coefficient represents an SFR, formed mass, mass fraction, luminosity fraction, or a model coordinate without a direct physical interpretation.

The forward model $\mathcal{F}_q$ maps the represented state into the measured data. It contains the physical SED-generation ingredients and the observational operator developed in Section 2. Methods using the same optimizer or sampler can therefore return different histories when their stellar population, nebular, dust, calibration, or instrumental models differ. Any explicit treatment of model discrepancy also belongs to this component.

The structure $\mathcal{C}_q$ contains hard physical constraints, regularization, prior distributions, and restrictions inherited from template libraries, simulations, or training populations. Non-negativity, smoothness, continuity priors, sparsity, energy balance, and simulation-manifold restrictions all act most strongly in directions that the likelihood constrains weakly. A machine-learning training distribution is included here because it determines which states are represented frequently, rarely, or not at all.

The inference mechanism $\mathcal{I}_q$ can be optimization, maximum-a posteriori inversion, posterior sampling, matrix factorization, data compression, neural regression, or conditional density estimation. Compression and inference should be distinguished: an operation can reduce the data dimension without reducing the historical state, reduce the historical representation without compressing the observations, or learn a lower-dimensional model set from an ensemble.

The uncertainty component $\mathcal{U}_q$ describes what uncertainty object is returned and how it is tested. It includes conditional covariance or posterior distributions, mock recovery, coverage, simulation-based calibration, prior sensitivity, posterior predictive assessment, held-out prediction, and robustness to model discrepancy. Its interpretation remains conditional on the preceding four components.

The tuple in Equation (33) is modular but not separable. Changing the representation changes the discretized response; changing the forward model rotates degeneracies; changing the prior changes weak-mode resolution; and changing the observing model changes which historical combinations are available to any inference algorithm. The taxonomy is therefore most useful for identifying which of the five components a methodological innovation actually changes. Here, combining components from different mathematical, statistical, or astrophysical traditions consistently means that their interfaces remain compatible: the representation must match the inputs expected by the forward model, the constraints or prior must be defined on that representation, the inference procedure must target the corresponding likelihood or simulator, and the uncertainty assessment must remain valid for the assembled configuration. For example, a PCA or NMF basis can define the representation, while MOPED can compress the observations used for inference. This combination is coherent only if the MOPED response derivatives and covariance are computed for the same forward model and basis coordinates, the constraints or prior apply to those coordinates, and the uncertainty assessment accounts for conditioning on both the learned basis and the compression construction.

5.2. Representation and Information Compression

Dimensional reduction can act on the historical state, on an ensemble used to define a basis, or on the measured data. These operations preserve different quantities and should not be described interchangeably as "information-preserving." Parametric and nonparametric histories primarily define $\mathcal{V}$ and $\mathcal{C}$; PCA learns a variance-efficient ensemble representation; MOPED constructs parameter-targeted summaries of the observations; and VESPA adapts temporal freedom to the information in an individual spectrum.

The methods in this subsection were selected as representative examples of dimensional reduction acting on different objects and at different stages of a reconstruction, rather than as an exhaustive list or as mutually exclusive end-to-end alternatives. PCA and MOPED are reusable operations that can be embedded within many pipelines: PCA constructs a low-rank ensemble representation, while MOPED compresses the measured data for a fixed parameterization and covariance model. VESPA is instead a reconstruction framework that adapts the dimension of the reported stellar-population history, and STECMAP represents the tradition of regularized full-spectrum inversion. Placing them together is therefore intended to expose these differences within the five-component taxonomy; other codes or methods can be classified by asking which component they primarily modify and which remaining components they inherit.

5.2.1. Parametric and Nonparametric Histories as Representation Choices

The parametric and nonparametric representations defined in Section 4.1 differ in flexibility and geometry rather than in whether they contain assumptions. Analytic SFH families reduce the dimension of the represented history before forward modeling and can stabilize broad quantities when the family is adequate. Their restricted geometry can also produce persistent bias and strongly induced priors on derived quantities [52].

Fixed-bin or basis-expanded histories enlarge the approximation space, but the data commonly constrain fewer combinations than the number of coefficients. Non-negativity, smoothness, continuity priors, and other structures then determine the weak directions. STECMAP, or STEllar Content via Maximum A Posteriori, is a representative regularized full-spectrum inversion framework rather than a generic dimensionality-reduction step [20]. In the five-component description, $\mathcal{V}_q$ contains a nonparametric stellar age distribution and, in the more general formulation, an age–metallicity relation; $\mathcal{F}_q$ contains the SSP-based spectral model and its treatment of extinction or continuum corrections; $\mathcal{C}_q$ includes positivity and smoothness; $\mathcal{I}_q$ is a penalized maximum-a posteriori inversion with the smoothing scale selected by generalized cross-validation; and $\mathcal{U}_q$ includes mock-based assessments of bias, dispersion, separability, and age resolution. STECMAP is included here because it makes the role of explicit regularization especially transparent, not because it is the only full-spectrum inversion framework. The relevant comparison is therefore between approximation error, observable resolution, and assumption sensitivity, not between "assumed" and "assumption-free" histories.

5.2.2. PCA: Variance-Preserving Ensemble Compression

Let $\boldsymbol{x}$ denote a spectrum, SED, history, or other vector drawn from a training ensemble with mean $\bar{\boldsymbol{x}}$. If the columns of $\boldsymbol{Q}_r$ are the leading r eigenvectors of the ensemble covariance, PCA defines

$$\boldsymbol{z}_{\mathrm{PCA}} = \boldsymbol{Q}_r^{\mathsf{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}), \qquad \hat{\boldsymbol{x}}_r = \bar{\boldsymbol{x}} + \boldsymbol{Q}_r \boldsymbol{z}_{\mathrm{PCA}}. \tag{34}$$

For a fixed rank and metric, this subspace minimizes the mean squared reconstruction error of the training ensemble.

PCA therefore preserves ensemble variance, not automatically the information required for a specified physical parameter or historical functional. A high-variance component can describe normalization, calibration, or a common population trend, whereas a low-variance feature can carry distinctive physical sensitivity. The result also depends on the training population, normalization, masking, wavelength weighting, and any whitening applied before the covariance is constructed [61].

PCA modifies $\mathcal{V}$ when its components form the represented spectral or historical basis and modifies $\mathcal{I}$ when it is used only to compress observations before another inference stage. PCA of spectra, model SEDs, and SFHs are consequently different methodological operations even though they share Equation (34).

5.2.3. MOPED: Fisher-Preserving Data Compression

MOPED is a parameter-targeted linear data-compression method that leaves the chosen parameterization unchanged and compresses a high-dimensional observation to one summary per inferred parameter [62]. We select it here as a canonical and analytically transparent example of local Fisher-preserving compression. Related approaches include compression by the score of a specified likelihood and information-maximizing neural-network summaries learned from simulations [63,64]. For MOPED, let $\boldsymbol{\mu}_{,i} = \partial\boldsymbol{\mu}/\partial\vartheta_i$ be evaluated at a fiducial state and let $\boldsymbol{C}$ be a parameter-independent covariance. The compression vectors and summaries can be constructed recursively as

$$\begin{aligned} \boldsymbol{b}_1 &= \frac{\boldsymbol{C}^{-1}\boldsymbol{\mu}_{,1}}{\left(\boldsymbol{\mu}_{,1}^{\mathsf{T}}\boldsymbol{C}^{-1}\boldsymbol{\mu}_{,1}\right)^{1/2}}, \\ \boldsymbol{b}_i &= \frac{\boldsymbol{C}^{-1}\boldsymbol{\mu}_{,i} - \sum_{j<i}\left(\boldsymbol{\mu}_{,i}^{\mathsf{T}}\boldsymbol{b}_j\right)\boldsymbol{b}_j}{\left[\boldsymbol{\mu}_{,i}^{\mathsf{T}}\boldsymbol{C}^{-1}\boldsymbol{\mu}_{,i} - \sum_{j<i}\left(\boldsymbol{\mu}_{,i}^{\mathsf{T}}\boldsymbol{b}_j\right)^2\right]^{1/2}}, \qquad x_i = \boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{y}. \end{aligned} \tag{35}$$

Under the assumptions of the construction and for nondegenerate derivative directions, the compressed statistics preserve the local Fisher information at the fiducial point.

This "lossless" property is local and model-conditional. It does not imply preservation of global multimodality, Bayesian evidence, arbitrary residual structure, parameter-dependent covariance information, or effects absent from the forward model. MOPED reduces the dimension of the observation, not an overflexible history representation, and acts primarily through $\mathcal{I}$ with essential dependence on $\mathcal{F}$ and the covariance model in $\mathcal{U}$.

5.2.4. VESPA: Adaptive Historical Resolution

VESPA addresses the dimension of the represented stellar-population history rather than the dimension of the observed spectrum. Starting from a maximum-resolution temporal model, it selects a coarser representation supported by an individual spectrum,

$$\mathcal{V}_{N_{\max}} \longrightarrow \mathcal{V}_{N_{\mathrm{gal}}}, \qquad N_{\mathrm{gal}} \leq N_{\max}, \tag{36}$$

where the recoverable complexity depends on the spectral information available for that galaxy [21].

This differs from fitting all maximum-resolution coefficients under a fixed smoothness penalty. The number and definition of the reported historical components change with the data, making temporal resolution part of the output. VESPA therefore acts mainly on $\mathcal{V}$ and $\mathcal{I}$, while its non-negativity, stellar population, dust, and uncertainty assumptions remain in $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{U}$.

The selected bins remain finite-resolution averages and are conditional on the adopted SSP model and selection procedure. VESPA is thus an important precedent for information-dependent historical resolution, but not a universal definition of adaptivity. Section 6 generalizes the principle to other bases, data types, and coupled states.

5.2.5. Three Distinct Domains of Dimensional Reduction

The three methods answer different questions. PCA asks which directions reconstruct a selected ensemble efficiently. MOPED asks which summaries preserve local sensitivity to a specified parameter vector. VESPA asks how much historical freedom an individual observation can support. Their primary preserved quantities are, respectively, ensemble variance, local Fisher information, and usable historical resolution.

These operations can be combined, but only when their conditional assumptions are consistent. A PCA or NMF basis can define candidate historical components, MOPED can compress the observations used to infer their amplitudes, and an adaptive rule can decide how many components should be reported. Every such construction should state which object was reduced, which metric defined the reduction, and which information or structure was intended to survive.

Consequently, differences between SFHs returned by two analyses cannot be attributed to PCA, MOPED, VESPA, STECMAP, or any other named operation alone unless the remaining components of $\mathfrak{M}_q$ are held fixed or their changes are tested separately. Different SSP libraries, dust models, constraints, priors, optimization targets, and uncertainty procedures can change the recovered history even when the same compression or representation strategy is used. Conversely, similar histories from two named methods need not imply that they preserve the same information.

5.3. Learned Representations and Amortized Inference

Learning can enter an integrated-light reconstruction at several distinct stages. At the representation level, NMF learns additive non-negative components from an ensemble of spectra, SEDs, or histories. At the inference level, a neural point estimator learns a direct map from observations or summaries to one loss-dependent estimate of the represented physical quantities. At the distributional-inference level, simulation-based inference uses simulated parameter–data pairs to approximate a conditional posterior or another likelihood-related inferential object. These operations answer different questions and can be combined within one analysis; they are not interchangeable alternatives merely because each uses a learned model.

With these roles established, the five-component taxonomy makes their assumptions explicit. NMF primarily modifies the representation $\mathcal{V}$ and the structural restrictions $\mathcal{C}$; a neural point estimator primarily modifies the inference mechanism $\mathcal{I}$ and can also involve a learned encoder or decoder in $\mathcal{V}$; and simulation-based inference primarily modifies $\mathcal{I}$ and the returned uncertainty object $\mathcal{U}$. Across these classes, the training distribution contributes to $\mathcal{C}$, the simulator and observational model supply $\mathcal{F}$, and calibration or domain-shift assessment belongs to $\mathcal{U}$. Learned methods therefore relocate assumptions into training ensembles, simulators, architectures, and loss functions rather than removing those assumptions.

5.3.1. NMF: Non-Negative Learned Representations

Non-negative matrix factorization (NMF) learns an additive basis and non-negative amplitudes from a non-negative training ensemble, so its immediate output is a representation of that ensemble rather than an SFH inferred from a new observation. For a training

matrix $\boldsymbol{X}$ containing spectra, SEDs, or histories, the learning problem can be written in the form

$$\left(\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{H}}\right) \in \underset{\boldsymbol{W},\boldsymbol{H}\geq\boldsymbol{0}}{\arg\min}\{\mathcal{D}(\boldsymbol{X}, \boldsymbol{W}\boldsymbol{H}) + \lambda_W \mathcal{R}_W(\boldsymbol{W}) + \lambda_H \mathcal{R}_H(\boldsymbol{H})\}. \tag{37}$$

The columns of $\boldsymbol{W}$ define additive basis components and $\boldsymbol{H}$ contains their non-negative amplitudes [65].

If $\boldsymbol{X}$ contains histories, the learned components define a constrained historical representation. If it contains spectra, the amplitudes are spectral coordinates and do not automatically constitute an SFH. The non-negative geometry prevents positive–negative cancellation and can improve interpretability, but it does not guarantee uniqueness or physical meaning. The factorization depends on the training ensemble, rank, loss, initialization, regularization, and normalization, and it need not align with the observable modes of a separate instrument and noise model.

NMF primarily modifies $\mathcal{V}$ and $\mathcal{C}$. Once its basis is fixed, the amplitudes can be inferred by conventional optimization, Bayesian sampling, or amortized methods. Uncertainty in learning the basis should be distinguished from uncertainty conditional on treating that basis as exact.

5.3.2. Deep Learning as Amortized Point Inference

A neural point estimator learns a rapid map from observations or summaries to a single vector of represented physical quantities, rather than returning a complete conditional distribution. Convolutional neural networks trained on simulated galaxy spectra to predict high-resolution SFHs provide a representative astronomical application [66]. For simulated training pairs $\{\boldsymbol{\vartheta}^{(\ell)}, \boldsymbol{y}^{(\ell)}\}$, a generic construction is

$$\widehat{\boldsymbol{\phi}} \in \underset{\boldsymbol{\phi}}{\arg\min}\, \frac{1}{N_{\mathrm{tr}}} \sum_{\ell=1}^{N_{\mathrm{tr}}} \mathcal{L}\left[\mathcal{N}_{\boldsymbol{\phi}}\left(\boldsymbol{s}(\boldsymbol{y}^{(\ell)})\right), \boldsymbol{\vartheta}^{(\ell)}\right], \qquad \widehat{\boldsymbol{\vartheta}} = \mathcal{N}_{\widehat{\boldsymbol{\phi}}}(\boldsymbol{s}(\boldsymbol{y})). \tag{38}$$

The target of the learned rule depends on the loss and the joint training distribution. Under idealized squared-error training, for example, the output approaches a conditional mean rather than a model-independent inverse of the forward operator.

This distinction is important for multimodal problems: a point estimator can average separated solution families into a history that represents none of them well. Physical admissibility must also be enforced through the output coordinates, decoder, projection, or loss. The training distribution acts as a population prior, and the simulator determines which physical and observational effects the network can learn. Fast evaluation does not protect against simulator mismatch or extrapolation beyond the training support.

A neural point estimator primarily modifies $\mathcal{I}$ and can also modify $\mathcal{V}$ through a learned encoder or decoder. Its scientific use requires test-set calibration across the intended galaxy and observational domains, together with explicit detection of out-of-distribution inputs.

5.3.3. Simulation-Based Inference

Simulation-based inference (SBI) applies when a forward simulator can generate parameter–data pairs even though an explicit likelihood is unavailable or impractical. Unlike a neural point estimator, it can target a complete conditional distribution and thereby retain asymmetric, correlated, or multimodal uncertainty under the simulator and prior. For neural posterior estimation, the target and its approximation can be written as

$$p_{\mathrm{sim}}(\boldsymbol{\vartheta} \mid \boldsymbol{y}) \propto p_{\mathrm{sim}}(\boldsymbol{y} \mid \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}), \qquad q_{\widehat{\boldsymbol{\phi}}}(\boldsymbol{\vartheta} \mid \boldsymbol{y}) \simeq p_{\mathrm{sim}}(\boldsymbol{\vartheta} \mid \boldsymbol{y}), \tag{39}$$

where the density estimator is trained on parameter–data pairs drawn from the prior or proposal and the simulator [67]. Related SBI methods estimate likelihoods, likelihood ratios, or informative summaries instead of the posterior directly.

SBI can represent asymmetric, correlated, or multimodal conditional uncertainty and can amortize inference over large samples. It does not remove ill-posedness: observationally equivalent histories should remain broad or multimodal under the simulator-defined posterior. A sharply concentrated result can instead reflect a restrictive prior, incomplete training support, density-approximation error, or miscalibration.

The returned distribution is conditional on the simulator, not automatically on the real galaxy-generating process. Simulation-based calibration and repeated coverage can test the inference algorithm under that simulator, while posterior predictive and out-of-distribution checks test aspects of simulator adequacy. SBI primarily changes $\mathcal{I}$ and $\mathcal{U}$; it leaves the choice of representation, physical simulator, and prior structure fully consequential.

5.3.4. Comparison of Learned and Amortized Approaches

NMF principally learns a constrained representation, neural point inference learns a loss-dependent decision rule, and neural posterior inference learns a conditional distribution. These operations are complementary: an NMF or physically designed basis can represent the history, an emulator can accelerate the SED generator, and SBI can infer the posterior over the resulting coefficients and nuisance parameters.

Machine learning is therefore not a separate alternative to inverse-problem analysis. It supplies flexible approximations to selected components of the same inferential structure. Its evaluation should focus on physical admissibility, training-support dependence, simulator discrepancy, calibration, and the stability of the scientific functionals ultimately reported.

5.4. Unified Comparison and Modular Synthesis

Table 1 summarizes representative method families within the five-component structure. The central columns provide a navigation map showing where each family principally acts and where it depends materially on other components. The entries describe methodological families rather than every option available in a particular software implementation, and the table is not a ranking.

The method families summarized in Table 1 are realized in contemporary SED-fitting frameworks through different combinations of the five components, and a software name alone does not specify one unique configuration.

ProSpect combines flexible parameterizations of the SFH and metallicity history with alternative stellar-population libraries, a Charlot–Fall attenuation prescription, infrared re-emission, nebular-line modeling, and a simple active-galactic-nucleus and hot-torus component, and it can be used for both forward generation and inversion [68]. Within Equation (33), choices of the represented SFH and metallicity histories primarily modify $\mathcal{V}_q$ and $\mathcal{C}_q$, whereas the stellar, nebular, dust, and active-galactic-nucleus prescriptions modify $\mathcal{F}_q$; the returned estimates remain conditional on their joint configuration.

The Dense Basis approach places its principal innovation in historical representation and its validation: candidate SFH basis families are assessed by comparing goodness of fit for mock SEDs with recovery of the generating histories, thereby targeting mass-assembly timing and multiple star-formation episodes that can be obscured by a single rigid analytic form [69]. It consequently acts mainly through $\mathcal{V}_q$, $\mathcal{I}_q$, and $\mathcal{U}_q$, while its historical information remains conditional on the stellar-population, dust, observational, and training assumptions used to construct and validate the basis.

Table 1. Representative method families classified within $\mathfrak{M}_q = (\mathcal{V}_q, \mathcal{F}_q, \mathcal{C}_q, \mathcal{I}_q, \mathcal{U}_q)$. The component map is navigational rather than exclusive: every complete implementation must specify all five components.

Method Family	Component Map					Preserved or Returned Quantity	Key Structural Assumptions	Principal Caution
	$\mathcal{V}$	$\mathcal{F}$	$\mathcal{C}$	$\mathcal{I}$	$\mathcal{U}$			
Parametric SED fitting [52]	●	○	●	○	○	Model-conditional parameters or posterior summaries	Functional family, parameter bounds, priors, and physical SED model	Narrow uncertainty can coexist with persistent SFH misspecification
Fixed-bin or regularized nonparametric fitting [17,20]	●	○	●	○	○	Regularized estimate or posterior over a flexible represented history	Grid or basis, normalization, non-negativity, smoothness, continuity, or other priors	Nominal bins can exceed effective resolution and become prior-sensitive
PCA [61]	●		○	●		Maximum retained ensemble variance for a fixed rank and metric	Training population, preprocessing, weighting, normalization, and retained rank	Variance need not align with physical observability
MOPED [62]	○	○		●	○	Local Fisher information for the specified parameterization under its assumptions	Fiducial state, parameter-independent covariance, derivative rank, and parameter ordering	Global likelihood structure and model discrepancy need not be preserved
VESPA [21]	●	○	●	●	○	A data-dependent finite-resolution stellar-population history	Candidate partitions, SSP and dust models, non-negativity, and selection criterion	Selected components remain averaged and introduce selection uncertainty
NMF [65]	●		●	○	○	Additive low-rank reconstruction of the training ensemble	Training set, rank, loss, normalization, initialization, and regularization	The factorization is nonunique and need not follow observable directions
Neural point inference	○	○	○	●	○	A loss-dependent conditional point summary	Simulator, training distribution, architecture, loss, and output constraints	Can collapse multimodality and fail outside the training support
Simulation-based inference [67]	○	○	○	●	●	Approximate conditional distribution under the simulator and prior	Simulator, prior or proposal, training coverage, density family, and optimization	Internal calibration does not establish simulator adequacy for real galaxies

Notation: ● marks a component that the method family principally modifies or supplies; ○ marks a material- or implementation-dependent coupling. An empty cell indicates that the component is not usually the defining operation of that family, not that it can be omitted from a complete analysis.

BAGPIPES provides rapid generation and Bayesian nested-sampling inference for configurable combinations of photometric and spectroscopic data, with flexible SFH components, metallicity, nebular emission, and dust prescriptions [16]. CIGALE emphasizes modular panchromatic modeling under energy balance, constructing large model grids from selected SFH, stellar, nebular, attenuation, dust-emission, radio, and active-galactic-nucleus modules and estimating physical properties from the resulting likelihood distributions [70]. These implementations illustrate that flexibility in the physical forward model, flexibility in the SFH representation, and the treatment of active-galactic-nucleus emission are separable design choices rather than fixed properties of a broad method label.

This distinction is empirically consequential: applying 14 widely used SED-fitting codes to common CANDELS data, Pacifici et al. [9] found closer agreement in stellar mass than in SFR and dust attenuation and traced important differences to choices in SFH, nebular, dust, and active-galactic-nucleus modeling. The comparison reinforces the use of the complete tuple $\mathcal{M}_q$ for interpreting code-to-code variation and complements recent broader syntheses of galaxy SED modeling [8].

Several conclusions follow from Table 1. First, no method label uniquely determines the scientific content of an analysis. A nonparametric model can be strongly prior-dominated, a neural method can use a rigid analytic SFH, and a Bayesian calculation can

employ either a minimal or highly coupled physical model. The complete tuple, not the most visible algorithmic label, determines the estimand and its uncertainty.

Second, the methods preserve different objects. PCA preserves ensemble variance; MOPED preserves local Fisher information under specified conditions; VESPA adapts usable historical resolution; NMF preserves a non-negative low-rank approximation; a neural point estimator targets a loss-dependent conditional summary; and SBI approximates a simulator-defined conditional distribution. Low reconstruction error, high explained variance, Fisher preservation, and posterior calibration are therefore not interchangeable claims of historical accuracy.

Third, the components can be combined constructively. A learned basis can be embedded in a physical SED generator, observations can be compressed before posterior sampling, and amortized inference can be applied to fixed or adaptive representations. Modularity is scientifically useful only when the population, physical, observational, and uncertainty assumptions passed between modules remain consistent.

A complete comparison should consequently address at least four questions: whether the representation can express the scientifically relevant histories; whether those histories generate distinguishable observations; whether the inference and uncertainty approximation are calibrated under the stated model; and whether the conclusions remain stable under credible forward-model discrepancy. Computational efficiency is important for survey-scale work, but it is a separate criterion from observability and physical reliability.

The taxonomy reveals a common limitation: the historical representation is often fixed before the information content of the individual observation is assessed. A low-dimensional model can be too restrictive for informative data, while a high-dimensional model can introduce many prior-controlled directions for weak data. Section 6 develops information-driven adaptive representation as a general principle for aligning reported historical complexity with the modes supported by the combined physical and observational forward problem.

6. Information-Driven Adaptive Representation

Sections 3 and 4 established two complementary facts. First, integrated-light data constrain only a limited set of historical modes at finite and generally nonuniform resolution. Second, the representation, regularization, and prior determine how the remaining weak directions appear in the reconstructed history. A fixed representation chosen without reference to this information can therefore fail in opposite ways: it can be too restrictive for informative data or unnecessarily detailed for observations that support only broad historical averages.

In practical terms, a high-quality spectrum with diagnostics of recent star formation may support several distinct recent-age intervals, whereas an old or lower-information SED may constrain only a broad early-versus-late mass split. An adaptive representation should allow the first result to remain detailed but report the second at the broader resolution actually supported, even if both fits use a fine internal grid.

Information-driven adaptive representation treats this mismatch as a design problem. Its central principle is that the dimension, localization, and geometry of the reported historical representation should be commensurate with the information transmitted through the physical and observational forward problem. The objective is not to maximize the number of fitted components, nor to enforce the same minimum complexity for every galaxy. It is to preserve scientifically relevant freedom where distinct responses exist and to avoid presenting independent structure where the observations constrain only combinations or averages.

This principle is state- and observation-dependent. The useful temporal resolution changes with the stellar population, metallicity, dust content, redshift, wavelength coverage, spectral resolution, covariance, and signal-to-noise ratio. Recent star formation can be localized by ultraviolet or nebular diagnostics in one system, while ancient activity may remain constrained only through a broad cumulative mass. The same input grid therefore need not correspond to the same inferential resolution across a survey.

The idea has important precedents. The information analysis of integrated spectra by Ocvirk et al. [20] showed that the number of nominal stellar-population components can substantially exceed the number of combinations that are stably recoverable. VESPA made temporal resolution data-dependent at the level of individual spectra, returning finer components only where they were supported by the observations [21]. Here, we formulate the broader principle that representation design should respond to observability across data types, basis families, and coupled physical states.

The discussion proceeds from the distinction between nominal and information-supported complexity, to the forms of evidence that can inform adaptive temporal or basis design, and finally to the scope, limitations, and validation requirements of any data-dependent representation.

6.1. Nominal Freedom and Information-Supported Complexity

Let $\mathcal{V}_\alpha$ denote a candidate historical representation, where the design index α may specify the number of coefficients, temporal boundaries, basis functions, the dimension of an internal representation, or physical state variables. Its nominal dimension counts the independent coordinates introduced by the analyst. It does not count how many combinations of those coordinates produce distinguishable changes in the data.

Within a specified representation, physical scaling, reference state, forward model, and uncertainty description, the distinction can be expressed conceptually as

$$p_{\mathrm{obs}}(\alpha; \boldsymbol{h}_0, \boldsymbol{\omega}, \mathcal{F}, \boldsymbol{C}_{\mathrm{eff}}) \lesssim p_{\mathrm{nom}}(\alpha), \tag{40}$$

where p_{nom} is the number of represented degrees of freedom and p_{obs} denotes, only at a conceptual level, the number of independently useful directions supported under the stated conditions. Equation (40) is not a parameterization-invariant definition or a universal threshold rule. Its operational interpretation depends on the normalization of perturbations, the scientific accuracy required, and the treatment of observational and model uncertainty.

This distinction also fixes the terminology used throughout this Review: nominal resolution belongs to the representation, whereas effective resolution belongs to the complete forward and inference procedure.

A simple temporal example illustrates the distinction. Dividing one age interval into two bins adds an independent coefficient, but it adds an independently recoverable historical quantity only if redistributing star formation between the bins produces a response distinguishable from noise, nuisance variation, and model discrepancy. If the corresponding SSP contributions are nearly indistinguishable after attenuation and observational projection, the data constrain their sum more directly than their contrast. Finer binning then improves the approximation space without providing equivalent temporal resolution.

This does not imply that every high-dimensional internal representation is inappropriate. A fine basis can reduce approximation error in forward predictions and can support accurate estimates of broad functionals. The important distinction is between internal computational freedom and reportable historical information. An analysis can use many internal model coefficients while reporting only cumulative mass fractions, temporally averaged SFRs, or other combinations whose resolution is supported by the data.

Observable complexity is also not determined by the length of the data vector. Thousands of neighboring spectral pixels can describe a comparatively small number of physical response patterns, whereas a small number of complementary measurements can separate an important degeneracy. The relevant diagnostics are the observable modes, effective null directions, and resolution behavior described in Section 3, rather than the number of bands, pixels, or fitted coefficients alone.

The same logic applies beyond the SFH. Adding independent histories for enrichment, dust, gas, or interstellar-medium conditions can increase nominal freedom much faster than the data-supported dimension. In a coupled model, the observable directions can combine several physical components. Adaptive design should therefore be applied to the complete represented galaxy-evolution state and to the scientific functionals derived from it, not only to the number of SFH bins.

6.2. Adaptive Temporal and Basis Design

A general adaptive design begins with a family of candidate representations and uses the effective forward response to determine which level of complexity is appropriate for the data and scientific objective. Schematically,

$$\{\mathcal{V}_\alpha\}_{\alpha\in\mathfrak{A}} \longrightarrow \mathcal{V}_{\mathrm{adapt}}\Big(\boldsymbol{y}, \mathcal{F}_{\mathrm{phys}}, \mathcal{O}_{\boldsymbol{\omega}}, \boldsymbol{C}_{\mathrm{eff}}, Q\Big), \tag{41}$$

where Q denotes the scientific quantity or class of quantities to be inferred. This relation states the relevant dependencies; it is not a specific selection algorithm.

Candidate evidence for adaptation. The evidence that can motivate refinement, retention, or coarsening of one candidate representation can be organized into four complementary classes. These classes identify possible sources of support for adaptation; they do not define a universal selection statistic, threshold, or stopping rule.

Response-based evidence. Singular values, local Fisher information, and related derivative-based diagnostics assess whether a proposed additional direction produces a sufficiently distinct response after physical modeling, observational projection, covariance weighting, and nuisance freedom are taken into account. Such evidence addresses the transmission of perturbations by the forward problem, but remains local, scaling-dependent, and conditional on the adopted model.

Resolution-based evidence. Resolution matrices and averaging kernels assess whether a reported component is localized at the intended temporal or physical scale. A subdivision can be available in the representation yet scientifically uninformative when its corresponding kernels are broad, strongly overlapping, displaced, or multilobed. Resolution-based evidence therefore connects internal computational freedom to the effective quantities that can be reported.

Predictive evidence. Held-out observables, posterior-predictive behavior, cross-wavelength prediction, and related predictive criteria can test whether additional representational freedom improves consequences that were not used to justify the refinement itself. Predictive improvement is especially informative when the proposed component represents a physical distinction expected to affect independent spectral features or wavelength regimes, although it does not by itself establish unique historical recovery.

Robustness-based evidence. Stability across reasonable representations, priors, nuisance models, forward-model variants, and repeated data realizations assesses whether a proposed feature survives changes in assumptions that act along weak directions. A refinement whose location or amplitude changes substantially under modest alternatives should be treated as assumption-sensitive even when it improves the fitted likelihood.

The four classes are complementary. Response- and resolution-based diagnostics describe local information transmission and estimator behavior, whereas predictive and

robustness diagnostics examine the consequences of adaptation beyond one fitted representation. A future adaptive method may combine them differently according to the target functional Q, the observing configuration, the computational setting, and the acceptable scientific error. The present Review does not promote any one combination as a universal decision rule; its calibration and repeated-sampling evaluation belong to the three-layer program of Section 8.

For temporal bins, the design principle is to retain a subdivision when the associated historical contrast produces a credible and scientifically useful response, and to combine intervals when only their aggregate contribution is stable. Recent epochs may support finer components because stellar spectra and nebular emission evolve rapidly, whereas old populations commonly require broader averages. These tendencies are qualitative rather than universal, because metallicity, dust, redshift, and wavelength coverage can change the response.

The selected bins should still be interpreted through their effective resolution. A component that remains in an adaptive partition is not automatically localized to its nominal boundaries. Its estimator can average the true history over a broader or asymmetric interval. Adaptive binning is most transparent when the reported components are broadly compatible with the resolution kernels or stable integrated quantities identified by the analysis.

Adaptive design is not limited to piecewise-constant histories. Splines, wavelets, multiresolution dictionaries, PCA or NMF components, and learned representations can all provide candidate bases. The singular vectors of a local response are useful diagnostics because they identify well-transmitted and weakly transmitted combinations. They are not automatically suitable as the final physical basis: they can be nonlocal, signed, state-dependent, and difficult to interpret under positivity constraints. Their most useful role is to reveal which contrasts in an interpretable basis are redundant or unsupported.

Multiresolution representations offer a natural compromise. Broad components describe the mass assembly that is robustly observable, while finer components represent departures from that coarse history where the data provide additional leverage. The hierarchy need not be uniform in time or across physical variables. An analysis may support localized recent star formation, a broad ancient stellar component, a mean enrichment history, and only an integrated dust-heating quantity.

Learned bases add information from a training population but must be assessed after propagation through the physical and observational operators. Two components that are distinct in simulation space can produce nearly identical SEDs, while a low-variance component can be observationally distinctive. Representation efficiency in the training ensemble and identifiability in the target data are separate criteria. Rare histories and out-of-distribution galaxies also remain conditional on the support of the training set.

The scientific target Q should guide the design. A representation optimized for total mass assembled before a broad epoch need not resolve the duration of a recent burst, and a representation intended to identify recent quenching need not allocate equal freedom to ancient times. Target-aware representation can therefore be more informative and more transparent than attempting uniformly fine reconstruction and deriving the scientific quantity afterward.

Adaptation can be performed for individual galaxies or within a population model. Object-level adaptation reflects galaxy-specific observability, but its coefficients are not automatically comparable across different partitions. Population studies should compare common functionals defined across all representations. Hierarchical or shared multiresolution models can provide common coordinates and borrow strength, but they introduce population assumptions and selection dependence that must be made explicit.

These possibilities are not competing definitions of adaptivity. They implement the same principle in different settings: representation freedom should be introduced where it contributes a distinguishable and interpretable response to the scientific question, and unsupported detail should be retained only as explicitly prior-controlled internal freedom or replaced by broader reportable quantities.

6.3. Scope, Limitations, and Validation Requirements

Adaptive representation does not remove the fundamental limitations of the inverse problem. It cannot recover an exact or effective null direction, correct an inadequate forward model automatically, or establish global uniqueness from a local response analysis. It can reduce the mismatch between nominal and observable complexity, but the resulting representation remains conditional on the adopted physical model, observational operator, covariance, candidate family, and scientific target.

Model discrepancy is a particularly important limitation. A refinement can appear informative because the adopted model predicts a distinctive response in a wavelength region that is not modeled accurately. The additional component can then absorb calibration error, stellar-library mismatch, or incomplete dust or nebular physics. Adaptivity must therefore be evaluated relative to model credibility and held-out prediction, not solely through an improved fit to the data used for selection.

The candidate family also matters. An adaptive procedure selecting among piecewise-constant histories cannot escape that family, and a learned dictionary cannot represent states absent from its training support. Different bases can describe the same observable subspace with different numbers of coefficients. The selected dimension is consequently an analysis-dependent measure of supported complexity, not a parameterization-independent measurement of the galaxy's intrinsic historical complexity.

Data-dependent representation introduces selection uncertainty. When the same observation determines both the representation and its coefficients, uncertainty conditional on the selected model does not automatically include the possibility that another noise realization would have selected a different partition or basis. General post-selection results make this issue clear [71], but the nonlinear, constrained, and physically coupled setting considered here requires dedicated methodological treatment.

Because the representation is selected from the data, its uncertainty cannot be assessed conditionally on one selected partition or basis alone. The complete adaptive pipeline must be rerun in each mock realization, and performance should be evaluated through scientific functionals defined consistently across candidate representations. Selection frequency and boundary stability may be reported as representation-specific diagnostics, while matched- and mismatched-model experiments, coverage analysis, held-out prediction, and general reporting requirements are developed in Section 8.

This Review deliberately stops at these general requirements. It does not define a universal split or merge statistic, information threshold, stopping rule, hyperparameter-selection procedure, post-selection correction, or production implementation. It also does not claim demonstrated gains in coverage, robustness, or real-data performance. Formal properties of data-supported complexity and concrete adaptive algorithms are reserved for dedicated foundational and implementation studies.

The central conclusion of this section is therefore limited but consequential: historical resolution is an emergent property of the galaxy state, physical forward model, observational operator, uncertainty description, representation, and inference. Information-driven adaptive representation seeks to align the reported complexity with that resolution. Section 7 extends this principle from the SFH to coupled galaxy-evolution states, for which the supported directions can combine several physical histories.

7. From SFHs to Coupled Galaxy-Evolution States

The SFH provides the clearest example of a continuous history inferred from integrated light, but it is not the only history that shapes a galaxy SED. Stellar populations inherit an evolving chemical composition; young stars affect and are affected by ionized gas; dust changes in abundance and properties; and attenuation, scattering, and re-emission depend on the distribution of stars and interstellar material. The physical roles of these processes and their mapping into observed data are developed in Section 2. The purpose of the present section is to examine the corresponding inferential consequence: how the target of reconstruction changes when these components are allowed to vary together.

The central point is that enlarging the represented galaxy-evolution state does not turn an integrated SED into a separate measurement of each physical history. It commonly exposes cross-component degeneracies that were hidden when metallicity, dust, gas, or geometry were fixed. The appropriate target is therefore a set of observable modes within a coupled galaxy-evolution state, together with physically interpretable functionals that remain stable under the allowed exchanges among its components.

This extension also changes the role of physical modeling. Coupling enrichment, dust, gas, and radiation to star formation can exclude inconsistent trajectories and improve predictions, but the resulting reduction in uncertainty is supplied partly by the model rather than by an additional observation. Integrated Galactic Archaeology must therefore distinguish information transmitted by the data from separation imposed through physical admissibility. The first subsection defines the coupled state and its reportable quantities, while the second describes physically admissible subsets and practical model hierarchies.

7.1. Beyond a Single Star Formation History

For the present discussion, the coupled galaxy-evolution history introduced in Equation (2) can be grouped into physical blocks,

$$\boldsymbol{h}_{\rm gal}(t) = \left[\boldsymbol{h}_{\psi}(t), \boldsymbol{h}_{Z}(t), \boldsymbol{h}_{\rm d}(t), \boldsymbol{h}_{\rm ISM}(t), \boldsymbol{h}_{\rm RT}(t)\right], \tag{42}$$

where Equation (42) regroups the components of $\boldsymbol{h}(t)$ rather than defining a different representation of the physical state. The block $\boldsymbol{h}_{\psi}$ contains the star formation history; $\boldsymbol{h}_{Z}$ contains the total and gas-phase chemical states and the compositions inherited by newly formed stars; $\boldsymbol{h}_{\rm d}$ contains the dust mass, composition, and grain-size distribution; $\boldsymbol{h}_{\rm ISM}$ contains quantities such as density, ionization state, pressure, and radiation-field intensity; and $\boldsymbol{h}_{\rm RT}$ contains the geometry and transfer state. The grouping is conceptual, and the blocks need not be dynamically or statistically independent.

Calibration, redshift, kinematic, and other observational quantities are most naturally treated as nuisance variables associated with the observational operator. They can nevertheless be included in the local parameter vector when their data-space responses overlap with those of the physical histories. Keeping this distinction explicit preserves the separation between the galaxy state and the observational description introduced in Section 2.

Equation (42) is a conceptual partition, not a prescription to fit every block as a freely varying function. A component can be represented explicitly, derived from another component through an evolution model, described by a small number of parameters, or marginalized as nuisance structure. For example, a stellar metallicity history can be fitted independently, generated from a chemical-evolution model, or replaced by a mean composition when the observations do not support temporal enrichment information. These choices define different inverse problems even when they use the same measured SED.

The blocks are physically connected. Star formation and stellar evolution alter the gas, metal, radiation, and dust states, while those states affect the radiation through which star formation is inferred. The detailed process models need not be repeated here; what matters

for the inverse problem is that a perturbation assigned to one block can be compensated by changes in another. Fixing the other blocks removes those compensating directions by assumption and can make an SFH appear more sharply determined than it would be in a joint analysis [6,7,23].

A coupled formulation should therefore distinguish among quantities required internally by the forward model, quantities inferred explicitly, and scientific quantities intended to be reported. These sets need not have the same dimension. A panchromatic model can require many internal state variables to generate a realistic SED while supporting only a few robust outputs, such as cumulative stellar-mass assembly, a broad enrichment timescale, obscured recent star formation, or an integrated dust-heating contribution.

Increasing physical completeness can either narrow or broaden the inferred SFH. A chemically consistent or energy-balanced model can exclude physically implausible combinations and reduce conditional uncertainty. Releasing a previously fixed dust, metallicity, or nebular component can instead reveal a wider family of histories that the data cannot separate. The latter is not necessarily a loss of inferential quality; it can be a more faithful account of the information in the observations.

The aim is consequently not to replace one SFH curve with several independently fitted curves. It is to construct a state representation in which important physical alternatives are explicit, unsupported freedom is controlled transparently, and the final scientific quantities can be related to the combinations that the data actually constrain.

7.2. Physically Admissible States and Model Hierarchies

A finite representation of the coupled state generally contains trajectories that are mathematically expressible but physically impossible. Star formation rates and component masses must remain non-negative; cosmic-age limits must be respected; and any adopted mass, metal, or energy relations must be satisfied. More detailed models can connect enrichment, gas flows, dust evolution, nebular conditions, and radiative transfer. These restrictions define the physically admissible subset of the represented state space.

Let $\mathcal{X}_{\rm gal}$ denote the ambient space of represented histories. We write the admissible subset schematically as

$$\mathcal{A}_{\rm phys}(\boldsymbol{\phi}_{\rm phys}) \equiv \left\{ \boldsymbol{h}_{\rm gal} \in \mathcal{X}_{\rm gal} \;\middle|\; \begin{matrix} \mathcal{E}_{\rm phys}[\boldsymbol{h}_{\rm gal};\boldsymbol{\phi}_{\rm phys}]=\boldsymbol{0} \\ \mathcal{G}_{\rm phys}[\boldsymbol{h}_{\rm gal};\boldsymbol{\phi}_{\rm phys}]\succeq\boldsymbol{0} \end{matrix} \right\}, \tag{43}$$

where $\mathcal{E}_{\rm phys}$ denotes equality relations or evolution equations, $\mathcal{G}_{\rm phys}$ denotes inequality and boundary constraints, and $\boldsymbol{\phi}_{\rm phys}$ contains the assumptions and parameters defining those admissibility relations. The latter may overlap with, but need not be identical to, the physical ingredients collected in $\boldsymbol{\theta}_{\rm phys}$ in the forward operator.

The content of Equation (43) is deliberately model-dependent. Positivity and the finite age of the Universe are comparatively general restrictions. Energy balance, a specified enrichment law, dust-growth and destruction prescriptions, or a geometry-dependent radiative-transfer calculation introduce progressively stronger physical structure [27,29–31,72]. These relations can improve coherence and prediction, but their contribution to the posterior should not be described as direct information from the SED.

Physical coupling can remove observationally acceptable but inconsistent solutions. It can also introduce new uncertainty through stellar yields, gas-flow efficiencies, mixing, grain processing, or geometry. A more elaborate model is therefore not guaranteed to make each represented physical component more identifiable. Its principal advantage is that it exposes the physical alternatives and allows them to be tested against additional observables rather than hidden within an SFH or attenuation parameter.

A practical analysis can use a hierarchy of forward models. Depending on the scientific target, one can move from stellar-population synthesis to stellar and nebular emission, spectrochemical coupling, panchromatic energy balance, dust evolution, and geometry-dependent radiative transfer. Section 2 contains the physical definitions and representative implementations of these stages. The methodological point here is that every stage changes both the admissible state and the observable degeneracies.

Within such a hierarchy, model complexity should be introduced in response to a demonstrated need. An additional component is consequential when it reduces structured residuals, changes a bias found in mismatched mocks, improves prediction of a held-out observable, or separates a scientifically important alternative. An added state variable that only improves the fitted likelihood through weakly constrained compensation can increase model dependence without improving the target inference.

The admissible set can be curved, bounded, nonconvex, or disconnected. Local mode analysis describes its tangent response near a reference state, but it does not characterize the complete global solution. Hard physical coupling can also create overly narrow results when the evolution model is incomplete. More uncertain relations should therefore be parameterized, tested across model variants, or accompanied by explicit process or observation-space discrepancy rather than enforced as exact truths.

The representation of the coupled state should itself follow the information-driven principle of Section 6. It is rarely justified to assign equally fine resolution to the SFH, enrichment, dust, gas, and geometry. The data may support a localized recent-SFH component, a broad composition summary, and only an integrated dust or gas quantity. Useful directions can also cross the boundaries between these blocks, so independently fine bases for every history can be less effective than a smaller set of interpretable joint functionals.

Simulation libraries and learned manifolds offer another form of admissibility by encoding correlations among galaxy properties. They can improve efficiency and realism within their support, but they inherit the subgrid physics, resolution, and population biases of the simulations. Out-of-distribution diagnostics are therefore necessary before such a manifold is treated as the complete space of plausible galaxies.

A reliable coupled-state analysis should state which variables are inferred explicitly, which are derived from physical relations, which are marginalized, and which conclusions change when those choices are varied. It should distinguish direct observational separation from model-assisted separation and should report scientific functionals at the effective resolution supported by the joint response.

The central conclusion is that an integrated SED constrains a limited set of combinations within a physically structured state space. The task is to identify which combinations are supported by the observations, which distinctions are supplied by physical assumptions, and which conclusions survive reasonable changes in both. Section 8 organizes the controlled experiments, inverse-problem diagnostics, and forward-model revisions required to make those distinctions reliable.

8. A Three-Layer Research Program

The preceding sections have identified the main elements of Integrated Galactic Archaeology: a physical and observational forward problem, a restricted set of observable historical modes, explicit representation and prior choices, information-driven control of model complexity, and a coupled galaxy-evolution state. The remaining question is how these elements should be developed and tested in practice. The answer requires a research program in which empirical validation, inverse-problem diagnosis, and forward-model construction inform one another.

We organize this program into three layers with distinct but interacting roles. Layer I uses controlled mock observations to determine where an inference succeeds or fails and returns an empirical map of failure modes. Layer II explains those outcomes using the observability, resolution, regularization, prior, and discrepancy concepts introduced in Sections 3–6 and returns a structural diagnosis. Layer III revises the state representation and forward-model hierarchy using the physical construction developed in Sections 2 and 7, producing new simulations and predictions for renewed testing. The three layers are not independent stages and do not define a particular software pipeline. They describe the evidence required to turn an apparently successful fit into a scientifically reliable reconstruction. The iterative relationship among them is summarized in Figure 5.

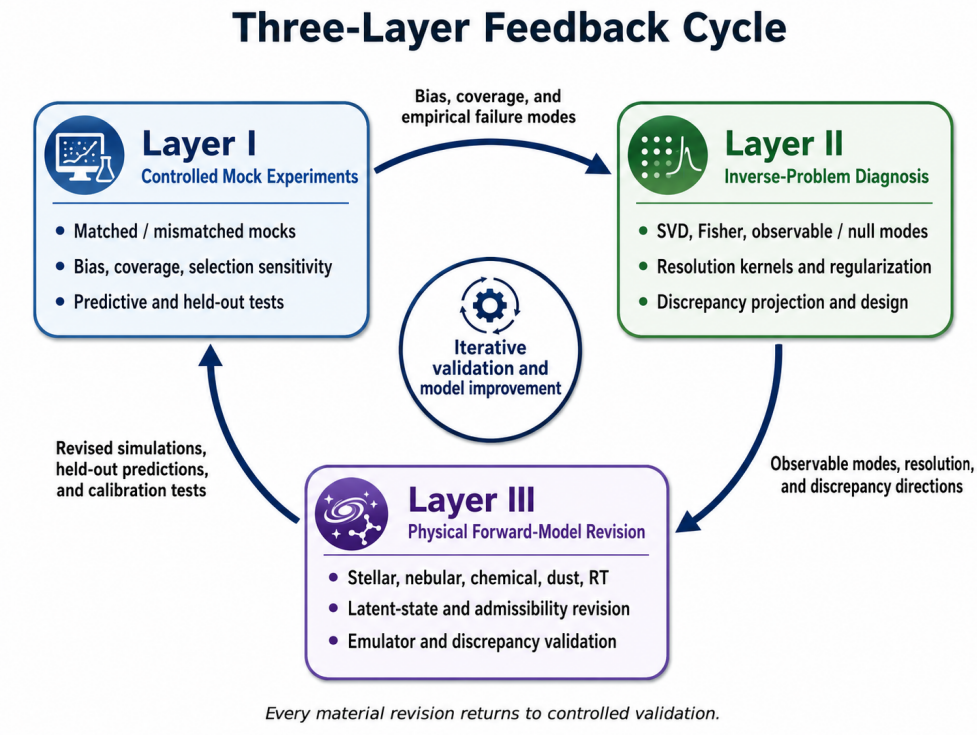


Figure 5. Three-layer feedback cycle for Integrated Galactic Archaeology. Layer I identifies empirical failure modes through controlled mock experiments, Layer II diagnoses their structural origins through inverse-problem analysis, and Layer III revises the state representation and forward-model hierarchy. Revised simulations and predictions then return to Layer I for renewed validation, so that every material change in the representation or forward model is tested within the complete cycle.

8.1. Layer I: Controlled Mock Experiments

Controlled mock experiments provide the empirical laboratory of the framework. A known galaxy history is propagated through a specified physical forward model, observational operator, and stochastic measurement process, after which the intended inference pipeline is applied without access to the generating history. Such experiments can measure bias, variance, effective resolution, coverage, prior sensitivity, and predictive performance under conditions that can be varied systematically. Existing SED-fitting frameworks provide practical tools for constructing these tests; BAGPIPES, for example, supports flexible mock spectra and photometry together with Bayesian recovery experiments [16].

The first requirement is to separate matched-model from mismatched-model tests. In a matched-model experiment, generation and inference use the same representation, physical forward model, observational operator, and stochastic data model. This tests computational correctness and conditional statistical behavior, but it describes an idealized

closed world. In a mismatched-model experiment, the generating and fitted descriptions differ in selected respects, such as the SFH family, stellar population model, abundance pattern, nebular prescription, attenuation geometry, dust emission, calibration covariance, or selection rule. These tests determine whether model inadequacy is exposed as lack of fit or projects into a biased history.

A useful mock suite must span both underlying galaxy states and observing configurations. Young, old, dusty, metal-poor, bursty, rejuvenating, and recently quenched systems need not support the same modes. The same intrinsic history can also yield different recoverability under broadband photometry, narrowband data, spectroscopy, or incomplete multiwavelength coverage. Validation should therefore map performance across the intended scientific domain rather than report one average error over a heterogeneous test set.

Comparisons should be made through scientific quantities that remain meaningful across representations. Examples include stellar mass formed over broad intervals, cumulative mass-assembly fractions, recent mean SFRs, mass-weighted ages, broad transition epochs, metallicity summaries, and predictions for held-out observables. Coefficients associated with different adaptive bins or learned bases are generally not commensurable, whereas common functionals can be evaluated even when the internal representations differ.

Every data-dependent stage must be repeated within each mock realization. If the basis, number of temporal components, regularization strength, nuisance-parameter model, or quality cut is selected from the data, fixing that choice before repeated-sampling evaluation omits selection uncertainty. Coverage and calibration should also be examined conditionally across galaxy type, redshift, signal-to-noise ratio, and other scientifically relevant regimes, because correct average behavior can conceal severe failures in particular subpopulations.

The principal output of Layer I is not a ranking of complete fitting codes. It is an empirical map of failure modes: where temporal structure is merged, where noise or priors create apparent detail, where model discrepancy projects into the inferred state, where uncertainty loses calibration, and which held-out predictions fail. The five-component taxonomy $\mathfrak{M}_q = (\mathcal{V}_q, \mathcal{F}_q, \mathcal{C}_q, \mathcal{I}_q, \mathcal{U}_q)$ of Section 5 allows these failures to be associated with the representation, forward model, added structure, inference mechanism, or uncertainty assessment rather than with a software name alone.

8.2. Layer II: Inverse-Problem Theory and Information Analysis

Layer II converts the empirical failure map from Layer I into a structural diagnosis. It uses the observable modes, effective null directions, resolution operators and averaging kernels, regularization and prior effects, and discrepancy projections already defined in Sections 3 and 4, together with the representation-design principles of Section 6. The purpose is not to derive those tools again, but to determine which mechanism produced a particular failure and which modification is likely to address it.

For example, alternating activity between neighboring old-age bins can indicate a weak contrast mode rather than genuine temporal structure. A broadened recovered burst can reflect the estimator's resolution rather than an optimization error. A narrow posterior along a weak response direction can arise from a prior or physical boundary. A coherent spectral mismatch aligned with an age- or dust-sensitive response can bias the inferred history while leaving a small residual. These cases require, respectively, coarser reporting, explicit resolution characterization, assumption-sensitivity analysis, or revision of the forward and discrepancy models.

The same analysis connects observational choices to historical information. Removing a wavelength interval, changing spectral resolution, or replacing spectroscopy with photometry changes the effective response and can erase or rotate particular modes. Conversely, an additional measurement is valuable when it introduces a response that is complementary to an existing degeneracy. Layer II therefore guides not only representation design but also the selection of wavelengths, spectral features, and scientific functionals discussed in Section 9.

These diagnostics remain conditional on the reference state, the representation and coefficient scaling, the physical and observational forward models, and the covariance or likelihood used in the analysis. They should be evaluated across representative galaxy states and supplemented by global posterior exploration when nonlinear or disconnected solution families are possible. Disagreement between a local prediction and mock behavior is itself informative: it can reveal active constraints, nonlinearity, an incorrect covariance, or numerical effects that are absent from the local analysis.

The output of Layer II is actionable rather than purely descriptive. It identifies which components should be combined, which quantities can be reported at a stable resolution, where prior sensitivity must be shown, which directions associated with the represented galaxy state or observational nuisance parameters should be included, and which additional observations are likely to separate an important degeneracy. Those recommendations guide revisions in Layer III, and the revised model and pipeline are then returned to Layer I for renewed testing.

8.3. Layer III: Physical Forward Modeling and State Construction

Layer III revises the state representation and the physical and observational components of the forward model. The relevant physical ingredients—stellar population synthesis, enrichment, nebular emission, dust processing, and radiative transfer—and the observational operator have been described in Sections 2 and 7. Here, the emphasis is on how the state description, $\mathcal{F}_{\mathrm{phys}}$, and, when required, $\mathcal{O}_{\omega}$ should be developed in response to evidence from the other two layers.

Greater physical complexity is not automatically an improvement. Adding an unconstrained component can reduce residuals by creating new compensation directions without making the historical interpretation more accurate. An additional process is scientifically useful when it reduces structured mismatch, improves predictions of independent observables, changes a bias identified in mismatched mocks, or supplies a physically justified relation among otherwise degenerate state variables. Conversely, a component that has negligible effect on the target quantities may be fixed, marginalized approximately, or omitted within a clearly stated domain of validity.

A hierarchy of forward models is therefore preferable to one presumed final model. A stellar-only description can be extended to include nebular emission, chemical coupling, panchromatic energy balance, dust evolution, or geometry-dependent radiative transfer as required by the scientific target and validation data. Comparing adjacent levels of the hierarchy distinguishes conclusions supported directly by the observations from precision supplied by physical coupling. If a quantity becomes narrow only after a strong evolution relation is imposed, that gain should be described as model-assisted.

Forward-model development must also specify uncertain ingredients and model discrepancies. Alternative stellar libraries, abundance prescriptions, nebular calculations, attenuation models, and radiative-transfer approximations provide empirical directions for sensitivity tests, although they cannot span all possible inadequacies. Held-out lines, wavelength ranges, spatial information, or external gas and dust measurements are especially

valuable because they test the added physics rather than merely reward additional flexibility in the fitted data.

Computational approximations should be assigned to the component that they modify. Interpolation and emulation approximate the forward model $\mathcal{F}$, whereas compression and simulation-based inference act primarily through the inference and uncertainty components $\mathcal{I}$ and $\mathcal{U}$. All can make physically rich models tractable, but their accuracy must be assessed in the response directions that matter for the inverse problem. Small average flux errors are not sufficient if an approximation distorts a weak spectral feature or derivative that carries much of the information about a target history.

The output of Layer III is not a universal forward model. It is a hierarchy of physically interpretable models with explicitly defined state variables, uncertain ingredients, discrepancy assumptions, validation observables, and domains of applicability. Every material revision defines a new inverse problem and therefore returns to Layer I for renewed testing and, where needed, to Layer II for structural diagnosis.

8.4. The Three-Layer Feedback Cycle

The preceding subsections describe the three layers individually. Their operational relationship is cyclic rather than linear:

$$\begin{aligned} \text{controlled mock experiments} &\rightarrow \text{empirical failure map} \\ &\rightarrow \text{inverse-problem diagnosis} \\ &\rightarrow \text{representation, state, or forward-model revision} \\ &\rightarrow \text{predictive and calibration tests} \\ &\rightarrow \text{renewed controlled mock experiments.} \end{aligned} \tag{44}$$

A revision should be tested under the conditions that exposed the original failure and under new conditions designed to reveal unintended consequences. The test suite should expand as the model becomes more ambitious, but highly realistic end-to-end simulations should not replace simple experiments that isolate individual mechanisms.

The cycle can recommend different responses to superficially similar failures. If adjacent ancient-SFH components are observationally inseparable, the appropriate response may be to coarsen the reported history rather than improve the optimizer. If an emission-line residual is orthogonal to the local response range, the nebular or calibration model should be revised rather than adding a burst. If an additional wavelength range rotates an age–dust degeneracy, the observational operator has expanded the useful subspace. If several physical models remain equally predictive, the appropriate result may be a model-conditional range rather than a forced choice.

Representation, inference, and uncertainty must be reassessed after every material change in the represented galaxy state, representation, or forward model. New physical variables create new degeneracies; a new discrepancy term changes the effective information; and an adaptive representation introduces selection uncertainty. Validation therefore applies to the complete $(\mathcal{V}, \mathcal{F}, \mathcal{C}, \mathcal{I}, \mathcal{U})$ configuration rather than to each module in isolation. For survey applications, the same cycle must also reproduce detection, targeting, redshift success, quality cuts, and follow-up selection before population trends are interpreted.

The cycle naturally produces a hierarchy of scientific claims. The strongest claims concern common functionals that are recovered with calibrated uncertainty, supported by identifiable responses, stable across reasonable representations and forward models, and successful in predicting independent observables. More model-sensitive conclusions can remain valuable when their dependence is stated explicitly. Fine structure that changes

under modest revisions should be treated as a conditional model realization rather than as an observed event.

This research program also defines the boundary of the present Review. This Review establishes the shared architecture and identifies the questions that a reliable method must answer. Formal properties of observable subspaces and data-supported complexity, concrete adaptive selection rules, post-selection inference, extensive mock comparisons, and production implementations require dedicated methodological studies. The three-layer cycle provides the structure within which those later results can be developed and evaluated without being claimed prematurely here.

Integrated Galactic Archaeology is thus a continually tested relationship among observations, physical and observational forward models, representations, and inference. Controlled experiments establish empirical behavior; inverse-problem analysis explains the information structure; and physical modeling determines whether the inferred histories correspond to credible galaxy evolution. The following section returns to the observational outlook and asks how future data and survey design can enlarge the set of evolutionary modes that can be reconstructed reliably.

9. Outlook and Conclusions

Integrated-light astronomy is entering a period in which increasingly deep spectroscopy, wide-area imaging, and panchromatic observations can constrain galaxy evolution over unprecedented ranges of redshift, environment, and physical state. These advances will increase both the statistical precision of individual measurements and the size of the galaxy samples available for archaeological inference. Their scientific effect, however, will depend on whether the added observations distinguish new evolutionary combinations rather than merely measure already observable directions more precisely.

The inverse-problem framework developed in this Review provides a natural basis for making this distinction. The value of an observation is determined by its response to the coupled galaxy-evolution state, its complementarity to existing data, and the credibility of the physical forward and observational operators used to interpret it. Additional measurements can enlarge the observable subspace, improve the effective resolution of selected scientific functionals, or test alternative physical explanations. They can also add redundant information, introduce new nuisance parameters, or expose model discrepancy that was hidden at lower precision.

The outlook for Integrated Galactic Archaeology therefore depends on coordinated progress in observation, forward modeling, representation, inference, and validation. No single future facility or computational method will eliminate the non-uniqueness of integrated-light reconstruction. Progress will instead come from designing observations that separate scientifically important degeneracies, constructing physical models capable of predicting the resulting data, and reporting conclusions at the resolution and level of model dependence that those observations support.

9.1. Expanding the Observable Subspace Through Multiwavelength Data

Multiwavelength observations are valuable because different spectral regimes respond to different combinations of stellar age, chemical composition, ionized gas, dust, and interstellar-medium conditions. The relevant gain is not simply a longer wavelength baseline. It is the addition of response directions that are sufficiently distinct from those already measured. Table 2 summarizes the principal leverage and persistent limitations of representative observational regimes.

Table 2. Principal galaxy-evolution sensitivities of representative observational regimes. The listed associations are not one-to-one correspondences: each observable remains conditional on the physical forward model, observing configuration, and other components of the coupled galaxy state.

Observational Regime	Principal Leverage	Persistent Degeneracies and Model Dependence	Representative Facilities or Data
Rest-frame ultraviolet	Recent star formation, short-lived massive stars, and low-mass fractions of young populations	Dust attenuation, metallicity, binary evolution, the upper initial mass function, and stochastic recent activity	Space-based ultraviolet data and rest-frame ultraviolet observations at higher redshift; JWST spectroscopy for appropriate redshift ranges [73]
Optical spectroscopy	Stellar ages, metallicities, abundance-sensitive features, nebular emission, and stellar kinematics	Age–metallicity coupling, abundance response, velocity broadening, continuum calibration, and incomplete stellar or nebular libraries	Ground-based and space-based spectroscopy, including redshifted rest-frame optical spectroscopy with JWST [73]
Near-infrared imaging and spectroscopy	Older stellar populations, redshifted rest-frame optical diagnostics, stellar-mass-sensitive continuum emission, and rest-frame optical morphology	Uncertain evolved stellar phases, molecular features, nebular contamination, dust, and mass-to-light-ratio systematics	JWST, Euclid, and Roman imaging or slitless spectroscopy [73–75]
Mid- and far-infrared emission	Absorbed energy, obscured star formation, dust heating, and grain emission	Heating by older populations, dust geometry, radiation-field distributions, grain properties, and active-galactic-nucleus contamination	JWST mid-infrared observations and complementary infrared facilities [73]
Millimeter and submillimeter data	Cold dust, molecular and atomic gas, interstellar-medium excitation, and spatially resolved material reservoirs	Dust temperature and emissivity, line-to-mass conversion factors, phase structure, chemistry, and aperture matching	ALMA continuum and line observations [76]
Wide-field and time-domain surveys	Population distributions, environmental trends, morphology, photometric-redshift constraints, variability, and statistical calibration samples	Strong compression by broad bands, blending, heterogeneous depth, selection, redshift uncertainty, and population-model misspecification	Euclid, Rubin's Legacy Survey of Space and Time, Roman, and cross-survey combinations [74,75,77]

The sensitivities in Table 2 should be interpreted through the response geometry introduced in Sections 3 and 7. Ultraviolet data can strengthen sensitivity to recent star formation, but their ability to separate recent activity from attenuation depends on complementary dust information. Optical spectral features can distinguish some age, metallicity, abundance, and nebular directions, but the separation remains conditional on spectral resolution and forward-model accuracy. Far-infrared and millimeter observations do not provide the same direct, age-sensitive diagnostics of stellar populations as photospheric spectral features, but their contribution to SFH inference is not limited to a single integrated recent-SFR estimate. Far-infrared dust emission constrains the absorbed luminosity and obscured recent star formation, subject to possible heating by older stars or an active nucleus, while millimeter continuum and line measurements constrain cold dust, gas reservoirs, and interstellar-medium conditions associated with ongoing and future star formation. When modeled jointly with ultraviolet, optical, and near-infrared data through energy balance or radiative transfer, these observables can reduce dust-SFH degeneracies and strengthen constraints on recent stellar-mass assembly. They do not, by themselves, time-resolve older stellar populations, and any detailed SFH inference from them remains conditional on the adopted heating, geometry, dust, gas, and interstellar-medium model.

A new observation therefore enlarges the observable subspace only when it adds a response that is not already contained, to adequate precision, in the existing data. Several neighboring broad bands can improve the measurement of one continuum direction while leaving the principal age–dust degeneracy largely unchanged. Conversely, one diagnostic

line, break, or dust constraint can be highly informative when its response is complementary to the alternatives permitted by the current model.

Complementarity is also state-dependent. Nebular lines can provide strong short-timescale information for a star-forming galaxy but little leverage for a passive system. Infrared observations can be decisive for a dusty galaxy while yielding primarily an upper limit for another. The historical information gained from a fixed filter set or spectral interval therefore varies with the galaxy state, redshift, signal-to-noise ratio, and the other data already available.

JWST and ALMA illustrate the value of combining physically different projections of the same system. Near- and mid-infrared spectroscopy can constrain stellar and nebular diagnostics at high redshift, while millimeter and submillimeter observations probe cold dust, gas, and interstellar-medium conditions [73,76]. Joint interpretation can connect past stellar assembly to obscured activity and current material reservoirs, but only when aperture, morphology, lensing where relevant, and the spatial association of the emitting components are included in the observation model.

Euclid, Rubin, and Roman provide a complementary survey-scale opportunity. Their wide-field optical and near-infrared data can map population distributions across large volumes and can supply homogeneous samples for hierarchical inference [74,75,77]. Large sample size does not remove individual-galaxy degeneracy. It instead increases the importance of calibration, selection modeling, population-level prior assumptions, and computational methods capable of preserving uncertainty across heterogeneous data quality.

The scientific leverage of multiple surveys will be greatest when their measurements are treated as components of a joint observational operator $\mathcal{O}_\omega$ rather than when independently derived catalog parameters are combined afterward. Joint modeling can preserve cross-survey covariance, aperture and blending information, non-detections, and the possibility that one dataset constrains nuisance quantities required to interpret another. It also makes explicit the selection process determining which galaxies receive each class of observation.

The computational scale of future surveys will require emulation, compression, amortized inference, or hierarchical approximations. These tools should be deployed within the modular structure of Section 5. Computational efficiency does not justify imposing one uniformly detailed historical representation on every galaxy. Information-driven representations can allocate greater complexity where the data support it while reporting common broad functionals across the full population.

Multiwavelength observations also provide tests of model adequacy. A history inferred from optical data can be required to predict ultraviolet emission, infrared luminosity, or spectral lines not used in the fit. Failure across wavelength regimes should not automatically be removed by adding flexible SFH structure. Such tension can identify missing stellar, nebular, dust, radiative-transfer, calibration, or aperture physics and should feed back into the three-layer program of Section 8.

The principal opportunity offered by future observations is therefore an expansion of the set of independently testable physical explanations. Progress should be assessed not only through narrower marginal intervals, but also through changes in historical resolution, degeneracy structure, sensitivity to priors, robustness across forward models, and successful prediction of independent observables.

9.2. Survey Design, Model Adequacy, and Calibrated Inference

The observable-subspace perspective turns observing strategy into part of the inverse-problem design. Filters, wavelength coverage, spectral resolution, exposure allocation, aperture, cadence, target selection, and follow-up should be chosen according to the

scientific functionals and physical alternatives that the observations are intended to distinguish. Maximizing the number of measurements or the total signal-to-noise ratio is not necessarily equivalent to maximizing the recoverable information about galaxy evolution.

An observing program can measure one already observable combination with exceptional precision while leaving the principal scientific degeneracy almost unchanged. A complementary measurement can be more valuable even when its individual signal-to-noise ratio is lower. The appropriate objective can be expressed through uncertainty in a target quantity, discrimination between evolutionary scenarios, predictive performance, or expected information gain, together with the observational cost [78,79].

The scientific target must be stated explicitly. A design optimized for recent mean SFR need not improve ancient mass assembly, and observations that constrain stellar metallicity can leave the temporal localization of star formation largely unchanged. Spectral resolution should be increased where it separates diagnostic features, while broader wavelength coverage can be preferable when it adds a qualitatively different physical response. These choices should be evaluated over a population of plausible galaxy states rather than around one universal fiducial SED.

This target dependence motivates sequential follow-up. Initial photometry or low-resolution spectroscopy can identify the dominant ambiguity for an individual galaxy, after which ultraviolet, line spectroscopy, infrared, or millimeter observations are assigned according to their expected leverage. Such adaptive allocation can use observing resources efficiently, but the follow-up rule becomes part of the survey selection function. Population inference must account for the fact that galaxies with unusual SEDs, strong lines, high predicted information gain, or better preliminary data may be observed preferentially.

Survey design must therefore address both individual-galaxy information and population identifiability. Deep fields can constrain detailed physical models and reveal discrepancies that remain hidden in shallower data. Wide surveys can determine population distributions over large volumes but generally provide coarser individual resolution. A tiered program is most powerful when the deep observations calibrate the forward, discrepancy, and population models applied to the wide component rather than serving as error-free labels for it.

Calibration observations and theoretical development are also part of the design problem. Once observational noise falls below uncertainty in stellar libraries, abundance responses, nebular calculations, attenuation geometry, or instrumental calibration, additional exposure in the same channel can yield diminishing physical return. Resources may be better spent on complementary wavelengths, repeated calibration, laboratory or stellar-model improvements, or observations that discriminate between alternative forward models.

This issue will become increasingly important as high-quality observations expose model discrepancy. The distinction between observational noise and model inadequacy introduced in Section 3.3 remains fundamental. A stochastic covariance contribution can represent some structured uncertainty, but a systematic or state-dependent mismatch can bias the inferred history rather than merely widen it. Flexible discrepancy models must also be constrained carefully because they can absorb real historical information [19,51].

Alternative stellar population, abundance, nebular, attenuation, dust, and radiative-transfer models provide practical sensitivity tests. Agreement among them strengthens the interpretation of common scientific functionals, whereas disagreement identifies model-assisted or model-dominated conclusions. Available models can share ingredients and systematic errors, so an ensemble of implementations should be viewed as a probe of known variation rather than as a complete probability distribution over physical truth.

Validation of future applications should follow the three-layer framework of Section 8, including matched- and mismatched-model experiments, posterior predictive assessment,

simulation-based calibration under the adopted simulator, and prediction of observables not used in the fit [80,81]. Held-out ultraviolet, infrared, emission-line, abundance-sensitive, gas, or dust measurements are especially valuable because they test the physical coupling rather than merely the flexibility of the fitted history. These tests distinguish computational correctness, calibration under a specified generator, and physical adequacy for real galaxies.

Survey-scale applications add three specific requirements: out-of-distribution assessment for amortized inference, response-aware validation of emulators, and end-to-end reproduction of the selection and population-inference pipeline. A neural estimator can return a precise result outside its training support, an emulator can distort a weak but informative feature despite small average flux error, and a restrictive population model can suppress rare histories. These checks determine whether computationally efficient inference remains calibrated across heterogeneous data quality and galaxy populations.

The most useful uncertainty statement will often concern a common scientific functional rather than every internal parameter. Broad cumulative mass fractions, recent mean SFRs, or coarse transition epochs can remain stable even when detailed histories vary across representations and forward models, whereas a narrow conditional interval for a fine temporal component can be scientifically weak when its location changes under modest changes in prior, basis, or physical model. Future analyses should therefore report the effective resolution and model dependence appropriate to the claim, identify the observational features that support it, and state which assumptions or alternative models change it materially. A large survey catalog of precise point estimates is not a substitute for calibrated historical inference if those distinctions are absent.

An inverse-problem-aware research program therefore allocates effort across observation, calibration, physical modeling, and statistical validation. The aim is not to obtain the densest possible SED for every galaxy, but to acquire and model the information that most directly separates scientifically consequential evolutionary alternatives. As observational precision increases, progress will depend increasingly on the credibility of that separation rather than on the formal sharpness of one model-conditional posterior.

9.3. Conclusions

Integrated spectra and SEDs extend galaxy archaeology far beyond the volume in which individual stars can be resolved. Present-day radiation retains information about stellar mass assembly, recent star formation, chemical enrichment, gas, dust, and interstellar-medium evolution, while the same spatial and temporal integration mixes contributions from different epochs and physical components. In this Review, we have used the term *Integrated Galactic Archaeology* for a unified framework that treats this reconstruction explicitly as an inverse problem. Its central conclusion is that an integrated SED does not generally reveal a unique, pointwise galaxy history; it constrains information-supported combinations that remain distinguishable after physical SED generation, observational projection, and statistical interpretation.

The recoverable object is conditional on the complete inferential structure. The physical forward operator $\mathcal{F}_{\mathrm{phys}}$ determines how histories generate radiation, the observational operator $\mathcal{O}_{\omega}$ determines which parts of that radiation enter the data, and the representation determines which histories can be expressed. Physical constraints, regularization, and priors complete weak directions, while model-discrepancy and validation assumptions determine whether the resulting uncertainty can support a physical claim. Observable modes and effective resolution are therefore more appropriate descriptors than nominal parameter number or bin width: nominal resolution belongs to the representation, whereas effective resolution belongs to the complete forward and inference procedure. A narrow

conditional interval does not by itself establish historical accuracy, and broad common functionals can remain robust even when fine bin-by-bin structure is assumption-sensitive.

The five-component structure $(\mathcal{V}, \mathcal{F}, \mathcal{C}, \mathcal{I}, \mathcal{U})$ organizes existing methods according to representation, forward model, added structure, inference mechanism, and uncertainty assessment. Within this taxonomy, PCA, MOPED, VESPA, NMF, conventional optimization, Bayesian sampling, deep learning, and simulation-based inference act on different components and preserve or return different quantities. They are therefore more productively viewed as potentially reusable modules than as mutually exclusive solutions to one algorithmic task.

Information-driven adaptive representation follows from the distinction between nominal freedom and information-supported complexity. A representation should be sufficiently flexible to express the histories relevant to the scientific target, but the detail reported from it should not substantially exceed the effective resolution supported by the data and model. This Review has presented that statement as a design principle rather than a completed algorithm. It has not introduced a universal selection statistic, threshold, hyperparameter prescription, post-selection correction, production implementation, or demonstrated performance improvement; those developments require dedicated foundational and validation studies.

The same perspective extends beyond the SFH. Galaxy SEDs depend on coupled histories of star formation, enrichment, gas, dust, nebular conditions, and radiative-transfer geometry, and the data can constrain mixed physical modes more directly than each underlying physical history separately. Physical evolution models and admissibility conditions can reduce the allowed state space, but the resulting precision is model-assisted and must be distinguished from information supplied directly by the observations.

Reliable development therefore requires the three-layer feedback cycle of controlled mock experiments, inverse-problem diagnosis, and physical forward-model revision. Mock observations reveal bias, resolution loss, calibration failure, and sensitivity to mismatch; structural diagnostics identify the responsible modes, assumptions, or discrepancy directions; and revised states, representations, and forward models generate a new inverse problem that must return to renewed validation. The objective is not convergence to one presumed universal model, but the identification of scientific conclusions that remain stable across credible physical, observational, and inferential descriptions.

Future multiwavelength observations can enlarge the observable subspace when they introduce responses complementary to those already measured. Their value should be assessed through the degeneracies they reduce, the target functionals they resolve, the physical models they test, and the selection functions they create. Increasing precision will also expose limitations in stellar, nebular, chemical, dust, radiative-transfer, and calibration models that were previously hidden by measurement noise, so complementary observations and forward-model improvement can become more valuable than additional exposure along an already measured direction.

The principal aim of Integrated Galactic Archaeology is consequently not to reconstruct the most detailed possible history. It is to determine which aspects of galaxy evolution are retained in integrated observations, at what effective resolution, and under which physical and statistical assumptions. Robust broad quantities should be separated from assumption-selected fine structure, model-assisted constraints from directly data-supported ones, and persistent ambiguity should be reported rather than removed through an unjustified preference for one history. By placing the galaxy state, physical and observational operators, representation, prior structure, inference, discrepancy, and validation within one framework, Integrated Galactic Archaeology provides a common language for these distinctions. Its success should be measured not by the visual complexity of reconstructed

histories, but by the reliability with which observable historical information is translated into testable statements about galaxy formation and evolution.

Author Contributions: Conceptualization, T.T.T.; methodology, T.T.T.; software, T.T.T., K.T.S., R.R.K. and S.A.M.; formal analysis, T.T.T.; investigation, T.T.T., K.T.S. and R.R.K.; writing—original draft preparation, T.T.T., K.T.S., R.R.K. and S.A.M.; writing—review and editing, T.T.T.; visualization, T.T.T., K.T.S. and R.R.K.; supervision, T.T.T.; funding acquisition, T.T.T. and R.R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS Grant-in-Aid for scientific research (24H00247) and by JST CREST Grant Number JP-MJCR24Q1. This work has also been supported in part by the Collaboration Funding of the Institute of Statistical Mathematics "Machine-Learning-Based Cosmogony: From Structure Formation to Galaxy Evolution".

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We are grateful to the referees for their useful and constructive comments and suggestions that improved this article much. We sincerely thank Shiro Ikeda, Satoshi Kuriki, and Kenji Fukumizu for providing the ideas that motivated this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Freeman, K.; Bland-Hawthorn, J. The New Galaxy: Signatures of Its Formation. *Annu. Rev. Astron. Astrophys.* **2002**, *40*, 487–537. [CrossRef]
2. Tolstoy, E.; Hill, V.; Tosi, M. Star-Formation Histories, Abundances, and Kinematics of Dwarf Galaxies in the Local Group. *Annu. Rev. Astron. Astrophys.* **2009**, *47*, 371–425. [CrossRef]
3. Helmi, A. Streams, Substructures, and the Early History of the Milky Way. *Annu. Rev. Astron. Astrophys.* **2020**, *58*, 205–256. [CrossRef]
4. Tinsley, B.M. Evolution of the Stars and Gas in Galaxies. *Fundam. Cosm. Phys.* **1980**, *5*, 287–388.
5. Bruzual, G.; Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **2003**, *344*, 1000–1028. [CrossRef]
6. Walcher, J.; Groves, B.; Budavári, T.; Dale, D. Fitting the integrated spectral energy distributions of galaxies. *Astrophys. Space Sci.* **2011**, *331*, 1–51. [CrossRef]
7. Conroy, C. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *Annu. Rev. Astron. Astrophys.* **2013**, *51*, 393–455. [CrossRef]
8. Iyer, K.G.; Pacifici, C.; Calistro-Rivera, G.; Lovell, C.C. The Spectral Energy Distributions of Galaxies. In *Encyclopedia of Astrophysics*; Mandel, I., Ed.; Elsevier: Oxford, UK, 2026; Volume 4, pp. 236–281. [CrossRef]
9. Pacifici, C.; Iyer, K.G.; Mobasher, B.; da Cunha, E.; Acquaviva, V.; Burgarella, D.; Calistro Rivera, G.; Carnall, A.C.; Chang, Y.Y.; Chartab, N.; et al. The Art of Measuring Physical Parameters in Galaxies: A Critical Assessment of Spectral Energy Distribution Fitting Techniques. *Astrophys. J.* **2023**, *944*, 141. [CrossRef]
10. Heavens, A.; Panter, B.; Jimenez, R.; Dunlop, J. The star-formation history of the Universe from the stellar populations of nearby galaxies. *Nature* **2004**, *428*, 625–627. [CrossRef] [PubMed]
11. Kennicutt, R.C., Jr.; Evans, N.J., II. Star Formation in the Milky Way and Nearby Galaxies. *Annu. Rev. Astron. Astrophys.* **2012**, *50*, 531–608. [CrossRef]
12. Worthey, G. Comprehensive Stellar Population Models and the Disentanglement of Age and Metallicity Effects. *Astrophys. J. Suppl. Ser.* **1994**, *95*, 107–149. [CrossRef]
13. Ocvirk, P.; Pichon, C.; Lançon, A.; Thiébaut, E. STECKMAP: STEllar Content and Kinematics from high resolution galactic spectra via Maximum A Posteriori. *Mon. Not. R. Astron. Soc.* **2006**, *365*, 74–84. [CrossRef]
14. Papovich, C.; Dickinson, M.; Ferguson, H.C. The Stellar Populations and Evolution of Lyman Break Galaxies. *Astrophys. J.* **2001**, *559*, 620–653. [CrossRef]
15. Pforr, J.; Maraston, C.; Tonini, C. Recovering galaxy stellar population properties from broad-band spectral energy distribution fitting. *Mon. Not. R. Astron. Soc.* **2012**, *422*, 3285–3326. [CrossRef]
16. Carnall, A.C.; McLure, R.J.; Dunlop, J.S.; Davé, R. Inferring the star formation histories of massive quiescent galaxies with BAGPIPES: Evidence for multiple quenching mechanisms. *Mon. Not. R. Astron. Soc.* **2018**, *480*, 4379–4401. [CrossRef]

17. Leja, J.; Carnall, A.C.; Johnson, B.D.; Conroy, C.; Speagle, J.S. How to Measure Galaxy Star Formation Histories. II. Nonparametric Models. *Astrophys. J.* **2019**, *876*, 3. [CrossRef]
18. Conroy, C.; Gunn, J.E.; White, M. The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *Astrophys. J.* **2009**, *699*, 486–506. [CrossRef]
19. Czekala, I.; Andrews, S.M.; Mandel, K.S.; Hogg, D.W.; Green, G.M. Constructing a Flexible Likelihood Function for Spectroscopic Inference. *Astrophys. J.* **2015**, *812*, 128. [CrossRef]
20. Ocvirk, P.; Pichon, C.; Lançon, A.; Thiébaut, E. STECMAP: STEllar Content from high-resolution galactic spectra via Maximum A Posteriori. *Mon. Not. R. Astron. Soc.* **2006**, *365*, 46–73. [CrossRef]
21. Tojeiro, R.; Heavens, A.F.; Jimenez, R.; Panter, B. Recovering galaxy star formation and metallicity histories from spectra using VESPA. *Mon. Not. R. Astron. Soc.* **2007**, *381*, 1252–1266. [CrossRef]
22. Maraston, C. Evolutionary population synthesis: Models, analysis of the ingredients and application to high-z galaxies. *Mon. Not. R. Astron. Soc.* **2005**, *362*, 799–825. [CrossRef]
23. Chevallard, J.; Charlot, S. Modelling and Interpreting Spectral Energy Distributions of Galaxies with BEAGLE. *Mon. Not. R. Astron. Soc.* **2016**, *462*, 1415–1443. [CrossRef]
24. Calzetti, D.; Kinney, A.L.; Storchi-Bergmann, T. Dust Extinction of the Stellar Continua in Starburst Galaxies: The Ultraviolet and Optical Extinction Law. *Astrophys. J.* **1994**, *429*, 582–601. [CrossRef]
25. Calzetti, D.; Armus, L.; Bohlin, R.C.; Kinney, A.L.; Koornneef, J.; Storchi-Bergmann, T. The Dust Content and Opacity of Actively Star-forming Galaxies. *Astrophys. J.* **2000**, *533*, 682–695. [CrossRef]
26. Charlot, S.; Fall, S.M. A Simple Model for the Absorption of Starlight by Dust in Galaxies. *Astrophys. J.* **2000**, *539*, 718–731. [CrossRef]
27. Silva, L.; Granato, G.L.; Bressan, A.; Danese, L. Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band. *Astrophys. J.* **1998**, *509*, 103–117. [CrossRef] [PubMed]
28. Saftly, W.; Baes, M.; De Geyter, G.; Camps, P.; Renaud, F.; Guedes, J.; De Looze, I. Large and Small-scale Structures and the Dust Energy Balance Problem in Spiral Galaxies. *Astron. Astrophys.* **2015**, *576*, A31. [CrossRef]
29. Asano, R.S.; Takeuchi, T.T.; Hirashita, H.; Inoue, A.K. Dust formation history of galaxies: A critical role of metallicity* for the dust mass growth by accreting materials in the interstellar medium. *Earth Planets Space* **2013**, *65*, 213–222. [CrossRef]
30. Nishida, K.Y.; Takeuchi, T.T.; Nagata, T.; Asano, R.S. A new galaxy spectral energy distribution model consistent with the evolution of dust. *Mon. Not. R. Astron. Soc.* **2022**, *514*, 2098–2115. [CrossRef]
31. Kano, R.R.; Takeuchi, T.T.; Kawamoto, E.R.; Asano, R.S.; Hagimoto, M.; Tamura, Y. Resolving the Dust Budget Crisis at $z \sim 8$ with Optically Thick, High-Density Molecular Clumps: MACS0416_Y1. *arXiv* **2026**, arXiv:2604.19928. [CrossRef]
32. Popescu, C.C.; Misiriotis, A.; Kylafis, N.D.; Tuffs, R.J.; Fischera, J. Modelling the spectral energy distribution of galaxies. I. Radiation fields and grain heating in the edge-on spiral NGC891. *Astron. Astrophys.* **2000**, *362*, 138–150.
33. Popescu, C.C.; Tuffs, R.J.; Madore, B.F.; Gil de Paz, A.; Völk, H.J.; Barlow, T.; Bianchi, L.; Byun, Y.I.; Donas, J.; Forster, K.; et al. A comparative study of the spatial distribution of ultraviolet and far-infrared fluxes from M101. *Astrophys. J.* **2005**, *619*, L75–L78. [CrossRef] [PubMed]
34. Popescu, C.C.; Tuffs, R.J.; Dopita, M.A.; Fischera, J.; Kylafis, N.D.; Madore, B.F. Modelling the spectral energy distribution of galaxies: V. The dust and PAH emission SEDs of disk galaxies. *Astron. Astrophys.* **2011**, *527*, A109. [CrossRef]
35. Baes, M.; Verstappen, J.; De Looze, I.; Fritz, J.; Saftly, W.; Vidal Pérez, E.; Stalevski, M.; Valcke, S. Efficient three-dimensional NLTE dust radiative transfer with SKIRT. *Astrophys. J. Suppl. Ser.* **2011**, *196*, 22. [CrossRef]
36. Viaene, S.; Baes, M.; Tamm, A.; Tempel, E.; Bendo, G.; Blommaert, J.A.D.L.; Boquien, M.; Boselli, A.; Camps, P.; Cooray, A.; et al. The Herschel Exploitation of Local Galaxy Andromeda (HELGA)—VII. A SKIRT radiative transfer model and insights on dust heating. *Astron. Astrophys.* **2017**, *599*, A64. [CrossRef]
37. Trayford, J.W.; Camps, P.; Theuns, T.; Baes, M.; Bower, R.G.; Crain, R.A.; Gunawardhana, M.L.P.; Schaller, M.; Schaye, J.; Frenk, C.S. Optical colours and spectral indices of $z = 0.1$ EAGLE galaxies with the 3D dust radiative transfer code SKIRT. *Mon. Not. R. Astron. Soc.* **2017**, *470*, 771–799. [CrossRef]
38. Camps, P.; Baes, M. SKIRT 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains. *Astron. Comput.* **2020**, *31*, 100381. [CrossRef]
39. Matsumoto, K.; Camps, P.; Baes, M.; De Ceuster, F.; Wada, K.; Nakagawa, T.; Nagamine, K. Self-consistent dust and non-LTE line radiative transfer with SKIRT. *Astron. Astrophys.* **2023**, *678*, A175. [CrossRef]
40. Baes, M.; Gebek, A.; Trčka, A.; Camps, P.; van der Wel, A.; Abdurro'uf; Andreadis, N.; Bokona Tulu, S.; Tassama Emana, A.; Fritz, J.; et al. The TNG50-SKIRT Atlas: Post-processing methodology and first data release. *Astron. Astrophys.* **2024**, *683*, A181. [CrossRef]
41. Baes, M.; Mosenkov, A.; Kelly, R.; Abdurro'uf; Andreadis, N.; Bokona Tulu, S.; Camps, P.; Tassama Emana, A.; Fritz, J.; Gebek, A.; et al. The TNG50-SKIRT Atlas: Wavelength dependence of the effective radius. *Astron. Astrophys.* **2024**, *683*, A182. [CrossRef]

42. Davies, J.I.; Baes, M.; Bianchi, S.; Jones, A.; Madden, S.; Xilouris, M.; Bocchio, M.; Casasola, V.; Cassara, L.P.; Clark, C.; et al. DustPedia: A definitive study of cosmic dust in the local Universe. *Publ. Astron. Soc. Pac.* **2017**, *129*, 044102. [CrossRef]
43. Clark, C.J.R.; Verstocken, S.; Bianchi, S.; Fritz, J.; Viaene, S.; Smith, M.W.L.; Baes, M.; Casasola, V.; Cassara, L.P.; Davies, J.I.; et al. DustPedia: Multiwavelength photometry and imagery of 875 nearby galaxies in 42 ultraviolet–microwave bands. *Astron. Astrophys.* **2018**, *609*, A37. [CrossRef]
44. Caplar, N.; Tacchella, S. Stochastic Modelling of Star-Formation Histories I: The Scatter of the Star-Forming Main Sequence. *Mon. Not. R. Astron. Soc.* **2019**, *487*, 3845–3869. [CrossRef]
45. Tacchella, S.; Forbes, J.C.; Caplar, N. Stochastic Modelling of Star-Formation Histories II: Star-Formation Variability from Molecular Clouds and Gas Inflow. *Mon. Not. R. Astron. Soc.* **2020**, *497*, 698–725. [CrossRef]
46. Iyer, K.G.; Tacchella, S.; Genel, S.; Hayward, C.C.; Hernquist, L.; Brooks, A.M.; Caplar, N.; Davé, R.; Diemer, B.; Forbes, J.C.; et al. The Diversity and Variability of Star Formation Histories in Models of Galaxy Evolution. *Mon. Not. R. Astron. Soc.* **2020**, *498*, 430–463. [CrossRef]
47. Iyer, K.G.; Speagle, J.S.; Caplar, N.; Forbes, J.C.; Gawiser, E.; Leja, J.; Tacchella, S. Stochastic Modeling of Star Formation Histories. III. Constraints from Physically Motivated Gaussian Processes. *Astrophys. J.* **2024**, *961*, 53. [CrossRef]
48. Iyer, K.G.; Starkenburg, T.K.; Bryan, G.L.; Somerville, R.S.; Alfonzo, J.P.; Anglés-Alcázar, D.; Cooray, S.; Davé, R.; Gabrielpillai, A.; Genel, S.; et al. How Does Feedback Affect the Star Formation Histories of Galaxies? *Astrophys. J.* **2025**, *994*, 174. [CrossRef]
49. Yung, L.Y.A.; Somerville, R.S.; Finkelstein, S.L.; Iyer, K.G. Investigating the Star Formation Histories of Galaxies from Cosmic Dawn to the Epoch of Reionization with the Santa Cruz SAM. *arXiv* **2026**, arXiv:2607.02650. [CrossRef]
50. Backus, G.; Gilbert, F. The Resolving Power of Gross Earth Data. *Geophys. J. Int.* **1968**, *16*, 169–205. [CrossRef]
51. Kennedy, M.C.; O'Hagan, A. Bayesian Calibration of Computer Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2001**, *63*, 425–464. [CrossRef]
52. Carnall, A.C.; Leja, J.; Johnson, B.D.; McLure, R.J.; Dunlop, J.S.; Conroy, C. How to Measure Galaxy Star Formation Histories. I. Parametric Models. *Astrophys. J.* **2019**, *873*, 44. [CrossRef]
53. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*, 1st ed.; Mathematics and Its Applications; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996; Volume 375, ISBN 978-0-7923-4157-4.
54. Hansen, P.C. *Discrete Inverse Problems: Insight and Algorithms*; Fundamentals of Algorithms; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2010; Volume 7. [CrossRef]
55. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [CrossRef]
56. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and Smoothness via the Fused Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 91–108. [CrossRef]
57. Stuart, A.M. Inverse Problems: A Bayesian Perspective. *Acta Numer.* **2010**, *19*, 451–559. [CrossRef]
58. Golub, G.H.; Heath, M.; Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **1979**, *21*, 215–223. [CrossRef]
59. MacKay, D.J.C. Bayesian Interpolation. *Neural Comput.* **1992**, *4*, 415–447. [CrossRef]
60. Kaipio, J.P.; Somersalo, E. *Statistical and Computational Inverse Problems*; Applied Mathematical Sciences; Springer: New York, NY, USA, 2005; Volume 160. [CrossRef]
61. Ronen, S.; Aragon-Salamanca, A.; Lahav, O. Principal component analysis of synthetic galaxy spectra. *Mon. Not. R. Astron. Soc.* **1999**, *303*, 284–296. [CrossRef]
62. Heavens, A.F.; Jimenez, R.; Lahav, O. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *Mon. Not. R. Astron. Soc.* **2000**, *317*, 965–972. [CrossRef]
63. Alsing, J.; Wandelt, B. Generalized Massive Optimal Data Compression. *Mon. Not. R. Astron. Soc. Lett.* **2018**, *476*, L60–L64. [CrossRef]
64. Charnock, T.; Lavaux, G.; Wandelt, B.D. Automatic Physical Inference with Information Maximizing Neural Networks. *Phys. Rev. D* **2018**, *97*, 083004. [CrossRef]
65. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [CrossRef] [PubMed]
66. Lovell, C.C.; Acquaviva, V.; Thomas, P.A.; Iyer, K.G.; Gawiser, E.; Wilkins, S.M. Learning the relationship between galaxies spectra and their star formation histories using convolutional neural networks and cosmological simulations. *Mon. Not. R. Astron. Soc.* **2019**, *490*, 5503–5520. [CrossRef]
67. Cranmer, K.; Brehmer, J.; Louppe, G. The Frontier of Simulation-based Inference. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 30055–30062. [CrossRef] [PubMed]
68. Robotham, A.S.G.; Bellstedt, S.; Lagos, C.D.P.; Thorne, J.E.; Davies, L.J.; Driver, S.P.; Bravo, M. ProSpect: Generating spectral energy distributions with complex star formation and metallicity histories. *Mon. Not. R. Astron. Soc.* **2020**, *495*, 905–931. [CrossRef]
69. Iyer, K.; Gawiser, E. Reconstruction of Galaxy Star Formation Histories through SED Fitting: The Dense Basis Approach. *Astrophys. J.* **2017**, *838*, 127. [CrossRef]

70. Boquien, M.; Burgarella, D.; Roehlly, Y.; Buat, V.; Ciesla, L.; Corre, D.; Inoue, A.K.; Salas, H. CIGALE: A python Code Investigating GALaxy Emission. *Astron. Astrophys.* **2019**, *622*, A103. [CrossRef]
71. Berk, R.; Brown, L.; Buja, A.; Zhang, K.; Zhao, L. Valid Post-Selection Inference. *Ann. Stat.* **2013**, *41*, 802–837. [CrossRef]
72. Dwek, E. The Evolution of the Elemental Abundances in the Gas and Dust Phases of the Galaxy. *Astrophys. J.* **1998**, *501*, 643–665. [CrossRef]
73. Gardner, J.P.; Abbott, R.; Mather, J.C.; Abell, J.S.; Abernathy, M.; Abney, F.E.; Abraham, J.G.; Abraham, R.; Abul-Huda, Y.M.; Acton, S.; et al. The James Webb Space Telescope Mission. *Publ. Astron. Soc. Pac.* **2023**, *135*, 068001. [CrossRef]
74. Mellier, Y. et al. [Euclid Collaboration]. Euclid. I. Overview of the Euclid Mission. *Astron. Astrophys.* **2025**, *697*, A1. [CrossRef]
75. Akeson, R.; Armus, L.; Bachelet, E.; Bailey, V.; Bartusek, L.; Bellini, A.; Benford, D.; Bennett, D.; Bhattacharya, A.; Bohlin, R.; et al. The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s. *arXiv* **2019**. [CrossRef]
76. Wootten, A.; Thompson, A.R. The Atacama Large Millimeter/Submillimeter Array. *Proc. IEEE* **2009**, *97*, 1463–1471. [CrossRef]
77. Ivezić, Ž.; Kahn, S.M.; Tyson, J.A.; Abel, B.; Acosta, E.; Allsman, R.; Alonso, D.; AlSayyad, Y.; Anderson, S.F.; Andrew, J.; et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.* **2019**, *873*, 111. [CrossRef]
78. Lindley, D.V. On a Measure of the Information Provided by an Experiment. *Ann. Math. Stat.* **1956**, *27*, 986–1005. [CrossRef]
79. Chaloner, K.; Verdinelli, I. Bayesian Experimental Design: A Review. *Stat. Sci.* **1995**, *10*, 273–304. [CrossRef]
80. Gelman, A.; Meng, X.L.; Stern, H. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Stat. Sin.* **1996**, *6*, 733–807.
81. Talts, S.; Betancourt, M.; Simpson, D.; Vehtari, A.; Gelman, A. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv* **2018**. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.