

On the origin of the fundamental metallicity relation and the scatter in galaxy scaling relations

John C. Forbes,^{1★} Mark R. Krumholz,¹ Andreas Burkert^{2,3†} and Avishai Dekel⁴

¹*Department of Astronomy & Astrophysics, University of California, Santa Cruz, CA 95064, USA*

²*University Observatory Munich (USM), Scheinerstrasse 1, D-81679 Munich, Germany*

³*Max-Planck-Institut fuer extraterrestrische Physik, Giessenbachstrasse 1, D-85758 Garching, Germany*

⁴*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel*

Accepted 2014 June 9. Received 2014 May 23; in original form 2013 November 6

ABSTRACT

We present a simple toy model to understand what sets the scatter in star formation and metallicity of galaxies at fixed mass. According to this model, the scatter ultimately arises from the intrinsic scatter in the accretion rate, but may be substantially reduced depending on the time-scale on which the accretion varies compared to the time-scale on which the galaxy loses gas mass. This model naturally produces an anticorrelation between star formation and metallicity at a fixed mass, the basis of the fundamental metallicity relation. We show that observational constraints on the scatter in galaxy scaling relations can be translated into constraints on the galaxy-to-galaxy variation in the mass loading factor at fixed mass, and the time-scales and magnitude of a stochastic component of accretion on to star-forming galaxies. We find a remarkably small scatter in the mass loading factor, $\lesssim 0.1$ dex, and that the scatter in accretion rates is smaller than that expected from N -body simulations.

Key words: galaxies: evolution – galaxies: fundamental parameters – galaxies: statistics.

1 INTRODUCTION

Large galaxy surveys in the past decade have taught us that star-forming galaxies fall on a main sequence (MS), a tight correlation between their stellar mass and star formation rates (SFRs; Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007). This correlation contains remarkably little scatter – the SFR varies by about ± 0.34 dex at fixed stellar mass (Whitaker et al. 2012; Guo, Zheng & Fu 2013).

The existence of the star-forming MS and its small scatter has substantially affected our understanding of galaxy evolution. The cosmological paradigm of the past few decades, Λ cold dark matter, predicts that dark matter haloes, and hence the galaxies occupying them, form hierarchically – large haloes are built from mergers of smaller haloes. The expectation was therefore that mergers between galaxies would be a major driver of their evolution over cosmological times, frequently triggering starbursts. The small scatter in the MS at multiple redshifts has shown, however, that most galaxies are not in fact experiencing any dramatic effects of major mergers (Rodighiero et al. 2011), and most stars form in ‘normal’ galaxies lying along this relation.

In addition to the MS, galaxies follow other scaling relations. Practically, this means that at a given point in cosmic history many

properties of galaxies are set by a single parameter associated with the galaxy’s mass. This fact has led to the development of a series of models dubbed ‘equilibrium’, ‘regulator’, or ‘bathtub’ models (Dekel, Sari & Ceverino 2009; Bouché et al. 2010; Cacciato, Dekel & Genel 2012; Davé, Finlator & Oppenheimer 2012; Genel, Dekel & Cacciato 2012; Feldmann 2013; Lilly et al. 2013), in which the properties of a galaxy are self-regulated near a stable equilibrium between accretion of gas, star formation, and outflows. The fundamental ingredients are an SFR that increases with increasing gas mass in the galaxy, and a time-scale on which the gas reservoir returns to its equilibrium value much shorter than the time-scale on which bulk properties of the galaxy (mass, accretion rate, outflow efficiency, star formation time-scale, etc.) vary.

Equilibrium models have enjoyed success in heuristically reproducing the average trends in galaxy evolution. However, such models are inherently incapable of quantifying higher order effects, including scatter in individual scaling relations and various fundamental metallicity relations (FMRs), found by numerous groups (Mannucci et al. 2010; Mannucci, Salvaterra & Campisi 2011; Yates, Kauffmann & Guo 2012; Bothwell et al. 2013; Lara-López, López-Sánchez & Hopkins 2013; Stott et al. 2013; Dekel & Mandelker 2014), typically parametrized as a quadratic function $Z(M_*, \text{SFR})$.

Many theoretical studies are focused on reproducing first-order relations, which is sufficiently difficult in and of itself that second-order relations are often neglected (however, for an exception, see Dutton, Bosch & Dekel 2010). In principle, however, to fully

★ E-mail: jcforges@ucsc.edu

† Max Planck Fellow.

understand galaxy evolution we should be able to understand not only average or median galaxy properties, but their full distribution. A significant advantage of the equilibrium models is that fitting them to any first-order relation is trivial, so extending them to understand the higher order relations is easier than with any other method.

In this work, we take one step beyond equilibrium models, allowing the mass accretion rate to vary with a fixed lognormal scatter. Such a scatter is expected based on N -body simulations (e.g. Neistein & Dekel 2008) and hydrodynamic simulations (e.g. Dekel et al. 2013). Depending on the time-scale on which the accretion rate varies, a population of these galaxies may never be in equilibrium. That is, their masses and metallicities may change substantially. However, they may still reach a statistical equilibrium, in which the full joint distribution of all galaxy properties becomes time invariant.

In Section 2, we introduce the basic formulation of this model and explore the implications for the origin of the width of the star-forming MS. We add metallicity to the model in Section 3, allowing us to examine the FMR and the scatter in the mass-metallicity relation (MZR, the correlation between stellar mass and gas-phase metallicity). The model we construct in these sections is independent of the first-order effects considered by most equilibrium models, which allows us to avoid uncertainty in the numerical values of numerous important parameters in galaxy evolution (e.g. the mass loading factor). In Section 4, we re-dimensionalize our model, taking a guess at the correct scalings and normalizations to use, and demonstrate the ability of this model to understand unknown physics based solely on the scatter in galaxy scaling relations. We discuss the limitations of our model, quantitative constraints it can place on the accretion process, and alternative explanations for the intrinsic width of these relations in Section 5, and conclude in Section 6.

2 A MINIMAL MODEL FOR SCATTER IN THE MS

To examine the origin of and scatter within the MS and FMR, we have constructed a minimal model which contains enough physics to produce such features. We do not aim to fully reproduce galaxy properties, tune parameters to match observations, or build upon dark matter merger trees. Instead we aim for simplicity and intuition. In our model, we describe the state of a galaxy by two numbers, its cold gas mass M_g and its metallicity $Z = M_Z/M_g$, where M_Z is the mass in metals. The gas mass evolves according to

$$\frac{dM_g}{dt} = \dot{M}_{\text{ext}}(t) - \frac{M_g}{t_{\text{loss}}}, \quad (1)$$

where t_{loss} is the characteristic time over which gas is lost to the system (through the formation of stars and the launching of galactic winds). Here we are implicitly assuming that a given galaxy ejects gas in galactic winds at a rate directly proportional to the SFR, the constant of proportionality being defined as the mass loading factor η . We also assume that all stellar evolution happens instantaneously so that a fixed proportion of all mass which forms stars is immediately returned to the gas reservoir of the galaxy, and a fixed fraction f_R is permanently locked in stellar remnants. We explicitly relate the loss rate with these quantities as follows:

$$\frac{M_g}{t_{\text{loss}}} = (f_R + \eta)\dot{M}_{\text{SF}} = (f_R + \eta)\frac{M_g}{t_{\text{dep}}}, \quad (2)$$

where \dot{M}_{SF} is the star formation rate and t_{dep} is the depletion time of all gas in the galaxy. This equation demonstrates that a great

deal of (poorly constrained) physics is hidden in t_{loss} . In this and the following section, however, we will simply scale the time coordinate to t_{loss} and consider ensembles of galaxies with fixed t_{loss} . This is a powerful technique, since our conclusions in these sections will be independent of the values and scaling relations of η , t_{dep} , and so forth.

The external accretion rate is \dot{M}_{ext} , which we parametrize as a lognormal distribution with fixed median and scatter,

$$\dot{M}_{\text{ext}}(t) = \exp(\mu + \sigma x(t)), \quad (3)$$

where $x(t)$ is a random variable distributed as a standard normal (zero mean, unit variance), with a new value drawn at a fixed time interval $t_{\text{coherence}}$. A more realistic model might have a spectrum of time-scales over which x would vary, but for simplicity, clarity, and analytical tractability we will use a single ‘coherence time’.

Equation (1) is simple enough that it may be solved analytically given a sequence of random numbers, i.e. $x(t)$. It is convenient to first non-dimensionalize by scaling the mass-loss rate to the median accretion rate,

$$\Psi = \frac{M_g}{t_{\text{loss}}} e^{-\mu}, \quad (4)$$

and the time to t_{loss} via $t = \tau t_{\text{loss}}$. Thus, τ represents time in units of mass-loss time-scales, and one may think of Ψ loosely as the SFR or the gas mass, though strictly speaking it is the mass-loss rate per median accretion rate. The evolution equation then becomes

$$\frac{d\Psi}{d\tau} = -\Psi + e^{\sigma x(\tau)}. \quad (5)$$

The value of μ has entirely dropped out, so the full distribution of Ψ is determined solely by the inherent scatter in the accretion rate, σ , and the number of mass-loss times over which the accretion rate remains constant, $\tau_c \equiv t_{\text{coherence}}/t_{\text{loss}}$.

This equation has exactly the same structure as the radiative transfer equation where accretion acts as the source term, time in units of mass-loss time-scales is similar to the optical depth, and the instantaneous value of Ψ is analogous to the intensity of radiation.

The mass-loss rate as a function of time can now be solved analytically if we are given a sequence of standard normals, x_k , $k = 0, \dots$, where each k corresponds to a new draw from the log-normal accretion rate distribution. In particular, suppose we know Ψ_k , the value of Ψ at the time that a new accretion rate is drawn. The value of Ψ from then up to the subsequent draw is given by

$$\Psi(\Delta) = \Psi_k e^{-\Delta} + e^{\sigma x_k} (1 - e^{-\Delta}) \quad \text{for } 0 \leq \Delta \leq \tau_c, \quad (6)$$

where we have defined $\Delta \equiv \tau - k\tau_c$, the time since the most recent switch in the accretion rate. This is simply a relaxation equation. We can see this more explicitly by defining $r_k = \Psi_k/e^{\sigma x_k}$, the mass-loss rate at the time of the switch in accretion rate relative to the newly chosen accretion rate – this is a measure of how far from equilibrium the system is immediately after a switch. Large values mean that the galaxy is losing much more mass than it is accreting, values less than unity mean gas is building up, and $r_k = 1$ means that the galaxy has equilibrated to its current accretion rate. Note that $r_k > 0$. Using r_k , we see that

$$\Psi(\Delta)/e^{\sigma x_k} = 1 + e^{-\Delta}(r_k - 1) \quad \text{for } 0 \leq \Delta \leq \tau_c. \quad (7)$$

In other words, the loss rate approaches the accretion rate exponentially on a mass-loss time-scale, with the deviation determined by the deviation at the time the accretion rate switched ($\Delta = 0$).

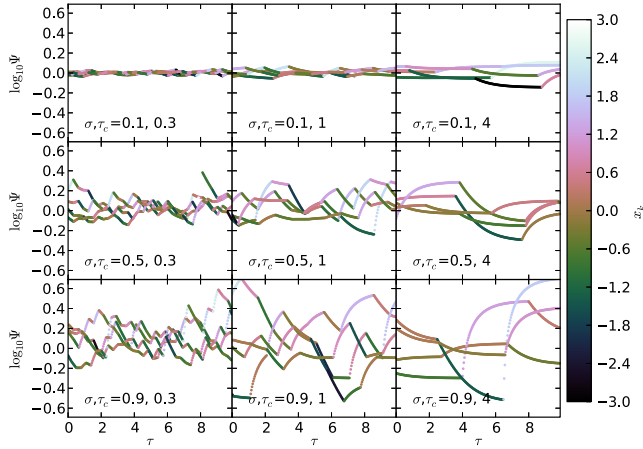


Figure 1. Ψ as a function of time. Recall that Ψ is the dimensionless mass-loss rate, which at fixed mass we take to be proportional to both gas mass and SFR. For each pair of σ (the intrinsic scatter, increasing top to bottom) and τ_c (the ratio of the coherence time to the mass-loss time, increasing left to right), we show the trajectories of five random galaxies over the course of a randomly selected 10 star formation times and coloured by the instantaneous value of $x(t)$ – lighter colours mean higher accretion rates. The galaxies exponentially approach $\Psi = e^{\sigma x_k}$, so when τ_c is large, the scatter in Ψ reflects the intrinsic scatter σ , but when τ_c is short, the many changes in the accretion rate never allow the galaxies’ mass-loss rates, Ψ , to reach the accretion rate $e^{\sigma x_k}$ – instead they remain near the average value.

Using equation (6), we can recursively compute Ψ_{k+1} from the previous Ψ_k by setting $\Delta = \tau_c$,

$$\Psi_{k+1} = \Psi_k e^{-\tau_c} + e^{\sigma x_k} (1 - e^{-\tau_c}). \quad (8)$$

The sequence of x_k ’s thereby determines a sequence of Ψ_k ’s, which are used as a scaffolding to construct the full solution $\Psi(\tau)$ for a given realization of the random variables. Examples of such realizations for various values of σ and τ_c are shown in Fig. 1.

We may also solve the recursion relation (equation 8) for Ψ_k explicitly, again given the sequence x_k ,

$$\Psi_k = (1 - e^{-\tau_c}) \sum_{i=0}^{k-1} e^{-(k-i-1)\tau_c} e^{\sigma x_i} + \Psi_0 e^{-k\tau_c}. \quad (9)$$

We see that the mass-loss rate at a switch in the accretion rate is therefore simply a sum of lognormally distributed random variables. Each draw from the distribution loses influence as it recedes into the past. In particular, $(k-i)\tau_c$ is simply the number of loss times since x_i was drawn. The prefactor of $1 - e^{-\tau_c}$ accounts for the fact that each individual draw of the accretion history matters less as the coherence time gets shorter. For example, even if the accretion rate is very large, if the coherence time is very short, the galaxy will only experience that accretion rate for a very short time.

In the limit that $\tau_c \gg 1$, only the most recent draw from the distribution matters and all previous draws are exponentially suppressed. In the opposite limit, $\tau_c \ll 1$, the leading factor $1 - e^{-\tau_c} \rightarrow \tau_c \approx N_{\text{loss}}^{-1}$, where N_{loss} is the number of draws from the accretion rate distribution in a given mass-loss time. For very short coherence times, Ψ_k becomes an average of the lognormal accretion rates over the most recent mass-loss time-scale, with more recent accretion rates weighted somewhat more heavily. The full probability distributions of Ψ and Ψ_k (which serves as an approximation to Ψ) for various values of τ_c and σ are shown in Fig. 2, along with the probability distributions of the five most recent draws from the appropriate lognormal distributions which are added together to give Ψ_k . As the

coherence times grow longer (panels from left to right), fewer draws from the accretion rate contribute to the current mass-loss rate. As the intrinsic scatter in the accretion rate increases (top to bottom), the probability densities of the draws from the accretion rate overlap more, meaning that it is increasingly likely that the most recent accretion rate is not the largest contributor to the current mass-loss rate.

The distribution of a random variable, i.e. Ψ_k which is the sum of lognormally distributed random variables, i.e. $(1 - e^{-\tau_c})e^{-j\tau_c + \sigma x_j}$, may itself be approximated as a lognormal random variable (Fenton 1960) with variance (in the log)

$$\sigma_{\Psi_k}^2 \approx \ln \left[1 + (e^{\sigma^2} - 1) \frac{1 - e^{-\tau_c}}{1 + e^{-\tau_c}} \right], \quad (10)$$

and median

$$\mu_{\Psi_k} \approx \sigma^2/2 - \sigma_{\Psi_k}^2/2. \quad (11)$$

Equation (10) represents a good analytical guess for the width of the star-forming MS. In the limit of large τ_c , $\sigma_{\Psi_k} \rightarrow \sigma$ as expected – when the galaxies are able to equilibrate to their accretion rates, the width of the MS is simply the intrinsic width in the accretion rate, σ . In this limit $\mu_{\Psi_k} \rightarrow 0$, i.e. the median mass-loss rate approaches the median accretion rate.

In the limit of short τ_c , the fraction $(1 - e^{-\tau_c})/(1 + e^{-\tau_c})$ reduces to $\tau_c/2$. Thus, when σ is also sufficiently small (namely $\sigma^2 \ll 1$), the variance of Ψ_k reduces to $\sigma^2 \tau_c/2$. Physically, when the accretion rate varies rapidly compared to the mass-loss time, any individual draw from the accretion rate distribution becomes unimportant and all that matters is the long-term average. The exception is when σ is large, in which case extremely large accretion rates become common and σ_{Ψ_k} again approaches σ , though we caution that the distribution becomes less lognormal as σ increases with small τ_c . Fig. 3 shows σ_{Ψ_k} computed with a Monte Carlo simulation (for details see Appendix A).

From this discussion, we see that τ_c and σ compete in setting the width of the MS – recall that $\Psi \propto \dot{M}_{\text{SF}}$ at fixed mass, so the logarithmic scatter in Ψ is the logarithmic scatter in the MS, modulo some complications that will be addressed in Section 5. Even after σ_{Ψ_k} has been scaled by the intrinsic accretion rate, both σ and τ_c enter the expression for the MS scatter. This suggests that σ may be interpreted as a third time-scale in the problem, namely the number of mass-loss times necessary to forget a typical accretion event.

3 INCLUDING METALLICITY

The metal content of the galaxy is evolved according to the instantaneous recycling approximation, in the spirit of Tinsley (1980) and Maeder (1992),

$$\frac{dM_g Z}{dt} = Z_{\text{IGM}} \dot{M}_{\text{ext}} + \dot{M}_{\text{SF}} (f_R y - f_R M_Z / M_g - \eta Z_w). \quad (12)$$

New metals are added along with accreting matter (the first term), and respectively produced by, locked up in the products of, and ejected from the galaxy by, star formation.

The SFR, taking into account galactic winds and stellar evolution (see equation 2), is $\dot{M}_{\text{SF}} = (f_R + \eta)^{-1} \dot{M}_g / t_{\text{loss}}$. The yield y is defined as the mass of metals produced during the course of stellar evolution per unit mass of gas locked in stellar remnants – if 1000 M_\odot of gas forms stars, a total of $y f_R 1000 M_\odot$ of metals will be produced by these stars and returned to the interstellar medium (ISM). For the galactic wind metallicity, the usual assumption throughout the literature on chemical evolution is that $Z_w = Z$, i.e. before gas

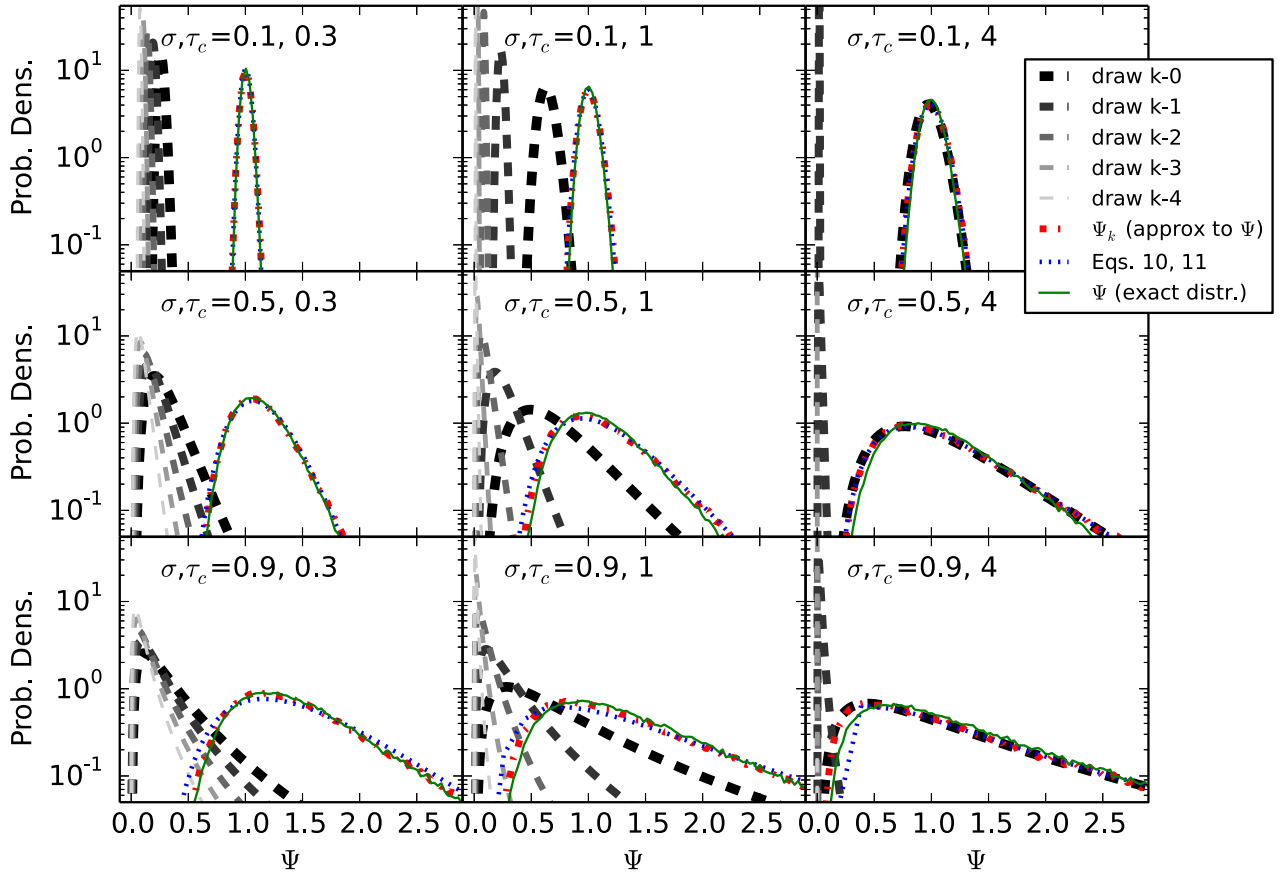


Figure 2. The construction of the distribution of Ψ , the dimensionless mass-loss rate. The red, blue, and green lines show the probability density of a galaxy with a given combination of σ and τ_c having, at a randomly selected time, a given value of Ψ . Green shows the full distribution, red shows the distribution of Ψ_k , namely Ψ at a switch in the accretion rate, and blue shows the analytic approximation to Ψ_k . These approximations are remarkably good, at least for this selection of σ and τ_c . The dashed grey lines show the lognormal distributions from which numbers are drawn and added together to compute Ψ_k . As τ_c increases, fewer draws from the accretion distribution contribute to Ψ , until eventually only the last draw matters and all the others are exponentially suppressed (rightmost panel). As the intrinsic width of the accretion distribution increases (top to bottom), the accretion rate distributions increasingly overlap, increasing the chances that the most recent accretion rate does not dominate in determining the current value of Ψ .

is ejected from a galaxy by stellar feedback, it is assumed to be perfectly well mixed with the ambient ISM (for recent exceptions, see Recchi et al. 2008; Peebles & Shankar 2011; Krumholz & Dekel 2012; Vogelsberger et al. 2013). This is a strong assumption because galactic winds are likely to be preferentially metal enriched. Physically, this is because metals are produced in the same places, sometimes by the same events, which are likely to cause galactic-scale outflows, namely sites of recent star formation, where massive stars emit ionizing radiation and end their lives as supernovae.

Following Forbes et al. (2014), we parametrize the wind metallicity Z_w in such a way that this assumption could be relaxed,

$$Z_w = Z + \xi y f_R / \max(\eta, 1 - f_R). \quad (13)$$

The assumption that $Z_w = Z$ corresponds in our model to $\xi = 0$ – the other extreme value is $\xi = 1$, corresponding to exactly no mixing between the metal-rich ejecta of stars and the ISM. This equation is derived by finding the maximum possible metallicity for a galactic wind and scaling that quantity by ξ . When η is small, the maximum Z_w is the metallicity of the stellar ejecta, $(y f_R + Z(1 - f_R))/(1 - f_R)$, but once η exceeds $1 - f_R$, the rate at which mass is lost in galactic winds exceeds the rate at which stars are ejecting mass, meaning some of the wind must be composed of material from the ambient ISM, reducing the maximum possible

wind metallicity and introducing a change in the behaviour of our equation at $\eta = 1 - f_R$.

Non-dimensionalizing as in the previous section, we arrive at

$$\frac{d\Psi Z}{d\tau} = -\Psi Z + Z_{\text{IGM}} e^{\sigma x(t)} + q\Psi, \quad (14)$$

where all of the factors associated with star formation and feedback can be collected into a single parameter

$$q \equiv \frac{y f_R}{f_R + \eta} \left(1 - \frac{\eta \xi}{\max(1 - f_R, \eta)} \right). \quad (15)$$

Much of the uncertainty in modelling the metallicity evolution of galaxies is encapsulated in q . We note that in the limit that η is large (probably the case for low-mass galaxies), $q \rightarrow y f_R (1 - \xi)/\eta$, while for small values of η , $q \rightarrow y$. In this sense, q may be considered an effective yield, and the more important galactic winds are, the more uncertain this parameter becomes.

Combining equations (14) and (5), we can obtain the evolution equation for metallicity,

$$\frac{dZ}{d\tau} = \frac{Z_{\text{IGM}} - Z}{\Psi / e^{\sigma x}} + q. \quad (16)$$

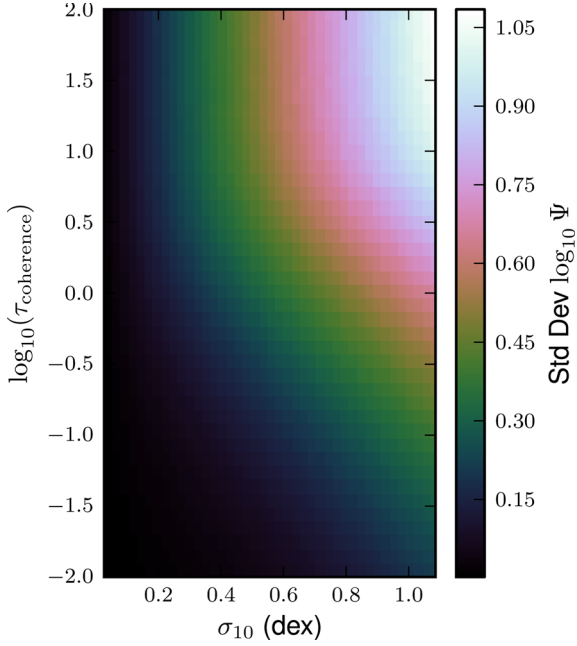


Figure 3. The width of the ‘MS’. Each pixel represents an ensemble of galaxies with fixed τ_c and σ , wherein the mass-loss rate Ψ was measured for each galaxy at a random time. The standard deviation of $\log_{10}\Psi$ in each bin is plotted above. Longer coherence times (upwards in the plot) allow the galaxies to equilibrate so that Ψ approaches the accretion rate, so the scatter in Ψ approaches σ . More precisely, the standard deviation of $\log_{10}\Psi$ approaches $\sigma_{10} = \sigma/\ln(10)$, the standard deviation of $\log_{10}\dot{M}_{\text{ext}}$. Coherence times shorter than the mass-loss time-scale, i.e. $\log_{10}\tau_c \lesssim 1$, lead to a reduction in the scatter roughly in proportion to τ_c – individual draws from the accretion rate matter increasingly less, and the galaxies do not have enough time for Ψ to approach the accretion rate before it changes.

This equation represents a competition between the effective yield of new metals formed during the course of stellar evolution and dilution of metals by accretion. In equilibrium, namely when the accretion rate is constant for many mass-loss times, $\Psi/e^{\sigma x} \rightarrow 1$, so for $dZ/d\tau = 0$, we find that the equilibrium metallicity is

$$Z_{\text{eq}} = q + Z_{\text{IGM}}. \quad (17)$$

The corresponding quantity for the relative mass-loss rate is $\Psi_{\text{eq}} = e^{\sigma x}$, where x is a standard normal. In this sense, metallicity is very different from Ψ – regardless of the accretion rate, the metallicity approaches a constant value, whereas the mass-loss rate approaches whatever random value of the accretion rate it is being fed.

The factor of q can be factored out of the metallicity evolution equation by defining

$$Z^\dagger = \frac{Z - Z_{\text{IGM}}}{q}, \quad (18)$$

the metallicity scaled to its equilibrium value by the effective yield – Z_{eq} corresponds to $Z^\dagger = 1$. The evolution equation for Z^\dagger then becomes independent of q , yielding

$$\frac{dZ^\dagger}{d\tau} = \frac{-Z^\dagger}{1 + e^{-\tau(r_k - 1)}} + 1. \quad (19)$$

The denominator is recognizable as the mass-loss rate normalized by the current accretion rate, $e^{\sigma x_k}$.

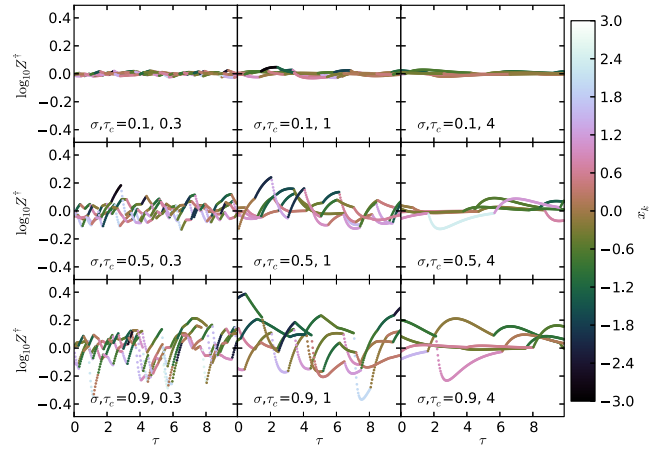


Figure 4. Z^\dagger as a function of time. For each pair of σ (the intrinsic scatter in the accretion rate, increasing top to bottom) and τ_c (the ratio of the coherence time to the mass-loss time, increasing left to right), we show the trajectories of five random galaxies over the course of a randomly selected 10 star formation times and coloured by the instantaneous value of $x(t)$ – lighter colours mean higher accretion rates. We can immediately see that high accretion rates tend to lead to low metallicities. We also see that for large coherence times (rightwards), each galaxy reaches an extremum in metallicity before returning to its equilibrium value ($Z^\dagger = 1$). Short coherence times mean that the accretion rate changes when the galaxy is near this extremum or before, meaning the ensemble width of $\log_{10}Z^\dagger$ will tend to peak near $\tau_c \sim 1$, and the relationship between metallicity and SFR will be strongest (most anticorrelated) for $\tau_c \lesssim 1$.

As with the star formation equation, we may solve equation (19) by elementary methods for a given fixed value of r_k , i.e. during the period when the accretion rate is constant. The solution is

$$Z^\dagger(\Delta) = \frac{Z^\dagger(0)r_k + e^\Delta - 1 + (r_k - 1)\Delta}{e^\Delta - 1 + r_k} \quad \text{for } 0 \leq \Delta \leq \tau_c. \quad (20)$$

In the limit of large Δ , we recover that $Z^\dagger \rightarrow 1$, i.e. the metallicity returns to its equilibrium value, though depending on how different the initial mass-loss rate is from the accretion rate, $\Psi_k/e^{\sigma x_k} = r_k$, the value of Z^\dagger may change dramatically over the course of a star formation time ($\Delta \sim 1$) before exponentially returning to unity. Examples of such trajectories are shown in Fig. 4. Notice that in the rightmost panels, despite the somewhat large value of τ_c , there are significant deviations from equilibrium immediately following a change in the accretion rate. For example, if the accretion rate increases, the ISM begins getting diluted immediately by the large accretion rate. The SFR, and hence the rate of metal production, takes some time to adjust to the new accretion rate while gas builds up in the ISM. It therefore takes some time before metal production can ‘pollute’ the large supply of low-metallicity infalling gas and return to the equilibrium $Z^\dagger = 1$, wherein for every unit of new clean gas added to the galaxy, star formation provides enough mass in metals to bring it to $Z = q + Z_{\text{IGM}}$. Similarly, if the accretion rate decreases suddenly, the rate at which metals are produced exceeds the rate necessary to pollute the new, smaller, supply of accreting gas, and the metallicity of the ISM increases until star formation and galactic winds can burn through the gas reservoir, allowing the galaxy to adjust to a lower rate of star formation (and hence metal production).

Immediately after a change in the accretion rate, i.e. in the limit that $\Delta \ll 1$,

$$Z^\dagger(\Delta) \approx Z^\dagger(0) + (1 - Z^\dagger(0))e^{\sigma x_k}/\Psi_k \Delta. \quad (21)$$

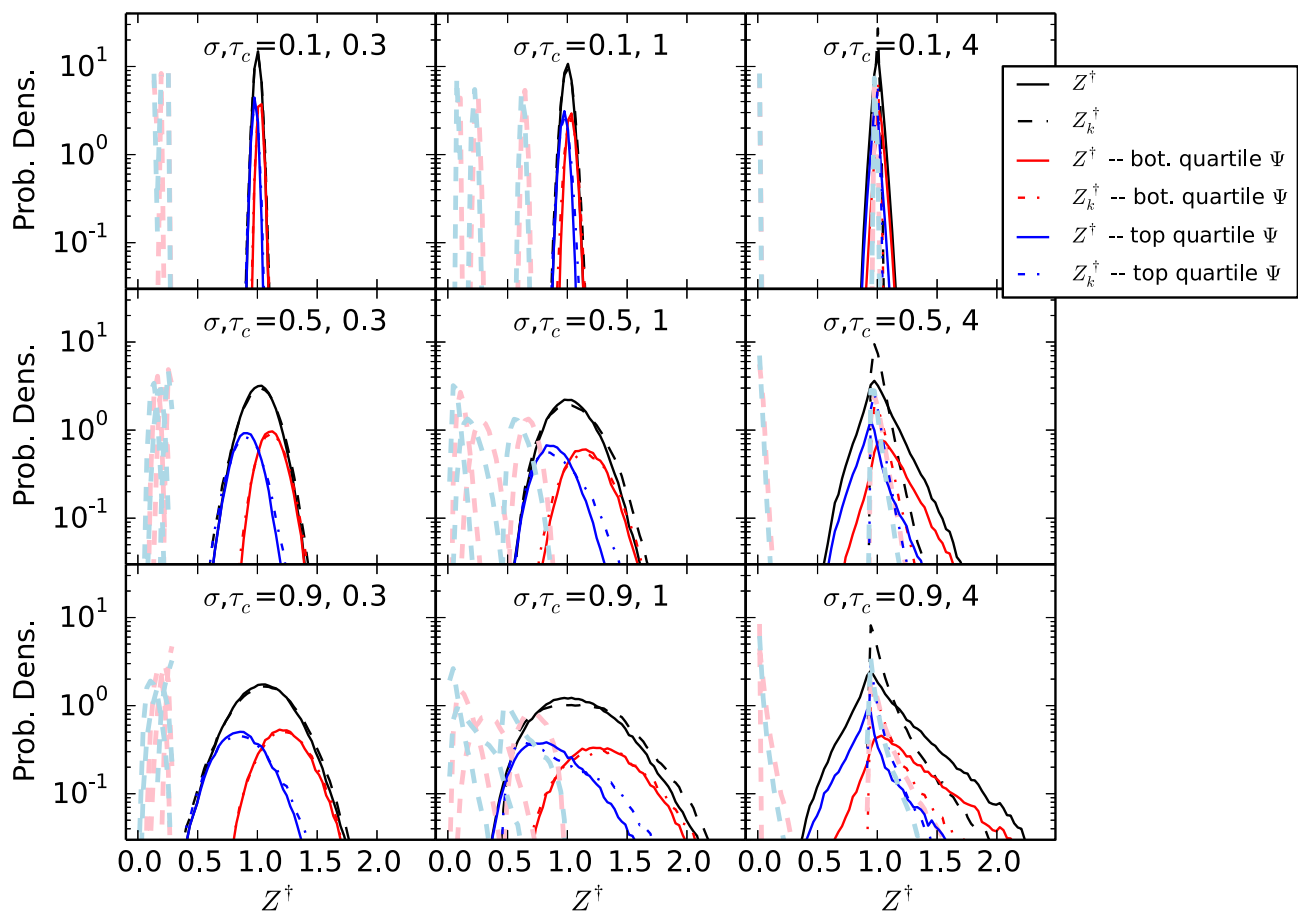


Figure 5. The construction of the distribution of the normalized metallicity, Z^\dagger . For the same parameters shown in the previous figures, we plot the PDF of Z^\dagger (solid line), the approximation to it using Z_k^\dagger (equation 23), and we also split the probability distribution into galaxies with high (low) SFRs in blue (red). We can immediately see that high-metallicity galaxies tend to have low SFRs, though the probabilities do overlap substantially. Light blue and red lines show the PDF of the contributions to Z_k^\dagger , split into the top and bottom quartiles in SFR, from old terms in the summation of equation (23). As with Ψ , the probability distribution of Z^\dagger may be thought of as the sum of a series of random draws wherein the influence of older draws is exponentially forgotten, although galaxies which end up in the blue (red) bin at the present time are likely to have had preferentially higher (lower) SFRs in the past, at least for $\tau_c \lesssim 1$. Unlike with Ψ , Z^\dagger becomes noticeably non-lognormal for larger values of τ_c , since galaxies sampled at a random time are likely to be near the equilibrium value $Z^\dagger = 1$, regardless of the accretion rate.

Thus, we can see that immediately following a change in accretion rate, whether the metallicity will increase or decrease depends only on the ratio $Z^\dagger(0)e^{\sigma\tau_c}/\Psi_k$. When this number is greater than unity, because the metallicity is high and/or the new accretion rate is larger than the present mass-loss rate, the metals will be diluted, whereas if this number is less than one, e.g. in a low-metallicity galaxy and/or one facing a sudden drop in accretion rate, stars will be forming fast enough to pollute the gas reservoir. This relation demonstrates the basic physical mechanism which gives rise to the mass–star formation–metallicity plane in our model, but also shows that this is a statistical relation only. A higher accretion rate (and hence an increasing SFR) is not guaranteed to produce a lower metallicity.

We can solve for the maximum deviation from $Z^\dagger = 1$ if the galaxy begins in equilibrium, namely $Z^\dagger(0) = 1$. In this case,

$$Z_{\text{extreme}}^\dagger = \frac{(r_k - 1)(1 + W_0((r_k - 1)/e))}{r_k - 1 + \exp(1 + W_0((r_k - 1)/e))}, \quad (22)$$

where $W_0(x)$ is the Lambert W function, namely the real solutions to $y = W_0(y)e^{W_0(y)}$.

As in the previous section, we can construct a recursion relation by setting $\Delta = \tau_c$ in equation (20). As with the mass-loss rate, we

see that the metallicity at the switches in accretion rate is given by a sum wherein the effects of long-past accretion are exponentially suppressed,

$$Z_k^\dagger = Z_0^\dagger \prod_{i=0}^{k-1} \frac{r_i}{r_i + e^{\tau_c} - 1} + \sum_{i=0}^{k-1} \frac{e^{\tau_c} - 1 + \tau_c(r_i - 1)}{e^{\tau_c} - 1 + r_i} \prod_{j=i}^{k-1} \frac{r_j}{e^{\tau_c} - 1 + r_j}. \quad (23)$$

Since $r_k > 0$ and $e^{\tau_c} > 1$, each factor in both of the products is guaranteed to be between 0 and 1 – for a median value of $r_k = 1$, each factor becomes $e^{-\tau_c}$. In that sense, this equation is very similar to equation (9) for Ψ_k . We show how Z^\dagger is constructed in Fig. 5 – in addition to the full distributions of Z^\dagger and Z_k^\dagger , we show the distributions for galaxies with Ψ in the top and bottom quartiles of galaxies with those values of σ and τ_c . As in Fig. 2, previous terms in the sum recede into irrelevance over the course of a few star formation times. There are, however, crucial differences between the construction of Z_k^\dagger and Ψ_k . To construct Ψ_k , independent lognormal variables were added together to get another quantity which was roughly lognormal. Here on the other hand, even though

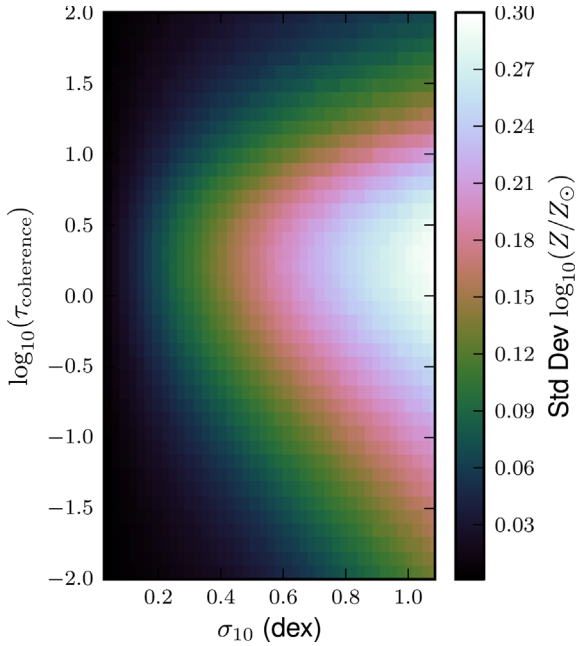


Figure 6. The width of the MZR. Each pixel represents an ensemble of galaxies with fixed τ_c and $\sigma_{10} = \sigma/\ln(10)$, wherein the metallicity Z was measured for each galaxy at a random time. The standard deviation of $\log_{10} Z/Z_\odot$ in each bin is plotted above. Longer coherence times (upwards in the plot) allow the galaxies to equilibrate so that Z approaches Z_{eq} (a constant value regardless of the accretion rate). Coherence times shorter than the mass-loss time-scale, i.e. $\log_{10} \tau_c \lesssim 0$, lead to a reduction in the scatter roughly in proportion to τ_c – individual draws from the accretion rate matter increasingly less. Larger intrinsic scatters (rightwards in the plot) negate this effect by making large accretion events typical.

an individual r_i is indeed roughly lognormally distributed (since $r_k = \Psi_k/e^{\sigma x_k}$ and Ψ_k is roughly lognormal), it is not the r_i 's being summed, but rather a more complicated function of r_i and τ_c . Moreover, the r_i are not independent of each other. Perhaps worst of all, since Z^\dagger does not monotonically approach its equilibrium value each time the accretion rate changes, but rather increases to $Z^\dagger_{\text{extreme}}$ before returning to $Z = Z_{\text{eq}}$, Z_k^\dagger becomes a bad approximation of the full distribution of Z^\dagger when $\tau_c \gtrsim 1$ [though eventually, for $\tau_c \gg 1$, the probability distributions of Z^\dagger and Z_k^\dagger both approach $\delta(Z^\dagger - 1)$].

We can also see that in this regime the distribution of Z^\dagger is not lognormal. We can none the less compute the standard deviation of $\log Z^\dagger$, shown in Fig. 6, for different fixed values of σ and τ_c . This has an observational counterpart in the scatter in the MZR, so we will use these predictions as a guide to understand how the observed scatter can constrain this model. The contours in Fig. 6 are substantially different from those in Fig. 3, so the scatter in the MS and the MZR may provide complementary constraints on the model. The physical reason is that for $\tau_c \gtrsim 1$, galaxies' metallicities are less and less affected by the intrinsic scatter σ for increasing τ_c , whereas for increasing τ_c the scatter in the SFR approaches the intrinsic scatter. The metallicities are unaffected because regardless of σ , $Z \rightarrow Z_{\text{eq}}$ over a few mass-loss times.

Returning to Fig. 5 we can also begin to see that Ψ and Z^\dagger are anticorrelated. Galaxies which have Ψ in the top quartile of their ensemble tend to have lower metallicities. This is a very simple physical effect, namely the competition between dilution of metals by new infall and pollution of metals by stellar evolution. If the galaxy is burning through excess gas from previous accretion events ($r_k > 1$), the metallicity will increase as the gas reservoir is

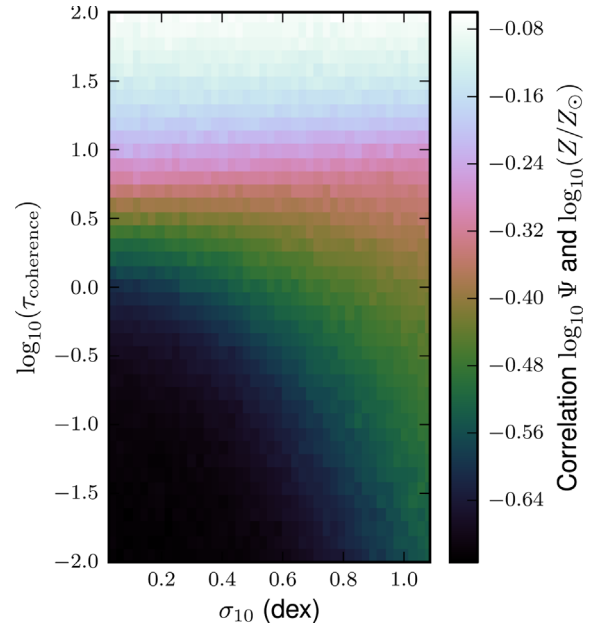


Figure 7. The correlation between Ψ and Z/Z_\odot . For small scatters and rapid variability in the accretion rate, the SFR and metallicity are substantially anticorrelated. Increasing the coherence time allows galaxies to return to their equilibrium Z regardless of the accretion rate, wiping out the anti-correlation.

polluted, whereas if the galaxy is accreting more gas than it is losing ($r_k < 1$), the SFR is slow enough that new gas is added faster than metals can be produced to pollute it. Using the same Monte Carlo simulations used to produce Fig. 3, we can explicitly show the correlation between $\log \Psi$ and $\log Z$ (Fig. 7), defined in general according to

$$\text{Corr}(X, Y) = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum_i (X_i - \bar{X})^2) (\sum_j (Y_j - \bar{Y})^2)}}, \quad (24)$$

where each sum is over all galaxies in the ensemble and \bar{X} indicates an average over the ensemble. The correlation is strongest in the ‘linear regime’, namely small rapid perturbations. When the coherence time exceeds a few star formation times, the correlation disappears – Z^\dagger is always close to unity regardless of the accretion rate. The correlation also weakens as the intrinsic scatter in the accretion rate increases. The correlation closely resembles the slope of the FMR, namely if we fit a linear relation to the set of points $(\log_{10} \Psi, \log_{10} Z^\dagger)$ produced for a fixed value of σ and τ_c , the resulting contours (Fig. 8) are similar to the contours in Fig. 7.

4 CONSTRUCTION OF GALAXY SCALING RELATIONS – AN INITIAL GUESS

The results we have derived thus far, namely the joint probability density of Ψ and Z^\dagger , are applicable only at fixed values of μ , σ , $t_{\text{coherence}}$, t_{loss} , Z_{IGM} , and q . We refer to an ensemble of galaxies with fixed values of these parameters as a simple (or stationary) galactic population (SGP). As we saw in the previous sections, the only variables which affect the joint distribution of Ψ and Z^\dagger are σ and $\tau_c = t_{\text{coherence}}/t_{\text{loss}}$. However, to compute physical quantities, i.e. the SFR, the metallicity, etc., one must specify the other variables.

The power of our approach using SGPs is that, to a reasonable approximation, the properties of a star-forming galaxy are set by

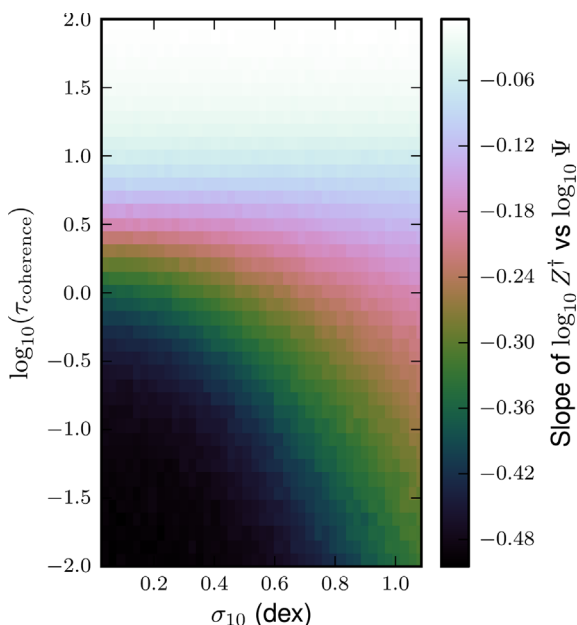


Figure 8. The ‘slope of the FMR’. For each τ_c and $\sigma_{10} = \sigma/\ln(10)$, we fit a linear model to the scatter plot of $\log_{10}\Psi$ versus $\log_{10}(Z/Z_\odot)$ of all galaxies drawn from the distribution, and plot the resulting slope here. We find uniformly that the slope is negative, though the slope flattens for large values of τ_c , and to a lesser degree, for large values of σ because the metallicity in each case becomes less sensitive to the current value of the accretion rate. Either it is approaching its fixed equilibrium value (large τ_c) or it is being more heavily influenced by past very dramatic accretion events (large σ_{10}).

a single parameter having to do with the size of the galaxy (e.g., stellar mass, halo mass, or K -band luminosity). One could therefore hope that, at a fixed stellar mass, the population of galaxies may be well described by a single SGP. Essentially, all of the scalings of mass loading factor, accretion rate, etc. which account for the slope and zero-points of galaxy scaling relations would be taken out, leaving only the intrinsic scatter.

Here we make an educated guess as to how to map the dimensionless model presented in the first few sections of this paper to observable galaxy scaling relations. Although alternative assumptions may be preferred by other practitioners, we hope that this exercise will be at least illustrative. In Section 4.1, we use results from N -body-only dark matter simulations to make guesses for the parameters of the accretion process: μ , σ , and $t_{\text{coherence}}$. Each of these parameters thereby has a predicted scaling with halo mass and redshift. In this procedure, we leave one free parameter f_ϵ , a constant less than unity, to describe how much the accretion rate is reduced below this guess. We then adopt the assumptions that $y = 0.054$, $f_R = 0.54$ [appropriate for a Chabrier (2005) initial mass function (IMF) with yields from solar-metallicity stars Maeder (1992) – see appendix A of Krumholz & Dekel (2012)], $Z_{\text{IGM}} \ll 0.02$, and $\xi = 0$. To fully specify the SGP model, the only remaining parameters are η , t_{dep} , and f_ϵ . In Section 4.2, we adopt a value of f_ϵ and power-law scalings of t_{dep} and η with halo mass such that we fit three galaxy scaling relations: M_* versus SFR, Z , and M_g/M_* . To do so, we need to assign a value of M_* to each halo mass, for which we take the Behroozi, Wechsler & Conroy (2013a,b) relations with a fixed lognormal scatter of 0.19, consistent with various observational constraints (Behroozi et al. 2013b; Reddick et al. 2013).

With all of these parameters specified as a function of halo mass and redshift, we can construct synthetic versions of these galaxy scaling relations, including their intrinsic scatter, and the higher order FMR, $Z(M_*, \text{SFR})$. Section 4.3 describes how we take the synthetic relations and extract three higher order quantities which we will use to constrain our model: the scatter in the MS, the scatter in the MZR, and the ‘slope’ of the FMR, namely the logarithmic derivative of the metallicity with respect to the SFR at fixed stellar mass. These three quantities can then be compared directly to observations, which allows us to rule out our initial guess, and non-trivially constrain σ , $t_{\text{coherence}}$, and other potential sources of scatter (Section 5.1). The parameters used throughout the paper are summarized in Table 1.

4.1 Baryonic accretion

As is often the case for modellers of galaxies, we will begin with the dark matter. For the purposes of this simple model, we will rely on the EPS-like (Press & Schechter 1974; Sheth & Tormen 1999) formalism presented in Neistein & Dekel (2008) and Neistein, Macciò & Dekel (2010). From this, we will derive approximations for μ , σ , and $t_{\text{coherence}}$. For a *Wilkinson Microwave Anisotropy Probe 5* cosmology, Neistein et al. (2010) can fit the cumulative mass function of haloes found in an N -body simulation if they construct halo accretion histories according to

$$\Delta S = \exp(\mu_p + x\sigma_p), \quad (25)$$

where x is a standard normal drawn at a fixed interval $\Delta\omega$, and

$$\begin{aligned} \mu_p = & (0.132 \log_{10} S + 2.404) \log_{10}(\Delta\omega) \\ & + 0.585 \log_{10} S - 0.436 \end{aligned} \quad (26)$$

$$\begin{aligned} \sigma_p = & (-0.333 \log_{10} S - 0.321) \log_{10}(\Delta\omega) \\ & + 0.0807 \log_{10} S + 0.622 \end{aligned} \quad (27)$$

In these equations, S is a measure of the amplitude of the dark matter power spectrum, and ω is a measure of time, similar to redshift.

Note that in general $\mu_p \neq \mu$ and $\sigma_p \neq \sigma$ because these refer to the mean and scatter in S – ω , rather than M – t space. To compute an estimate of the accretion rate, we take a grid of M_h and z – for each M_h we can compute S and draw a large number of ΔS values, from which we can compute ΔM_h , the change in halo mass. For each z we can compute Δt , the time between the current redshift and a time $\Delta\omega$ earlier. We then approximate each baryonic accretion rate as

$$\dot{M}_{\text{ext}} \approx f_b \epsilon \frac{\Delta M_h}{\Delta t}. \quad (28)$$

This estimate has three components – $\Delta M_h/\Delta t$ is an estimate of the dark matter accretion rate, f_b is the cosmic baryon fraction, and ϵ is a factor by which we further reduce the accretion rate to account for effects on the baryonic accretion rate from baryonic physics, e.g. the buildup of hot haloes, which reduces the rate at which gas accretes on to galaxies.

From our ensemble of ΔM_h , we can approximate μ and σ as the mean and standard deviation of $\ln \dot{M}_{\text{ext}}$. We note that the distribution of \dot{M}_{ext} is not guaranteed to be lognormal, and this procedure produces distributions with non-zero skew and kurtosis in $\ln \dot{M}_{\text{ext}}$. We neglect this non-Gaussian component and approximate \dot{M}_{ext} as lognormal, keeping in mind that this is merely a guess at the true baryonic accretion rate. We take the efficiency factor ϵ to be

$$\epsilon = f_\epsilon \min(0.31(M_h/10^{12} M_\odot)^{-0.25}(1+z)^{0.38}, 1.0), \quad (29)$$

Table 1. Important parameters used in this study.

Parameter	Equations	Description
<i>The accretion process</i>		
μ	3, 28	The log base e of the median baryonic accretion rate
σ	3, 28	The lognormal scatter in the (DM and baryonic) and accretion rate
$\Delta\omega$	30	The scale-free time step taken to generate accretion histories
$t_{\text{coherence}}$	3, 30	The time over which the accretion rate is constant before a new random value is drawn
ϵ	28, 29	The fraction of $f_b \dot{M}_{\text{DM}}$ which reaches the gas reservoir
f_ϵ	29	A fixed fraction by which the baryonic accretion rate may be reduced – fixed by first-order scaling relations
Z_{IGM}	12	Metallicity of infalling baryons, fixed at 2×10^{-4}
<i>Star formation</i>		
\dot{M}_{SF}	2	The star formation rate
M_g	1, 2	The gas mass available to form stars
η	2	The ratio of mass lost in galactic winds to stars formed – fixed by first-order scaling relations
t_{dep}	2	The depletion time, M_g/\dot{M}_{SF} – fixed by first-order scaling relations
f_R	2	The remnant fraction – fraction of mass not returned to the ISM from a simple stellar population, fixed at 0.54
t_{loss}	1, 2	The loss time, i.e. the time-scale on which the reservoir is depleted, $t_{\text{dep}}/(\eta + f_R)$
y	12	Mass of metals returned to the ISM per mass locked in stellar remnants, fixed at 0.054
ξ	13	Metallicity enhancement of galactic winds, fixed at 0 (perfect mixing)
q	15	A combination of y , f_R , ξ , and η we call the effective yield
<i>SGP parameters</i>		
τ_c	5	$t_{\text{coherence}}/t_{\text{loss}}$ – an input to the SGP
σ_{10}	3, 28	This is the same σ as above, but divided by $\ln 10$ – an input to the SGP
Ψ	4, 5	The ratio of the mass-loss rate (M_g/t_{loss}) to the median accretion rate e^μ – an output of the SGP
Z^\dagger	18, 19	The metallicity of the gas reservoir, subtracting out Z_{IGM} and normalizing to q – an output of the SGP
x_k	3	For a given galaxy, the k th draw from a standard Gaussian, which sets the accretion rate
<i>Other sources of scatter (see especially Section 5.1)</i>		
$\sigma_{\log_{10} M_*}$	–	Lognormal scatter added to $M_*(M_h)$, assumed uncorrelated with anything else
$\sigma_{\log_{10} t_{\text{dep}}}$	–	Lognormal scatter added to the depletion time
$\sigma_{\log_{10} \eta}$	–	Lognormal scatter added to the mass loading factor η
σ_μ	–	Gaussian scatter in μ . This is like including a stochastic component whose coherence time $\gg t_{\text{loss}}$

where we leave $f_\epsilon < 1$ a free parameter to be fitted in the next section. The remaining factors come from the fitting formula of Faucher-Giguère, Kereš & Ma (2011), which accounts for the suppression of accretion in high-mass haloes presumably due to hot virialized gas.

We estimate $t_{\text{coherence}}$ by the fact that in the dark matter simulations, the merger trees become non-Markov for $\Delta\omega \lesssim 0.5$, indicating that the accretion rates over time intervals shorter than that are correlated (Neistein & Dekel 2008). We therefore use

$$t_{\text{coherence}} \sim \left| \left(\frac{d\omega}{dz} \right)^{-1} \frac{d\tau}{dz} \Delta\omega \right|, \quad (30)$$

with $\Delta\omega = 0.5$. Obviously, this is only a rough estimate, since a real accretion history is likely to have more structure in Fourier space than the single period we assume here. We encourage those with cosmological simulations to measure this quantity, both in dark-matter-only and baryonic simulations. Another plausible value for $t_{\text{coherence}}$ might be the dynamical time of the halo or some other time-scale related to the baryon cycle. This choice might be appropriate if the primary supply of gas is re-accreting winds.

4.2 Fitting the first-order relations

To compute τ_c and re-dimensionalize the SGP, we still need to know the mass-loss time-scale, i.e. the depletion time and η , as well as ξ . These values are sufficiently uncertain that it is worth digressing to discuss how they may be chosen to fit the first-order galaxy scaling

relations, namely the star-forming MS, the MZR, and the stellar mass–gas mass relation.

To compare our model with many galaxy scaling relations, computed as a function of stellar mass M_* , we must pick an M_* . Since our model is purely equilibrium based, we have no way to compute integrated quantities like M_* besides appealing to other empirical relations. We employ the Behroozi et al. (2013a,b) model of the relation between stellar mass and halo mass, including a scatter in stellar mass at a fixed halo mass of 0.19 dex (Reddick et al. 2013). Another common approach is to simply use M_* rather than M_h (Lilly et al. 2013) as the parameter by which to scale the SGP parameters (η , ξ , etc.).

In a standard equilibrium model (Davé et al. 2012), the ‘centre’ of a galaxy scaling relation may be determined by setting time derivatives to zero. For instance, setting $dM_g/dt = 0$ yields an equilibrium star formation rate

$$\dot{M}_{\text{SF,eq}} = \dot{M}_{\text{ext}}/(\eta + f_R). \quad (31)$$

Immediately we can see the critical importance of two unknown pieces of physics – anything which displaces the baryonic accretion rate \dot{M}_{ext} away from the naive estimate $f_b \dot{M}_{\text{DM}}$ (e.g. preheating, halo quenching, AGN heating), and anything which removes gas from the star-forming gas reservoir of the galaxy (supernovae, radiation pressure, or cosmic rays).

Interestingly, the depletion time-scale does not enter into $\dot{M}_{\text{SF,eq}}$. This is because the gas mass is free to adjust to whatever it needs to be so that inflowing gas is balanced by sinks for the gas – old stellar

remnants and galactic outflows. For instance, a longer t_{dep} would simply mean a larger gas mass would be required to reach the same equilibrium between inflows and sinks,

$$\dot{M}_{\text{g,eq}} = \dot{M}_{\text{SF,eq}} t_{\text{dep}} = \dot{M}_{\text{ext}} t_{\text{loss}}. \quad (32)$$

Thus, we can see that SFRs can only be affected by physics which alters the star formation time-scale (e.g. H_2 regulation; Krumholz & Dekel 2012) to the degree that the galaxies are out of equilibrium. These physical considerations do, however, affect the equilibrium size of the gas reservoir.

If we make the additional restriction that $dZ/dt = 0$, we arrive at the equilibrium metallicity

$$Z_{\text{eq}} = Z_{\text{IGM}} + q = Z_{\text{IGM}} + \frac{y f_{\text{R}}}{f_{\text{R}} + \eta} \left(1 - \frac{\eta \xi}{\max(1 - f_{\text{R}}, \eta)} \right). \quad (33)$$

The metallicity is a particularly powerful probe of feedback physics because Z_{eq} is independent of the accretion rate, meaning that ‘preventive feedback’ or any other considerations which affect \dot{M}_{ext} do not affect the first-order MZR – essentially all that matters are η , the mass loading factor, and ξ or some other measure of the mixing between ejecta and the ISM. The drawback of using the MZR is that it is bedevilled by large systematic uncertainties in converting characteristics of metal emission lines into actual gas-phase metal abundances (Kewley & Ellison 2008).

Given these equilibrium relations, we attempt to find power-law scalings of η and t_{dep} , and a constant value of f_{e} , which will roughly fit the observed galaxy scaling relations. This leads us to take $f_{\text{e}} = 0.5$, $\eta = (M_{\text{h}}/10^{12} M_{\odot})^{-2/3}$, and $t_{\text{dep}} = 3 \text{ Gyr } (M_{\text{h}}/10^{12} M_{\odot})^{-1/2}$, as described below. These fits were done by hand, which is reasonable given the uncertainties in the mean relations.

For the mean relations themselves, we take the summaries of this data from Peebles & Shankar (2011) – these are plotted as the coloured lines and the points with error bars in Fig. 9. The data themselves largely come from the Sloan Digital Sky Survey (SDSS), with the exception of the gas masses. The fits use stellar masses from Tremonti et al. (2004) and Kauffmann et al. (2003), and SFRs from Brinchmann et al. (2004), both of which include aperture corrections. Metallicities are estimated by Kewley & Ellison (2008) using a variety of strong line metal calibrations from the literature, and then corrected by Peebles & Shankar (2011) to use a Chabrier (2003) IMF. The gas fraction data come from McGaugh (2005), Leroy et al. (2008), and Garcia-Appadoo et al. (2009).

To approximately fit the centres of the observed $z = 0$ MS, we had to use a steep scaling of the mass loading factor with halo mass, namely $\eta = (M_{\text{h}}/10^{12} M_{\odot})^{-2/3}$. This is a result of the short cooling times and/or cold streams in low-mass haloes which efficiently supply cold gas at a rate proportional to the dark matter accretion rate, which is much larger than the observed SFR in these galaxies. That is, ϵ is large for low-mass galaxies in our initial guess. If low-mass galaxies are in statistical equilibrium, these large values of η are necessary. It is possible that some other physical process is decreasing ϵ at these masses, or that the galaxies are not in equilibrium (Krumholz & Dekel 2012; Kuhlen et al. 2012; Kuhlen, Madau & Krumholz 2013). In the latter scenario, η may vary quite weakly, leaving the mass-loss time-scales at low masses to be comparable to the depletion times, themselves comparable to or much longer than the age of the Universe. We discuss this possibility further in Section 5.2, but for now we adopt $\eta \propto M_{\text{h}}^{-2/3}$.

We do not fit the MZR particularly well in this simple model (see Fig. 9). That is, most calibrations of the MZR are not well described by the power law in M_{h} we assume here. This is not extremely concerning given the systematic uncertainties in the observations.

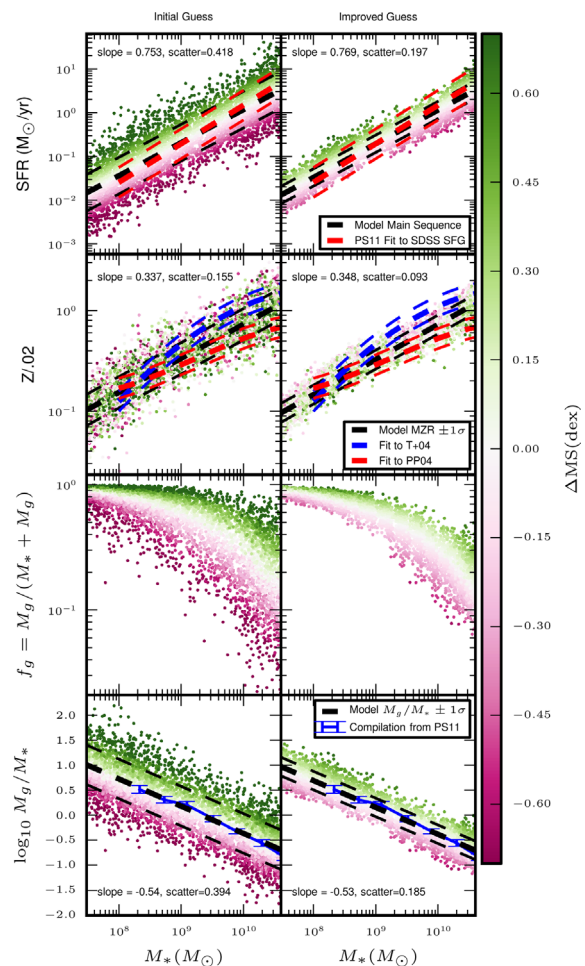


Figure 9. Scaling relations. We compare observed galaxy scaling relations and their scatters (coloured lines) with scaling relations (black lines) calculated by drawing galaxies (coloured points) from a sequence of SGPs using reasonable guesses of how the SGP parameters scale with halo mass. The left-hand panel is our initial guess, while the right-hand panel shows an improved guess which better fits the observational constraints on the scatter in these relations. In particular, the initial guess yields a model MS with intrinsic width wider than the observed scatter, while the improved guess has a much smaller intrinsic scatter. The colours indicate offset from the MS, namely the black dashed line fitted in the top panel. PS11 refers to Peebles & Shankar (2011), and T+04 and PP04 refer to metallicity calibrations used by Tremonti et al. (2004), and the $[\text{N II}]/\text{H}\alpha$ calibration from Pettini & Pagel (2004).

The scaling of η needed to fit the star-forming MS does leave the MZR in the right neighbourhood, without adjusting the fiducial values of $Z_{\text{IGM}} = 2 \times 10^{-4} \approx Z_{\odot}/100$, $\xi = 0$, or f_{e} and η .

In contrast, the gas fraction data are fitted remarkably well by our fiducial scalings. Given the reasonable fit of the SFR– M_{h} relation, we were left with one parameter to vary to fit $M_{\text{g}}/M_{\text{h}} - M_{\text{h}}$, namely t_{dep} . Despite the (lack of) trends from the GASS (Schiminovich et al. 2010) and COLD GASS (Saintonge et al. 2011) surveys in the total gas depletion time with mass, we find that we need $t_{\text{dep}} = 3 \text{ Gyr } (M_{\text{h}}/10^{12} M_{\odot})^{-1/2}$. Since η scales even more steeply, the mass-loss time-scale actually shortens with decreasing halo mass in this scenario, scaling as $t_{\text{loss}} \propto M_{\text{h}}^{1/6}$.

The result of applying these scalings is shown in Fig. 9. For each M_{h} below $10^{12.3} M_{\odot}$ provided in the Behroozi et al. (2013a,b) $M_{\text{g}}(M_{\text{h}})$ relation, we apply the scalings above to compute the

Table 2. The two reference models we use to generate synthetic galaxy scaling relations and FMRs.

	Initial guess ^a	Improved guess ^b
σ_{10} (dex)	0.45	0.225
$\Delta\omega$	0.5	0.2
$\sigma_{\log_{10} M_*}$ (dex)	0.19	0.07
$\sigma_{\log_{10} t_{\text{dep}}}$ (dex)	0	0
$\sigma_{\log_{10} \eta}$ (dex)	0	0
σ_{μ} (dex)	0	0.08

^aBased on N -body simulations.^bAdjusted to fit both the conservative observational constraints and the much tighter $-2 < \partial \log Z / \partial \log \dot{M}_{\text{SF}} < -1$.

necessary physical parameters to specify the SGP, draw 300 samples from the SGP, and assign each one a stellar mass according to $M_*(M_h)$, plus a fixed scatter, $\sigma_{\log_{10} M_*} = 0.19$ dex. The resultant population of galaxies, as designed, fits various observational constraints at $z = 0$, represented in each figure by the various coloured lines (which include representative scatters) from Peeples & Shankar (2011). The black lines show power-law fits and the computed $\pm 1\sigma$ scatter of the population. The observational fits are not necessarily precise, and in particular the MZR famously has many different fits depending on which calibration is used (Kewley & Ellison 2008). The two columns in Fig. 9 have the same scalings of t_{dep} , η , and q , but different values of σ , $\Delta\omega$, and $\sigma_{\log_{10} M_*}$. The left column shows the initial guess discussed here and in the previous section, while the right column has values of the accretion process parameters more in line with observations (the ‘improved guess’ model – see Table 2 and Sections 4.3 and 5.1). We have chosen the range of M_* to extend somewhat below the minimum stellar masses used in the various summaries of the SDSS data from Peeples & Shankar (2011), up to the stellar mass associated with the quenching of star-forming galaxies and hence their departure from the star-forming MS (Cheung et al. 2012).

It is worth emphasizing that fitting or not fitting the observed relations should be construed neither as success nor failure for our model – the equilibrium relations (equations 31, 32, and 33) have enough free parameters that fitting the observations is not challenging. However, it is encouraging that relatively little tuning was required for reasonable fits, and that we end up with models with a reasonable physical basis: energy-driven winds for $\eta \propto M_h^{-2/3}$ (see Murray, Quataert & Thompson 2005 for the basics of energy- versus momentum-driven winds), long depletion times at low masses owing to low H_2 fractions, η of order unity, and t_{dep} of the order of 3 Gyr at high masses.

4.3 Information in the scatter

Now that we have a reasonable fit for the redshift zero–median relations, we can return to our main goal – understanding the higher order relationships in the data. In particular, we would like to understand the scatter in the MS and the MZR, and the (negative) slope implied by the FMR of metallicity with respect to SFR at fixed stellar mass. Now that we have synthetic data, we can follow a simple procedure to fit a synthetic MS, MZR, and FMR. For the former two, we simply fit a linear model with least-squares regression to (the log base 10) of SFR versus M_* and Z versus M_* . For each synthetic galaxy, we can then subtract off the linear fit for SFR or Z at that galaxy’s M_* to find its residual. Finally, the scatter is calculated as the sample standard deviation of the residuals.

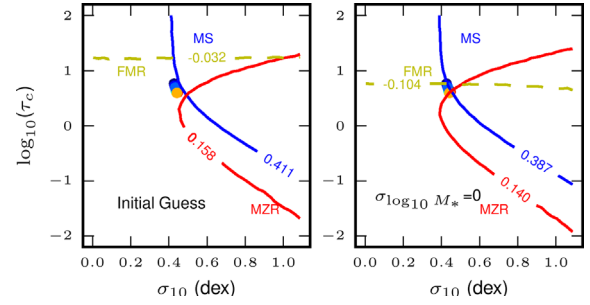


Figure 10. Information in the scatter of the MS and MZR. The coloured points near (0.45, 0.8) represent the values of τ_c and σ_{10} in the ‘initial guess’ model as a function of a wide range of halo masses. With a random realization of the initial guess model, we extract the scatter in the MS and MZR (quoted in fig. 9), and the linear slope $\partial \log Z / \partial \log \dot{M}_{\text{SF}}$ of the FMR. We can then find the contours in Figs 3, 6, and 12 corresponding to these values, i.e. the set of σ_{10} and τ_c the SGP would predict if we provide it with these two scatters and the FMR slope. These contours are shown as the coloured lines – the numbers for each contour are the scatters in the MS and MZR (blue, red), and the slope in the FMR (yellow) which we feed to the contour maps. When we do not include any uncorrelated scatter in the stellar mass (right-hand panel), the lines intersect at the input value, but including a reasonable scatter in M_* (left-hand panel) increases the scatters in the MS and MZR and flattens the FMR slope.

Meanwhile, we fit a synthetic FMR both with a linear model

$$\log_{10} Z = b_0 + b_1 \log_{10} M_* + b_2 \log_{10} \dot{M}_{\text{SF}}, \quad (34)$$

and, as is common practice, a quadratic model¹

$$\log_{10} Z = a_0 + a_1 m + a_2 s + a_3 m^2 + a_4 ms + a_5 s^2. \quad (35)$$

Here $m = \log_{10} M_*/M_\odot - 10$ and $s = \log_{10} \dot{M}_{\text{SF}}/(M_\odot \text{ yr}^{-1})$. From the linear fit, we can read off a value for $\partial \log_{10} Z / \partial \log_{10} \dot{M}_{\text{SF}}$ at fixed M_* , namely b_2 . For the quadratic fit, the slope is a function of both mass and \dot{M}_{SF} .

For all three of these quantities – the two scatters and the (logarithmic) slope of Z versus \dot{M}_{SF} at fixed stellar mass – we can compare both to observations and to predictions from the full joint distribution of Ψ and Z^\dagger by a single SGP (Figs 3, 6, and 8). The scatters in the MS, MZR, and FMR are not identical to that in an SGP because the relationship between stellar and halo mass also has some scatter. Thus, a population of galaxies at fixed M_* represents a weighted sum of SGPs.

Thus, there will be additional scatter in these synthetic observations, as compared to the SGP predictions. This is illustrated in Fig. 10 – we compare the input values of τ_c and σ_{10} (the coloured points, nearly on top of each other) to the values of τ_c and σ_{10} the SGP model would predict given the values of the synthetic scatters. That is, we construct a synthetic scaling relation, measure its scatter, and then plot the appropriate contour from Fig. 3, 6, or 8. When we artificially set $\sigma_{\log_{10} M_*} = 0$, these contours converge at the input values, but including the scatter in M_* flattens the FMR slope and increases the scatter in the MS and MZR, leaving a general region in τ_c – σ_{10} space where the contours are close to each other, but no trivial way to recover the input values.

This at once shows both the promise and the difficulty of using observed second-order relations to predict these parameters. The

¹ Note that the quadratic fits reported in the literature are typically fits to $12 + \log_{10}(\text{O}/\text{H})$, whereas here we are fitting to the metallicity. This only makes a difference in the a_0 term, and not to the slope with respect to the SFR at fixed stellar mass, on which we will concentrate.

constraints are largely independent of each other, so one could hope to non-trivially constrain the acceptable values of σ_{10} and τ_c . However, additional sources of scatter not included in the simple dimensionless SGP predictions can make it difficult to recover the values of these parameters simply by reading off where the contours intersect in this diagram.

None the less, this diagram (Fig. 10) and the associated SGP predictions (Figs 3, 6, and 8) provide some heuristic guidance. We see that in this parameter space, the input values of σ_{10} and τ_c are nearly independent of halo mass. This means that the fiducial model would predict no change in the scatters of the MS or MZR, nor any change in the slope of the FMR, which is roughly consistent with observations (Whitaker et al. 2012; Zahid et al. 2013). We also see that to reduce the synthetic scatter to below the observed scatter, $\lesssim \pm 0.34$ dex for the MS and $\lesssim \pm 0.1$ dex (Kewley & Ellison 2008, both of which may be regarded as upper limits on the intrinsic scatters), we could reduce the input accretion scatter, σ , or dramatically reduce the coherence time (and therefore τ_c). Our ‘improved guess’ model adjusts the initial guess to reduce the two scatters and steepen the FMR slope. In particular, we reduce σ by a factor of 2 and $\Delta\omega$ in our estimate of $t_{\text{coherence}}$ to 0.25 (see the next section).

5 DISCUSSION

In the previous section, we set up a fiducial set of assumptions to map the SGPs of Sections 2 and 3 into observable parameters. We were easily able to match the first-order relations, but our first guess produced scatters in our synthetic MS and MZR that were too large. In this section, we examine in more detail the full range of SGP-based models that are consistent with the observed constraints on the scatters in the MS and MZR and the slope of the FMR, and the range of halo masses and redshifts over which SGP-based models are valid in principle.

5.1 A more general model – do all galaxies at a fixed M_h correspond to one SGP?

The analysis of SGPs in Sections 2 and 3 explicitly assumes that a given galaxy has had the same values of μ , σ , t_{loss} , $t_{\text{coherence}}$, Z_{IGM} , and q for eternity. This is clearly false – galaxies increase their mass over time, moving them along any presumed scaling relations in e.g. t_{dep} or η , while other quantities likely depend explicitly on time, e.g. μ and $t_{\text{coherence}}$. The statistical equilibrium model we have proposed here, and other simpler models, may still be successful in describing galaxies because these quantities plausibly vary slowly relative to the internal time-scales of the galaxy, i.e. the loss time. Whereas the typical equilibrium model assumes this of the accretion rate, our model relaxes that particular assumption and allows the accretion rate to vary, possibly very quickly, relative to other time-scales.

Our model was constructed with the goal of understanding the scatter in galaxy scaling relations by examining the role of a known (and significant) scatter in dark matter accretion rates among galaxies at a given mass. However, it is also plausible that the mass loading factor, the depletion time, or some other quantity may vary between galaxies near a fixed mass, or within a given galaxy on relatively short time-scales. The former situation may be handled by our model by having multiple SGPs with different values of e.g. η at the same mass. The latter situation cannot be handled by SGPs as we have formulated them.

The scenario in which scatter in η , t_{dep} , or μ is responsible for the scatter in galaxy scaling relations has several distinct predictions compared to the stochastic accretion model we have presented.

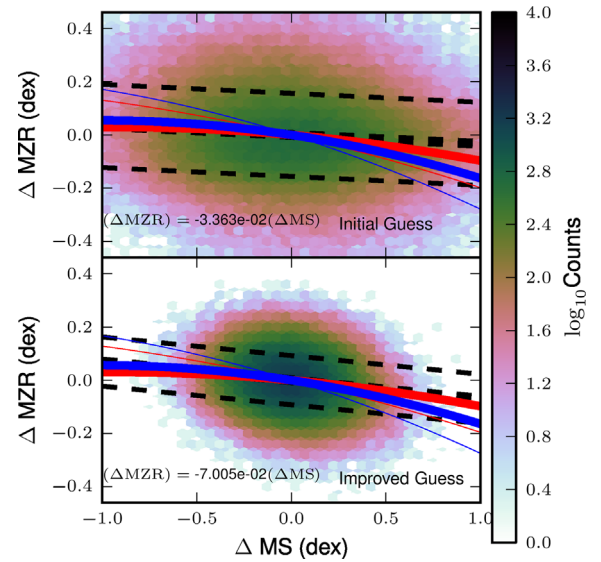


Figure 11. Displacements from first-order scaling relations. Here we show the offset of galaxies, in both our initial guess and improved guess models, from fits to their MS and MZR. The black lines show a linear fit to these data points, and the 1σ scatter of galaxies in Δ MZR around this fit. The coloured lines show predictions of these quantities from two different fits to the $z = 0$ FMR using different metallicity calibrations, with the thin lines corresponding to $M_* = 10^8 M_\odot$ and the thick to $10^{11} M_\odot$. Here we can see the slight, but significant observed and predicted anticorrelation between SFR and metallicity. Note that to fill out this histogram we drew a sample of 100 times the number of galaxies shown in Fig. 9.

In particular, the scatter, rather than being stochastic, would be constructed from several nearly parallel, slightly offset, equilibrium relations. One could likely find acceptable values for the scatter in the SGP parameters which reproduced the observed scatters, since each equilibrium relation depends on a different combination of the SGP parameters. Thus, one might expect to be able to have e^μ vary by ~ 0.34 dex and Z_{IGM} to vary by ~ 0.1 dex.

This model would indeed produce, more or less, the observed scatters in the MS and MZR, but it would not account for the decreasing metallicity with increasing SFR at fixed stellar mass (see Fig. 11), i.e. what we call the ‘slope’ of the FMR. In particular, the equilibrium metallicity is independent of the accretion rate, and the SFR is independent of Z_{IGM} (at least in this simple model), so there would be no slope in the FMR. The situation gets even worse with a scatter in η , since both the equilibrium SFR and Z are inversely related to η , which would tend to create a positive slope in the FMR. Similarly, a scatter in M_* at fixed halo mass tends to induce a positive slope – at a given M_* , galaxies from higher halo mass and lower halo mass will be present, and because both the MS and MZR have positive slopes, the higher (lower) halo mass galaxies will have higher (lower) SFRs and Z s, again leading to a positive slope in the FMR.

Given the difficulty in obtaining a negative FMR by adding scatters in the parameters which enter the equilibrium relations, compared with the natural way the negative slope arises in our statistical equilibrium model, via a time-varying accretion rate, it certainly seems that no alternative model is needed. In fact, by enforcing the requirement that $\partial \log Z / \partial \log \dot{M}_{\text{SF}} < 0$, we may be able to obtain limits on the scatter in parameters (M_* , η) which tend to make the slope in the FMR positive. We caution, however, that the uncertainty in the metallicity determinations is great enough that for some methods of computing the metallicity, the slope becomes positive at high

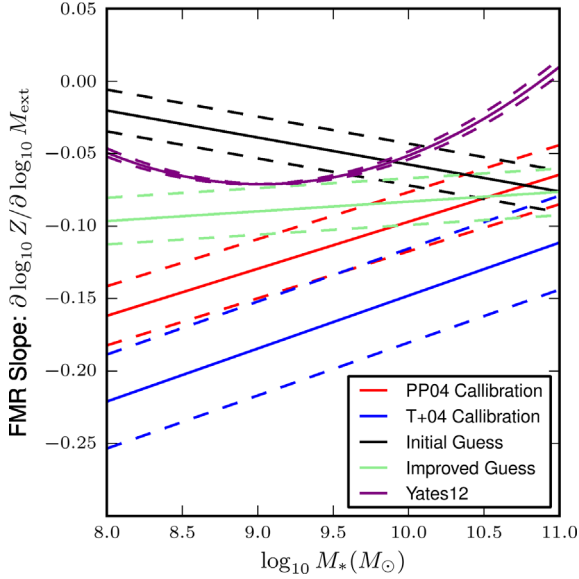


Figure 12. Slope in the FMR. For both our improved and initial guesses, as well as three fits to the $z = 0$ FMR using different metallicity calibrations, we show the slope in the metallicity with respect to the SFR, as a function of stellar mass. Since the FMR fits depend explicitly on \dot{M}_{SF}^2 , we must also choose an SFR at which to evaluate this quantity. We choose the MS value at that mass (as determined by our fit to the MS of our initial guess model), which we plot as the solid lines, and ± 0.3 dex, the dashed lines. The three different calibrations predict substantially different values, so to be somewhat conservative we have chosen to interpret the observational constraint as $\partial \log Z / \partial \log \dot{M}_{\text{SF}} < 0$, a fact on which most calibrations agree most of the time.

masses (Yates et al. 2012). For now we keep the weak, but perhaps not weak enough, restriction that $\partial \log Z / \partial \log \dot{M}_{\text{SF}} < 0$ and see what constraints we can derive.

To accomplish this, we set up a six-dimensional grid of models. Each point in the grid corresponds to a choice of σ_{10} , $\log_{10} \Delta\omega$, scatter in M_* , scatter in t_{dep} , scatter in η , and scatter in e^μ . For each point in the grid, we simulate a full set of galaxies – 200 per value of $M_h < 10^{12.3} M_\odot$, and compute the scatter in the MS and MZR and the slope in Z versus SFR at fixed M_* . We then compare each of these pieces of information to the observations in a maximally conservative way. We treat the observed scatters in the MS and MZR as upper limits on the intrinsic scatters, and make no assumptions about the purely observational scatter. Although there is an observationally known value of the slope of Z versus SFR at fixed M_* , we make no strong assumptions about the probability distribution function (PDF) of that parameter – we merely require that it is negative (see Fig. 12). Thus, for each point in the grid, we can say how many constraints that model violates: 0, 1, 2, or 3.

In Fig. 13, we project these scores, again in a maximally conservative way. For each point in the 2D projection, we look up all models in the full 6D space which have the two values under consideration in our projection, and we find the model which violated the fewest constraints. Thus, the figure shows the minimum number of constraints violated by *any* model with that combination of values. If any model with those coordinates is allowed by the constraints, the pixel is shown in blue. Each darker shade of red means every model with those coordinates violated at least one more constraint, up to all three.

This exercise demonstrates that even this conservative interpretation of the observed scatters as upper limits, combined with the

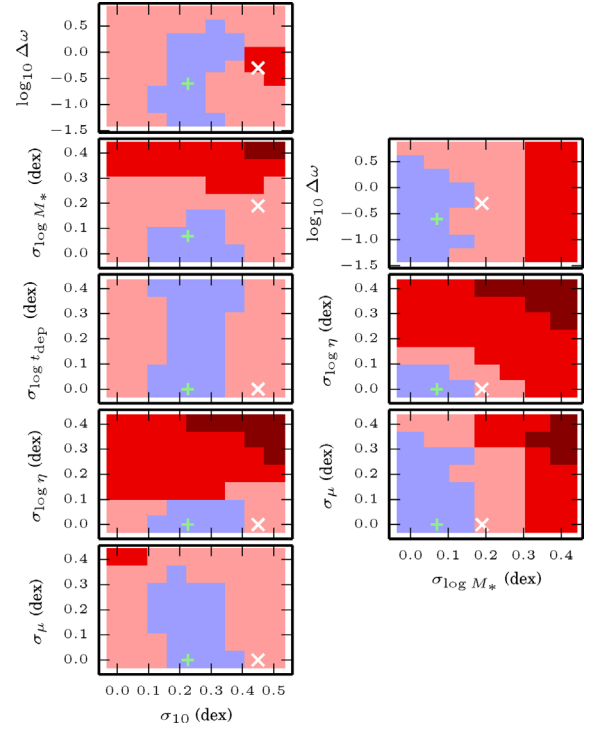


Figure 13. The set of parameters conservatively allowed by the observations. In addition to the parameters of the accretion process (σ_{10} and $\Delta\omega$), we include a variable lognormal scatter in η , M_* , t_{dep} , and the median accretion rate e^μ . These lognormal scatters have medians equal to the values used in Section 4.2 to fit the first-order relations. Blue pixels indicate that at least one model with that pair of parameters is consistent with the data. Each darker shade of red means the model which violates the fewest constraints for that pair of parameters violates one more constraint, up to all three. The white cross shows our initial guess, while the green ‘+’ shows our ‘improved guess’.

weak requirement that Z decrease with increasing SFR at fixed stellar mass, yields non-trivial constraints on the parameters. In particular, $\sigma_{10} \lesssim 0.35$ dex, smaller than that in our fiducial model. This may point to a smoothing out of the baryonic accretion rate relative to the scatter in dark matter accretion rates implied by the Neistein et al. (2010) formula. Perhaps even more interesting is that there is a minimum σ_{10} implied by our observational restrictions, $\sigma_{10} \gtrsim 0.1$ dex, which comes from the requirement that the FMR have a negative slope. In particular, if σ_{10} is too small, the subtle feature in the $M_*(M_h)$ relation causes both the SFR and metallicity of galaxies with $M_* \sim 10^9 M_\odot$ to be higher than the MS and MZR, and galaxies at other masses to be below those relations, generating a positive correlation between SFR and Z . However, other versions of the M_*-M_h relation do not necessarily include this feature.

The scatter in M_* for a given SGP must be $\lesssim 0.1$ dex. This is a bit at odds with the observational constraints from Reddick et al. (2013) (0.20 ± 0.03 dex at fixed maximum circular velocity), although we note that our constraint is on scatter in M_* that is *uncorrelated* with everything else in the SGP, whereas in reality it is quite plausible (and in fact predicted by the SGP – see Appendix A) that, at a fixed halo mass, M_* is correlated with both SFR and Z . This number can also be comparable to the 0.09 dex uncertainty in stellar mass reported by Tremonti et al. (2004), though again our constraint is only on the scatter in M_* that is uncorrelated with all other observables.

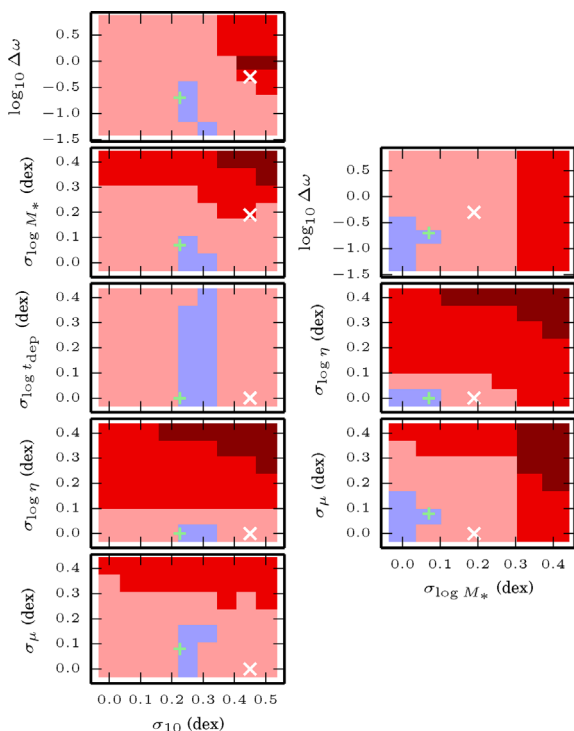


Figure 14. A plausible set of constraints on model parameters. Here we make the plausible but uncertain assumption that the FMR slope, $\partial \log Z / \partial \log \dot{M}_{\text{SF}}$, is between -2 and -1 , in addition to the constraints on the widths of the MS and MZR. As one might expect, narrowing the allowed range of FMR slopes dramatically reduces the allowed regions of parameter space. One should not take these regions to be genuine constraints, but rather to demonstrate the power of the slope of the FMR in constraining these parameters.

Another interesting constraint is that the scatter in η must be $\lesssim 0.1$ dex. This is surprisingly small, considering the great deal of theoretical uncertainty as to the actual values and scalings of η in the first place. In our models, this comes from the aforementioned effect that in the equilibrium relations, both Z and \dot{M}_{SF} are inversely related to η , so scatter in η tends to reverse the negative slope in the FMR.

Unsurprisingly, there is virtually no constraint on the scatter in t_{dep} . This is simply because the equilibrium relations for \dot{M}_{SF} and Z are independent of t_{dep} – to constrain this scatter one would need constraints on the scatter in the $M_{\text{SF}} - M_{\text{g}}/M_{\text{SF}}$ relation, although if such galaxies also had SFR measurements, they would have directly measured depletion times anyway. There is also relatively little constraint on scatter in the median accretion rate, e^{μ} . Essentially this is because, in the equilibrium relations, only the SFR is affected by this scatter, so as long as the scatter in e^{μ} is smaller than the scatter in the MS, there is no problem.

With these constraints in mind, we have altered our initial guess, simply by reducing σ_{10} , $\Delta\omega$, and $\sigma_{\log_{10} M_{\text{SF}}}$, and slightly increasing σ_{μ} (see Table 2). We label this model the ‘improved guess’ model. Purely for demonstration, we have also computed scores for the grid of models where we include not only an upper limit on the scatters, but also a much narrower range of acceptable slopes of the FMR, namely $-2 < \partial \log Z / \partial \log \dot{M}_{\text{SF}} < -1$. With these stronger restrictions, we get the projections shown in Fig. 14. Our ‘improved guess’ model is engineered to adhere to this much stronger constraint, though there are plenty of models which would be ruled out by this strict scoring that are still consistent with the observa-

tions. Unsurprisingly, the stronger constraint dramatically narrows the range of acceptable models in most of the projections. Particularly striking is that the allowed range of $\Delta\omega \sim \Delta z$, the interval in redshift over which galaxies have constant accretion rates in our model, is narrowed substantially from $\Delta\omega \lesssim 3$ to $\Delta\omega \lesssim 0.4$.

5.2 Domain of applicability

Under what circumstances might a real population of galaxies be in statistical equilibrium? We know that for a constant accretion rate, the SFR and the metallicity will equilibrate on the mass-loss time-scale. A standard equilibrium model therefore requires that t_{loss} be much less than the time-scale on which any parameter entering into the equilibrium relations, namely q (i.e. η , ξ , and f_{R}), Z_{IGM} , t_{dep} , and \dot{M}_{ext} , changes. The success of these models in understanding the first-order trends in galaxy scaling relations suggests that these requirements, while seemingly numerous, are at least marginally satisfied.

Our statistical equilibrium model relaxes one of these restrictions by splitting \dot{M}_{ext} into a (hopefully) slowly evolving median e^{μ} and a (potentially) rapidly varying stochastic component $e^{\sigma x(t)}$. Our formulation of this component introduces two time-scales, $t_{\text{coherence}}$ – the time between new draws from the lognormal distribution, and σt_{loss} – the time for a 1σ accretion event to be forgotten by the galaxy.

Figs 2 and 5 show graphically the exponential suppression of old draws of the accretion distribution in their influence on the full distribution of Ψ and Z^{\dagger} . In logarithmic space, the separation between the centres of the distribution of each draw is just $\tau_c = t_{\text{coherence}}/t_{\text{loss}}$, while the width of each distribution is σ . When $\sigma \lesssim \tau_c$, the distributions are well separated, and we conclude that galaxies may be in statistical equilibrium so long as t_{loss} is appreciably less than the time-scale on which any parameters of the SGP change (explicitly μ , σ , $t_{\text{coherence}}$, t_{loss} , Z_{IGM} , and q). Note that $t_{\text{coherence}}$ itself may well be shorter than or comparable to t_{loss} .

When $\sigma \gtrsim \tau_c$, the contributions from previous draws begin to matter significantly for the distribution of Ψ . In this case, the number of draws which are important increases from ~ 1 to $\sim \sigma/\tau_c$, so rather than t_{loss} being short, we need σt_{loss} to be short. At least in our initial guess, shown in the previous section, this is a minor effect since $\sigma \sim 1$. Therefore, the region of $M_{\text{h}}-z$ space where the statistical equilibrium model is valid should be comparable to the region where an ordinary equilibrium model is capable of reproducing the first-order galactic scaling relations, which in turn is set by the scaling of t_{loss} with halo mass and redshift.

In our fiducial model presented in the previous section, star-forming galaxies at every halo mass were in equilibrium at $z = 0$. As we mentioned, this is not necessarily the case in the real Universe – the mass loading factor may well scale weakly with halo mass, in which case low-mass galaxies, with their long depletion times, would be unable to equilibrate to their baryonic accretion rate even in a Hubble time. Either scenario is currently perfectly consistent with observations, since the mass-loss time-scale is unknown, owing to its dependence on η . In Fig. 15, we show regions of $M_{\text{h}}-z$ space where the equilibrium assumptions are valid – the bluer the colour, the better satisfied the condition that the Hubble time be much longer than t_{eq} , the maximum of the mass-loss time (t_{loss}), the coherence time ($t_{\text{coherence}}$), and the burn-through time (σt_{loss}).

To construct these diagrams, we also need to make an assumption regarding how input parameters vary not only with halo mass, but with redshift. We assume that the mass loading factor and the factor by which we reduce the efficiency, f_{ϵ} , are independent of

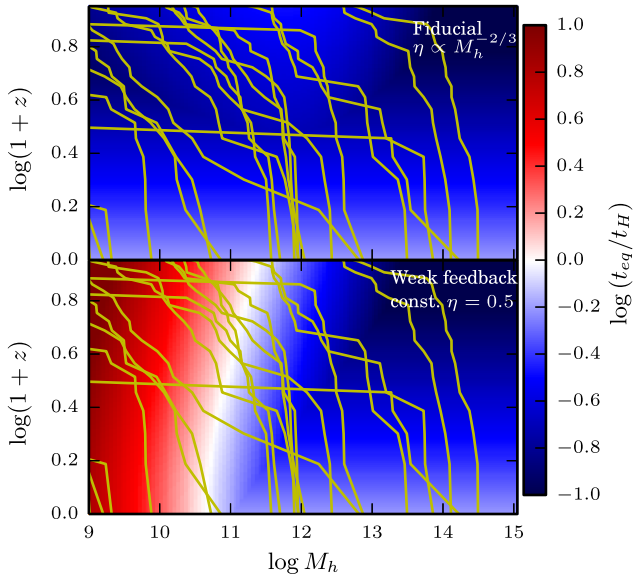


Figure 15. The validity of the statistical equilibrium model. Here we show the ratio of t_{eq} to the instantaneous Hubble time, as a function of halo mass and redshift. The bluer the colour, the more satisfied the condition that $t_{\text{eq}} < t_H$, required for the statistical equilibrium model to be valid. We show two cases for the scaling of the mass loading factor with halo mass: $\eta \propto M_h^{-2/3}$, which we use in our fiducial model, and $\eta = 0.5$. Both are consistent with direct observations of η . For reference, we overplot the trajectories of 20 haloes as they grow stochastically, as calculated by equations (25)–(27) with $\Delta\omega = 0.5$. With a strong scaling for η in halo mass and t_{dep} in redshift, the model is reasonable for haloes probed in galaxy surveys, but it is certainly plausible that low-mass galaxies are not in equilibrium.

redshift, but that the depletion time-scales as $t_{\text{dep}} \propto (1+z)^{-1}$, a somewhat weaker scaling than if the depletion time scaled with the dynamical time (Davé et al. 2012). This scaling is consistent with recent observational results from CO observations at high redshift (Tacconi et al. 2010, 2013; Saintonge et al. 2013), though of course there are large uncertainties. Moreover, these observations span a very limited range of mass and redshift compared to that shown in these diagrams. We therefore emphasize that these plots represent plausible assumptions, not predictions.

5.3 Evolution with redshift

Since it is plausible that a statistical equilibrium model may be used successfully at higher redshift, one may consider extending our analysis beyond the $z \approx 0.1$ data we have considered. There are however both theoretical and observational challenges. First, a great deal of poorly justified assumptions are required to scale the fiducial model to higher redshift, including the strong scaling of depletion time with redshift discussed in the previous section. Observationally, while there are measurements of the MS, MZR, and FMR at higher redshift, the metallicity measurements in particular are fraught with complications arising from the changing set of lines visible from ground-based telescopes and the uncertainty of converting line characteristics to physical metallicities in the substantially different environments of high-redshift galaxies (see Kewley et al. 2013a,b). The high-order quantities we discuss in this paper are therefore both difficult to predict and measure.

We can none the less make the basic point under the assumption that the SFR and metallicity have reached their equilibrium values,

the metallicity will be $Z_{\text{eq}} = Z_{\text{IGM}} + q$, where q is a combination of f_R , η , and ξ , none of which are expected to change dramatically with redshift. Thus, the MZR should not change with redshift if galaxies are in equilibrium. Observational studies tend to find at least some evolution (e.g. Mannucci et al. 2009; Yuan, Kewley & Richard 2013). However, given the difficulty in calibrating the zero-point of the MZR even at low redshift, and the possibility that samples of high-redshift galaxies are biased high in SFR and therefore low in Z (Stott et al. 2013), we do not believe that equilibrium models have been ruled out at higher redshifts.

Similarly, the slope of the FMR, which one may read off from the quadratic fit to the FMR (equation 35),

$$\partial \log Z / \partial \log \dot{M}_{\text{SF}} = a_2 + a_4 m + 2a_5 s \quad (36)$$

should be independent of the SFR. We therefore predict that in general $a_5 \approx 0$, although since m and s are highly correlated, the fact that many best-fitting models do not yield $a_5 = 0$ is not necessarily an indication that this prediction is wrong.

We can also describe how the fiducial model, which does not quite fit the data at $z = 0$, would scale to higher redshift. The predicted value of σ is not dependent on redshift, at least for the dark matter. Given our results suggesting that baryonic processes likely play an important role in smoothing the accretion, it is unclear how this smoothing process would evolve, so this constant value of σ is highly uncertain. Meanwhile $t_{\text{coherence}}$ will evolve strongly with redshift, but the quantity which sets the second-order scatter we consider is $\tau_c = t_{\text{coherence}}/t_{\text{loss}}$. If we assume that η at fixed halo mass does not change, then t_{loss} simply scales as the depletion time, which likely does decrease significantly towards higher redshift. If they decrease at a similar rate, near $1/(1+z)$, then τ_c is unlikely to evolve very much. In this scenario, we would predict that the scatter in the MS and MZR, and the slope in the FMR at fixed stellar mass would be roughly independent of redshift (and as evidenced in Fig. 10, independent of halo mass).

5.4 Relationship to other work

Our model bears a resemblance to several recent papers on equilibrium models (Davé et al. 2012; Lilly et al. 2013). Our model reduces to a slightly simpler version of the Davé et al. (2012) model when $\tau_c \rightarrow \infty$ and $\sigma \rightarrow 0$, i.e. the upper left of the $\sigma_{10}-\tau_c$ diagrams we showed in Sections 2 and 3.

One of the criticisms of the Davé et al. (2012) model has been its explicit assumption that $dM_g/dt = 0$. Simple toy models of the growth of galaxies under various star formation laws, (e.g. Feldmann 2013), point out that for many galaxies at redshift zero, $dM_g/dt < 0$. Indeed, in our recent work on the radially resolved evolution of disc galaxies since $z = 2$ (Forbes et al. 2014), we find that much of the galactic disc for many galaxies tends to be moderately out of equilibrium between local sources and sinks, with star formation being somewhat higher than the (local) accretion rate.

The equilibrium model of Lilly et al. (2013) attempts to address this issue by allowing part of the incoming accretion to build up in the gas reservoir. The price they pay is that the SFR becomes an input to their model rather than an output. Perhaps even more worrying is that they assume $dZ/dt = 0$ always, even when $dM_g/dt \neq 0$. Note that Feldmann (2013) makes a similar argument. As shown by our sample trajectories in Section 3, it actually takes Z longer to equilibrate than the SFR, since the metallicity can only equilibrate once the SFR catches up (or falls off) to the accretion rate. It is therefore odd that they assume that the metallicity is in equilibrium while the SFR is not. It is only through this oddity that they are

able to fit a second-order relation, i.e. the FMR, with their model. We consider our model to be both more self-consistent and more powerful, in that we can generate a scatter in the SFR and metallicity, and about the FMR itself. Pipino, Lilly & Carollo (2014) have recently examined the validity of the equilibrium approximations made in these models and the resulting relationship between specific SFR and metallicity, and find reasonable agreement between their numerical model and Lilly et al. (2013)'s approximation.

Another important result, and to our knowledge the only previous theoretical attempt to address the scatter in the MS, is Dutton et al. (2010). They use a rather sophisticated semi-analytic model, including cooling from virial shock-heated gas in a dark matter halo and star formation as a function of specific angular momentum (i.e. radius) in the disc, although they do not include any way for the gas to change its angular momentum (as we do in Forbes, Krumholz & Burkert 2012; Forbes et al. 2014). They find a significant but small scatter in their model star-forming MS arising from variation in halo concentration, which in turn causes differences in the mass accretion histories between different galaxies of the same halo mass. Their model therefore resembles ours in the limit that $\tau_c \rightarrow \infty$, but $\sigma \neq 0$. Our model's more flexible treatment of the accretion process and other model parameters (e.g. the mass loading factor) gives us somewhat more insight on the problem of scatter not only in the MS but also in the MZR and FMR, although of course our model is far simpler in terms of its treatment of star formation, and we can make no predictions regarding other important and interesting quantities (Dutton 2012, i.e. galaxy sizes and rotation curves).

6 CONCLUSION

In the past few years, a new view has emerged as a useful way of understanding galaxies. In this picture, galaxies are in a slowly evolving equilibrium between accretion, star formation, and galactic winds regulated by the mass of cold gas in their ISM. To the degree that the parameters controlling this balance are well-defined functions of the mass of a galaxy and its redshift, this sort of model may be used to understand the connection between galaxy scaling relations and these physical parameters, which are not known from first principles.

In the spirit of these equilibrium models, we have presented a simple model which relaxes a key assumption in the equilibrium model, namely that the rate at which baryons enter the gas reservoir varies slowly. A population of galaxies in our model has been fed by the same stochastic accretion process for eternity, or at least long enough that the full joint distribution of all galaxy properties has become time invariant. We therefore refer to our picture as a statistical equilibrium model, since the individual galaxies are not in equilibrium, but the population is.

With this model, we study a number of second-order relationships about the well-known galaxy scaling relations between the stellar mass and the SFR (the star-forming MS), and the stellar mass and metallicity (the MZR). We look at the scatter at fixed stellar mass in both of these quantities, as well as the (anti)correlation between star formation and metallicity at fixed stellar mass. Our main conclusions are as follows.

(i) Including a stochastic scatter in the accretion rate at the level expected from N -body cosmological simulations naturally produces an anticorrelation between SFRs and metallicities at fixed stellar mass, as observed, and a scatter in both the star-forming MS and MZR somewhat larger than the observed scatters.

(ii) Neglecting the scatter in model parameters (i.e. the mass loading factor, the depletion time, the scatter in stellar mass at fixed halo mass, etc.), all second-order quantities (the scatter in the MS, the scatter in the MZR, and the slope in metallicity with respect to the SFR at fixed stellar mass) are determined by only two parameters: the scatter in the accretion rate and the ratio of the time-scale on which the accretion varies to the time-scale on which the galaxy loses gas mass.

(iii) Using a maximally conservative interpretation of the available data, we are able to constrain these two parameters as well as a number of additional parameters, namely the scatter in the mass loading factor at fixed halo mass and the uncorrelated scatter in M_* at fixed halo mass. We find that the lognormal scatter in the baryonic accretion rate σ_{10} is between about 0.1 and 0.4 dex, moderately smaller than what we would have predicted based on N -body simulations and assuming that the baryons follow the dark matter. This may point to some process in the haloes of galaxies which smooths out variations in the baryonic accretion rate, or a substantial amount of baryon cycling, which has the effect of averaging out the accretion rate over a longer time period. We find that the scatter in the mass loading factor at fixed halo mass is less than 0.1 dex, remarkably small considering the theoretical uncertainty in the details of the physics of feedback. Our constraint on the time-scale over which the accretion rate varies is much weaker, with the allowed time period exceeding a change in redshift of 1, but it could be narrowed considerably by stronger constraints on the FMR.

The statistical equilibrium framework we have presented here is a novel approach to understanding higher order properties of galaxy scaling relations analytically. This sort of analysis may be plausibly extended to other galaxy properties beyond what we have considered here, or to other astrophysical applications involving a self-regulated process being driven stochastically. For our particular application of this method, we hope that our analysis motivates new development in both theory and observation.

On the theory side, $t_{\text{coherence}}$ (or a more sophisticated quantity describing the time-scales on which accretion varies) and σ may be measured with some confidence in both dark-matter-only and baryonic cosmological simulations. Semi-analytic models may be altered to include the appropriate level of variability in baryonic accretion rates. Meanwhile, we have shown that observable quantities, e.g., the scatter in the star-forming MS, can provide significant constraints on properties of the baryonic accretion process and the galaxy-to-galaxy variability of the mass loading factor. Our inferences are, however, limited by our limited certainty on the intrinsic scatter in the scaling relations we have considered and the true parameters of the FMR. Pinning down these quantities observationally at a variety of masses and redshifts may substantially improve our understanding of the details of baryonic accretion and feedback.

ACKNOWLEDGEMENTS

JCF is supported by the National Science Foundation Graduate Research Fellowship under Grant Nos DGE0809125 and DGE1339067. MRK acknowledges support from the NSF through CAREER grant AST-0955300 and by NASA through AFTP grant NNX13AB84G. AB acknowledges support from the Cluster of Excellence 'Origin and Structure of the Universe'. AD was supported by ISF grant 24/12, by GIF grant G-1052-104.7/2009, by a DIP grant, and by the I-CORE programme of the PBC and the ISF grant 1829/12. The Monte Carlo simulations detailed in Appendix A were carried out on the UCSC supercomputer Hyades, which is supported

by NSF grant AST-1229745. We would like to thank Erica Nelson, Pieter van Dokkum, Peter Behroozi, and Eyal Neistein for helpful conversations, and the anonymous referee for a helpful report.

REFERENCES

- Behroozi P. S., Wechsler R. H., Conroy C., 2013a, *ApJ*, 762, L31
 Behroozi P. S., Wechsler R. H., Conroy C., 2013b, *ApJ*, 770, 57
 Bothwell M. S., Maiolino R., Kennicutt R., Cresci G., Mannucci F., Marconi A., Ciccone C., 2013, *MNRAS*, 433, 1425
 Bouché N. et al., 2010, *ApJ*, 718, 1001
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
 Cacciato M., Dekel A., Genel S., 2012, *MNRAS*, 421, 818
 Chabrier G., 2003, *PASP*, 115, 763
 Chabrier G., 2005, in Corbelli E., Palte F., eds, *Astrophysics and Space Science Library*, Vol. 327, *The Initial Mass Function 50 Years Later*. Springer, Dordrecht, p. 41
 Cheung E. et al., 2012, *ApJ*, 760, 131
 Daddi E. et al., 2007, *ApJ*, 670, 156
 Davé R., Finlator K., Oppenheimer B. D., 2012, *MNRAS*, 421, 98
 Dekel A., Mandelker N., 2014, preprint ([arXiv:1402.2283](https://arxiv.org/abs/1402.2283))
 Dekel A., Sari R., Ceverino D., 2009, *ApJ*, 703, 785
 Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, *MNRAS*, 435, 999
 Dutton A. A., 2012, *MNRAS*, 424, 3123
 Dutton A. A., van den Bosch F. C., Dekel A., 2010, *MNRAS*, 405, 1690
 Elbaz D. et al., 2007, *A&A*, 468, 33
 Faucher-Giguère C., Kereš D., Ma C., 2011, *MNRAS*, 417, 2982
 Feldmann R., 2013, *MNRAS*, 433, 1910
 Fenton L., 1960, *IRE Trans. Commun. Syst.*, 8, 57
 Forbes J., Krumholz M., Burkert A., 2012, *ApJ*, 754, 48
 Forbes J. C., Krumholz M. R., Burkert A., Dekel A., 2014, *MNRAS*, 438, 1552
 Garcia-Appadoo D. A., West A. A., Dalcanton J. J., Cortese L., Disney M. J., 2009, *MNRAS*, 394, 340
 Genel S., Dekel A., Cacciato M., 2012, *MNRAS*, 425, 788
 Guo K., Zheng X. Z., Fu H., 2013, *ApJ*, 778, 23
 Kauffmann G. et al., 2003, *MNRAS*, 341, 33
 Kewley L. J., Ellison S. L., 2008, *ApJ*, 681, 1183
 Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013a, *ApJ*, 774, L10
 Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013b, *ApJ*, 774, 100
 Krumholz M. R., Dekel A., 2012, *ApJ*, 753, 16
 Kuhlen M., Krumholz M. R., Madau P., Smith B. D., Wise J., 2012, *ApJ*, 749, 36
 Kuhlen M., Madau P., Krumholz M. R., 2013, *ApJ*, 776, 34
 Lara-López M. A., López-Sánchez A. R., Hopkins A. M., 2013, *ApJ*, 764, 178
 Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, *ApJ*, 136, 2782
 Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, *ApJ*, 772, 119
 Maeder A., 1992, *A&A*, 264, 105
 Mannucci F. et al., 2009, *MNRAS*, 398, 1915
 Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, 408, 2115
 Mannucci F., Salvaterra R., Campisi M. A., 2011, *MNRAS*, 414, 1263
 McGaugh S. S., 2005, *ApJ*, 632, 859
 Murray N., Quataert E., Thompson T. A., 2005, *ApJ*, 618, 569
 Neistein E., Dekel A., 2008, *MNRAS*, 383, 615
 Neistein E., Macciò A. V., Dekel A., 2010, *MNRAS*, 403, 984
 Noeske K. G. et al., 2007, *ApJ*, 660, L43
 Peeples M. S., Shankar F., 2011, *MNRAS*, 417, 2962
 Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, L59
 Pipino A., Lilly S. J., Carollo C. M., 2014, *MNRAS*, 441, 1444
 Press W. H., Schechter P., 1974, *ApJ*, 187, 425
 Recchi S., Spitoni E., Matteucci F., Lanfranchi G. A., 2008, *A&A*, 489, 555
 Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
 Rodighiero G. et al., 2011, *ApJ*, 739, L40
 Saintonge A. et al., 2011, *MNRAS*, 415, 61
 Saintonge A. et al., 2013, *ApJ*, 778, 2
 Schiminovich D. et al., 2010, *MNRAS*, 408, 919
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
 Stott J. P. et al., 2013, *MNRAS*, 436, 1130
 Tacconi L. J. et al., 2010, *Nature*, 463, 781
 Tacconi L. J. et al., 2013, *ApJ*, 768, 74
 Tinsley B. M., 1980, *Fundam. Cosm. Phys.*, 5, 287
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031
 Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, *ApJ*, 754, L29
 Yates R. M., Kauffmann G., Guo Q., 2012, *MNRAS*, 422, 215
 Yuan T.-T., Kewley L. J., Richard J., 2013, *ApJ*, 763, 9
 Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, *ApJ*, 771, L19

APPENDIX A: DETAILS OF THE MONTE CARLO SIMULATIONS

Throughout this paper, we have presented heatmaps of various quantities as a function of $\tau_c = t_{\text{coherence}}/t_{\text{loss}}$, and $\sigma_{10} = \sigma \log_{10}(e)$. Computing each of these quantities for the model is typically a non-trivial task which requires a Monte Carlo simulation, in which a large ensemble of galaxies are sampled at random times to sample the underlying true distribution of the quantity in question for galaxies in this model. Here we describe the details of these simulations.

For each pixel in these grids of τ_c versus σ_{10} , we sample an ensemble of 30 000 galaxies. Each galaxy is started at a time $\tau = 0$ with initial values $\Psi = Z^\dagger = 1$, the equilibrium values for those quantities in the limit $\sigma \rightarrow 0$ and $\tau_c \rightarrow \infty$. The galaxies are then evolved for long enough that, for all practical purposes, they forget their initial conditions (formally our model assumes that the galaxy population has been undergoing the same stochastic accretion process for eternity, but this is obviously impractical computationally). To determine ‘long enough’, we use the analytic results derived in Sections 2 and 3 which show that galaxies forget their initial conditions with an e-folding time of t_{loss} . We also note that for our distribution to represent the true long-term steady-state distribution, as discussed in Section 5.2, galaxies with large scatters in their accretion rate need to experience enough draws from the accretion rate distribution that even the tail of the probability distribution of past events has no influence on the present distribution.

We therefore define a time-scale $\tau_{\text{long}} = 1 + \tau_c + \sigma$, i.e. a time guaranteed to be of the order of the longest time-scale in the problem for any choice of τ_c and σ . We then calculate the number of draws from the accretion distribution necessary to simulate each galaxy out to $15\tau_{\text{long}}$, namely $k = 15\tau_{\text{long}}/\tau_c$. We then draw a pseudo-random number uniformly distributed between 0 and 1, and compute the galaxy’s properties (Ψ, Z^\dagger) at the time $\tau_{\text{obs}} = (k + r)\tau_c$, where r is the random number. This samples the full distribution of these quantities for the population of galaxies in steady state. In general, the computational cost is just proportional to k , since to compute Ψ and Z^\dagger , we must first draw k random numbers and compute the sequence Ψ_1, \dots, Ψ_k and $Z_1^\dagger, \dots, Z_k^\dagger$ before we can calculate those quantities at τ_{obs} . Thus, the models with short coherence times, i.e. $\tau_c \ll 1$, are the most expensive.

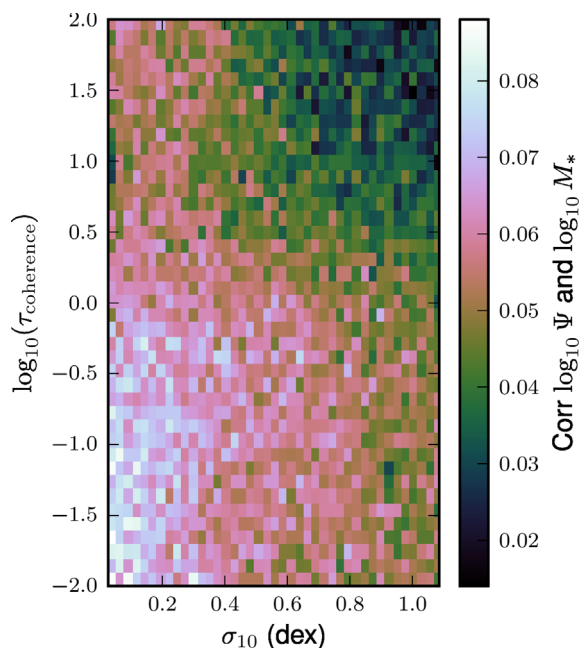


Figure A1. The correlation between SGP versions of the SFR and stellar mass. The non-zero correlation shows that scatter in stellar mass at fixed halo mass can drive galaxies along the MS, rather than being merely uncorrelated.

With the 30 000 samples for each of the 41 by 41 points in the grid of σ and τ_c , we can then compute each of the quantities shown in this paper – the standard deviation of Z^\dagger and Ψ (independently), the correlation between the two quantities, and the linear slope. We also record a quantity which may be regarded as a proxy for stellar mass, defined as

$$\mathcal{M}_* = \int_0^{\tau_{\text{obs}}} \Psi(\tau) d\tau. \quad (\text{A1})$$

Naturally the magnitude of this quantity is, on average, proportional to the amount of time we let the simulations run, which is chosen subjectively to be $\gtrsim 15\tau_{\text{long}}$. However, one may expect that any statistical properties which remove the mean may be physically relevant. In Figs A1 and A2, we show the correlation between our analogues to star formation and metallicity, and stellar mass.

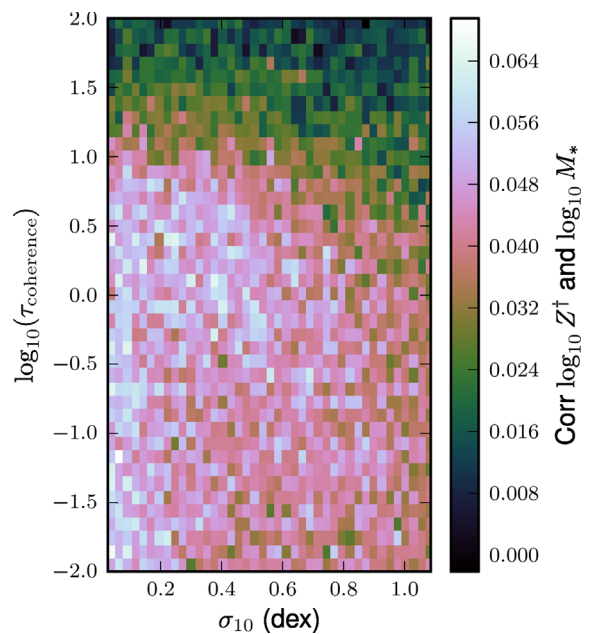


Figure A2. The correlation between SGP versions of the metallicity and the stellar mass. Just as with the SFR, the correlation is positive everywhere, meaning again that scatter in stellar mass at fixed halo mass scatters galaxies along the MZR.

Clearly, over much of parameter space there is a small but appreciable correlation between each quantity and \mathcal{M}_* , meaning that the tension between our result that $\sigma_{\log M_*} \lesssim 0.15$ dex and the observational constraint that the scatter in stellar mass at fixed halo mass be 0.19 dex (Reddick et al. 2013) is not a large concern. This is because our constraint is on scatter in M_* that is uncorrelated with any other quantity, whereas in reality, as in the SGP, the stellar mass may well be positively correlated with the quantities in question, in which case at least some of the scatter in M_* will be along the first-order scaling relations, and therefore would not contribute to the scatter in the relation at fixed stellar mass.

This paper has been typeset from a \LaTeX file prepared by the author.