

8-21-2018

Classifying galaxy spectra at $0.5 < z < 1$ with self-organizing maps

S. Rahmani

The University of Western Ontario

H. Teimoorinia

National Research Council Canada

P. Barmby

The University of Western Ontario, pbarmby@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/physicspub>

Citation of this paper:

Rahmani, S.; Teimoorinia, H.; and Barmby, P., "Classifying galaxy spectra at $0.5 < z < 1$ with self-organizing maps" (2018). *Physics and Astronomy Publications*. 60.

<https://ir.lib.uwo.ca/physicspub/60>

Classifying galaxy spectra at $0.5 < z < 1$ with self-organizing maps

S. Rahmani^{1,2*}, H. Teimoorinia³, P. Barmby^{1,2}

¹*Department of Physics & Astronomy, Western University, London, ON N6A 3K7, Canada*

²*Centre for Planetary Science & Exploration, Western University, London, ON N6A 3K7, Canada*

³*NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada*

9 September 2021

ABSTRACT

The spectrum of a galaxy contains information about its physical properties. Classifying spectra using templates helps elucidate the nature of a galaxy’s energy sources. In this paper, we investigate the use of self-organizing maps in classifying galaxy spectra against templates. We trained semi-supervised self-organizing map networks using a set of templates covering the wavelength range from far ultraviolet to near infrared. The trained networks were used to classify the spectra of a sample of 142 galaxies with $0.5 < z < 1$ and the results compared to classifications performed using K-means clustering, a supervised neural network, and chi-squared minimization. Spectra corresponding to quiescent galaxies were more likely to be classified similarly by all methods while starburst spectra showed more variability. Compared to classification using chi-squared minimization or the supervised neural network, the galaxies classed together by the self-organizing map had more similar spectra. The class ordering provided by the one-dimensional self-organizing maps corresponds to an ordering in physical properties, a potentially important feature for the exploration of large datasets.

Key words: galaxies: high-redshift, galaxies: spectra, methods: observational, methods: statistical, methods: data analysis

1 INTRODUCTION

Nearly all information we can obtain from a galaxy is encapsulated in the light it emits; every observable phenomenon in a galaxy leaves a footprint on its spectral energy distribution (SED). We can determine some physical properties of a galaxy by properly modeling various features observed in its SED, and the general shape of the SED can be used as an identifier of the morphological type. Considering the main features of SEDs, galaxies can be categorized into two main groups: quiescent or star-forming. Each group has its own characteristic features and can in turn be divided into many sub-branches.

Several attempts have been made to create a complete set of galaxy spectral templates using observations of nearby galaxies (e.g. Kinney et al. 1993, Kinney et al. 1996, hereafter K96, Bershady et al. 2000, Mannucci et al. 2001). Based on their usage, these templates are restricted to certain wavelengths. In the near-infrared (NIR) to ultraviolet (UV) wavelengths (the region where the energy output of stars peaks), the spectrum¹ contains information about the

main physical properties of galaxies, e.g., age, star formation rate (SFR), stellar mass, metallicity, and interstellar medium.

One approach to categorizing observed galaxy spectra involves matching to templates. This approach can help to answer the question of whether galaxy spectra form a continuous distribution or can be separated into discrete groups. Templates are also helpful in computing *K*-corrections and discovering new classes of objects as outliers (Folkes et al. 1996). A commonly used method of categorization via template matching is chi-squared minimization: the quantity $\chi^2 = \sum_{\lambda} [(\text{obs} - \text{templ})/\text{uncert}]^2$ is calculated between each observation and each template and the best-matched template is that with the lowest χ^2 . (For spectra, the observations and templates are typically made to match at a set/range of particular wavelengths, since the spectral shape is usually more important than the normalization.) This technique is often used in computing photometric redshifts for galaxies (e.g. Hutchinson et al. 2016; Delchambre 2016; Ilbert et al. 2006). In this study of galaxy spectral classification, we use chi-squared matching as a reference, since it

* E-mail: srahma49@uwo.ca

¹ In the remainder of this paper, we are concerned with a limited

range in wavelength, and so we refer to spectra rather than spectral energy distributions.

is both simple and familiar to astronomers. However, chi-squared matching has its limitations: it assumes that the templates cover the full range of galaxy properties, and does not account for the possibility that a spectrum may share features with more than one template. Uncertainties which are non-Gaussian or not correctly estimated may cause chi-squared matching to fail.

Other approaches to classifying galaxy spectra include artificial neural networks, K-means clustering, and principal component analysis (PCA) (e.g. [Karamelas et al. 2012](#); [Allen et al. 2013](#); [Ordovás-Pascual & Sánchez Almeida 2014](#); [Shi et al. 2015](#)). These methods can make use of templates, but generally do so in a more flexible manner than chi-squared matching, for example by giving the probability that a galaxy belongs to each of several categories. Compared to chi-squared matching, these methods can provide additional insights about observed spectra, for example about their distances and/or ordering in feature space (neural networks, K-means) or their underlying dimensionality (PCA). However, they have the disadvantage that they do not usually account for measurement uncertainties.

Neural networks can be trained using two methods: supervised and unsupervised. In supervised methods, a neural network is trained using input data based on a desired outcome. These methods are very useful for classification of data with specific target values. In unsupervised methods there is no prediction of output: these methods classify data based on their underlying structures and hidden patterns. Unsupervised methods are very helpful for data exploration and finding hidden patterns when the underlying structure of the data is not well established. A Kohonen self-organizing map (also called self-organizing map, or SOM) is a (semi)-supervised neural network.

The goal of this work is to investigate the use of self-organizing maps in classifying galaxy spectra against templates, and compare them to other methods. The questions we want to answer are: (1) What is the degree of agreement between classifications produced by χ^2 matching, one- and two-dimensional self-organizing maps, K-means clustering and artificial neural networks? (2) Are certain spectral types more (or less) consistently classified by the different methods? (3) Which classification methods produce classes with the highest internal similarity? (4) Does the ordering of classifications provided by one-dimensional self-organizing maps correspond to an ordering in physical properties?

We analyze the same sample of 142 galaxies with $0.5 < z < 1$ as [Teimoorinia \(2012, hereafter T12\)](#), who classified the spectra using a supervised (multi-layer feed-forward) neural network method with the [K96](#) templates. As a base for comparing the methods' performance, we find the best chi-squared match among the 12 templates for each of the 142 galaxies. We use this to compute a chi-squared agreement score for each method. To measure how well each classification method groups similar objects together, we calculate the silhouette scores and dispersions in properties between galaxies classified in the same groups. An important limitation to this analysis is the fact that *a priori* classifications for the observed galaxy spectra are not available. This means that many classification metrics, which rely on knowledge of “true” labels, are not applicable.

This paper is organized as follows. In [Section 2](#), we present the data used to train the networks and the data

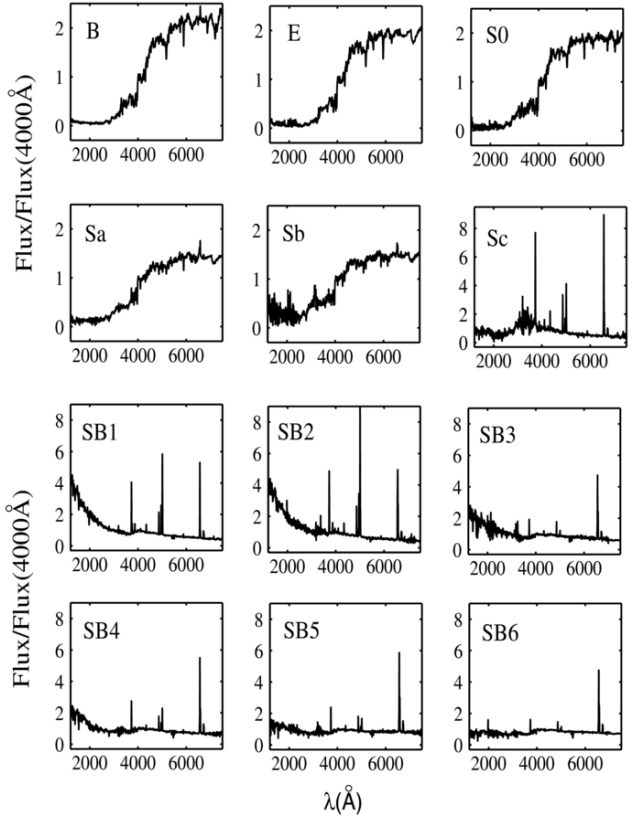


Figure 1. Kinney et al. (1996) spectral templates for 12 types of galaxies, from fig. 1 in [Teimoorinia \(2012\)](#). The type of each template is shown in each frame. Plots B, E, S0, Sa, Sb and Sc show spectra that belong to quiescent galaxies. Starburst galaxy spectra are indicated with SB1 to SB6. Higher numbers represent more intrinsic extinction. Note that the vertical axis label from [Teimoorinia \(2012\)](#) is incorrect: the quantity plotted in the vertical axis is flux density, not flux.

classified using the trained networks. We describe the SOM and K-means clustering methods in [Section 3](#). The results of the spectral classifications and comparison with previous studies are presented in [Section 4](#). In [Section 5](#), we summarize our results and discuss potential future work in this subject.

2 DATA

We use spectral templates from [K96](#) (in 10 \AA wavelength bins) to train neural networks. We classify the spectra of 142 galaxies at $0.5 < z < 1$ from [T12](#) using the trained networks, and use their physical properties to test the new classifications. Following the [T12](#) work, we chose these two sets of data not only to show the application of SOMs in spectral clustering, but also to easily compare supervised and unsupervised methods.

2.1 Kinney spectral templates

[K96](#) used ultraviolet-optical spectra of 70 star-forming and quiescent nearby galaxies to produce a set of templates that contained 12 types of spectral templates. These templates

have been widely used in many studies to determine morphological type of galaxies or properties of specific types of galaxies (e.g. Shakouri et al. 2016; Paiano et al. 2016; Laporte et al. 2016; Holden et al. 2016). K96 stated that these templates can also be used to classify the spectra of high-redshift galaxies.

The 12 templates are divided based on their morphological types for quiescent galaxies or their extinction for starburst galaxies (Fig. 1). The quiescent group of galaxies includes Bulge (B), Elliptical (E), S0, Sa, Sb, and Sc galaxies. The bulge group represents galaxies similar to M31 and M81, whose UV and optical spectra are dominated by their bulge stellar populations. The starburst galaxies are divided into six groups (SB1 to SB6) based on their intrinsic extinctions ($E(B - V)$). As Fig. 1 shows, SB1 galaxies have lower internal extinctions ($E(B - V) \simeq 0.05$), while SB6 galaxies have the highest amount of extinction ($E(B - V) \simeq 0.65$) among starburst galaxies. In the quiescent (B to Sb) templates, the spectrum is redder; strong absorption lines and the 4000 Å break are distinguishable. The SEDs of starburst galaxies are flatter in the optical and near-infrared region than those of the quiescent ones and show strong emission lines. For more details on each spectral type, we encourage readers to see K96 and references therein. The K96 spectra span from ~ 1200 Å to 10000 Å with a resolution of ~ 10 Å. However, in this work we only use observations in the rest-frame wavelength range $\sim 1200 < \lambda < 8000$ Å to train our networks; this wavelength range was chosen due to the availability of flux information in those wavelengths for all 12 templates.

2.2 SED and Properties of the Sample Galaxies

T12 selected 142 galaxies from the spectroscopic campaign of the ESO GOODS-South field (Vanzella et al. 2005, 2006, 2008). The 142 galaxies were selected based on the availability of photometry from HST/ACS, VLA/ISAAC, and Spitzer/MIPS and IRAC (10–13 filters with $\sim 0.4 < \lambda < 24$ μm in the observed frame). Data from these instruments was necessary in order to have a complete picture of stellar population and star formation rate. For each galaxy, a robust spectroscopic redshift and photometric measurements from the GOODS-MUSIC catalogue (Santini et al. 2009) were available. T12 used point-spread-function matched photometry from Santini et al. (2009) as input to the Code Investigating GALaxy Emission (CIGALE); (Noll et al. 2009, hereafter N09) to generate the best-fit SED for each galaxy. T12 produced the best SED match, with a wavelength interval of 910 Å to ~ 80 cm, for each galaxy². Assuming decreasing SFR and visual attenuation (τ) model, Salpeter initial mass function (Salpeter 1955), and an old stellar population with age of ~ 10 Gyr, he derived physical properties of the galaxies such as age and stellar mass. Some of these properties are shown in Table 1. In Section 4.1, we study these properties for each category. More details on creating SEDs and extracting information about galaxy properties using CIGALE can be found in N09 and T12.

The basis for our classification is the SEDs that were

Table 1. Description of the properties of Teimoorinia (2012) galaxies; the output result of CIGALE

Par.	Unit	Description
f_{burst}	—	mass fraction of young single population (SP) model
t_{oSP}	Gyr	age of old SP model
t_{ySP}	Gyr	age of young SP model
t_{D4000}	Gyr	D4000-related age
M_{star}	M_{\odot}	total stellar mass
SFR	M_{\odot}/yr	instantaneous SFR
A_{FUV}	mag	attenuation at 1500 Å

produced by T12. Both the raw data used to create the SEDs,³ and the SEDs themselves, in the form of flux per rest frame wavelength, are publicly available (see Appendix A and the Supplementary Information). Since we have chosen the T12 SEDs to be classified by the trained network, we only used the part of the SEDs that have the same wavelength range as the K96 spectral templates.

3 METHOD

In this section we briefly review the use of K-means clustering, artificial neural networks, and self-organizing maps in astronomy and describe their algorithms. As K-means is both simpler and more well-known, we introduce it briefly first before discussing artificial neural networks in general and the special case of self-organizing maps in detail.

3.1 K-means clustering

K-means clustering (MacQueen 1967) is an unsupervised method used in many astronomical studies (e.g. D’Abrusco et al. 2012; Ordovás-Pascual & Sánchez Almeida 2014; Boersma et al. 2014; Tammour et al. 2016). The K-means method partitions a data set x_i , having n points in d -dimensional feature space, into K clusters, in such a way that each data point belongs to the cluster with the nearest mean value. The number of clusters $K \leq n$ must be decided by the user in advance, and the cluster centroids in feature space are located at positions μ_k . The goal of the algorithm is to minimize the sum

$$J = \sum_{i=1}^n \sum_{k=1}^K \min_k (\|x_i - \mu_k\|^2) \quad (1)$$

where $\|z\|$ indicates the distance measurement in d -dimensional space. In this work the Euclidean distance is used, but other metrics are also possible. The algorithm is initialized by randomly choosing K points as cluster centroids. Each data point is assigned to the cluster with the closest centroid, and the centroids are re-calculated. The steps of re-assigning cluster membership and re-calculating the cluster centroids are repeated until a stopping criterion is reached (a maximum number of iterations or no change

² This wavelength interval is the default output of the CIGALE code.

³ See this ESO webpage.

in cluster assignments). We used the MATLAB K-means library (Seber 1984; Spath 1985), written for MATLAB2015B, to perform K-means clustering. The default maximum number of iterations (100) was used.

3.2 Artificial neural networks

Artificial neural networks (ANNs; Bishop 2006), which are inspired by the way neurons in a human brain route and process data, are very powerful tools that are used in data processing and pattern recognition problems. An ANN contains many interconnected units (nodes or neurons) which process data and work together to solve problems. It uses a training method to learn about nonlinear and complex relations between input and output data, and how to apply these relations to new sets of data.

One example of such a training method is back propagation, an error-correction learning method in which the corrections are made in the last layer and propagate toward the first layer. The signal for the corrections comes from a loss function (e.g., a mean square error function). ANNs have many uses in statistics including as an alternative to conventional regression models; Marquez et al. (1991) found that ANNs “perform best [in terms of mean absolute percentage error] under conditions of high noise and low sample size,” conditions which often occur in astronomy.

ANNs have seen a number of uses in extragalactic astronomy. These include object detection and star/galaxy classification (Andreon et al. 2000), estimation of photometric redshifts (Vanzella et al. 2004), predicting morphological classifications, spectral types and redshifts for galaxies in the Sloan Digital Sky Survey (Ball et al. 2004), ranking quenching parameters in central galaxies (Teimoorinia et al. 2016), and predicting galaxy infrared luminosities from optical observations (Ellison et al. 2016). Comparing an ANN-based technique to χ^2 minimization for estimating reddening of O and B stars, Gulati et al. (1997) found that the two methods showed good agreement, with the ANN method showing faster performance on a test set. Vanzella et al. (2004) found that, compared to other photometric redshift estimation techniques, their ANN-based technique performed better in terms of both redshift precision and computational speed. In this work we are comparing to the results of T12 who used a two-layer feed-forward ANN with a mean-squared error performance function, minimized via the Levenberg-Marquardt algorithm. For further details of the ANN classification we refer the reader to T12.

3.3 Self-organizing maps

A Kohonen self-organizing map (also called self-organizing map, or SOM) is a (semi)-supervised neural network for mapping and visualizing a complex and nonlinear high dimension data introduced by Kohonen (1982). It shows simple geometrical relationships in non-linear high dimensional data on a map (Kohonen 1998). The training of a self-organizing map is fully unsupervised; a competitive learning method is used in which the neighborhood function serves to find the topological behaviour of the input data. Using the resulting map as a template to classify other data requires labelling the output, which is why some groups consider self-organizing maps to be a (semi)-supervised method.

The utilization of the SOM in astronomy dates back to the 1990s, with Odewahn et al. (1992), Hernandez-Pajares & Floris (1994), and Murtagh (1995) among the first to use SOMs in their studies. Geach (2012) used COSMOS data to demonstrate two of the main applications of SOMs: object classification and clustering, and photometric redshift estimation. The latter has been the subject of many other studies (e.g. Carrasco Kind & Brunner 2014). From classifying quasars’ spectra to star/galaxy classifications, from gamma-ray burst clustering to classification of light curves, this method has proved to be useful in various fields of astronomy (e.g. Maehoenen & Hakala 1995; Miller & Coe 1996; Andreon et al. 2000; Balastegui et al. 2001; Rajaniemi & Mähönen 2002; Brett et al. 2004; Scaringi et al. 2009).

Large spectroscopic surveys have made available integrated spectra of millions of galaxies. These integrated spectra combine the light of billions of individual stars and nebulae within a galaxy, and finding patterns and common characteristics between galaxies can be a complex task. in der Au et al. (2012) introduced a new clustering tool based on the SOM method for analyzing these large datasets. They used ~ 60000 spectra from the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) to test their tool, and created very large SOMs to analyze the type of spectra/objects. They also generated SOMs from quasars’ spectra in order to find unusual types of spectra. Later, Meusinger et al. (2016) used these SOMs and updated data from SDSS and other surveys to find a new class of quasars. The other application of SOMs is to find outliers or errors in the data. Fustes et al. (2013) produced a package based on SOM to classify spectra from the GAIA survey that were previously classified as “unknown” by the SDSS pipeline. This package can distinguish an astronomical object from instrumental artifact, and then classify the object based on its spectrum.

The self-organizing map is a clustering method which reduces the dimensionality of the data, usually to one or two dimensions (1D or 2D), while preserving topological features of the original dataset. The result of an SOM is a set of nodes (neurons) that are arranged in a 1D or 2D arrays (Kohonen 1998). To first order, SOM and K-means work similarly: both algorithms operate on the distances between n objects, each represented by d features. The SOM algorithm also considers distances between objects in the low-dimensional map, as controlled by a neighbourhood function. When the radius of the neighbourhood function goes to zero, the SOM algorithm loses its ordering power and acts like K-means clustering. Each node may contain one or more samples from the input data. The distance between the nodes represents similarity or dissimilarity of the underlying samples, i.e., similar data are closer together in the array and the distance between two nodes is related to the dissimilarity of their samples. A weight vector “ \mathbf{W} ” with the same dimension as the input data is associated with each node and will be varied during the training process. This vector is the key factor in determining the position of the nodes in a map. The set of weight vectors $\mathbf{W} \in \mathbb{R}^{m \times d}$ where m is the number of nodes. Geach (2012) presented the application of the self-organized map and discussed its algorithm in detail.

3.3.1 SOM algorithm

Assume we have a dataset which contains vectors, $\mathbf{V} \in \mathbb{R}^{nd}$, and we want to map them on an S1 by S2 map. Sizes of SOMs are arbitrary and there are no rules regarding choosing one over the other. Vesanto (2005) suggested that a total number of $5\sqrt{d}$ neurons is a sufficient size, but users usually choose the size of the grids based on their dataset and their application of the results.

We start by creating $S1 \times S2 = m$ empty neurons. The arrangement of these neurons depends on the map’s topology provided by the user. In the case of 1D maps, since each neuron has two immediate neighbours, the topology of the map does not have any effect on the final result and any topology can be chosen. However, in 2D maps, the shape of the neurons specifies the number of immediate neighbours for each neuron and it is up to user to choose the most suitable shape based on the data. In this paper, we choose hexagonal topology, which gives each neuron six neighbours, and provides more interactions between neurons. Initially a random weight vector, $w \in \mathbf{W}$, will be assigned to each node. The process of creating SOM happens over a series of N iterations (indexed by timestep t), where N is set by the user. During each iteration the weight vectors might change according to the Kohonen learning rule (equation 3). In each iteration, the SOM code:

- (i) chooses a random vector from the dataset ($v \in V$).
- (ii) calculates the Euclidean distance in \mathbb{R}^d space for each node, j , as $D_j^2 = \sum_{i=0}^{i=d} (v_i - w_{ji})^2$, and finds a neuron with minimum D_j , ($D_{j_{\min}}$). This neuron is the winning node and is called the Best Matching Unit (BMU).
- (iii) computes the radius of the neighbourhood of the BMU to find nodes within this radius. The weight vectors of these nodes will be affected in the next steps. The radius of the neighbourhood is arbitrary and can be set to be as high as half of the SOM size. It then decays exponentially over each iteration as

$$r_{\text{BMU}}^t = r_{\text{BMU}}^0 e^{(-t/\tau)} \quad (2)$$

where τ is a decay constant and is usually set to be the same as the number of iterations, N . r_{BMU}^0 and r_{BMU}^t are the radii of the neighbourhood at 0th and t th iteration, respectively.

- (iv) changes the weight vectors of the BMU and all the nodes within r_{BMU}^t as:

$$w(t+1) = w(t) + L(t) \times R(t) \times (v(t) - w(t)) \quad (3)$$

where $L(t) = L_0 e^{(-t/\tau)}$ is the learning factor, which prevents the divergence of the SOM and $R(t) = \exp(-\frac{D_j^2}{2r_{\text{BMU}}^t})$ is the influence rate. $R(t)$ determines how the weight of each node in the neighbourhood of BMU will change.

These steps are then repeated N times.

3.3.2 Creating self-organizing maps

In order to create SOMs, we use the MATLAB neural network toolbox (NNT, Beale et al. 2013), written for MATLAB2015B. An SOM in NNT can be created by the NEWSOM or SELF-ORGMAP libraries, both of which work in two phases, an “ordering phase” and a “tuning phase”. The first phase is called the “ordering phase” and starts with maximum neighbourhood distance and an initial high learning factor (usually

0.9) is provided by the user. The ordering phase continues for a requested number of iterations. During the iterations, the learning factor decreases to the tuning phase learning factor and the neighbourhood distance reaches that of the tuning phase as well. Both the learning factor and the neighbourhood of the tuning phase are set by the user. The amount by which these two factors change in each iteration depends on the number of iterations.

In the second, or “tuning” phase, the neighbourhood distance is kept at the user-defined minimum. The learning factor, however, decreases gradually. The gradual change in the learning factor helps to fine-tune the topology results, leading to a more stable SOM. To allow the fine tuning, the number of iterations in this phase must be much larger than the that of the ordering phase. We chose the number of epochs in the tuning phase to be 3 times the number of epochs in the ordering phase.

We create the final SOMs with initial values for number of iterations in ordering phase, ordering phase learning factor, tuning phase learning factor, and tuning phase neighbourhood distance of 1000, 0.9, 0.02, and 1, respectively. To present our results, we use NNT’s built-in plotting tool. Specifically, we use two of the plots in this tool: a hits map, which shows the number of times each neuron has become the winner (hits), and a distance map, which shows the distance between those neurons. In the maps, the coloured hexagonal shapes represent the neurons. The distances in a distance map are shown by the grey cycle colours: the darker the colour, the larger the distance between neurons. In the hit map, neurons with zero hits are left empty.

As mentioned in the introduction, one of the main advantages of the SOM method is that the resulting networks can be used to cluster new datasets with no additional training required. New data with the same dimensionality as the original input data can be presented to the already trained network: the SOM algorithm finds the best matching unit considering the weight of nodes in the trained network and each vector from the new data set. The winning node determines the place of each new vector on the SOM. The location of the new vectors on the SOM allows them to be compared with the original data set.

3.4 Classification and clustering metrics

As discussed in the Introduction, we compare the classification methods described above to the chi-squared template matching method. We compute the value of chi-squared between each galaxy spectrum and each template spectrum as

$$\chi^2 = \sum_i (\text{spectrum} - \text{template})^2 / \text{template} \quad (4)$$

where the i are all of the tabulated wavelengths in spectra and templates. This treatment is the equivalent of assuming that the K96 template spectra are photon-noise dominated, since they do not have tabulated uncertainties. Each galaxy spectrum is assigned to the template class with the lowest chi-squared value.

We use several metrics to measure the degree of agreement between classification methods. We compute a “chi-squared agreement” score as follows. If a method sorted a galaxy into the same cluster or neuron as its chi-square

matched template, this was considered a match and scored 1 point. For the larger-sized self-organized maps, if a galaxy was classified between two nodes, one of which contained its best-fit chi-square template, this scored 0.5 point. Otherwise, the chi-squared agreement score for that galaxy was zero points. Summing the chi-squared agreement score over all galaxies in the sample indicates how well each technique matched the results of chi-squared fitting. Summing the chi-squared agreement score over all methods for a single galaxy indicates how well the methods agreed for a particular spectrum.

Classification and clustering are related techniques, and we make another comparison between the classification methods with a clustering metric, the silhouette score (Rousseeuw 1987). This score is a metric used to describe cluster compactness and isolation, given by:

$$S = \frac{1}{n} \sum \frac{b - a}{\max(a, b)} \quad (5)$$

where a is the mean distance between a point and the other points in its cluster, and b is the mean distance between a point and the nearest cluster of which that point is not a member. A higher silhouette score corresponds to a partition with better defined clusters. For a given dataset, the score generally declines with the number of clusters and is sometimes used to determine the optimal number of clusters. In this work we use the silhouette score to compare classifications with a comparable number of classes, as a measure of similarity within the classes. Mean values for each classification were calculated using the SCIKIT-LEARN Python package (Pedregosa et al. 2011).

4 RESULTS AND DISCUSSION

In this section we show the results of the neural networks and K-means clustering trained using the K96 template spectra. Training with K96 templates results in networks or clusters that have regions corresponding to galaxies of known morphological type. The trained networks or clusters can then be used to categorize other galaxies.

The first step in categorizing galaxies is choosing the number of classes. In the K-means method empty clusters are not allowed; hence a $K = 12$ clustering would produce one-object clusters as in Fig. 1. We did not want to have too dense or too sparse clusters. Therefore we chose the number of clusters in the K-means clustering method to be $K = 4$. Since empty neurons are allowed in SOMs, finding an optimal number of neurons for SOM networks is more subtle than just specifying a number of clusters. In order to find a sufficient size for the trained neural networks, we created maps with sizes ranging from 1×2 to 50×50 . Varying the grid size of the maps helps us to monitor whether tighter grouping of galaxies is due to their similar properties or a lack of map space to separate them. Based on the size of the data and SOM results, we found the optimum grid size to be 1×22 and 12×12 in 1D and 2D maps, respectively. For each grid we created different SOMs with different learning factors, neighbourhood distances, and number of iterations to find the optimum.

We started our analysis by creating 1D SOMs and K-means clustering. First, we created SOMs with only two neu-

rons (1×2 map), and then increased the number of neurons one at a time (Section 4.1.1). We performed K-means clustering on the K96 templates (Section 4.1.2) and compared the results with the results of the 1×4 -sized 1D SOMs (Section 4.1.2). For 2D networks (Section 4.2), we found the optimum size for the network to be 12×12 and created two types of SOM to show different applications of 2D SOMs. In Section 4.3 we compare the results of the different classification in several ways. For each generated neural network, we compare the results with the K96 categorization. We also use these networks to classify the T12 galaxy sample, and compare this classification with that from the supervised networks in T12 and with K-means clustering.

4.1 One-dimensional self-organizing maps

4.1.1 Training the networks

To start our classification, we assumed that galaxies can be divided into only two general types, quiescent and starburst. This corresponds to a network with only two neurons. We increased the size of the map gradually until the 12 input samples divide into the 12 neurons.

Figs. 2 – 4 show the results of the training networks. The upper panels of the figures show the neurons and the relative distances between their weights. As mentioned in Section 3, an increase in the darkness of colours between neurons represents an increase in relative distance between the neurons. The lower panels of the figures show the number of K96 templates that are placed in each neuron.

In the upper map in Fig. 2, the dark colour between two neurons indicates that the relative distance between these two groups is high, and these two groups are distinguishable. In the lower part of Fig. 2, we see that the templates are divided into two groups of 5 and 7. Although we know from K96 and T12 that 6 of the templates are quiescent and the other 6 are starbursts, the SOM results show 5 of the galaxies in one group and the other 7 in the second group. In this method, the Sc template has been categorized as starburst due to the relatively higher disk stellar population relative to that of the bulge. According to K96, Sc galaxies are considered to be late Hubble type galaxies which have flatter spectra compared to other quiescent galaxies.

Fig. 3 shows the results of the training in a 1×3 network. In these plots we force the galaxies to be categorized in a maximum of three groups. If the templates in Fig. 2 truly belonged in two groups, they would be grouped into two groups in this network, even when we try to cluster them into three. However, in the lower panel of Fig. 3, we can see that the middle node contains two templates. These two templates, which are separated from the group of starburst templates in the lower part of Fig. 2, are the SB5 and SB6 types. In the upper plot of Fig. 3, the colour between two right neurons is black and the colour between two left neurons is white. The black colour indicates that the left neuron has the most differences with the other two groups, while the white colour shows that the two right neurons are very similar to one another.

Comparing Fig. 2 to Fig. 3 shows that the starburst templates are divided into two groups. Based on the colours between these two groups, we conclude that they are more similar to each other than to the ones in the left neurons;

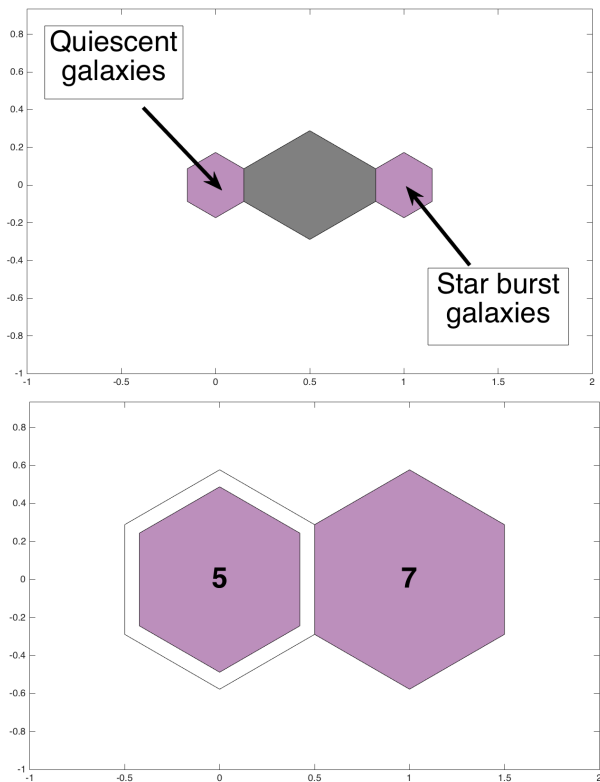


Figure 2. Results of training network in 1×2 grid. The upper panel is a distance map and the lower panel is a hit map. In this network, 5 of the templates from Kinney et al. (1996) are categorized as quiescent galaxies and the rest are starbursts. Because of their strong emission lines, Sc galaxies are moved towards the starburst ones.

both groups are starbursting and have strong emission lines. On the other hand, the SB5 and SB6 templates have the highest internal extinctions; this causes the spectra to become flatter at shorter wavelengths. The flatter UV spectra make these two templates more similar to quiescent galaxies than to other starbursts. Therefore, in the networks, SB5 and SB6 types are grouped close to the quiescent galaxy templates.

We increased the size of the maps gradually until the templates were divided into twelve groups (Figs. B1 to B11 in Appendix B and Figure 4). Since there are more nodes in higher grid SOMs, the algorithm pays more attention to small differences between groups. If the templates from K96 had completely distinct spectral types, then a 1×12 -sized SOM would be expected to show 12 different groups each containing a single template. However, in Fig. B9, we see that three of the neurons contain two templates and three of them are empty. It is evident that there was no template in the K96 sample that can fill those empty neurons. In Fig. B9, from left to right, templates with types B and E, SB3 and SB4, and SB1 and SB2 are the ones grouped together. The SB3 and SB4 grouping breaks when we increase the size of the network to 1×15 (Fig. B10). The SB1 and SB2 templates, however, remain in the same neuron until the size of the map is increased to 1×20 (Fig. B11). The separation between templates B and E only happens when the size of the SOM exceeds 1×22 (Fig. 4).

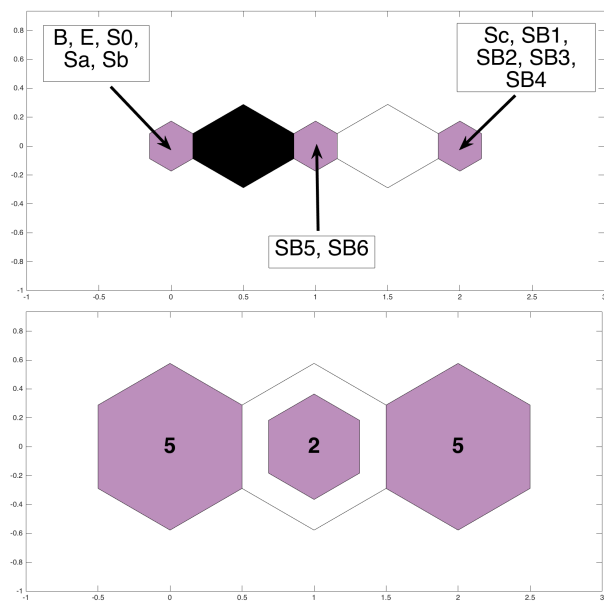


Figure 3. The same as Fig. 2 but showing the results of training network in 1×3 grid. In this network again, 5 of the Kinney et al. (1996) templates are categorized as quiescent and 7 as starbursts. However, this time 2 templates (SB5 and SB6) are separated from the starburst groups.

In Fig. 4, we can see twelve distinct groups. Five groups in the left-side neurons are separated from the rest of the groups with two dark-grey colours between them. Five groups in the right-side neurons are separated from the rest by the black colour, and Sc and SB6 galaxies are in two separate blocks. This figure is a good example of the problem of using one-dimensional SOMs to explore the data. Since the templates are forced to be arranged in a linear array, the SB6 and Sc groups are isolated and not connected to any other groups. However, in a two-dimensional map, due to the connection of each neuron to 6 more neurons, template galaxies have more degrees of freedom to find their relation with other neurons. As a big picture, Fig. 4 shows that even though the galaxies are clustered in twelve groups, there still remain only two main distinct groups. The closest occupied neuron in the starburst side of the SOM, belongs to the SB6 type. This template has the most extinction and the general shape of its spectrum has more similarity to quiescent galaxies than to other starburst types.

The fact that the templates need to have at least 22 different neurons to be divided into 12 groups shows that the differences between SB1 and SB2, and B and E, templates are very small. Because of their similarities, they tend to stay in the same group until the network becomes large enough to pay attention to the smallest particularity.

4.1.2 One-dimensional self-organizing maps and K-means clustering: templates

We performed K-means clustering on the K96 template spectra using 4 clusters. Fig. 5 shows the clustering of the template spectra into 4 groups using both K-means and SOMs. Both methods separated the most starburst-like and quiescent galaxies in a similar manner. However, they show dis-

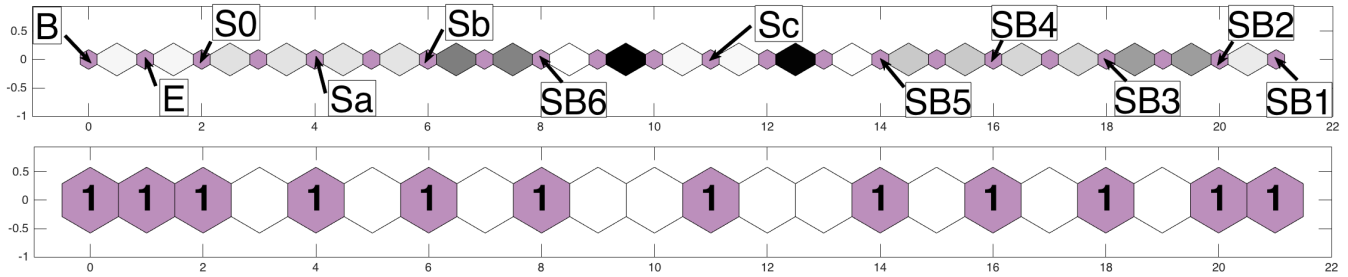


Figure 4. The same as Fig. 2 but this time the figure shows results of training network in 1×22 grid.

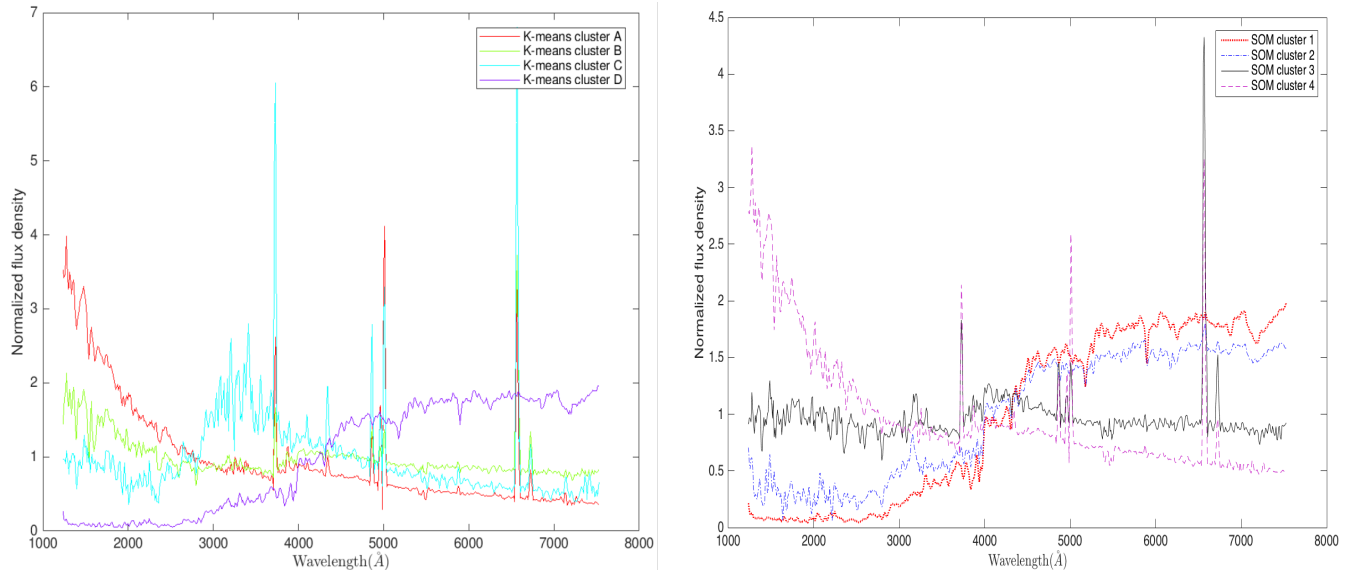


Figure 5. Clustering the K96 template spectra using K-means clustering (left) and SOM (right). Note that both methods randomly assign the initial values for their analysis, therefore the cluster labels are only used for distinguishing cluster membership.

crepancies for spectral types between these extreme cases. Table 2 shows the group assignments of the K96 templates in the two methods. Templates B, E, S0 and Sa are grouped in the same cluster/neuron in both K-means and 1×4 -sized SOM. K-means grouped the Sb template with the above templates, but the SOM network classified it as a separate group, possibly because it is much more noisy in the ultraviolet range. The K-means method clustered the Sc template as a separate group, likely due to its distinct features in the 3000–4000 Å range, while the SOM network grouped it with SB5 and SB6. Classifying the Sc template in the group with SB5 and SB6 obscures the peak feature in the Sc template. However, when the number of SOM nodes is increased (e.g. in Fig. 4), the Sc template separates and occupies a single neuron. Another noticeable difference is that K-means clustered the SB1 and SB2 templates in one group and SB3 to SB6 in another, while the SOM network grouped the SB1 to SB4 templates in one neuron and SB5 and SB6 in another. SB1 and S2 have similar strong emission lines in the blue, while SB3–6 show H α emission but only weaker emission lines at shorter wavelengths; the SB5 and SB6 templates are flatter with a less-pronounced 4000 Å break than the remaining SB templates. By comparing the membership of different clusters in Table 2 we can see that K-means cluster-

K96 template	K-means group	SOM group
B	C	1
E	C	1
S0	C	1
Sa	C	1
Sb	C	2
Sc	D	3
SB1	B	4
SB2	B	4
SB3	A	4
SB4	A	4
SB5	A	3
SB6	A	3

Table 2. Group assignments for K96 template galaxies, using K-means and 1×4 -sized SOM methods

ing emphasizes the input templates’ similarities, while SOM emphasizes the templates’ differences.

4.1.3 Classifying the galaxy sample

After training with the templates, we used both K-means and self-organizing maps to classify the 142 galaxies in the

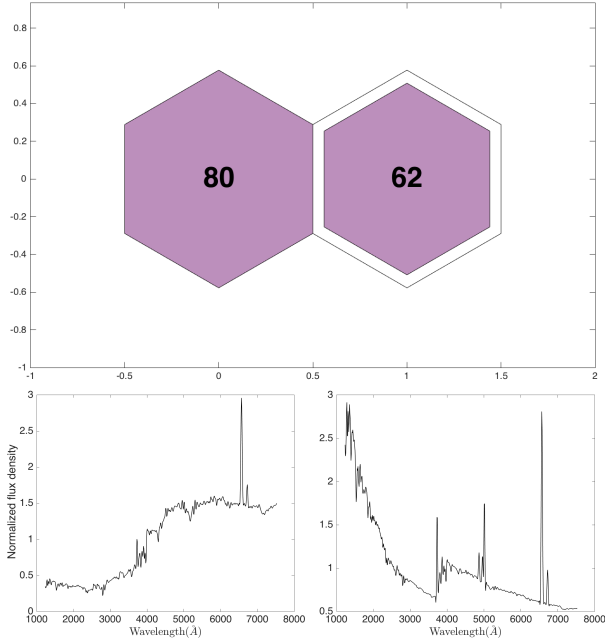


Figure 6. Classification of fitted galaxy spectra from Teimoorinia (2012) using the 1×2 network trained from the Kinney et al. (1996) templates (Fig. 2). Upper panel: a hit map with the number in each node representing the number of galaxies belonging to that group. In this case 80 means that 80 of the 142 spectra are classified into the quiescent group while the 62 of the 142 spectra are categorized as starburst galaxies. Lower panel: median spectra of the galaxies in each group (flux densities are normalized by dividing spectra by flux density at 4000\AA), quiescent on the left and starburst on the right.

T12 sample. K-means classification was performed by assigned each galaxy to the K-means cluster with the closest centroid and the results are discussed in Section 4.3. This section examines the results of one-dimensional SOM-based classification in detail.

The upper panel in Fig. 6 shows the result of this classification using the 1×2 network from Fig. 2. Eighty of the galaxies have spectral types similar to those of the quiescent galaxies and the spectral types of the rest are similar to the starburst galaxies. The lower left panel shows the median spectrum of the 80 galaxies that were classified as quiescent. These galaxies are similar to the ones in the left node in the upper panel of Fig. 2: the spectra clearly show the 4000\AA break, one of the signatures of quiescent galaxies. The $H\alpha$ emission in the spectrum could be from galaxies with similar spectral types to Sa galaxies, but with stronger emission lines. The median spectrum in the lower right panel of Fig. 6 shows strong emission lines and bright ultraviolet continuum, indications of a high star formation rate.

Since in this network galaxies were forced to be divided into a maximum of two groups, the strongest feature in a galaxy’s spectrum predominantly decided to which group the galaxy belonged. Galaxies with a weak 4000\AA break, strong emission lines and ultraviolet upturn are categorized as starbursts while galaxies with a strong 4000\AA break are categorized as quiescent. Increasing the size of the SOM helps solve the problem of galaxies which have features that are common in both groups.

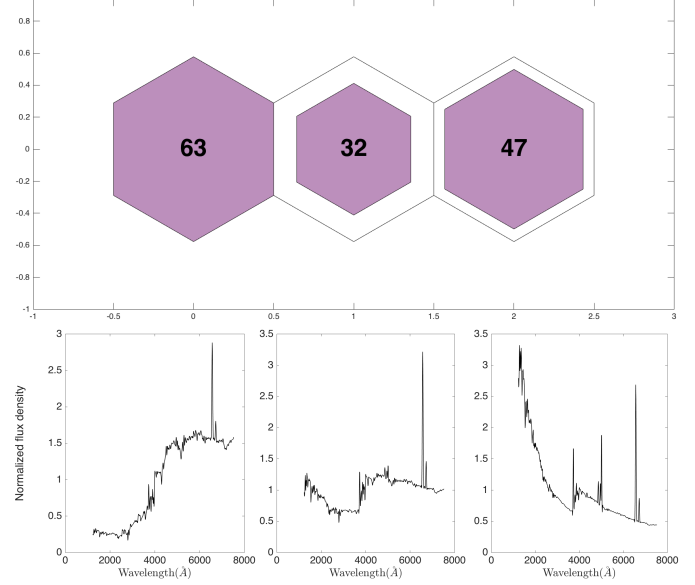


Figure 7. Same as Fig. 6, but in this figure, we used a network with size of 1×3 (Fig. 3) to classify the sample galaxies.

Fig. 7 presents the result of classifying the spectra of the galaxies using the 1×3 network (from Fig. 2): 66 of the galaxies belong to the quiescent group, and 47 belong to the starburst group. However, 29 of the galaxies are similar to starbursts in some, but not all, of their features. Galaxies in this group have strong emission lines and are ultraviolet-bright, but they also have a strong 4000\AA break, which makes them cluster closer to the quiescent galaxies (middle panel in the lower part of Fig. 6).

In Fig. 8, we use the 1×22 network to classify the sample galaxies. As mentioned in Section 4.1.1, in this network size we observed the first separation of the K96 templates into 12 different neurons. As in Figs. 6 and 7, the upper panel of Fig. 8 shows the number of galaxies (out of the 142) belonging to each neuron in the 1×22 SOM. The lower panels present the median spectra of the galaxies in each neuron. Since galaxies had more space to separate, some of the neurons are left empty. Thus, instead of having 22 median spectra in the lower part of Fig. 8, there are only 14.

Comparing the upper panel of Fig. 8 with the lower panel of Fig. 4 shows that the occupied neurons are not necessarily the same. If a cluster from the T12 sample fills the same neuron as a K96 template, we can conclude that the spectra of galaxies in the cluster are very similar to those of the template. Otherwise, we can conclude that the spectra of galaxies in the cluster are (dis)similar to two of the K96 templates. For the latter case, the colours establish which template has greater similarity to the T12 cluster.

The first two neurons in the upper plot in Fig. 8 are empty, while these neurons were occupied by galaxies B and E in the trained network. We therefore conclude that there are no galaxies in the T12 sample with spectral type similar to that of type B or E. There are 3 galaxies in the third neuron, similar to the S0 type. 19 of the galaxies are in the fifth neuron and similar to the Sa type, while the 6 galaxies in the fourth neuron have spectral type similar to both S0 and Sa galaxies. There are 8 Sb type galaxies and 19

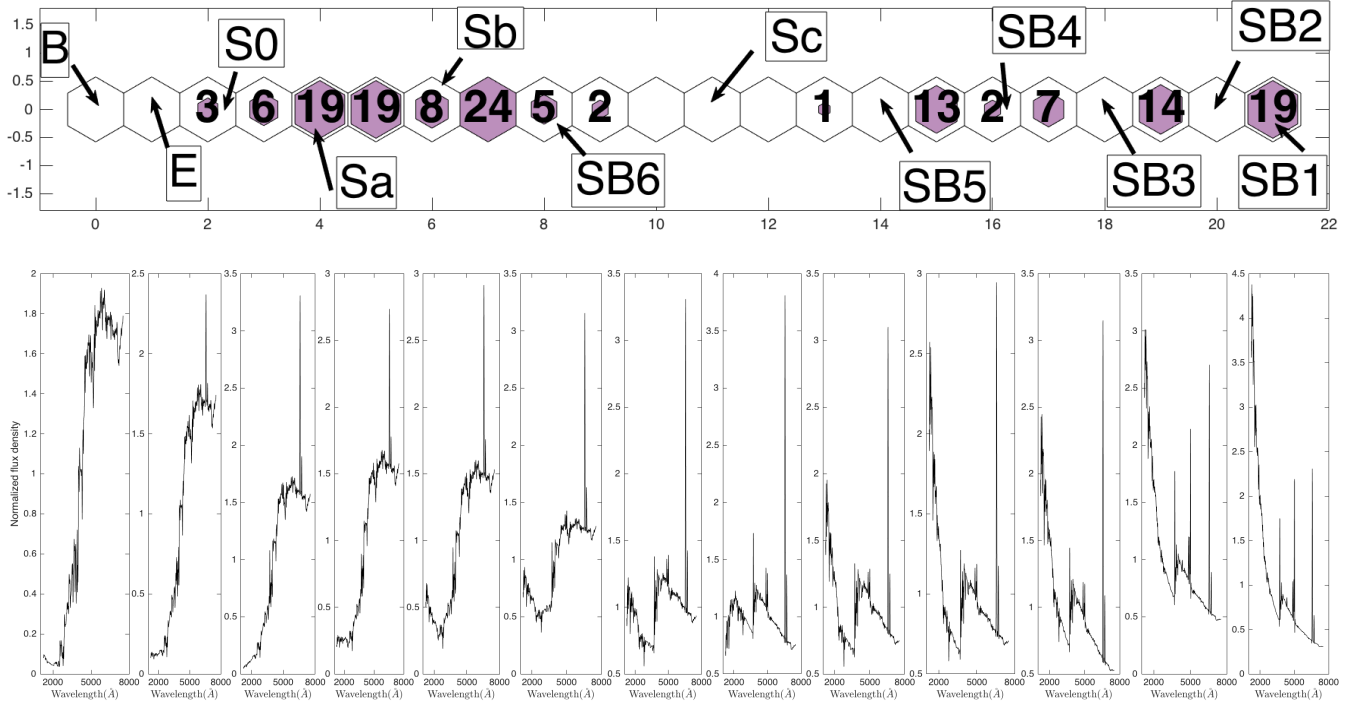


Figure 8. Same as Fig. 6, but in this figure, we used a network with size of 1×22 (Fig. 4) to classify the sample galaxies. In the median spectra of each neuron, we can clearly see the changes from quiescent galaxies to the starburst ones. The 4000\AA break weakens from left to right and the emission line strengths increase.

galaxies with spectral types similar to both Sa and Sb. The following two neurons (eighth and ninth neurons from the left) are on the edge of the quiescent and starburst galaxies. In the upper panel in Fig. 4, the colour between these two neurons is black, which shows that these two are very different. Therefore, 24 galaxies in the eighth neuron have similar spectra to Sb type galaxies, and the spectra of the 5 galaxies in the ninth neuron are similar to the SB6 galaxies. The far right neurons in the network belong to galaxies with spectra comparable to types SB1 and SB2.

4.2 Two-dimensional self-organizing maps

The one-dimensional networks are great tools to categorize the spectral types of galaxies and monitor the changes of properties of galaxies with category. However, each neuron in these networks is limited to connecting to one other neuron in each direction. In a 2D network, each neuron can have more than two immediate neighbours, providing a more complete picture of relations between spectral types. One of the other key features of SOMs compared to supervised classification is the ability to both interpolate between and extrapolate outside of training set classes. This permits a fuller exploration of dataset properties and detection of outliers.

As described in Section 3, one of the main advantages of the SOM is that when the weight of one neuron is adjusted after finding the best matching unit, the weight of the whole map will be changed. This quality of the SOMs provides a unique opportunity to analyse 2D networks using two approaches. At first, as in the 1D SOMs, we assumed that all the galaxies must have spectra similar to any of

the 12 templates in Fig. 1 (i.e. no outliers). In this case, we can categorize spectra of other galaxies based on those 12 types. In the second approach, we trained networks with the assumption that there might be completely different types of galaxies not encountered before. This approach can be used to trace and recognize the outliers, which could include galaxy types that are not present in the K96 templates (e.g. dwarf galaxies, luminous infrared galaxies, or active galactic nuclei). To execute the second approach, we used the SELF-ORGMAP library in MATLAB. Unlike the NEWSOM library in which the ordering step neighbourhood distance is set by default (although in an SOM it should be a free parameter), in SELF-ORGMAP we can change this value. Using the SELF-ORGMAP code, based on the size of a map and the ordering step neighbourhood distance, we can arrange that the data spread all over the SOM or sit close together, allowing room for outliers.

To get the best result from both methods, we should have SOMs with “enough” neurons to give any new types of galaxies a chance to find their places in the map. “Enough neurons” is a vague statement, but as mentioned in Section 3 the size of SOMs is arbitrary and depends on the size of the input data and the kind of information needed from the SOMs. Since the training dataset (the K96 templates) has 12 members, we assumed a SOM with size of 8×8 would be a good start. We then increased the size of the SOM to find the highest size suitable for the training sample. We noticed that very few changes occurred after the SOMs were grown to a size greater than 12×12 , so we used this as the maximum size.

The upper left panel in Fig. 9 shows the SOM results

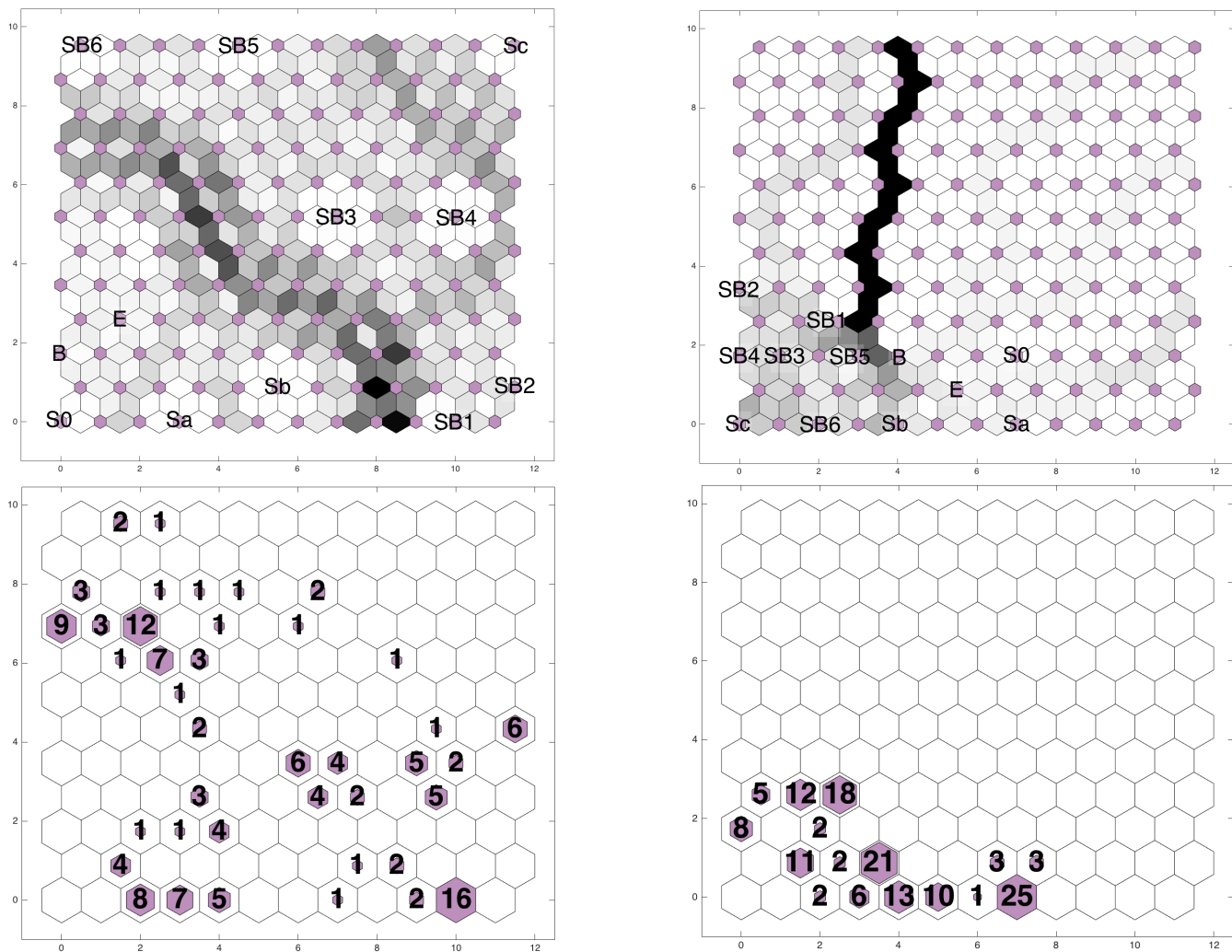


Figure 9. 12×12 2D SOM results. Upper panel: SOMs trained using the Kinney et al. (1996) templates. For training 2D SOMs two different approaches were considered: either only 12 types of galaxies exist (left) or not (right). Lower panel: classifying the galaxies from Teimoorinia (2012), using the trained networks shown in the upper panels. From the lower right panel, we can see that there are no outliers in the galaxies from T12, and we can use the map in the lower left panel as a final clustering result for the T12 galaxies.

from the first approach. Since we considered that spectra of all galaxies can be categorized using the K96 templates, the galaxies were placed all over the map. Using this network to categorize any set of spectra forced the spectra to be either in the same neurons as the K96 templates or in the neurons between them. In the map shown in the upper right panel in Fig. 9, the K96 templates are in a small region. This provides enough freedom for the spectra of galaxies to be placed everywhere in the map, even far away from the templates.

In the upper left part of the map in Fig. 9, although galaxies have more ways to be separated, they were separated into two main groups. There is a distinguishable strip of grey, dark grey and in some cases black colour in the map: this strip separates quiescent galaxies from starburst ones. The lower map shows that 5 of the neurons on the left side of the strip are full. These five neurons are the same as the ones on the left hand side of Fig. 2 (quiescent galaxies), the only difference being that in this map they have more space to be separated from each other. When we use this network

to categorize the spectra of galaxies, any galaxy placed on the left hand side of the strip is a quiescent galaxy and any one placed on the right side of the strip is a starburst one. The decision about what sub-type of quiescent or starburst galaxy is based upon its position relative to each type in the SOM. T12 showed that by masking emission lines (i.e., by ignoring details in spectra) two types of galaxies, Sb and SB6, can be misclassified as one another. As mentioned before, few-node, one-dimensional SOMs are a good method to obtain a big picture of the problem under study with less emphasis on fine detail. For example, the SB6 category is placed near Sb galaxies in the 1D network in Fig. 4. However, in the upper left panel of Fig. 9, we can see that the Sb and SB galaxies are completely separated. This shows that more details of spectral shapes are considered by 2D networks, which is one of the main advantages of using 2D maps over 1D ones. In this case, the two dimensions of the self-organizing map correspond approximately to quiescent versus star-forming (left to right) and less to more extinction (bottom to top).

As in the other map, in the upper right panel of Fig. 9, galaxies are generally divided into two main groups. The border between the quiescent galaxies and the starburst ones is the black strips in the middle of the map, ending with the bright grey colour at the bottom of the map in the fifth neuron. In this network, neurons in the right side of the strip represent the quiescent galaxies and the neurons in the left side represent the starburst ones. When categorizing a new set of spectral types, if the new spectra are similar to the K96 templates all of them will be placed in the bottom of the map, but if the new spectra represent different types of galaxies, they would sit in any other neurons in the map. In large datasets, one can easily use this network to figure out whether there is any of new type of spectra (or any outliers) in the datasets. Since the networks are already available, this procedure should be quick (from a few seconds to a few minutes) and easy for big datasets.

We used both 2D networks to categorize the T12 galaxies and the results are shown in the lower panels of Fig. 9. In the lower right map in Fig. 9, all galaxies are placed in the bottom part of the map and we can conclude that in the T12 sample there are no outliers or spectra with very different from the K96 templates. The lower left map of Fig. 9 shows the T12 galaxies categorized based on the network in the upper left of Fig. 9. Comparing this categorization with the 1D one from Fig. 8, we can see that only 23 galaxies correspond exactly to K96 types. Using 2D maps results in categorizing galaxies into more intermediate types than in 1D maps. In both lower maps in Fig. 9, most galaxies are in the quiescent side of the SOM, which was predictable from the results in Section 4.1.3. This does not necessarily imply that in general there are more quiescent galaxies at higher redshift but could be a selection effect for galaxies that had more reliable redshift estimates.

Although for ease of presentation this paper first discusses 1D networks and continues to 2D networks, we suggest that users of SOMs for galaxy spectral classification should use 2D networks first. These can be used as a first step with networks similar to the upper left map in Fig. 9 to identify outliers in the sample. After finding and removing outliers in the input data, the use of 1D and 2D maps is complementary. One-dimensional maps are helpful for exploring the general and ordered behaviour of the data, while 2D maps can show a more detailed picture.

4.3 Comparing classification methods

The one and two-dimensional self-organizing map classifications are compared briefly above. The 2-D maps allow galaxies to be members of more intermediate classes, and this is a general advantage of SOMs compared to the artificial neural network method used by T12: in the SOM method galaxies do not need to exactly match template spectra. T12 found that the ANN method could only match 105 out of the 142 galaxies directly to the K96 templates; classifying the remaining 37 galaxies required combining template spectra. Below we compare various other classification methods with the results of the 1-D self-organizing maps using the metrics described in Section 3.4. Because the 2-D SOMs subdivide the sample into a larger number of groups, each containing a small number of objects, they are not directly comparable to the other methods and are not included here.

Method	chi-squared agreement score
K-Means	124
1 × 4-sized SOM	103
1 × 12-sized SOM	111.5
1 × 22-sized SOM	79
supervised ANN	52

Table 3. Chi-squared agreement score for different methods

Method	Silhouette score
K-Means	0.549
1 × 4-sized SOM	0.371
1 × 12-sized SOM	0.302
1 × 22-sized SOM	0.249
chi-squared fitting	0.221
supervised ANN	0.142

Table 4. Silhouette scores for different methods

To compare results from SOM, K-means, and supervised ANN, we first consider how well these methods matched a simple chi-squared fitting classification. Matching galaxy spectra to templates using chi-squared, we found that not all of the templates had matches among the galaxy sample: no galaxy spectra were best-matched to the B, E, or Sc templates. This is consistent with the results from the self-organized maps. Summing the chi-squared agreement score over all galaxies in the sample allows us to compare how well each technique matched the results of chi-squared fitting.

The results, given in Table 3, show that the K-means method and 1 × 12-sized SOM have the most similarity with minimum chi-square fitting.

Summing the chi-squared agreement score over all methods for a single galaxy shows how well the methods agree for a particular spectrum. For example, if the total score of a galaxy is 5, it means that all techniques classified the galaxy into the same group. Comparing the scores, we found that 12 per cent of galaxies scored 5, 36 per cent of galaxies had a score greater than 4, and 56 per cent had a total score of 3 or more. The score-5 galaxies classified similarly by all methods were primarily (60 per cent) in the Sa template group. The score ≥ 3 galaxies were predominantly SB1, Sa, and Sb type (32, 28, 18 per cent, respectively). The different methods show the most agreement for galaxies similar to the SB1 and Sa templates, with less agreement on the high-extinction starburst types (SB5, SB6).

A final comparison between the classification methods is made with the silhouette score (Rousseeuw 1987), given in Table 4. The highest score among our classification was for K-means; however as discussed in Section 3.4, the silhouette score usually declines as the number of groups increases, so we cannot conclude from score alone that K-means is the superior method. However, the 1 × 12-sized SOM, chi-squared fitting, and supervised neural network all have either 8 or 9 classes. In this case we can compare scores directly and conclude that the self-organized map yields groups with more internal similarity. This is in contrast to the comparison between K-means and self-organized maps for the templates alone (see Section 4.1.2) where K-means clustering produced

groups more similar to each other, and emphasizes that the exact data used in a clustering process can strongly affect the outcome.

4.4 Properties of spectral groups

To check whether the classifications described above are meaningful, in this section we examine the relations between median properties of the galaxies in each group. Fig. 10 shows the median of stellar mass, age, specific star formation rate (sSFR; star formation rate per stellar mass), and f_{burst} of the galaxies in each neuron in the 1×2 , 1×4 , and 1×8 self-organized map networks. In all plots, the horizontal axis is the number of the neurons divided by the size of the network. As shown in Fig. 10, in all three networks, stellar mass and age decrease while sSFR and f_{burst} increase as the type changes from quiescent galaxies to starburst. Separating galaxies based on spectral types also leads to a separation in properties derived (via *CIGALE*) from the SEDs, as expected since the spectral types are also based on *CIGALE* fitting.

CIGALE has various models to derive the properties of galaxies. Through these models, some of the properties are already known to be correlated with each other, e.g., stellar mass and star formation rate. N09 studied other relations between properties with no direct correlation in the models in a sample of SINGS galaxies. They found a tight correlation between sSFR and t_{D4000} , which suggests that younger stellar population correlates with higher SFR. They also found correlations between stellar mass and SFR, and stellar mass and t_{D4000} . Since in the *CIGALE* code, stellar mass is a free parameter, N09 argued that any stellar-mass-related correlation must be astrophysically meaningful. They also studied relations between the attenuation at 1500 \AA (A_{FUV}) and sSFR, age, and stellar mass and did not find any correlation. T12 replicated the upper plots in Fig. 11, and found a tighter correlation than in the N09 results.

To compare our classifications with previous work, Fig. 11 shows plots similar to those shown by both N09 and T12. The points represent the median values of the properties of the galaxies in each group, with different symbol types representing classification method and error bars indicating the standard deviation of those properties. The colour scale represents neuron number in the SOM method, shading from quiescent to starburst. The medium grey (purple in the online version) point in the middle of the black (blue in the online version) points in Fig. 11 corresponds to galaxies with spectra similar to type Sc: the shape of their spectra indicates that they are old galaxies with high sSFR.

Overall, Fig. 10 and the upper panels in Fig. 11 show the same trends as noted by N09 and T12: galaxies with lower mass tend to be younger and more active. The median group properties as derived by all of the methods show high Pearson correlation coefficients ($r^2 \gtrsim 0.9$) between the quantities shown. The ordering of groups in the SOM maps also corresponds to an ordering in age: from older to younger galaxies, the colour of the points changes from light grey to black (red to blue in the online version).

N09 studied relations between A_{FUV} and other properties of galaxies and found that the attenuation had no dependence on the specific star formation or age. In contrast to N09, we find a correlation between A_{FUV} and these

two parameters, shown in the lower panels of Fig. 11. Although weaker than the correlations above, Pearson correlation coefficients are higher for the SOM groups than for other methods, particularly for the 1×12 -sized SOM. The slightly different distribution of this particular classification (see Fig. B9) classifies the most quiescent objects as similar to slightly more active galaxies, lessening the turnover in properties as A_{FUV} increases.

The general sense of the correlation between A_{FUV} and sSFR in Fig. 11 is similar to the trend found by Dale et al. (2007). They used IR luminosity to UV luminosity ratio ($L_{\text{IR}}/L_{\text{FUV}}$) as a measure of A_{FUV} and compared it with sSFR for all 75 galaxies in the SINGS survey. They found that for quiescent galaxies (E, S0 and S0/a), $L_{\text{IR}}/L_{\text{FUV}}$ (or A_{FUV}) correlates with sSFR, and for spiral galaxies, there is an anticorrelation between $L_{\text{IR}}/L_{\text{FUV}}$ and sSFR. Since our sample has no E-type galaxies, we cannot confidently claim the same relation between A_{FUV} and sSFR. However, Fig. 11 does indicate an increase in A_{FUV} with increasing sSFR for the quiescent types and a decrease with increasing sSFR for the starburst types. Both Dale et al. (2007) and N09 argued that these apparent trends can be a result of the dependence of star formation history on $L_{\text{IR}}/L_{\text{FUV}}$. Whether this dependence is real or not, our results here show that using self-organizing maps can separate spectra of galaxies in such a way that the characteristics of each of these groups are in agreement with the general picture of galaxy evolution.

One of the key features of SOMs is the ability to not only classify objects but to provide an *ordered* classification. K-means clustering and supervised ANNs can sort galaxy spectra into classes but do not naturally order them. Ordering is important in the case where the physical causes of the differences between classes/clusters are unknown. Average properties of objects within clusters (possibly from data not used for classification) can be used to test hypotheses about the physical differences. Figures 10 and 11 show that the SOM ordering also corresponds to the ordering of the physical properties of the galaxy models used to make the classified spectra. The median properties of the spectral groups change on going from one end of the SOM to the other. Therefore, by using the SOM method, we can study not only how galaxies group into spectral clusters, but the trends between the clusters as well.

5 SUMMARY

Self-organizing maps can be used to classify celestial objects (e.g. stars, quasars, spectra of galaxies, light curves, etc.). In this work we presented a detailed comparison between different types of self-organizing maps and other methods used to classify galaxy spectra. Based on our experience we suggest some general guidelines for use of SOMs in galaxy spectral classification. If a broad and general classification is required, networks can have one dimension with a few neurons. If one needs more detailed classifications, a higher number of neurons should be used. Since self-organizing maps do not include the uncertainty of input parameters, sometimes too much attention to detail can cause problems in classifications. When using the SOM method, one should consider whether small differences between objects are phys-

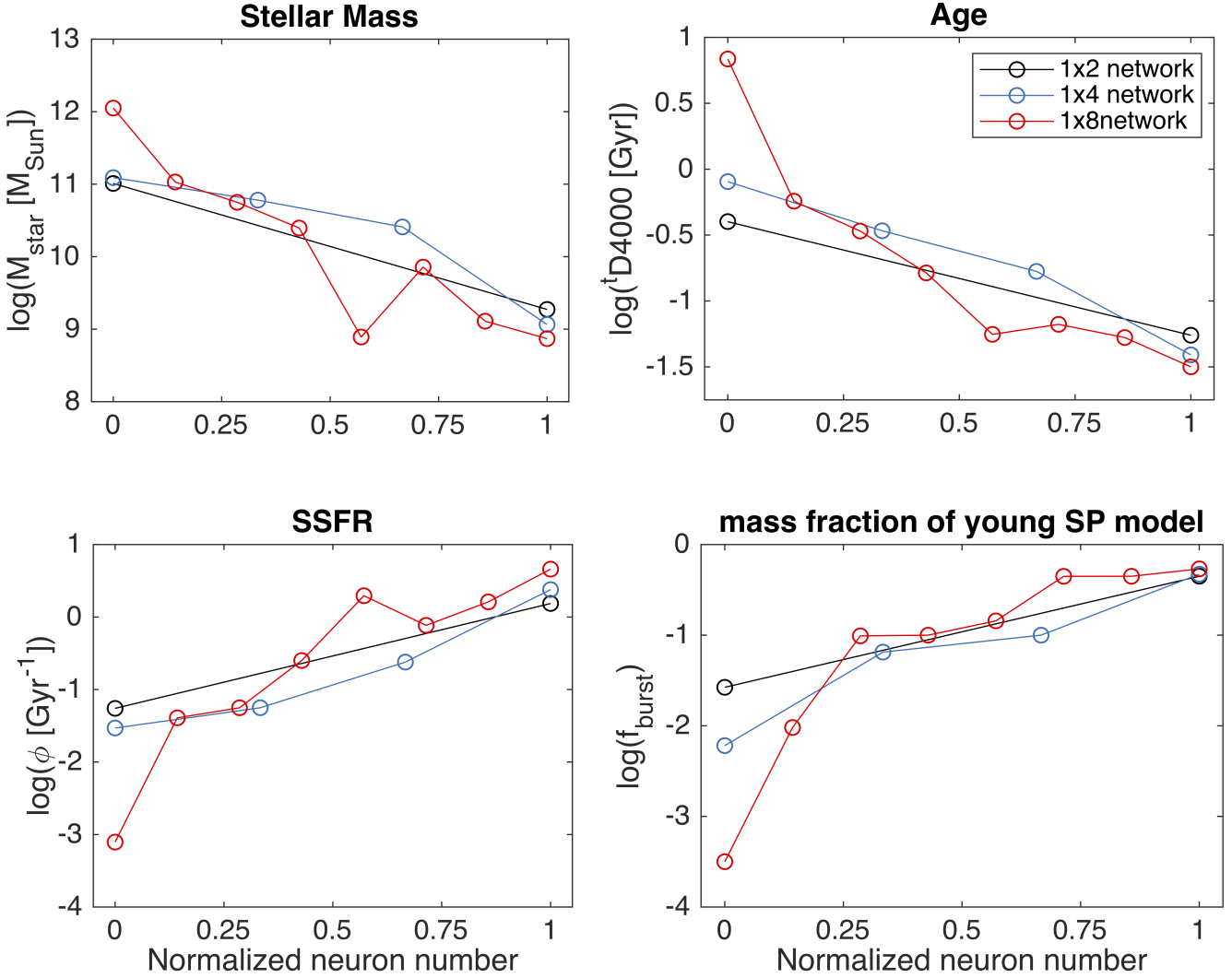


Figure 10. Comparing the median of four properties of galaxies in each node in 1×2 (black circles), 1×4 (blue diamonds), and 1×8 (red triangles) networks.

ically meaningful when separating two groups from each other.

We used SOMs to classify the template spectra of K96, made from galaxies with known morphological type, and created networks with different uses. By varying the size of the networks, we found the relative similarity between the K96 template classes, which can be roughly ordered in one dimension by their amount of star formation. A one-dimensional network with 22 neurons was needed in order to separate all 12 K96 spectra; we concluded that K96 types B and E, and types SB1 and SB2, are very similar to each other. Two-dimensional networks allow more freedom in the galaxy spectral classification. Training 12×12 neurons with the K96 spectra, we found that the two dimensions of the resulting self-organized maps corresponded roughly to strength of star formation and amount of extinction. The one-dimensional self-organizing map ordering combined these two properties, with the highest-extinction template SB6 appearing between two less-extincted templates (Sb and Sc).

A sample of 142 high-redshift galaxy spectra from T12 was classified by the trained networks. This particular sam-

ple of high-redshift galaxies is well-described by the range of the K96 templates although many of the spectra fall in between template classes. In the two-dimensional network, only 23 galaxies occupied exactly the same neurons as the K96 template spectra (in one-dimensional networks this number was 56). The freedom of having in-between types is one of the main differences between supervised and unsupervised artificial neural networks: the supervised training method used by T12 could not classify 37 out of 142 spectra in galaxy sample as matching one of the same set of templates.

Comparing different classification methods, we found that spectra classified similarly by all methods were well-matched to quiescent templates (K96 Sa and Sb). Spectra with emission lines or ultraviolet emission showed more variable classification results. As measured by a chi-squared agreement score, classification produced by self-organized maps showed closer correspondence to the classification from a chi-squared match than results from the supervised ANN method. One-dimensional self-organized-map based classification resulted in a higher silhouette score than classifica-

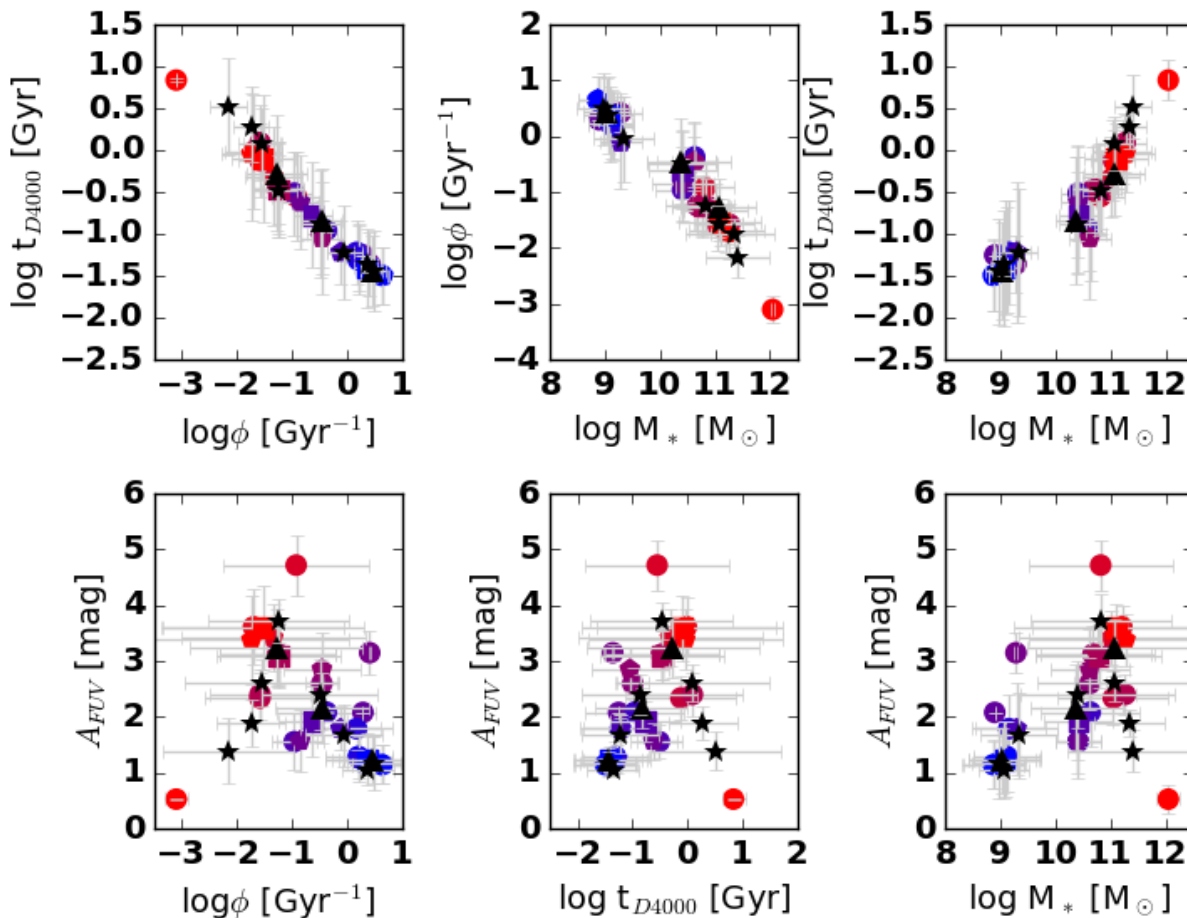


Figure 11. From top to bottom: relations between age (t_{D4000}) and sSFR (ϕ), sSFR and stellar mass, age (t_{D4000}) and stellar mass, and A_{FUV} with sSFR, age (t_{D4000}) and stellar mass. Circles, pentagons, and squares show the median values of these properties of galaxies in the 1×22 , 1×12 and 1×4 SOM networks; triangles show the median value of properties for each K-means cluster (using $K = 4$) and stars show median values for groups from T12. Error bars show standard deviations of the group properties. Colours show the galaxy types in the SOM classification. The lightest grey (most pure red in the online version) colour shows E type galaxies and the black (most blue in the online version) indicates the SB1 galaxies. The shading of other colours between light grey and black (red and blue in the online version) measures the location of galaxies along the SOM from quiescent to starburst.

tion using chi-squared minimization or the supervised neural network, indicating that the galaxies classed together by the SOM had more similar spectra.

The relations between group-averaged properties of the sample galaxies using the SOM-based classification were found to be consistent with those from previous studies. The self-organizing map method can order the spectra in such a way as to also match the order along the age/star formation rate and specific star formation rate-stellar mass relations. Such an ordered classification does not naturally arise from other methods such as K-means clustering and supervised artificial neural networks. Ordered classification can be used to investigate underlying physical causes of changes in properties between groups and for this reason we conclude that self-organized maps can be a highly useful tool in exploratory analysis of astronomical spectra.

ACKNOWLEDGMENTS

The authors thank the referee for thorough and insightful comments which helped us to improve the work. The authors thank S. Lianou, A. Tammour, S.C. Gallagher, R.G. Abraham, M. Daley, M. Gorski and A. Sigut for their useful comments. S.R. and P.B. also acknowledge research support from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., et al. 2009, *ApJS*, 182, 543
- Allen J. T., Hewett P. C., Richardson C. T., Ferland G. J., Baldwin J. A., 2013, *MNRAS*, 430, 3510

- Andreon S., Gargiulo G., Longo G., Tagliaferri R., Capuano N., 2000, *MNRAS*, 319, 700
- Balastegui A., Ruiz-Lapuente P., Canal R., 2001, *MNRAS*, 328, 283
- Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, *MNRAS*, 348, 1038
- Beale M., Hagan M., Demuth H., 2013, *MATLAB Neural Network Toolbox User's Guide*, MathWorks, Version 8.0.1 (Release 2013a)
- Bershady M. A., Jangren A., Conselice C. J., 2000, *AJ*, 119, 2645
- Bishop C. M., 2006, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA
- Boersma C., Bregman J., Allamandola L. J., 2014, *ApJ*, 795, 110
- Brett D. R., West R. G., Wheatley P. J., 2004, *MNRAS*, 353, 369
- Carrasco Kind M., Brunner R. J., 2014, *MNRAS*, 438, 3409
- D'Abrusco R., Fabbiano G., Djorgovski G., et al. 2012, *ApJ*, 755, 92
- Dale D. A., Gil de Paz A., Gordon K. D., Hanson H. M., Armus L., Bendo G. J., Bianchi L., Block M., Boissier S., et al. 2007, *ApJ*, 655, 863
- Delchambre L., 2016, *MNRAS*, 460, 2811
- Ellison S. L., Teimoorinia H., Rosario D. J., Mendel J. T., 2016, *MNRAS*, 455, 370
- Folkes S. R., Lahav O., Maddox S. J., 1996, *MNRAS*, 283, 651
- Fustes D., Manteiga M., Dafonte C., Arcay B., Ulla A., Smith K., Borrachero R., Sordo R., 2013, *A&A*, 559, A7
- Geach J. E., 2012, *MNRAS*, 419, 2633
- Gulati R., Gupta R., Singh H., 1997, *PASP*, 109, 843
- Hernandez-Pajares M., Floris J., 1994, *MNRAS*, 268, 444
- Holden B. P., Oesch P. A., González V. G., Illingworth G. D., Labbé I., Bouwens R., Franx M., van Dokkum P., Spitler L., 2016, *ApJ*, 820, 73
- Hutchinson T. A., Bolton A. S., Dawson K. S., Allende Prieto C., Bailey S., Bautista J. E., Brownstein J. R., Conroy C., Guy J., Myers A. D., Newman J. A., Prakash A., Carnero-Rosell A., Seo H.-J., Tojeiro R., Vivek M., Ben Zhu G., 2016, *AJ*, 152, 205
- Ilbert O., et al., 2006, *A&A*, 457, 841
- in der Au A., Meusinger H., Schalldach P. F., Newholm M., 2012, *A&A*, 547, A115
- Karamelas A., Kontizas M., Rocca-Volmerange B., Bellas-Velidis I., Kontizas E., Livanou E., Tsalmantza P., Dapergolas A., 2012, *A&A*, 538, A38
- Kinney A. L., Bohlin R. C., Calzetti D., Panagia N., Wyse R. F. G., 1993, *ApJS*, 86, 5
- Kinney A. L., Calzetti D., Bohlin R. C., et al. 1996, *ApJ*, 467, 38
- Kohonen T., 1982, *Biological Cybernetics*, 43, 59
- Kohonen T., 1998, *Neurocomputing*, 21, 1
- Laporte N., Infante L., Troncoso Iribarren P., Zheng W., Molino A., et al. 2016, *ApJ*, 820, 98
- MacQueen J., 1967, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: Statistics Some methods for classification and analysis of multivariate observations. University of California Press, Berkeley, Calif., pp 281–297
- Maehoenen P. H., Hakala P. J., 1995, *ApJL*, 452, L77
- Mannucci F., Basile F., Poggianti B. M., Cimatti A., Daddi E., Pozzetti L., Vanzi L., 2001, *MNRAS*, 326, 745
- Marquez L., Hill T., Worthley R., Remus W., 1991, in *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences Vol. 4*, Neural network models as an alternative to regression. pp 129–135
- Meusinger H., Schalldach P., Mirhosseini A., Pertermann F., 2016, *A&A*, 587, A83
- Miller A. S., Coe M. J., 1996, *MNRAS*, 279, 293
- Murtagh F., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E., eds, *Astronomical Data Analysis Software and Systems IV Vol. 77 of Astronomical Society of the Pacific Conference Series*, Unsupervised Catalog Classification. p. 264
- Noll S., Burgarella D., Giovannoli E., Buat V., Marcellac D., Muñoz-Mateos J. C., 2009, *A&A*, 507, 1793
- Odehahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, *AJ*, 103, 318
- Ordovás-Pascual I., Sánchez Almeida J., 2014, *A&A*, 565, A53
- Paiano S., Landoni M., Falomo R., Scarpa R., Treves A., 2016, *MNRAS*, 458, 2836
- Pedregosa F., Varoquaux G., Gramfort A., et al., 2011, *Journal of Machine Learning Research*, 12, 2825
- Rajaniemi H. J., Mähönen P., 2002, *ApJ*, 566, 202
- Rousseuw P. J., 1987, *Journal of Computational and Applied Mathematics*, 20, 53
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Santini P., Fontana A., Grazian A., Salimbeni S., Fiore F., Fontanot F., Boutsia K., Castellano M., Cristiani S., de Santis C., Gallozzi S., Giallongo E., Menci N., Nonino M., Paris D., Pentericci L., Vanzella E., 2009, *A&A*, 504, 751
- Scaringi S., Cottis C. E., Knigge C., Goad M. R., 2009, *MNRAS*, 399, 2231
- Seber G., , 1984, *Multivariate analysis*
- Shakouri S., Johnston-Hollitt M., Dehghan S., 2016, *MNRAS*, 458, 3083
- Shi F., Liu Y.-Y., Sun G.-L., Li P.-Y., Lei Y.-M., Wang J., 2015, *MNRAS*, 453, 122
- Spath H., 1985, *The cluster dissection and analysis theory FORTRAN programs examples*. Prentice-Hall, Inc.
- Tammour A., Gallagher S. C., Daley M., Richards G. T., 2016, *MNRAS*, 459, 1659
- Teimoorinia H., 2012, *AJ*, 144, 172
- Teimoorinia H., Bluck A. F. L., Ellison S. L., 2016, *MNRAS*, 457, 2086
- Vanzella E., Cristiani S., Dickinson M., Giallisco M., Kuntschner H., Haase J., Nonino M., GOODS Team 2008, *A&A*, 478, 83
- Vanzella E., Cristiani S., Dickinson M., GOODS Team 2005, *A&A*, 434, 53
- Vanzella E., Cristiani S., Dickinson M., GOODS Team 2006, *A&A*, 454, 423
- Vanzella E., Cristiani S., Fontana A., Nonino M., Arnouts S., Giallongo E., Grazian A., Fasano G., Popesso P., Saracco P., Zaggia S., 2004, *A&A*, 423, 761
- Vesanto J., , 2005, *SOM implementation in SOM Toolbox*, <http://www.cis.hut.fi/projects/somtoolbox/documentation/somalg.shtml>(accessed MARCH 2, 2016)

APPENDIX A: INPUTS TO GALAXY SPECTRAL CLASSIFICATION

The spectra classified in this work are available through the Supplementary Information to this paper. The data are contained in a single machine-readable file with wavelength as the rows and individual galaxy spectra as the columns. Table A1 gives the properties of the best-fit models used to create these spectra.

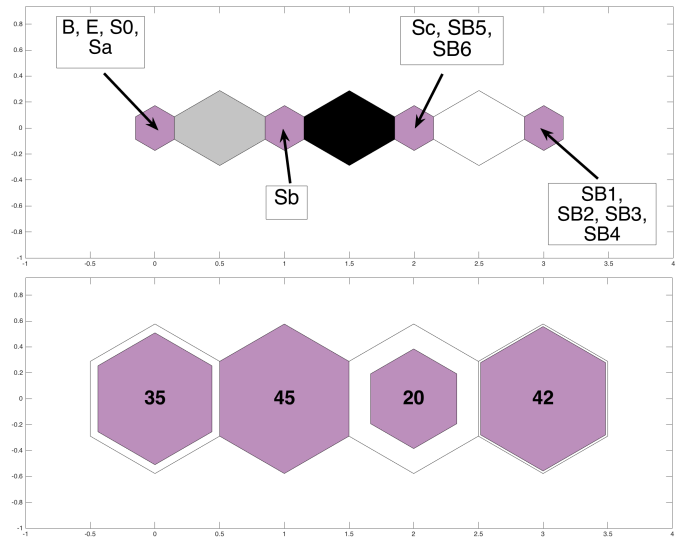


Figure B1. Results of training network in 1×4 grid.

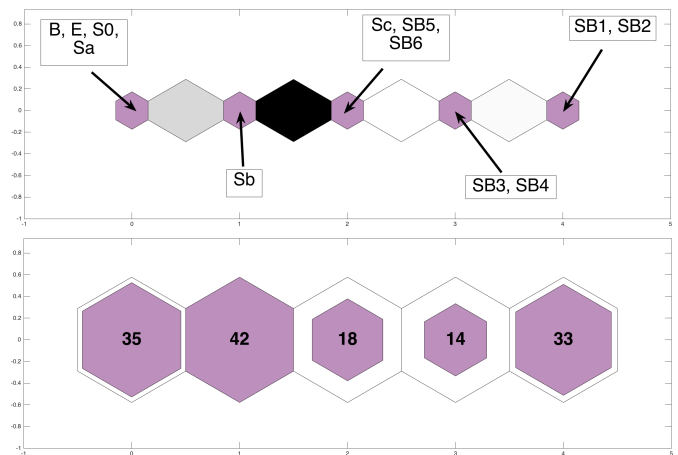


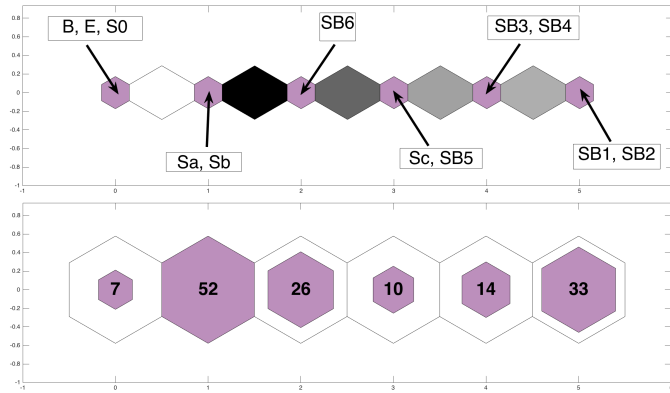
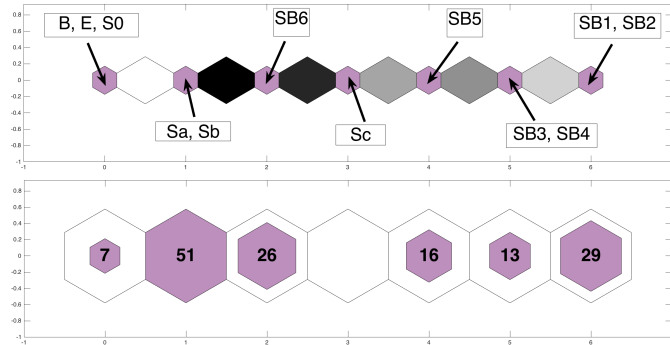
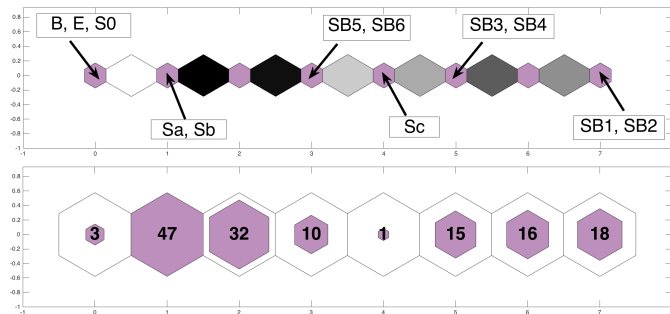
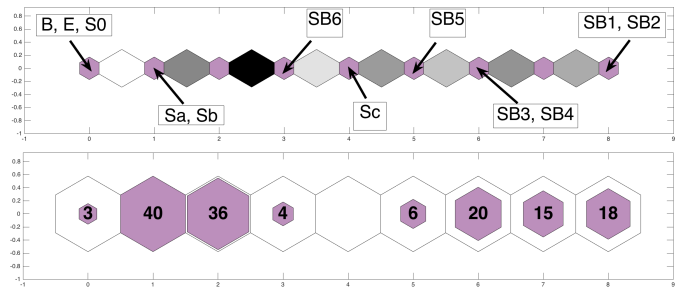
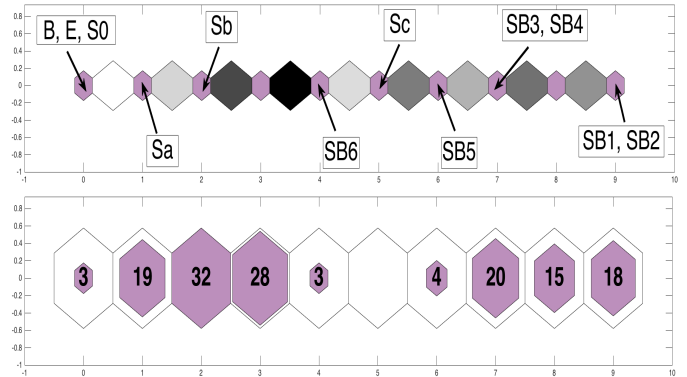
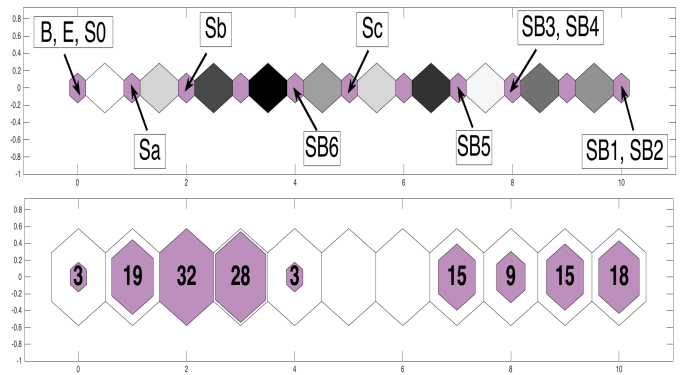
Figure B2. Results of training network in 1×5 grid.

APPENDIX B: RESULTS OF TRAINING FOR ONE-DIMENSIONAL SELF-ORGANIZING MAPS

As mentioned in Sec. 4.1, we changed the size of the SOM from 1×2 to 1×22 . In that section we show some example of the results; here we show the rest of the maps to monitor the changes in the map over various sizes.

Table A1. Properties of best-fit models created by T12. The full table is available online.

ID	GOODS ID	redshift	$\log(M_{\text{star}})$ [M_{\odot}]	t_{D4000} [Gyr]	$\log(\phi)$ [Gyr^{-1}]	$\log(f_{\text{burst}})$	A_{FUV}	T12 class
G1	GDS-J033214.59-274913.4	0.468	9.09 ± 0.29	-1.42 ± 0.1	0.30 ± 0.32	-0.32 ± 0.45	0.70 ± 0.45	SB-1
G2	GDS-J033227.88-275140.4	0.521	11.11 ± 0.06	0.66 ± 0.05	-2.36 ± 0.19	-3.19 ± 0.41	0.83 ± 0.29	E
G3	GDS-J033244.7-274922.8	0.522	10.56 ± 0.05	-0.49 ± 0.01	-1.26 ± 0.07	-1.00 ± 0.04	1.56 ± 0.09	—
G4	GDS-J033213.21-274715.7	0.523	8.78 ± 0.31	-1.18 ± 0.34	-0.05 ± 0.49	-0.70 ± 0.73	2.26 ± 0.95	—
G5	GDS-J033230.23-274519.9	0.523	9.12 ± 0.28	-1.27 ± 0.2	0.07 ± 0.33	-0.35 ± 0.46	1.34 ± 0.57	SB-4
G6	GDS-J033232.18-274534.9	0.523	11.23 ± 0.04	0.41 ± 0.04	-1.70 ± 0.09	-3.00 ± 0.09	0.88 ± 0.04	—

**Figure B3.** Results of training network in 1×6 grid.**Figure B4.** Results of training network in 1×7 grid.**Figure B5.** Results of training network in 1×8 grid.**Figure B6.** Results of training network in 1×9 grid.**Figure B7.** Results of training network in 1×10 grid.**Figure B8.** Results of training network in 1×11 grid.

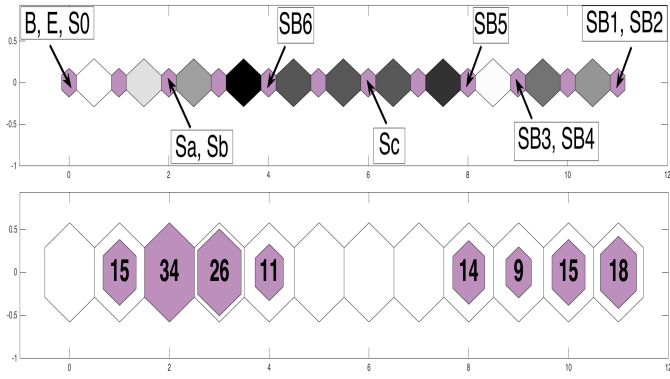


Figure B9. Results of training network in 1×12 grid.

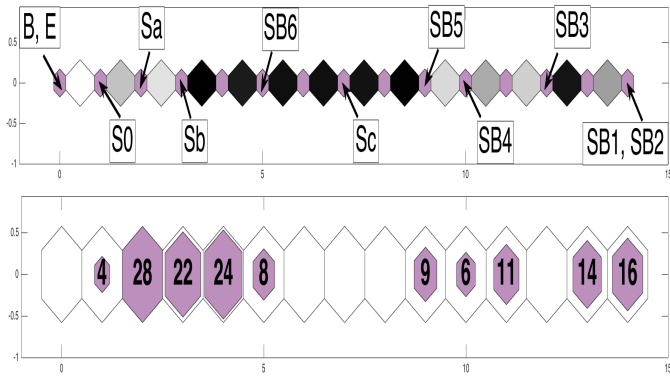


Figure B10. Results of training network in 1×15 grid.

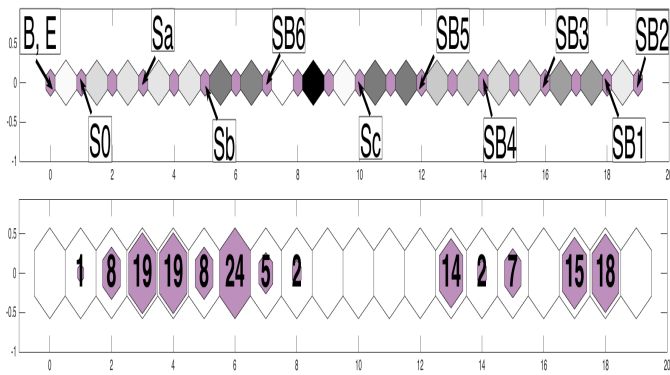


Figure B11. Results of training network in 1×20 grid.