

# Sensitivity analysis of a galaxy formation model

Piotr Oleśkiewicz<sup>1</sup>★ and Carlton M. Baugh<sup>1,2</sup>

<sup>1</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>2</sup>*Institute for Data Science, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted 2019 December 3. Received 2019 November 24; in original form 2019 October 4

## ABSTRACT

We present the first application of a variance-based sensitivity analysis (SA) to a model that aims to predict the evolution and properties of the whole galaxy population. SA is a well-established technique in other quantitative sciences, but is a relatively novel tool for the evaluation of astrophysical models. We perform a multiparameter exploration of the GALFORM semi-analytic galaxy formation model, to compute how sensitive the present-day *K*-band luminosity function is to varying different model parameters. The parameter space is scanned using a low-discrepancy sampling technique proposed by Saltelli. We first demonstrate the usefulness of the SA approach by varying just two model parameters, one that controls supernova feedback and the other the heating of gas by active galactic nucleus. The SA analysis matches our physical intuition regarding how these parameters affect the predictions for different parts of the galaxy luminosity function. We then use SA to compute Sobol' sensitivity indices varying seven model parameters, connecting the variance in the model output to the variance in the input parameters. The sensitivity is computed in luminosity bins, allowing us to probe the origin of the model predictions in detail. We discover that the SA correctly identifies the least important and most important parameters. Moreover, the SA also captures the combined responses of varying multiple parameters at the same time. Our study marks a much needed step away from the traditional 'one-at-a-time' parameter variation often used in this area and improves the transparency of multiparameter models of galaxy formation.

**Key words:** methods: numerical – galaxies: statistics.

## 1 INTRODUCTION

Galaxy formation is a complex process, which we are only now just starting to understand through a combination of observations, numerical simulations, and analytical modelling. Two main theoretical techniques are used to model the formation and evolution of galaxies: semi-analytical modelling (SAM) and hydrodynamic simulations (for a review, see Somerville & Davé 2015). SAMs use physically motivated, simplified mathematical relations to describe the evolution of baryons in growing dark matter haloes (Baugh 2006; Benson 2010). Hydrodynamic simulations, on the other hand, tend to make fewer assumptions and approximations than SAMs and solve the fluid equations governing the dynamics of baryons. Nevertheless, in hydrodynamic simulations many processes, such as star formation, remain 'subgrid' due to the finite numerical resolution of the simulation and our inability to write down the precise equations describing some processes (Crain et al. 2015; Ludlow, Schaye & Bower 2019). In the absence of a complete mathematical description, physical processes are described in both SAMs and hydrodynamic simulations by approximate equations

that contain parameters. Values have to be chosen for these parameters to specify a model. Here, we present a new application of an established statistical method to assess the impact of changes in model parameters on the output of a model.

The past few years have seen tremendous breakthroughs in the hydrodynamic simulation of galaxy formation for significant galaxy populations in cosmological volumes (Schaye et al. 2014; Vogelsberger et al. 2014; Pillepich et al. 2018). Nevertheless, SAMs remain an attractive and valuable complement to hydrodynamical simulations due to their flexibility and speed. These properties of SAMs mean that they can be used to build intuition about physical processes, by running thorough investigations of the impact of varying model parameters [e.g. see the comprehensive exploration of perturbations around the fiducial model presented by Lacey et al. (2016)]. Also, SAMs remain the method of choice to populate large-volume *N*-body simulations using a physical galaxy formation model: the fiducial simulation volumes used in SAMs are around 100 times bigger than those used in the current state-of-the-art hydrodynamical simulations. The predictions of SAMs have reached an impressive level of maturity through careful comparisons between the predictions of different groups and techniques (e.g. Contreras et al. 2013; Knebe et al. 2015; Guo et al. 2016; Mitchell et al. 2018).

\* E-mail: piotr.oleskiewicz@durham.ac.uk

Nevertheless, some scepticism remains regarding SAMs, much of which can be traced to the way in which the model parameters are set. Traditionally, models have been calibrated by developing physical intuition about how the model responds to changes in selected parameter values, such as those which control the mass loading of winds driven by supernovae, and then varying one parameter at a time to hone in on a best-fitting model. Often the quality of the model reproduction of the calibration data is judged by eye and compromises are made in order to match multiple data sets; these steps are hard to quantify and therefore difficult to reproduce. The 'best-fitting' model is reported as a single choice of parameter set that defines the model. The primary motivation for producing a single model is the desire to build mock catalogues for galaxy surveys (Baugh 2013). However, users often want to know the uncertainty on the model predictions and how the predictions respond to changes in the input parameters.

The range of processes modelled by SAMs lends them the flexibility to predict varied observation but at the cost of having to specify a number of parameters, which complicate model optimization or calibration. A number of techniques have been devised to reduce the complexity or dimensionality of the parameter space and to perform efficient searches of the parameter space: principal component analysis (hereafter PCA; Benson & Bower 2010), Bayesian emulators (Bower et al. 2010; Gómez et al. 2012), particle swarm optimizer (hereafter PSO; Ruiz et al. 2015), Markov Chain Monte Carlo (Henriques et al. 2009; Lu et al. 2011, 2012; Henriques et al. 2013; Mutch, Poole & Croton 2013; Martindale et al. 2017), and Latin-hypercube sampling (Bower et al. 2010; Rodrigues, Vernon & Bower 2017).

Here, we apply sensitivity analysis (SA) to quantify the dependence of the model output on the variation in the values of the model input parameters. The analysis of Gómez et al. (2014) using the CHEMTREEN SAM of Tumlinson (2009) is similar in scope to our work. They use an analysis of variance (ANOVA) technique for variance decomposition instead of sensitivity indices, and Gaussian processes for model fitting. Here, we use the GALFORM SAM effectively as a black-box model, and evaluate the sensitivity of the model outputs to the variation of the input parameters. A SAM is an ideal candidate for sensitivity analysis, as the interactions between parameters are complex enough to develop a black-box-like behaviour (becomes easier to experiment with than to understand; Golovin et al. 2017); however, many parameters have a natural physical interpretation, and hence it will be straightforward to develop intuition about how sensitive the model outputs should be to changing the inputs. Many parameters also have either physically motivated bounds, or at least a plausible range of possible values.

A criticism often aimed at SAMs is that they contain too many free parameters. This is usually rebuffed with the insistence that the parameters are physical, not statistical. Model fitting alone is therefore insufficient for interpreting how well a SAM is performing. A different research question, one this study tries to address, is how sensitive the model is to the parameter variation – in other words, how well do we understand the impact of the physical processes and their interactions on the model predictions?

Sensitivity analysis (Fisher 1918; Sobol' 1993, 2001; Saltelli et al. 2010) is an area of statistical modelling that analyses how the variance of the output of a model is affected by variance in the model inputs. It is closely related to uncertainty analysis and model optimization, and can be used to test the robustness of the model predictions to uncertainty in the input parameters, quantify dependence of the outputs of a model on different parameters, identify model non-linearities, and guide subsequent model optimization.

This addresses a common criticism of black-box models, namely that after adding sufficiently many free parameters they can be fine-tuned to match any observations, and provide a single set of predictions. While model optimization can be used to compute confidence intervals, SA is uniquely positioned to quantify model responses and the relative importance of the inputs. This addresses the complaint about SAMs listed above, that providing a spread of model predictions is preferable to fitting to the observations. Using SA, we will be able to not only tell how much model predictions vary for individual outputs but also quantify how much of this variance can be attributed to individual model inputs (or their combinations).

There are several SA techniques, not all of which are suitable for analysing non-linear models with a high-dimensional parameter space. With a few exceptions,<sup>1</sup> SA is done in three stages:

- (i) sampling of the parameter space
- (ii) model evaluation in the parameter space
- (iii) computation of sensitivity indices

Here, we use a variance-based SA, which adopts the improvement introduced by Saltelli et al. (2017) over the Sobol' indices. Variance-based methods aim to decompose the variance of the model output into the contributions from individual parameter variances, as well as the combined variances of the interactions of multiple combinations of parameters changing at once. In order to avoid a computational penalty for evaluating all possible parameter combinations, input parameters are treated as probability distributions, and the sensitivity of the model output is estimated approximately. Moreover, a number of numerical optimizations have been introduced into the sampling and index calculation techniques, to improve the convergence of the indices and average over the values that are too difficult to compute efficiently.

This work diverges from previous studies in two important ways: first, we narrow the scope of this investigation to computing only sensitivity indices, and we do not attempt to provide the best-fitting values for a galaxy formation model. We believe that SA is not the best tool for this task, as it investigates model responses at the extreme values of input parameters, and often for unusual combinations of inputs, where the model no longer reproduces the observable values. Second, we do not limit ourselves to measuring responses of the model to individual parameters and their linear combinations. Instead, we use sensitivity indices to capture both individual and combined impacts of parameters. Lastly, this study focuses exclusively on one observable, the *K*-band luminosity function (LF), calculated using the GALFORM SAM, and probes how this specific model reacts to changes in the input parameters. Our scope is narrower, but also deeper than any previous study in this area.

The layout of this paper is as follows. In Section 2, we set out the theoretical background, introducing the GALFORM model and, for completeness, giving the equations for the processes that we vary (Section 2.1). We then discuss variance-based sensitivity analysis (Section 2.2), the concept of low-discrepancy sampling (Section 2.3), the exploration of parameter space using Saltelli sampling (Section 2.4), and we define the sensitivity indices (Section 2.5) and illustrate these ideas with a toy model (Section 2.6). Our results using GALFORM are presented in Section 3 and our conclusions are given in Section 4.

<sup>1</sup> Some methods, such as Gaussian processes, use parameter exploration to simultaneously measure model sensitivity and maximize goodness-of-fit for model output(s).

## 2 THEORETICAL BACKGROUND

Here, we set out the theoretical ideas used in the paper. Section 2.1 gives a brief overview of the GALFORM semi-analytical model, introducing the processes that are varied in the sensitivity analysis. Section 2.2 introduces variance-based sensitivity analysis, Section 2.3 discusses the sampling of a model parameter space, and Section 2.4 covers Saltelli sampling. Section 2.5 defines the sensitivity indices and Section 2.6 illustrates their use with a toy model. Section 2.7 discusses the use of GALFORM output in the sensitivity analysis.

### 2.1 GALFORM

GALFORM is a SAM that aims to predict the properties of galaxies starting from dark matter halo merger histories that are typically extracted from an N-body simulation (Cole et al. 2000; Baugh 2006; Bower et al. 2006; Lacey et al. 2016). GALFORM models the processes that shape the galaxy population using a set of physically motivated, non-linear differential equations that track the exchange of mass, energy, and angular momentum between the different components of a galaxy. The processes modelled are

- (i) the merger histories of dark matter haloes
- (ii) the heating and cooling of gas and the formation of galactic discs
- (iii) quiescent star formation in galactic discs
- (iv) bursts of star formation triggered by galaxy mergers or dynamically unstable discs
- (v) feedback driven by supernovae (SNe), which can eject cold gas from a galaxy
- (vi) heating by an active galactic nucleus (AGN), which can prevent gas cooling
- (vii) chemical enrichment of stars and gas

These processes are in many cases modelled by equations that contain parameters. A GALFORM model corresponds to a set of parameters whose values have been chosen so that the model reproduces a particular set of observations. Some of these parameters govern different choices for processes in the model, such as the radial density profile assumed for the hot gas within a halo or the stellar initial mass function (IMF) that describes the number of stars of different masses produced in episodes of star formation. For example, the Gonzalez-Perez et al. (2014) model assumes a universal, solar neighbourhood IMF whereas the Lacey et al. (2016) model invokes a top-heavy IMF in bursts of star formation and a solar neighbourhood IMF in quiescent star formation. Even though these two models are implemented in the same N-body simulation, the choices made regarding the IMF and the slightly different emphasis on which observations the model should reproduce most closely mean that there are several differences in the values of the parameters that define these galaxy formation models.

Here, we use the recalibration of the Gonzalez-Perez et al. (2014) model introduced by Baugh et al. (2019) for the Planck Millennium N-body simulation, which we refer to as GP14.PMILL. The Planck Millennium N-body simulation (hereafter the PMILL simulation) adopts the Planck cosmology (Planck Collaboration XVI et al. 2014; see Table 1) and has superior mass resolution and halo merger histories that are better sampled in time compared with earlier N-body simulations into which GALFORM was implemented (see Table 1). Below, we review the processes that we vary in the sensitivity analysis. A full description of GALFORM can be found in Lacey et al. (2016).

**Table 1.** Planck Collaboration XVI et al. (2014) cosmology used in the P-Millennium simulation; the last two rows give the simulation box length and the number of particles used.

Parameter	Value
$\Omega_\Lambda$	0.693
$\Omega_M$	0.307
$\Omega_{\text{baryon}}$	0.04825
$h$	0.6777
$\sigma_8$	0.8288
$n$	0.967
$L[h^{-1} \text{ Mpc}]$	542.16
$N_p$	5040 <sup>3</sup>

#### 2.1.1 Star formation rate

The GP14.PMILL model uses an empirically motivated star formation law that was introduced by Blitz & Rosolowsky (2006) and implemented in GALFORM by Lagos et al. (2011). The star formation rate is given by

$$\Sigma_{\text{SFR}} = \nu_{\text{SF}} \times f_{\text{mol}} \times \Sigma_{\text{gas}}, \quad (1)$$

where  $\Sigma_{\text{SFR}}$  is the star formation rate per unit area,  $\Sigma_{\text{gas}}$  is the surface density of gas,  $\nu_{\text{SF}}$  is the inverse of the star formation time-scale, and  $f_{\text{mol}}$  is the ratio of the surface densities of the molecular and total gas masses,  $\Sigma_{\text{mol}}/\Sigma_{\text{gas}}$ .

#### 2.1.2 Supernova feedback

Supernova feedback in GALFORM is modelled as a process that ejects cold gas from a galaxy to a reservoir of mass  $m_{\text{res}}$ , at a rate of

$$\dot{m}_{\text{out}} = \beta \psi, \quad (2)$$

where  $\psi$  is the star formation rate and  $\beta$  is a mass loading factor defined as

$$\beta = \left( \frac{V_c}{V_{\text{hot}}} \right)^{-\gamma_{\text{SN}}}. \quad (3)$$

Here,  $V_{\text{hot}}$  and  $\gamma_{\text{SN}}$  are model parameters and  $V_c$  is the effective circular velocity of the disc or bulge (for starbursts) at the half-mass radius. Note that these equations are applied to quiescent and burst star formation. Different values can be adopted for the  $V_{\text{hot}}$  parameters for the disc and burst contributions to star formation.

Gas is returned from this reservoir to the hot gas halo at the rate of

$$\dot{m}_{\text{ret}} = \alpha_{\text{ret}} \times \frac{m_{\text{res}}}{\tau_{\text{dyn}}}, \quad (4)$$

which is controlled by the free parameter  $\alpha_{\text{ret}}$ ;  $\tau_{\text{dyn}} = r_{\text{vir}}/V_{\text{vir}}$  is the dynamical time of the halo, where  $r_{\text{vir}}$  is the virial radius of the halo and  $V_{\text{vir}}$  is the effective circular velocity at this radius.

#### 2.1.3 AGN feedback

Supermassive black holes (SMBHs) grow in the centres of galaxies, and inject energy into the gas reservoir of a halo following accretion, which disrupts the cooling process (see Fanidakis et al. 2011 and Griffin et al. 2019 for descriptions of the treatment of SMBHs in GALFORM). In GALFORM AGN heating occurs when two conditions

**Table 2.** The GALFORM parameters analysed in this work. The parameter ranges have been taken from previous analyses (Bower et al. 2010; Rodrigues et al. 2017).

Process	Parameter	Min.	Max.
Star formation	$\nu_{\text{SF}}$ [ $\text{Gyr}^{-1}$ ]	0.2	1.2
Supernova feedback	$\gamma_{\text{SN}}$	1.0	4.0
	$\alpha_{\text{ret}}$	0.2	1.2
	$V_{\text{hot, disc}}$ [ $\text{km s}^{-1}$ ]	100	550
	$V_{\text{hot, burst}}$ [ $\text{km s}^{-1}$ ]	100	550
AGN feedback	$\alpha_{\text{cool}}$	0.2	1.2
Disc instabilities	$f_{\text{stab}}$	0.61	1.1

are met: (i) the hot gas halo is in quasi-hydrostatic equilibrium, defined in terms of the ratio of the cooling time,  $\tau_{\text{cool}}$ , to the free-fall time,  $\tau_{\text{ff}}$ :

$$\frac{\tau_{\text{cool}}(r_{\text{cool}})}{\tau_{\text{ff}}(r_{\text{cool}})} > \frac{1}{\alpha_{\text{cool}}}, \quad (5)$$

where  $\alpha_{\text{cool}}$  is a parameter, and (ii) the AGN power required to balance the radiative cooling luminosity  $L_{\text{cool}}$  is below a fraction  $f_{\text{Edd}}$  of the Eddington luminosity  $L_{\text{Edd}}$  of the SMBH of mass  $M_{\text{BH}}$ :

$$L_{\text{cool}} < f_{\text{Edd}} L_{\text{Edd}}(M_{\text{BH}}). \quad (6)$$

### 2.1.4 Disc instabilities

Galaxies can also undergo morphological transformations and starbursts as a result of disc instabilities. Galaxy discs that are dominated by rotational motions are unstable to bar formation when they are sufficiently self-gravitating. We assume that discs are dynamically unstable to bar formation if (Efstathiou, Lake & Negroponte 1982)

$$F_{\text{disc}} = \frac{V_c(r_{\text{disc}})}{\sqrt{1.68 G M_{\text{disc}}/r_{\text{disc}}}} < f_{\text{stab}}, \quad (7)$$

where  $M_{\text{disc}}$  is the total disc mass (i.e. stars plus cold gas),  $r_{\text{disc}}$  is the disc half-mass radius, and the factor 1.68 relates this to the exponential scale length of the disc.

The quantity  $F_{\text{disc}}$  measures the contribution of disc self-gravity to its circular velocity, with larger values corresponding to less self-gravity and so greater disc stability. Efstathiou et al. (1982) found a stability threshold  $F_{\text{stab}} \approx 1.1$  for a family of exponential stellar disc models. Note that a completely self-gravitating stellar disc would have  $F_{\text{disc}} = 0.61$ , which is therefore the minimum value allowed for this parameter.

### 2.1.5 Parameter selection

We consider the relative importance of the processes described in Sections 2.1.1–2.1.4 by performing an SA on the parameters that describe these phenomena. The parameters and the ranges over which they are varied are listed in Table 2. In some instances, the parameter range is reasonably well defined, such as  $f_{\text{stab}}$ , as discussed above in Section 2.1.4. In other cases, the choice of range of parameter values is less well defined. For example, using simple conservative arguments,  $\gamma_{\text{SN}}$  could take on values of 1 and 2 in the momentum and energy conserving phases of the wind evolution (Ostriker & McKee 1988; Lagos, Lacey & Baugh 2013). Numerical simulations of winds have suggested different values of  $\gamma_{\text{SN}}$ . The other parameters defining the GALFORM model beyond those listed in Table 2 are held fixed.

### 2.1.6 Model output

After the formation and evolution of galaxies are calculated over the merger history of the dark matter haloes in the PMILL simulation, galaxy luminosities can be obtained from the predicted star formation rate and metallicity of the stars produced using a stellar population synthesis model. Dust extinction is calculated in post-processing, based on the size and gas metallicity of each galaxy (Gonzalez-Perez et al. 2014; Lacey et al. 2016). The model output that we focus on here is the *K*-band luminosity function at  $z = 0$ .

## 2.2 Variance-based sensitivity analysis

The SA method we use here closely follows those used by Sobol' (2001) and Saltelli et al. (2017), which are designed to decompose variance in the model output into the variances of the input parameters and their interactions using as few model evaluations as possible.

Many SA approaches suffer from a number of shortcomings, which make them unsuitable for analysing non-linear models. By non-linear models, we mean here those that are characterized by interactions between the inputs<sup>2</sup> and which therefore cannot be analysed effectively using regression or one-at-a-time (OAT) parameter variation techniques (Morris 1991).

Unlike other methods, variance-based SA allows a full exploration of the input space, and therefore accounts for the interactions between parameters and non-linear responses of the model. It follows that variance-based methods are able to evaluate the total-effect indices (see below) and rank the parameters in order of their influence on the output (Chan, Saltelli & Tarantola 1997; Sobol' 2001; Saltelli et al. 2010).

Finally, we note that all SA methods assume that the model inputs are independent, which might not hold in general for complex models. For instance, correlations between inputs, or unphysical combinations of their values, cannot be recognized by SA techniques. Similarly, variance-based SA currently assumes that the model output is a scalar. This means that the model outputs are independent of one another; for example, in the case of the luminosity function, the model prediction in a luminosity bin is considered to be independent of the results in other bins and from other outputs, e.g. other galaxy statistics. Even if the output of the model is multidimensional, and even if it is correlated across one or more dimensions,<sup>3</sup> each of the outputs must be analysed in isolation from the others. Unfortunately, at the time of writing, there are *no* well-established techniques that quantify or alleviate these two shortcomings. However, these limitations do not apply to GALFORM: the input parameters can all be varied independently and freely across the entire parameter space, and our outputs will be quantized and analysed independently.

## 2.3 Sampling parameter space

Sampling the high-dimensional parameter space of a complex model requires a trade-off between the accuracy of the sampling and

<sup>2</sup>Interactions between inputs occur, for example, when varying two or more input parameters produces a significantly different response from the model than would be expected from summing the change produced by varying the parameters independently.

<sup>3</sup>We know that this is the case in galaxy formation models because if the luminosity function changes in a given bin, this will lead, for example, to a change in the luminosity–circular velocity relation. Benson (2014) argued that correlations between bins in the observed luminosity function are important in setting model parameters.

computational expense. The accuracy of the sampling describes how well the space is probed – have any potentially interesting regions of the parameter space been overlooked because too few points have been sampled or because the method used has left gaps in the space?

The accuracy of a sampling scheme can be assessed formally in terms of its ‘discrepancy’. The lowest discrepancy sampling possible is a regular grid. However, this is subject to ‘aliasing’ or a lack of resolution due to the fixed gaps in the parameter space between the model evaluations; interesting model behaviour could be hidden in the unsampled parts of the parameter space. The convergence of the exploration of the parameter space is slow with a regular grid. The aliasing can be reduced and the convergence rate sped up by using a random sequence to sample the parameter space, which leads to a higher density of sampling in some parts of parameter space compared to a regular grid. The drawback in this case is that some regions of the parameter space will be more sparsely sampled than they were using a regular grid. A random sequence is formally described as the highest discrepancy sampling. Ideally, for a fixed number of sampling points, we want to strike a balance between avoiding the regular sampling achieved using a grid and leaving big gaps unsampled in the parameter space, as happens with random sampling.

Several quasi-random techniques have been proposed to generate sequences that approach this ideal of ‘low-discrepancy’ sampling, and which also ensure fast convergence of the uncertainties in the sensitivity indices. A quasi-random sequence is one designed to generate points in  $d$ -dimensions, which appear random but which are generated deterministically to have certain desired properties. Unlike pseudo-random and truly random sequences, successive points in a quasi-random sequence fill the gaps left by the previous points in the parameter space. The ‘random’ part of the name is technically a misnomer, as the sequence is fully deterministic, but yields a uniform distribution when projected on to any dimension of the parameter space.

A quasi-random sequence can be designed to minimize its discrepancy. For a low-discrepancy sequence, all of its subsequences also have low discrepancy. If a given sequence is uniform, its discrepancy tends to be zero as its length increases. For these reasons, quasi-random low-discrepancy sequences are used to maintain a balance between rapid convergence of numerical algorithms, a thorough coverage of the parameter space, and a high uniformity of a resulting sample along all dimensions of the parameter space (Press et al. 2007, section 7.8). Quasi-random sequences are therefore an attractive replacement for pseudo-random sequences in many applications that require a high-quality sampling.

Sampling based on low-discrepancy sequences, such as the recurrent additive sequence (Ulam 1960), Halton sequence (Halton 1964), Latin hypercube (Stein 1987), or Sobol’ sequence (Sobol’ 1967; Levitan et al. 1988), can be used in numerical integration and model optimization and have been shown to outperform schemes based on truly random, or pseudo-random number generators, while achieving significantly faster convergence rates (Sobol’ 1993). The advantage of these sequences over truly random and pseudo-random sequences can be attributed to the fact that the low-discrepancy property guarantees gapless sampling over the entire parameter space.

The low-discrepancy quasi-random sequence typically used in SA is the Sobol’ sequence (Sobol’ 1967). It can be efficiently calculated, and produces a sample that quickly converges to the correct set of sensitivity indices, as verified by checking against analytically calculated values for test models. Even though it is impossible to estimate the required number of model evaluations

prior to running the SA, there exists a natural convergence criterion – the sum of the first-order indices, defined in equation (11), has to add up to unity. Moreover, even if the SA did not converge after the initial run, additional evaluations can be easily added (see the example in the next subsection).

## 2.4 Saltelli sequence sampling

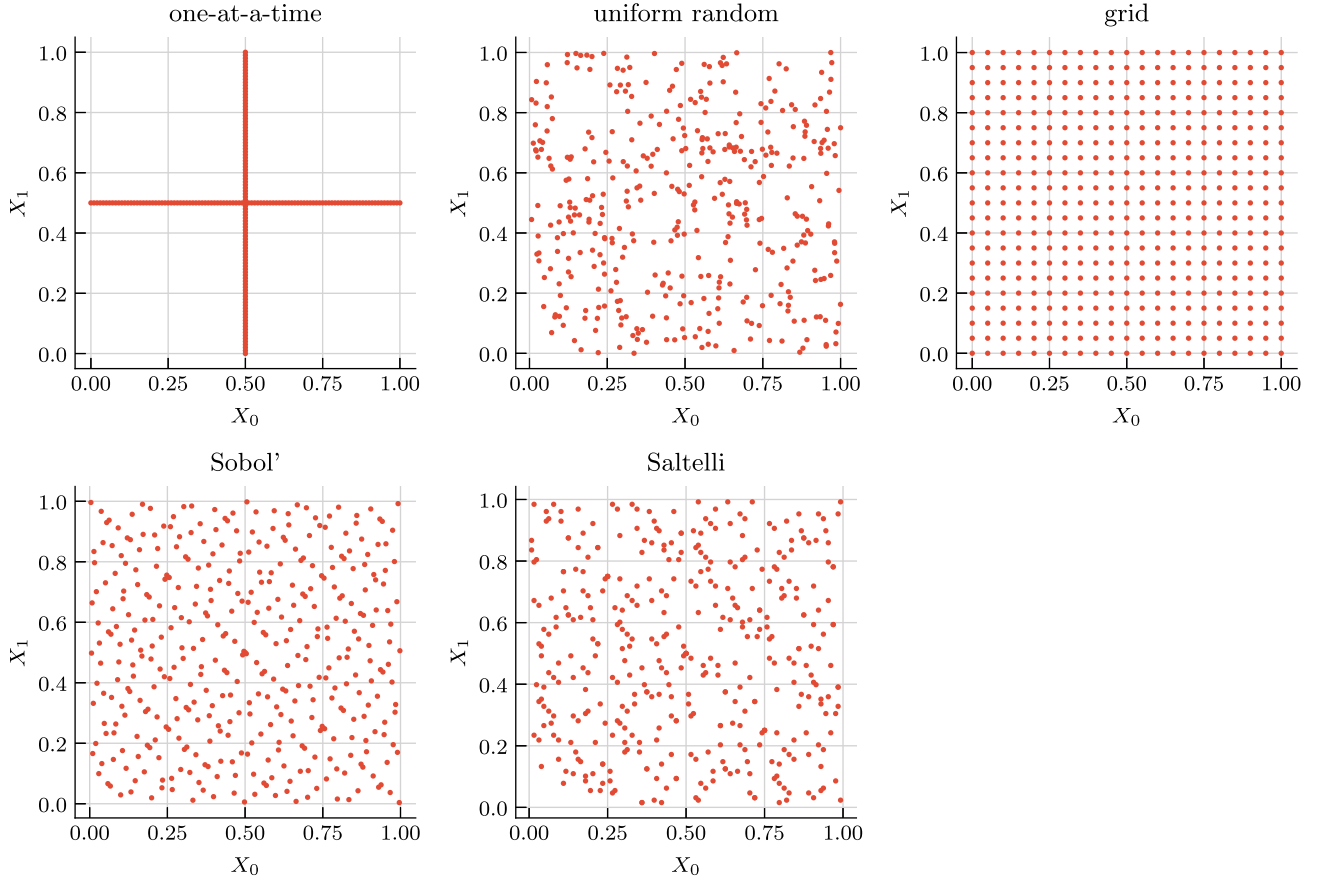
The Sobol’ sequence was originally proposed as a method of improving the convergence of numerical integration (Sobol’ 1967). Antonov & Saleev (1979) developed an efficient computational method to implement Sobol’ sampling. Saltelli et al. (2010) combined multiple Sobol’ sequences to further reduce the number of points required for the estimation of the sensitivity indices, improving the convergence rate.

Hereafter, we refer to the Sobol’ sequence as an  $N$  by  $d$  matrix, where  $N$  is the number of points of a  $d$  dimensional parameter space.

The Saltelli sequence is obtained as follows: first, we generate an  $N$  by  $2d$  Sobol’ sequence, (as demonstrated for the case of  $N = 4$ ,  $d = 3$  in the first line of equation 8). Let the first  $d$  columns be called submatrix **A** (blue), and the last  $d$  submatrix **B** (red), color text is available online. The values in the matrices indicate the locations in parameter space at which the model is to be evaluated, for parameters which can take on values over the range 0 to 1. We next construct a number  $d$  of  $N$  by  $d$  matrices  $\mathbf{A}_B^{(i)}$ , for  $i \in \{1, 2, \dots, d\}$ , such that for each  $\mathbf{A}_B^{(i)}$  the  $i^{\text{th}}$  column is taken from matrix **B**, while the remaining columns come from matrix **A**. The matrices **A**, **B**, and  $\mathbf{A}_B^{(i)}$  specify all the points of the parameter space at which the model is to be evaluated (one point per row), giving a total of  $N \times (2 + d)$  evaluations that are required to calculate the first-order sensitivity indices.

$$\begin{aligned} \text{Sobol}(4, 3) &= \begin{bmatrix} 0.500 & 0.500 & 0.500 & 0.500 & 0.500 & 0.500 \\ 0.250 & 0.750 & 0.250 & 0.750 & 0.250 & 0.750 \\ 0.750 & 0.250 & 0.750 & 0.250 & 0.750 & 0.250 \\ 0.125 & 0.625 & 0.875 & 0.875 & 0.625 & 0.125 \end{bmatrix} \\ \mathbf{A}_B^{(1)} &= \begin{bmatrix} 0.500 & 0.500 & 0.500 \\ 0.750 & 0.750 & 0.250 \\ 0.250 & 0.250 & 0.750 \\ 0.875 & 0.625 & 0.875 \end{bmatrix} \\ \mathbf{A}_B^{(2)} &= \begin{bmatrix} 0.500 & 0.500 & 0.500 \\ 0.250 & 0.250 & 0.250 \\ 0.750 & 0.750 & 0.750 \\ 0.125 & 0.625 & 0.875 \end{bmatrix} \\ \mathbf{A}_B^{(3)} &= \begin{bmatrix} 0.500 & 0.500 & 0.500 \\ 0.250 & 0.750 & 0.750 \\ 0.750 & 0.250 & 0.250 \\ 0.125 & 0.625 & 0.125 \end{bmatrix} \end{aligned} \quad (8)$$

A visual impression of the different sampling approaches is given by Fig. 1, which shows five commonly used types of sampling: OAT, uniform pseudo-random number generator, uniform grid sampling, a two-dimensional Sobol’ sequence, and Saltelli sampling. The OAT approach is often used with far fewer evaluations than shown here, which makes it computationally cheaper than the other approaches. The drawback of this method is clear from the vast areas of the parameter space that are left unexplored. This problem is only exacerbated on increasing the dimensionality of the parameter space. The pseudo-random number generation suffers from poor convergence, as randomness often results in over- and undersam-



**Figure 1.** Comparison of selected parameter space sampling strategies. Each panel contains 400 points sampled between  $[0,1]$  in the  $X_0$  and  $X_1$  dimensions using different methods, as labelled in each panel.

pling of many regions. The Sobol' and Saltelli sequences uniformly sample the parameter space and achieve the low-discrepancy target at a reasonable computational cost.

## 2.5 Sensitivity indices

Given a scalar model  $Y$  with independent inputs, we can define the first-order effect of the variance in the input  $X_i$  as

$$E_i = E_{\mathbf{X}_{-i}}(Y|X_i) = \int Y(X_i) \text{pdf}(X_i) \prod_{i \neq j} dX_j \quad (9)$$

$$V_i = \text{Var}_{X_i}(E_i) = \int (E_i - E(Y))^2 \text{pdf}(X_i) dX_i, \quad (10)$$

where  $X_i$  is  $i^{\text{th}}$  model input,  $V_i$  is the variance integrated in  $X_i$  space over dimension  $i$ , and  $E_i$  is the mean  $Y$  value, integrated over the  $d$ -dimensional  $X$  space in all dimensions except  $i$ . Since  $V_i$  can only take values between 0 and  $\text{Var}(Y)$ , the total variance in the model output, we can define normalized first-order sensitivity indices  $S_i$  of each input parameter as

$$S_i = \frac{V_i}{\text{Var}(Y)}, \quad (11)$$

which measures the effect that varying the input  $X_i$  has on the output, averaged over variations of all other inputs. If  $S_i = 1$ , all variance in  $Y$  comes from the variance in  $X_i$ , whereas if  $S_i = 0$ , none of it does, and  $Y$  is independent of  $X_i$ .

In order to measure the interactions between model parameters, we can define higher order indices. For second-order interactions, the combined variance is

$$V_{ij} = \text{Var}_{X_{ij}}(E_{\mathbf{X}_{-ij}}(Y|X_i, X_j)) - V_i - V_j, \quad (12)$$

from which  $S_{i,j}$  can be calculated analogously to  $S_i$ .

It should now be obvious from the definition of the model variance why the OAT methods are inappropriate for complex models – they do not consider the full contribution to the model variance given by equation (9) (which averages over *all* values of the other inputs, instead of being measured only at a designated slice, as shown in the relevant panel of Fig. 1), nor does OAT treat the combined variance of two (equation 12) or more variables correctly.

For a deterministic model, the only source of variance in the output is the variances of the inputs. Therefore, from variance decomposition it follows that

$$\sum_{i=1}^d V_i + \sum_{i < j}^d V_{ij} + \dots + V_{12\dots d} = \text{Var}(Y), \quad (13)$$

which we normalize to obtain the sensitivity indices of all orders

$$\sum_{i=1}^d S_i + \sum_{i < j}^d S_{ij} + \dots + S_{12\dots d} = 1. \quad (14)$$

A direct consequence is that, in order to analytically decompose the total variance of the model, one needs to compute variances of  $2^d - 1$  variables, which can be computationally expensive for complex

models. However, if we assume that the indices decrease as their order increases (which is correct for the model of interest here), we might be less interested in the precise values of higher order contributions, and focus instead on the total higher order response of a given variable. In this case, it is convenient to combine the higher order terms into a total-order index

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}(\text{Var}_{X_i}(Y|\mathbf{X}_{\sim i}))}{\text{Var}(Y)} = 1 - \frac{\text{Var}_{\mathbf{X}_{\sim i}}(E_{X_i}(Y|\mathbf{X}_{\sim i}))}{\text{Var}(Y)}, \quad (15)$$

containing all terms of the decomposed output variance, which include  $X_i$ . Unlike the first-order indices, the  $S_{Ti}$  do not have to add up to 1, as they include all the input interactions.<sup>4</sup>

Higher order effects can also be calculated in simpler analyses, such as ANOVA (Fisher 1918), high-dimensional model representations (Sobol' 1993), or derivative-based methods. However, the total indices are a unique feature of the variance-based SA, and are a major advantage of this methodology, as they allow for a direct comparison of the linear and non-linear impacts of the input parameters.

Following Jansen (1999), Sobol' (2001), and Saltelli et al. (2010), we can use the approximate forms of the first- and total-order sensitivity indices, based on the sampling matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{A}_B^{(i)}$

$$\text{Var}_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i)) \approx \frac{1}{N} \sum_{j=1}^N f(\mathbf{B})_j \left( f(\mathbf{A}_B^{(i)})_j - f(\mathbf{A})_j \right) \quad (16)$$

$$E_{\mathbf{X}_{\sim i}}(\text{Var}_{X_i}(Y|\mathbf{X}_{\sim i})) \approx \frac{1}{2N} \sum_{j=1}^N \left( f(\mathbf{A})_j - f(\mathbf{A}_B^{(i)})_j \right)^2, \quad (17)$$

where  $f(\mathbf{X})$  is the model  $f$  evaluated at point  $X$ .

## 2.6 Illustrative sensitivity analysis of a toy model

The performance of the sensitivity analysis estimator can be demonstrated using a toy model. The Ishigami function is an example of such a model, and is commonly used to test the predictions of sensitivity analysis because it contains non-linear interacting terms. Nevertheless, the sensitivity indices can be calculated analytically and compared to the estimated values.

The Ishigami function is defined by equation (14) of Ishigami & Homma (1991) as

$$Y(X_1, X_2, X_3) = \sin(X_1) + a \sin^2(X_2) + b X_3^4 \sin(X_1), \quad (18)$$

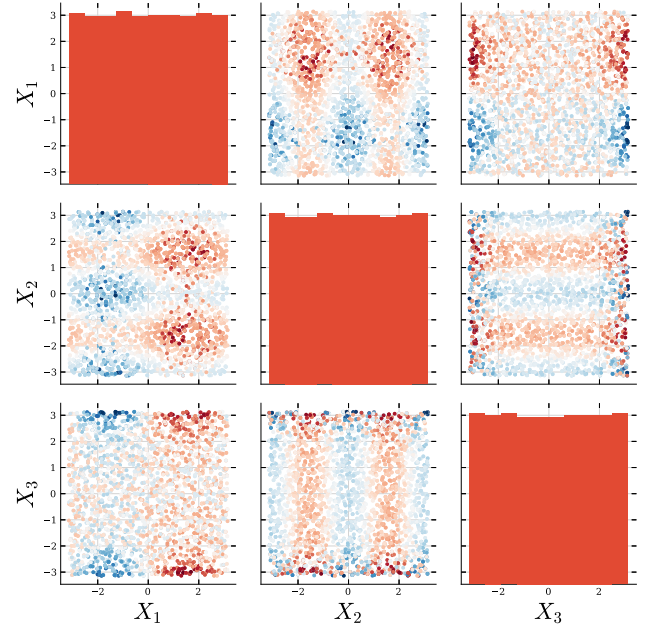
where the  $X_i$  are random variables uniformly distributed between  $-\pi$  and  $\pi$ , such that  $\text{pdf}(X_i) = U(-\pi, \pi)$ , and  $a, b$  are numerical constants, here chosen to be 7 and 0.1, respectively.

The SA was carried out by running the model on inputs generated by a three-dimensional Sobol' sequence for 500 realizations, which resulted in  $4000 = 500 \times (2 + 2 \times 3)$  values of  $X_i$  (as explained in Section 2.4). Next, equation (18) was evaluated at each  $X_i$  point, giving a vector  $Y$  of length 4000. Finally, the vector  $Y$  was analysed using the SALib Python package (Herman & Usher 2017).

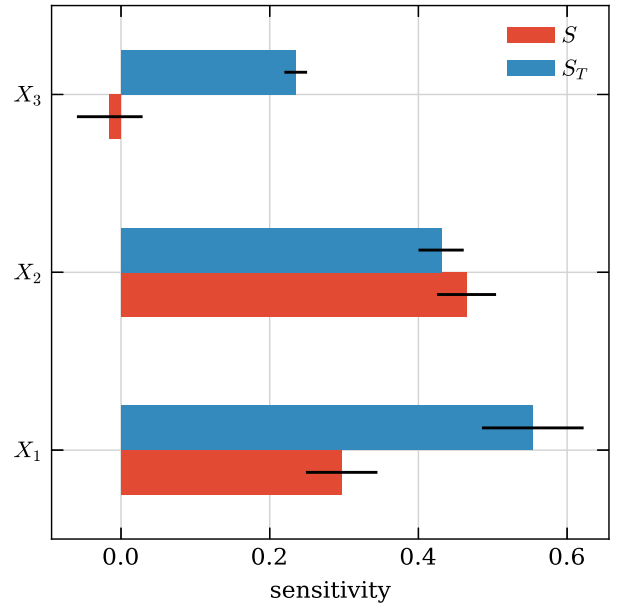
The evaluations of equation (18) are shown in Fig. 2, and the first- and total-order sensitivity indices of the three input parameters are shown in Fig. 3. It is interesting to draw some qualitative observations from Fig. 2:

(i) varying  $X_1$  and  $X_2$  in isolation results in large changes in  $Y$ ; this is reflected by large values for  $S_1$  and  $S_2$ .

<sup>4</sup>In this case, the whole is literally more than the sum of the parts.



**Figure 2.** The 4000 evaluations of the Ishigami function (equation 18). On-diagonal histograms show distribution of the  $X_i$  parameters. Off-diagonal scatter plots show pairs of parameters and are colour-coded to show the value of output  $Y$ .



**Figure 3.** First-order (red) and total (blue) sensitivity indices of the three input parameters of the Ishigami function, with  $1\sigma$  confidence bars (black).

(ii) varying  $X_1$  and  $X_2$  together has a large effect on  $Y$ ; this is reflected by large values for  $S_{T1}$  and  $S_{T2}$ .

(iii) varying  $X_3$  for mid-range values produces little effect, but varying other parameters at extreme  $X_3$  values produces a large change in  $Y$ ; correspondingly,  $S_3$  is nearly zero, but  $S_{T3}$ , which captures the global response of  $Y$  to  $X_3$ , is larger.

(iv)  $S_1$  is negative, despite being defined in terms of non-zero variance (equation 11); this is the result of using a numerical approximation instead of an analytical formula to estimate  $S_1$ ;

however, note that the confidence interval includes the origin, and so the value of  $S_1$  is consistent with zero.

A more complete SA would involve computing second-order indices, and comparing sensitivity indices for different versions of the Ishigami function, such as with different  $a, b$  parameters, or over different  $X_i$  ranges. However, this more complete analysis is beyond the scope of this section, as it is only meant for demonstration purposes. The source code used to reproduce this analysis has been made public: <https://github.com/oleskiewicz/sensitivity/releases/tag/v1.0>.

## 2.7 GALFORM output used in the sensitivity analysis

When applying SA to a model with a multidimensional output, it is necessary to select the most interesting outputs manually. The Sobol' index method assumes, and can only be calculated for, separate one-dimensional output vectors  $Y$ . From the formal standpoint, this is problematic as the sensitivity indices contain no information about any correlations between various model outputs. However, in practice, one could perform model runs that follow the Saltelli sampling and then carry out separate sensitivity analyses for any desired number of model outputs, since running the model is more time consuming than calculating the Sobol' indices.

Here, we focus on the prediction of GALFORM for the  $K$ -band LF at  $z = 0$ , calculated as described in Section 2.1.6. We have chosen to consider this statistic due to the well-understood influence of the model parameters on the form of the luminosity function [see the extensive discussion in Lacey et al. (2016)]. Varying the parameters around the values used in the fiducial model shows that the bright and faint ends of the luminosity function are regulated by different physical processes. Therefore, the sensitivity indices could be easily verified for errors, and we will be able to quantify our intuition regarding the relative importance of the different feedback modes on the abundance of galaxies at different luminosities.

We have elected to perform the analysis on the model output values normalized by the observational data (see equation 20) instead of on the model output itself. This way, the values we focused on were close to the ones typically used for model optimization, and had a reduced dynamic range, being effectively normalized by the observational values. Analysing a SAM independently of the observational constraints, while interesting in its own merit, is outside the scope of this work.

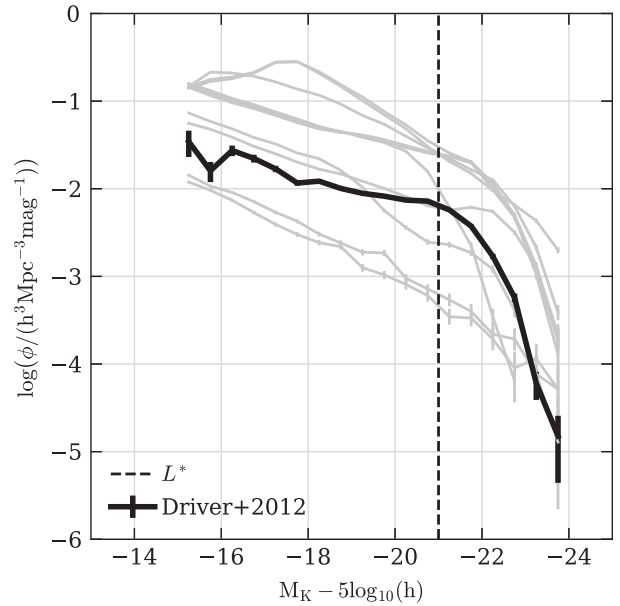
## 3 RESULTS

### 3.1 Sensitivity analysis experiments

We have carried out two separate sensitivity analyses using GALFORM: (i) 600 GALFORM model runs varying two parameters ( $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$ ), and (ii) 1600 model GALFORM runs varying seven parameters (see Table 2).

For both series of runs, an SA was carried out on the  $K$ -band LF at  $z = 0$ , with two different binnings of the LF used to compute the sensitivity indices, as explained below.

In the first instance, we performed a simple analysis by splitting the LF into two broad luminosity bins, one covering a range of luminosities brighter than  $L^*$  and the other luminosities fainter than  $L^*$  (see Fig. 4). For each run, we calculated two model outputs covering the bright and faint ends of the LF,  $d_{\text{faint}}$  and  $d_{\text{bright}}$ , defined by summing the normalized differences between the observed and



**Figure 4.** The  $K$ -band LF at  $z = 0$  in the AB magnitude system. Grey lines represent 10 GALFORM model realizations randomly chosen from the 1600-model run series. The black line represents the observational data from Driver et al. (2012). The black vertical line is drawn at  $L = L^*$ , and separates the bright and the faint ends of the LF.

predicted values of the luminosity function for luminosities brighter and fainter than  $L^*$ , e.g.

$$d_{\text{faint}} = \sum_{L < L^*} \frac{\log_{10}(\phi) - \log_{10}(\hat{\phi})}{\log_{10}(\phi)}, \quad (19)$$

with  $d_{\text{bright}}$  defined analogously for  $L > L^*$ . The observed luminosity function  $\hat{\phi}$  is taken from Driver et al. (2012). Unlike a traditional measure of model fitness, we *do not* take the absolute value or square of the distance between the model prediction and the data. This is because the sign of the output (i.e. the sense of the discrepancy) is valuable information for the sensitivity indices, as it contains the direction of the model response.

This coarse analysis is quantitatively identical to measuring the LF using only two broad luminosity bins. This exercise has two goals: (i) to verify that SA produces explainable results that can be interpreted in accordance with our physical intuition about the galaxy formation model, and (ii) to check the convergence of the sensitivity indices and their confidence intervals, which can be estimated as explained in Section 2.2.

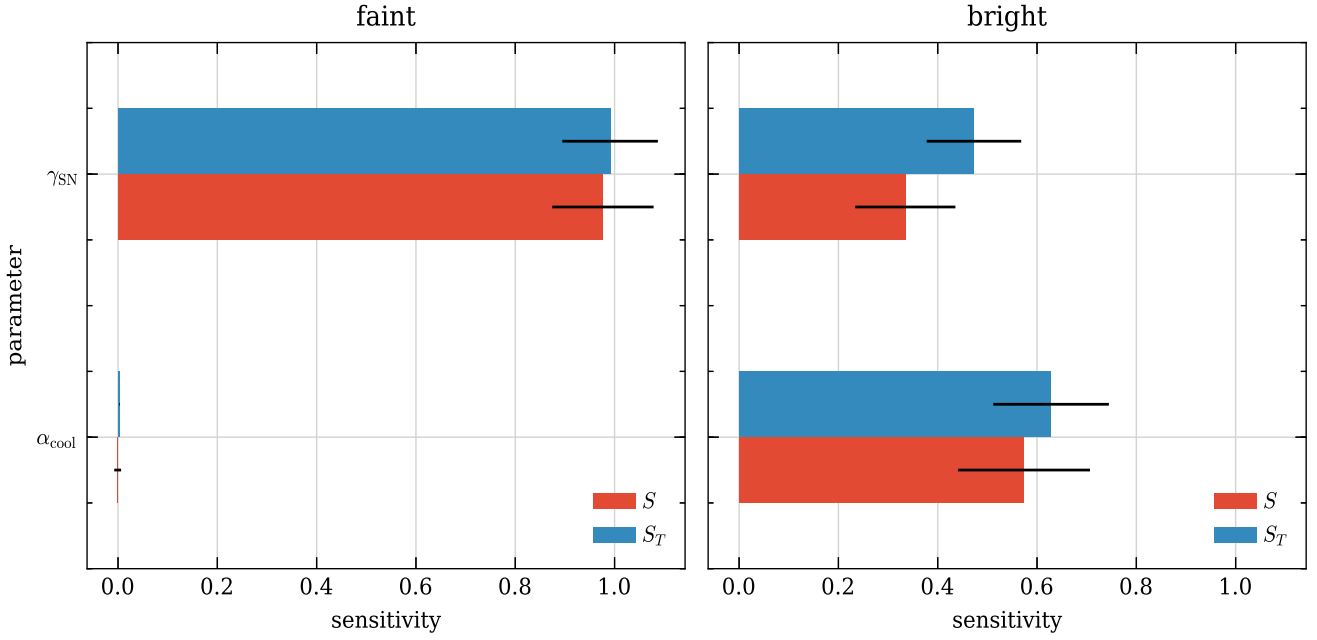
After this coarse two-bin analysis, in the second case we calculate sensitivity indices for each of the 18 luminosity bins,  $L_i$ , using the quantity

$$d_i = \frac{\log_{10}(\phi_i) - \log_{10}(\hat{\phi}_i)}{\log_{10}(\phi_i)}. \quad (20)$$

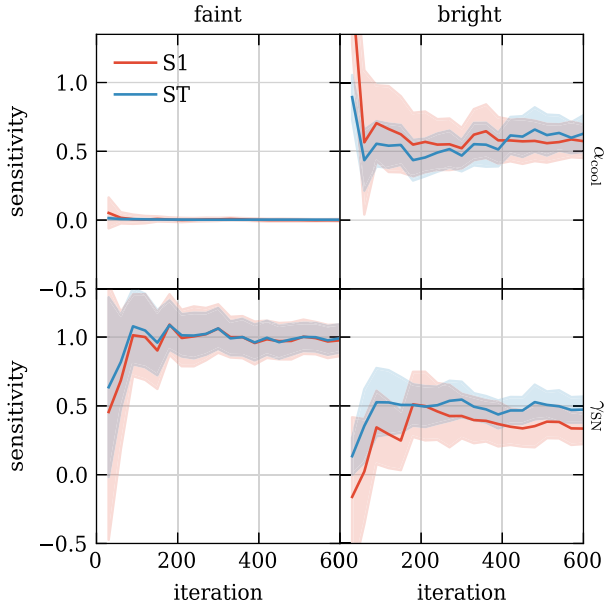
This serves as a fine-grained analysis, which can quantify the relative impact of different parameters on the individual segments of the LF, as well as uncovering interactions between model parameters.

### 3.2 Feedback processes and the luminosity function

The first series of runs, which analysed the effects of changing two of the parameters that specify different feedback processes in



**Figure 5.** Sensitivity indices for the first series of runs, when varying two parameters in GALFORM:  $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$ , for the  $K$ -band luminosity function measured in two coarse luminosity bins. The colours of the bars indicate different indices, first (blue) (equation 11) and total (red) (equation 15) order for a given variable. The left-hand panel shows indices for  $L < L^*$ , and the right for  $L > L^*$  (see equation 19). The black bars show the  $1\sigma$  confidence interval for the sensitivity indices.



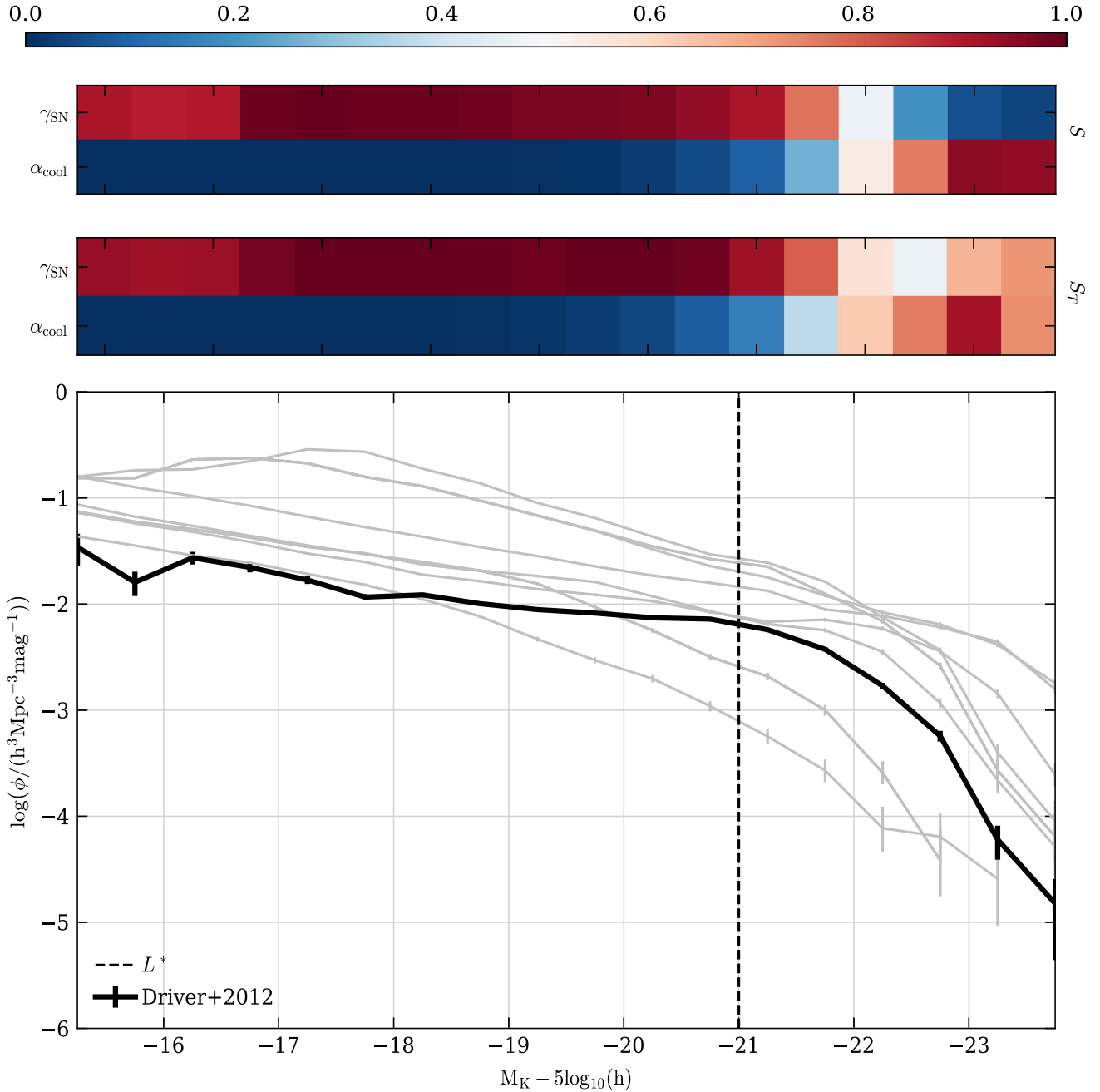
**Figure 6.** Convergence of the first- and total-order sensitivity indices for the first series of runs, when varying two parameters of GALFORM,  $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$ , as a function of a number of samples. The sensitivity indices in this case are computed in each of two broad luminosity bins, covering, respectively, the faint and bright ends of the luminosity function. Individual subplots show the results for the faint and bright ends of the  $K$ -band LF (columns, labelled on the top), and the  $\alpha$  parameters (rows, labelled on the right). Solid lines correspond to the values of the indices, and the shaded regions to the  $1\sigma$  confidence band of the values, both colour-coded by the order of the indices as labelled in the legend.

GALFORM,  $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$ , was carried out to verify the usefulness of the SA and to evaluate its effectiveness, given our physical intuition, regarding the expected impact on the LF of varying these model parameters. Only two parameters were allowed to vary to speed-up the analysis and allow for an easier interpretation of results:  $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$  (see Table 2 for the range of parameter values considered). Recall that  $\gamma_{\text{SN}}$  controls the mass loading of SNe-driven winds and  $\alpha_{\text{cool}}$  determines the halo mass above which AGN heating shuts down the cooling flow.

Fig. 5 shows the first- and total-order sensitivity indices calculated from 600 GALFORM model runs for the coarse-binned analysis of luminosity function using two bins, one fainter and one brighter than  $L^*$ , as prescribed by equation (19). The results are striking, but not unexpected: it is clear that  $\gamma_{\text{SN}}$  is the dominant parameter out of the two in shaping the model output for galaxies fainter than  $L^*$  (and hence, that such galaxies are mainly affected by SNe feedback) and that both parameters have similar significance for galaxies brighter than  $L^*$  (albeit  $\alpha_{\text{cool}}$  is slightly more important), and so bright galaxies are affected by SNe feedback and AGN heating. Moreover,  $S_1$  and  $S_T$  are comparable in all cases, which means that the model response to varying these parameters is mostly linear.

Fig. 6 shows the convergence of the indices from Fig. 5 as a function of the number of samples  $N$ . The indices do not change substantially after 100 GALFORM runs.

Fig. 7 shows the first- and total-order sensitivity indices for the fine-grained analysis of the LF using multiple luminosity bins, using equation (20) as model output. We can see that  $L^*$  is close to coinciding with the bin at which AGN heating starts to become important, which explains the results shown in Fig. 5. We also learn that while SNe feedback does not interact with the AGN heating at the faint end of the LF, its influence over the bright end is strongly correlated.



**Figure 7.** First- and total-order sensitivity indices (equation 11) for the first series of runs, when varying two parameters in GALFORM:  $\alpha_{\text{cool}}$  and  $\gamma_{\text{SN}}$ . Bottom panel:  $K$ -band luminosity function at  $z = 0$  like in Fig. 4; grey lines correspond to 10 randomly chosen runs; black line is the observational data from Driver et al. (2012); dashed vertical line corresponds to  $L^*$ . Top panels: first- and total-order (as labelled on the right) sensitivity indices of two variables ( $y$ -axis) for 18 individual magnitude bins ( $x$ -axis), colour-coded by value between 0 (not sensitive) and 1 (most sensitive) as labelled by the colour bar at the top.

We did not consider the best-fitting model for this two-parameter case, since we perform a rudimentary estimate of the best-fitting parameter set in the next section, when varying more GALFORM model parameters at the same time.

### 3.3 Sensitivity analysis over a multidimensional parameter space

The design of the second experiment, in which seven GALFORM parameters are varied simultaneously (Table 2), is inspired by

the work on parameter optimization using Bayesian emulators by Bower et al. (2010) and Rodrigues et al. (2017). For comparison, we use the same parameter ranges adopted in their studies. This exercise requires significantly more model realizations than the first one, since we sample a higher dimensional parameter space and aim to observe interactions between more parameters.

Fig. 8 shows the parameter space and its sampling, colour-coded by the goodness-of-fit measure

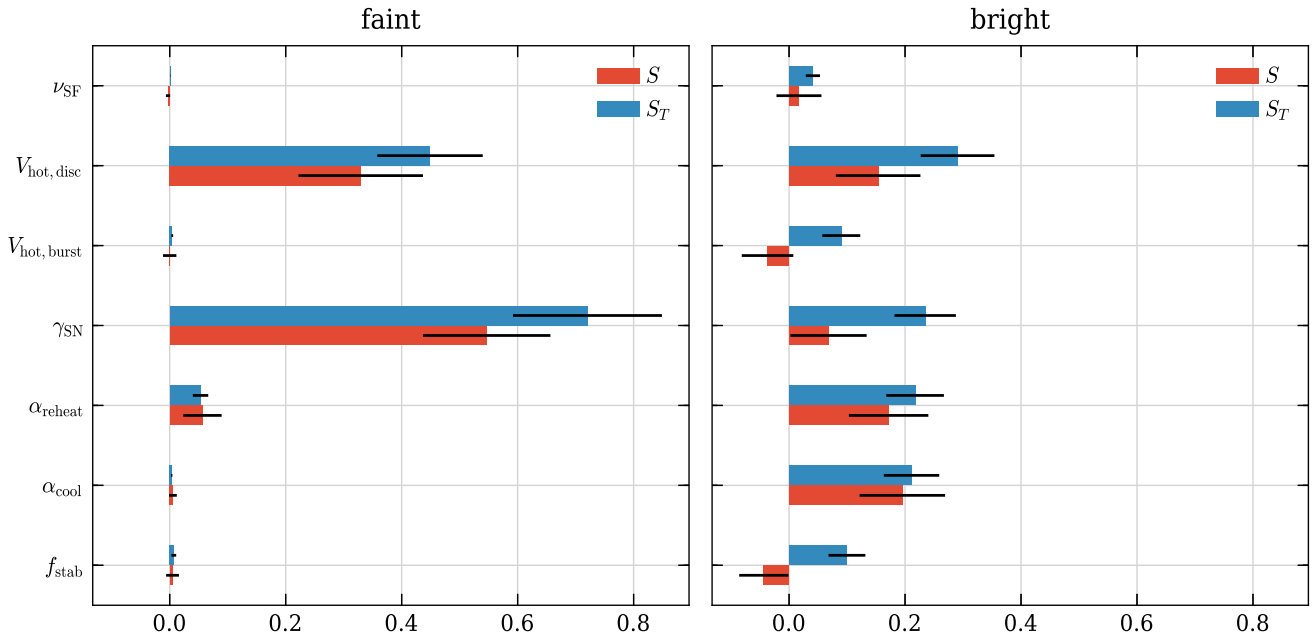
$$\chi^2 = \sum_i \frac{(\log_{10}(\phi_i) - \log_{10}(\hat{\phi}_i))^2}{\log_{10}(\phi_i)}, \quad (21)$$



**Figure 8.** The parameter space of the second GALFORM experiment in which seven parameters are varied over 1600 realizations of the model. On-diagonal histograms show the nearly uniform distribution of the individual parameters, as expected for Saltelli sampling. Off-diagonal scatter plots show the parameter space for pairs of parameters, colour-coded by the goodness-of-fit  $\chi^2$  (equation 21) of the model prediction for the  $K$ -band luminosity function using 18 luminosity bins (Fig. 4), to the observational estimate from Driver et al. (2012); blue points correspond to runs with low values of  $\chi^2$ , as labelled by the colour bar.

where the sum is carried out over all luminosity bins and low values of  $\chi^2$  are blue. While  $\chi^2$  is not a robust model output for SA, as it does not contain information about the *direction* of the model response as explained in Section 3, it is still a useful measure of a global model response or ‘quality of fit’. The on-diagonal histograms indicate that the Saltelli sampling produces a nearly

uniform sampling of parameter space, as expected from a low-discrepancy sequence. The off-diagonal scatter plots give a first indication of some of the first-order index results: the  $\chi^2$  of the model LFs is sensitive to variation of  $\gamma_{\text{SN}}$ , is degenerate in the  $\gamma_{\text{SN}}-v_{\text{hot disc}}$  plane (which follows directly from equation 3), and depends only weakly on other parameters.



**Figure 9.** Sensitivity indices for the second series of 1600 GALFORM runs, varying seven model parameters (Table 2), computed using the coarse two-bin description of the luminosity function. The bar colours indicate the values of different indices, the first index (equation 11) ( $S_1$ , red) and total-order index (equation 15) ( $S_T$ , blue) for each parameter. The left-hand panel shows indices for galaxies in the luminosity bin fainter than  $L^*$ , and the right-hand panel for galaxies in the bin brighter than  $L^*$  (equation 19). The black bars show the  $1\sigma$  confidence intervals for the sensitivity indices.

Fig. 9 shows the first- and total-order sensitivity indices for the coarse-binned analysis of the LF (equation 19). Since sensitivity indices are derived from the normalized variance (equations 11 and 15), the values should always be between 0 and 1, which they are (including the confidence interval). Similarly to Fig. 5, in Fig. 9 we see two different types of behaviour of the GALFORM model: the faint end is dominated by SNe feedback, while the bright end has a mixed, non-linear response to many parameters. Interestingly, while AGN feedback (via  $\alpha_{\text{cool}}$ ) has the highest first-order sensitivity index ( $S_1$ ) for the bright end of the LF, the total-order indices ( $S_T$ ) of SNe feedback processes dominate. Of particular interest are the  $f_{\text{stab}}$  and  $\nu_{\text{hot,burst}}$  parameters. These parameters have nearly zero first-order response indices (which means that their impact cannot be detected by an OAT analysis), but their combined higher order responses are significant.

It is instructive to see the origin of the values reported in Fig. 9, by inspecting how the sensitivity changes bin-by-bin (equation 20) in Fig. 10. The results are consistent with Section 3.2, and together provide an interpretation of the behaviour of the GALFORM model. Moreover, displaying the model output together with model sensitivity can be of use when manually tweaking the model, allowing for a fine, manual control over the precise details of the LF (or, indeed, other outputs).

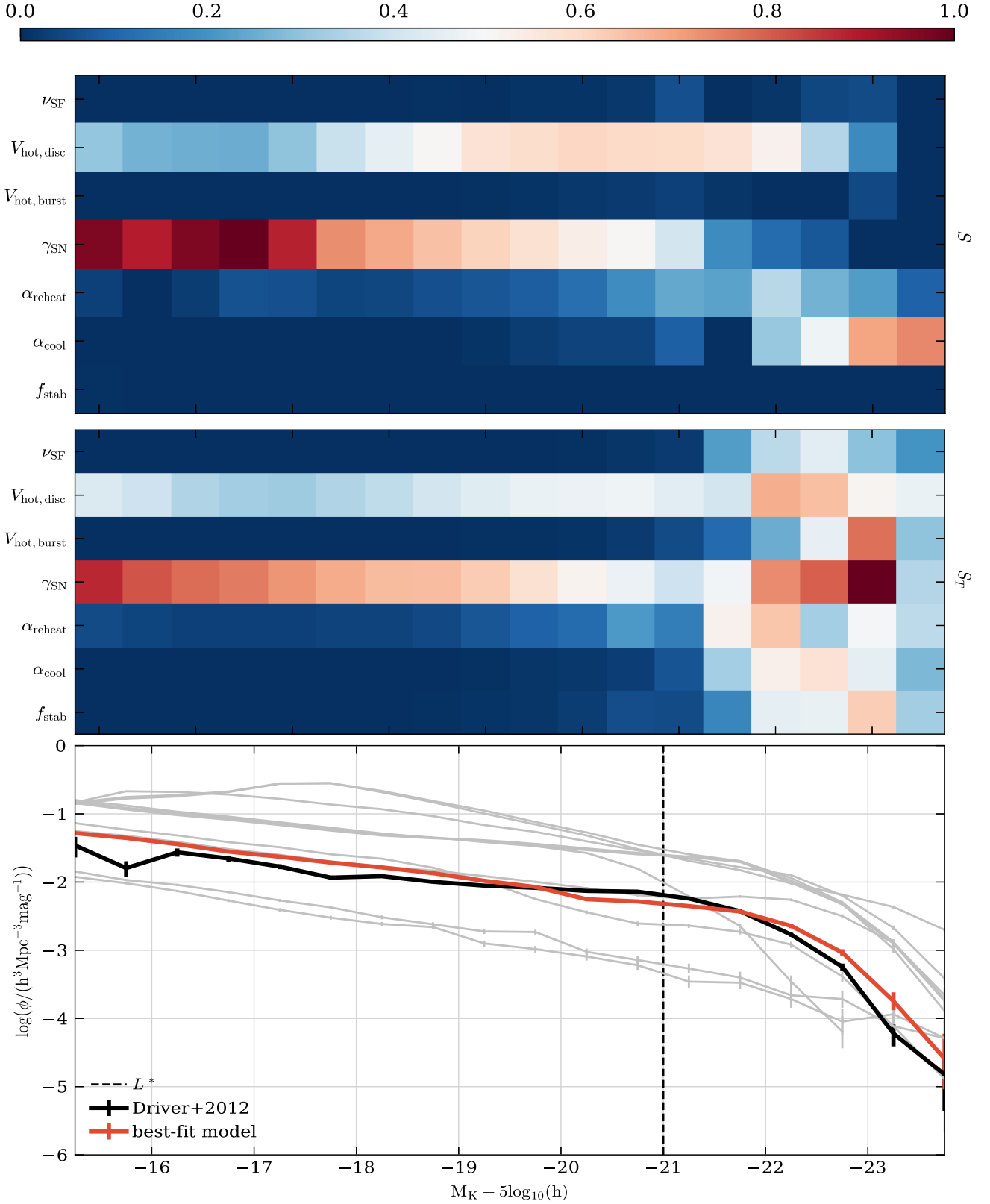
Finally, we note that Fig. 10 also shows the LF for the best-fitting model, as determined by the smallest value of equation (21). This can be considered an additional benefit of running SA – requiring so many model realizations naturally finds one that is likely to be close to a global optimum. The best-fitting parameter values are reported in Table 3. Note that the values diverge from those reported in Lacey et al. (2016), due to different fitting method and the fact that this study only considered the  $K$ -band LF, whereas Lacey et al. (2016) took into account multiple observations in a manual parameter tuning. Of particular interest is the value of  $V_{\text{hot,disc}}$ , which is over 20 per cent larger than in the previous calibration of this

GALFORM model. We attribute this difference to the fact that, as discussed in Section 3.3 and shown in Fig. 9, the combined total-order sensitivity index of  $V_{\text{hot,disc}}$  outweighs the first-order index for both ends of the  $K$ -band LF. This suggests that the optimal value of this parameter could be missed by OAT model fitting. The differences in the other parameter values are not as significant as they might seem – the variables with the highest sensitivity match the previously reported values pretty closely (e.g.  $\gamma_{\text{SN}}$  is within 7 per cent), and the variables with low sensitivity that diverge by a significant margin by definition of the sensitivity indices do not have significant impact on the  $K$ -band LF.

## 4 CONCLUSIONS

We have used variance-based sensitivity analysis to analyse the sensitivity of the  $K$ -band luminosity function predicted using the GALFORM semi-analytical model of galaxy formation to the variation of the model parameters. We have shown that sensitivity analysis is a useful tool, which goes beyond simple model fitting and OAT parameter variation, and we have demonstrated that it can be applied to a challenging problem in computational astrophysics. Variance-based sensitivity analysis is perhaps particularly useful for the semi-analytic modelling of galaxy formation modelling, due to the computational expense of searching a multidimensional parameter space and the non-linearity of the model. These features have led some to view such models as black boxes. Part of the aim of the sensitivity analysis presented here is to make the behaviour of the model and how it responds to parameter changes more transparent.

In its present form, sensitivity analysis can only deal with one-dimensional outputs of a model, which on the one hand means that it cannot be used to resolve correlations in model outputs (such as between the predictions in different bins of the luminosity function or between the luminosity function in different bands; see Benson 2014), yet on the other hand this feature gives the



**Figure 10.** First- and total-order sensitivity indices (equation 11) for the second series of runs, when varying seven GALFORM parameters (Table 2). Bottom panel: K-band luminosity function at  $z = 0$  as in Fig. 4; grey lines show 10 randomly chosen GALFORM models; the black line connects observational data from Driver et al. (2012); dashed vertical line corresponds to  $L^*$ ; the solid red line shows the best-fitting model. Top panels: first- and total-order (as labelled on the right) sensitivity indices of two variables (y-axis) for individual magnitude bins (x-axis), colour-coded by value between 0 (not sensitive) and 1 (most sensitive) as labelled by the colour bar at the top.

**Table 3.** The best-fitting GALFORM parameters found in this work, in relation to Driver et al. (2012).

Parameter	Value
$\nu_{\text{SF}}$ [ $\text{Gyr}^{-1}$ ]	0.46
$\gamma_{\text{SN}}$	3.45
$\alpha_{\text{reheat}}$	0.74
$V_{\text{hot,disc}}$ [ $\text{km s}^{-1}$ ]	332.69
$V_{\text{hot,burst}}$ [ $\text{km s}^{-1}$ ]	392.90
$\alpha_{\text{cool}}$	0.58
$f_{\text{stab}}$	0.77

scientist performing the study unlimited flexibility in choosing and parametrizing the outputs they find the most important. Here, we have elected to perform the sensitivity analysis using the model predictions in luminosity bins cast in terms of the difference between the computed and measured  $K$ -band luminosity function at  $z = 0$ . Our motivation for this was that by choosing an established observable with a well-understood connection to the underlying physical processes and their description in terms of GALFORM parameters, we could make a convincing case for the usefulness of the sensitivity analysis.

With this in mind, future work on SA might want to examine the variance of the outputs of the semi-analytic model alone, independently of the corresponding measured observable values. There are three main reasons for such an approach: (i) Using the full dynamic range of the predictions: normalizing the model output by observations flattens the dynamic range, and while SA works equally well for small and large values, by only analysing a flat version of the model predictions we effectively take the regions in which the model gives a flat or steep response (for instance, the faint and bright ends of the LF, respectively) and make them look the same. (ii) Independence of post-processing: by comparing to data, we had to make a choice about the norm of the discrepancy between the model output and observations – do we retain the sense of the discrepancy or square it? A different SA study could have chosen differently, altering the results. By analysing model outputs independently of the observations, these choices are no longer necessary. (iii) Data independence: SA results could change if a different data set is used with the same model.

Moreover, the  $K$ -band luminosity function is just one possible output and there are many others, which a successful semi-analytic model should reproduce accurately. Analysing all of these is outside the scope of this study, but we hope to have shown that SA is a promising avenue of research.

Finally, we note that while correctly estimating model sensitivity can be useful in guiding model optimization and improving the physical interpretation of the parameters of the galaxy formation models, one must remember that even the most rigorous sensitivity analysis can only provide the answers with regard to the model, not the underlying physical system itself (Taleb & Douady 2013). Therefore, the relationship between the structure of the model and that of the physical system remains open to discussion.

## ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Facilities Council (ST/L00075X/1). PO acknowledges an STFC studentship funded by STFC grant ST/N50404X/1. The inspiration for this project came from a placement by PO at Atom Bank, which was supported by the STFC/Durham University Centre for Doctoral

Training in Data Intensive Science, supported by ST/P006744/1. This work used the DiRAC Data Centric system at the Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1, and the Durham University. DiRAC is part of the National E-infrastructure.

## REFERENCES

- Antonov I., Saleev V., 1979, *USSR Comput. Math. Math. Phys.*, 19, 252
- Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101
- Baugh C. M., 2013, *PASA*, 30, e030
- Baugh C. M. et al., 2019, *MNRAS*, 483, 4922
- Benson A. J., 2010, *Phys. Rep.*, 495, 33
- Benson A. J., 2014, *MNRAS*, 444, 2599
- Benson A. J., Bower R., 2010, *MNRAS*, 405, 1573
- Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, *MNRAS*, 407, 2017
- Chan K., Saltelli A., Tarantola S., 1997, ACM Press, Proc. of the 1997 Winter Simulation Conf., Sensitivity Analysis of Model Output: Variance-Models Make the Difference, Atlanta, GA, p. 8
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, *MNRAS*, 432, 2717
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Driver S. P. et al., 2012, *MNRAS*, 427, 3244
- Efstathiou G., Lake G., Negroponte J., 1982, *MNRAS*, 199, 1069
- Fanidakis N., Baugh C. M., Benson A. J., Bower R. G., Cole S., Done C., Frenk C. S., 2011, *MNRAS*, 410, 53
- Fisher R. A., 1918, *Trans. R. Soc. Edinburgh*, 52, 399
- Golovin D., Solnik B., Moitra S., Kochanski G., Karro J. E., Sculley D., eds, 2017, Google Vizier: A Service for Black-Box Optimization, Available at: <http://www.kdd.org/kdd2017/papers/view/google-vizier-a-service-for-black-box-optimization>
- Gómez F. A., Coleman-Smith C. E., O'Shea B. W., Tumlinson J., Wolpert R. L., 2012, *ApJ*, 760, 112
- Gómez F. A., Coleman-Smith C. E., O'Shea B. W., Tumlinson J., Wolpert R. L., 2014, *ApJ*, 787, 20
- Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *MNRAS*, 439, 264
- Griffin A. J., Lacey C. G., Gonzalez-Perez V., Lagos C. D. P., Baugh C. M., Fanidakis N., 2019, *MNRAS*, 487, 198
- Guo Q. et al., 2016, *MNRAS*, 461, 3457
- Halton J. H., 1964, *Commun. ACM*, 7, 701
- Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, *MNRAS*, 396, 535
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, *MNRAS*, 431, 3373
- Herman J., Usher W., 2017, *J. Open Source Softw.*, 2, 97
- Ishigami T., Homma T., 1991, Proceedings. First International Symposium on Uncertainty Modeling and Analysis. IEEE Comput. Soc. Press, College Park, MD, p. 398
- Jansen M. J., 1999, *Comput. Phys. Commun.*, 117, 35
- Knebe A. et al., 2015, *MNRAS*, 451, 4029
- Lacey C. G. et al., 2016, *MNRAS*, 462, 3854
- Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011, *MNRAS*, 416, 1566
- Lagos C. D. P., Lacey C. G., Baugh C. M., 2013, *MNRAS*, 436, 1787
- Levitay Y., Markovich N., Rozin S., Sobol' I., 1988, *USSR Comput. Math. Math. Phys.*, 28, 88
- Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, *MNRAS*, 416, 1949

- Lu Y., Mo H. J., Katz N., Weinberg M. D., 2012, *MNRAS*, 421, 1779
- Ludlow A. D., Schaye J., Bower R., 2019, *MNRAS*, 488, 3663
- Martindale H., Thomas P. A., Henriques B. M., Loveday J., 2017, *MNRAS*, 472, 1981
- Mitchell P. D. et al., 2018, *MNRAS*, 474, 492
- Morris M. D., 1991, *Technometrics*, 33, 161
- Mutch S. J., Poole G. B., Croton D. J., 2013, *MNRAS*, 428, 2001
- Ostriker J. P., McKee C. F., 1988, *Rev. Mod. Phys.*, 60, 1
- Pillepich A. et al., 2018, *MNRAS*, 473, 4077
- Planck Collaboration XVI et al., 2014, *A&A*, 571, A16
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. Cambridge Univ. Press, Cambridge
- Rodrigues L. F. S., Vernon I., Bower R. G., 2017, *MNRAS*, 466, 2418
- Ruiz A. N. et al., 2015, *ApJ*, 801, 139
- Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M., Tarantola S., 2010, *Comput. Phys. Commun.*, 181, 259
- Saltelli A. et al., 2019, *Environmental Modelling & Software*, 114, 29
- Schaye J. et al., 2014, *MNRAS*, 446, 521
- Sobol' I. M., 1967, *Ž. Vyčisl. Mat. Mat. Fiz.*, 7, 784
- Sobol' I. M., 1993, *Math. Model. Comput. Exp.*, 1, 407
- Sobol' I., 2001, *Math. Comput. Simul.*, 55, 271
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Stein M., 1987, *Technometrics*, 29, 143
- Taleb N. N., Douady R., 2013, *Quant. Finance*, 13, 1677
- Tumlinson J., 2009, *ApJ*, 708, 1398
- Ulam S., 1960, *Problems in Modern Mathematics*. John Wiley & Sons, New York
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.