SEARCH FOR EXTREMELY METAL-POOR GALAXIES IN THE SLOAN DIGITAL SKY SURVEY. II.
HIGH ELECTRON TEMPERATURE OBJECTS

J. SÁNCHEZ ALMEIDA1,2, E. PÉREZ-MONTERO3, A. B. MORALES-LUIS1,2, C. MUÑOZ-TUÑÓN1,2, R. GARCÍA-BENITO3,
S. E. NUZA4, AND F. S. KITAUERA

1 Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain; jos@iac.es
2 Departamento de Astrofísica, Universidad de La Laguna, Tenerife, Spain
3 Instituto de Astrofísica de Andalucía, CSIC, P.O. Box 3004, E-18080 Granada, Spain
4 Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany

Received 2015 June 22; accepted 2015 December 31; published 2016 March 3

ABSTRACT

Extremely metal-poor (XMP) galaxies are defined to have a gas-phase metallicity smaller than a tenth of the solar value (\(12 + \log(O/H) < 7.69\)). They are uncommon, chemically and possibly dynamically primitive, with physical conditions characteristic of earlier phases of the universe. We search for new XMPs in the Sloan Digital Sky Survey (SDSS) in a work that complements Paper I. This time, high electron temperature objects are selected; metals are a main coolant of the gas, so metal-poor objects contain high-temperature gas. Using the algorithm \(k\)-means, we classify 788,677 spectra to select 1281 galaxies that have particularly intense \([O\text{ III}]\lambda5007\) with respect to \([O\text{ III}]\lambda4363\) with a proxy for high electron temperature. The metallicity of these candidates was computed using a hybrid technique consistent with the direct method, rendering 196 XMPs. A less restrictive noise constraint provides a larger set with 332 candidates. Both lists are provided in electronic format.

Supporting material: machine-readable tables

1 INTRODUCTION

Extremely metal-poor (XMP) galaxies are defined to have a gas-phase metallicity 10 times smaller than the Sun (e.g., Kunth & Östlin 2000). Even though the exact threshold is arbitrary, this definition selects chemical unevolved galaxies, thus providing a gateway to study physical processes in conditions that are characteristic of the early universe, which are now unusual. XMPs have been used to determine the primordial He abundance produced during the Big Bang (Peimbert et al. 2010; Steigman 2010), study the star formation in conditions of low metallicity (Shi et al. 2014; Filho et al. 2016), infer the formation of dust in the early universe (Fisher et al. 2014), analyze properties of a primitive interstellar medium (ISM; Izotov & Thuan 2007), or constrain the properties of the first stars (Thuan & Izotov 2005; Kehrig et al. 2015). XMP galaxies are not primeval, but rather galaxies with an underlying evolved stellar population (e.g., Aloisi et al. 2007; Corbin et al. 2008; Pérez-Montero et al. 2010; Annibali et al. 2013) that look young because they may have recently accreted metal-poor gas (Östlin et al. 2001; Ekta & Chengalur 2010; Sánchez Almeida et al. 2013b, 2014b, 2015), similar to the cold-flows predicted by the numerical simulations of galaxy formation (Dekel et al. 2009; Schaye et al. 2010; Nuza et al. 2014b; Sánchez Almeida et al. 2014a). In this case, the gas-forming stars in XMPs could trace the cosmic web in their immediate surrounding, providing a method to study and characterize this elusive component of the cosmic web.

Given the general interest and the potential to use XMPs as astronomical tools, there have been many attempts to look for new candidates, i.e., to enlarge the list of XMPs with the possibility of finding the least metallic object in the local universe. However, despite the unquestionable success in finding new objects, the club of XMPs is still quite exclusive. The review paper by Kunth & Östlin (2000) contains only 31 targets, Kniazev et al. (2003) add 8 new targets from the early data release of Sloan Digital Sky Survey (SDSS), the search in SDSS-DR6 by Guseva et al. (2009) yields 44, and the systematic bibliographic search for all XMPs in the literature carried out by Morales-Luis et al. (2011; hereinafter Paper I) renders 140. (Paper I includes targets found by Izotov et al. 2004, 2006; Kniazev et al. 2004; and Izotov & Thuan 2007.)

New local metal-poor objects have been discovered since the publication of Paper I (e.g., Izotov et al. 2012; Skillman et al. 2013; Guseva et al. 2015; James et al. 2015), but the situation has not changed in essence. Moreover, even though the number of members of the club has increased since iZw18 was discovered 45 years ago (Sargent & Searle 1970; Searle & Sargent 1972), its low metallicity has not been superseded; the present record-breaking object SBS 0335-052W (Izotov et al. 2009) has a metallicity close to that of iZw18 and around 2% of the solar value. The existence of such a lower limit for the metallicity of the XMPs has challenged interpretation for decades (e.g., Kunth & Lebouteiller 2011).

The low number of observed XMPs poses a problem because XMPs are expected to be the most common galaxies in the local universe. Because of the luminosity–metallicity relationship (e.g., Skillman et al. 1989; Lee et al. 2006), all galaxies fainter than absolute magnitude \(M_B \approx -12.5\) are expected to...
be XMPs (Berg et al. 2012). The extrapolation of the observed galaxy luminosity function to the faint end implies that these ultra-faint XMPs should be the most numerous galaxies in the local universe (e.g., Blanton et al. 2005; Loveday et al. 2015). The reason why these objects are not detected in large quantities seems to be an observational bias associated with the fact that they are both faint and of low surface brightness. Only when they undergo a starburst phase, and their brightness and surface density increase, do they become detectable in typical magnitude-limited optical surveys (see Skillman et al. 2013; James et al. 2015, and references therein). In order to explain the supposed lack of XMPs, Skillman et al. (2013) proposed the existence of two classes of XMPs: (1) the faint, low surface brightness, quiescent XMPs, which are very numerous but usually undetected; and (2) the bursting XMPs, which go through a transit starburst that allows them to appear in optical surveys like the SDSS analyzed here (Stoughton et al. 2002). Unless otherwise explicitly stated, this paper refers to the latter, even though they are often called XMPs for simplicity. We will mention the class of quiescent XMPs in connection with the number density estimate in Section 4.4.

With a few exceptions, the observed XMPs are bursting XMPs, which makes them blue and compact when the star-forming region is large with respect to the size of the galaxy (Kunth & Östlin 2000). Thus, bursting XMPs tend to be blue compact dwarf (BCD) galaxies. Their optical spectrum is characterized by intense emission lines (e.g., Terlevich et al. 1991). Many of these XMPs have cometary or tadpole morphology (Papaderos et al. 2008, Paper I). XMP galaxies are gas rich, with a typical gas-mass to stellar-mass ratio in excess of 10 (Filho et al. 2013). They tend to be found in relative isolation, although some show loose companions. They also prefer low density environments, identified as voids and sheets in cosmological numerical simulations (Filho et al. 2015).

One of the difficulties of enlarging the sample of XMPs is related to the determination of their chemical abundances, which critically depends on the available optical emission lines. The most accurate method relies on measuring electron temperatures, and uses weak and near-UV emission lines that are often unavailable. In addition, tailored data-mining techniques are required to identify them because XMPs appear as rare objects in very large surveys. Here, we report a new systematic search for XMP galaxies in the Sloan Digital Sky Survey (SDSS-DR7; Abazajian et al. 2009) that overcomes the two aforementioned difficulties. The philosophy remains very much like in Paper I. We carry out an automated classification of the SDSS spectra using k-means (e.g., Sánchez Almeida et al. 2010; Ordovás-Pascual & Sánchez Almeida 2014) in a range of wavelengths that are particularly sensitive to metallicity. One or a few classes have spectra characteristics of low-metallicity galaxies. Then we individually measure the metallicity of the objects in these classes so as to confirm or discard the XMP candidates as true members of the XMP party. The difference between Paper I and the present work is in the wavelength region chosen to classify the spectra. In Paper I, we used the region around Hα, where the two [N II] lines tend to disappear when the metallicity goes to zero (e.g., Pettini & Pagel 2004; Sánchez Almeida et al. 2012). However, even objects with significant [N II] can be XMP (Morales-Luis et al. 2014), because depending on the evolutionary state and chemical history, galaxies may have enhanced N/O and present large [N II] for a given O (e.g., Pérez-Montero & Contini 2009). Here we use the region between the lines [O III]λ4363 and [O III]λ5007, which is known to depend on the temperature of the H II region producing the emission-line spectrum. Because metals are a main coolant agent of the gas (e.g., Pagel et al. 1979), metal-poor objects are also high-temperature objects, yielding a relationship between electron temperature and metallicity. This approach is similar to the one already used by Ly et al. (2015), except that they employ it to search for galaxies at a redshift around 0.8, so that the wavelength region of [O III] is observed in the near-IR.

The paper is organized as follows. First, we describe the automated search for XMP candidates using k-means (Section 2). It leads to 1281 candidates. Then, we use the SDSS spectra of the candidates to measure their integrated metallicity, which narrows down the list to 196 XMPs (Section 3). Emission line fluxes and abundances are calculated in Sections 3.1 and 3.2, respectively. The final list with 196 XMPs is worked out in Section 3.3 (Table 1), where we also include a second list (Table 2) with 332 potential XMPs that are selected under less restrictive noise constraints. The global properties of the XMPs are analyzed in Section 4, explicitly, stellar mass (Section 4.2), reddening and dust mass (Section 4.1), absolute magnitude and color (Section 4.3), number density (Section 4.4), metallicity threshold (Section 4.5), morphology (Section 4.6), underlying stellar population (Section 4.7), and large-scale environment (Section 4.8). The results are summarized and discussed in Section 5. The number of false positives and false negatives in Table 1 is estimated in the Appendix.

2. SEARCH FOR XMP CANDIDATES CLASSIFYING THE SPECTRUM AROUND [O III]λ4363

The ratio between the emission of [O III]λ4363 and the pair [O III]λλ4959, 5007 strongly depends on the metallicity of the emitting gas. The reason is well known and rather straightforward. Metals are a main coolant of the gas (e.g., Pagel et al. 1979), so metal-poor gas is also high-temperature gas. The ratio [O III]λ4363 to [O III]λλ4959, 5007 is used to measure [O III] electron temperature in H II regions (e.g., Hägele et al. 2008), so it is also a good proxy for metallicity. The three [O III] lines are excited by collisions with free electrons, and their decay through photon emission renders the observed lines. The excitation potential of the upper level of [O III]λ4363 (5.35 eV) is significantly higher than the upper level shared by [O III]λλ4959 and [O III]λ5007 (5.35 eV), so as the temperature increases [O III]λλ4363 increases relative to the other two lines, that is to say, as the metallicity of the gas decreases. The relative fluxes change by more than one order of magnitude when the electron temperature changes in the range of the typical temperatures in H II regions, for example, between 5 and 25 kK (e.g., Osterbrock 1974).

We take advantage of the sensitivity of the spectrum on metallicity to search for XMP candidates. Using the automatic classification algorithm k-means (Sánchez Almeida et al. 2010; Ordovás-Pascual & Sánchez Almeida 2014), we are able to classify the full SDSS-DR7 spectral database according to the shape of the spectrum of the galaxies in the spectral region of interest, from 4200 to 5200 Å, which contains [O III]λλ4363, [O III]λλ4959, and [O III]λ5007. As in Paper I, the signal to be classified, S(λ), is the observed spectrum in restframe wavelength, f(λ), after removing the continuum intensity, If(λ), and after the subsequent normalization to the emission at
the reference wavelength (4363 Å in this case), explicitly,

\[ S(\lambda) = \frac{I(\lambda)}{I(4363 \text{ Å})} - I(4363 \text{ Å}) ]. \]

We start from the full set of 788,677 galaxy spectra from SDSS-DR7 with a redshift smaller than 0.25 (Sánchez Almeida et al. 2010). The classification renders 17 types or classes. The two with the largest [O iii]λ4363, which include 24,200 objects, are reclassified to obtain 12 subclasses. Some of these subclasses collect noisy spectrum, and those are discarded so that k-means acts as a noise filter. Four of the subclasses show particularly intense [O iii]λ4363, as depicted in Figure 1. (Note that, even in these cases, the flux of [O iii]λ4363 is much smaller than the flux of [O iii]λ4959, [O iii]λ5007—the spectra in Figure 1 have [O iii]λ5007 some 50 times larger than [O iii]λ4363.) Classes 8 and 11 in Figure 1 are discarded because the mean spectrum of the class appears in the region of active galactic nuclei (AGNs) of the BPT diagram (Baldwin et al. 1981; Sánchez Almeida et al. 2012). Therefore, their high excitation is produced by hard AGN radiation, rather than by lack of efficient coolants in a stellar-radiation excited H\textsc{ii} region. The AGN nature of the discarded classes is corroborated by the large width of their emission lines (cf. AGN classes, 8 and 11, with star-forming classes 4 and 6), a property also characteristic of AGNs. After the classification and trimming, we are left with 1281 bursting XMP candidates.

We compared these 1281 candidates with the 32 XMPs selected in Paper I after classifying the spectral region around H\textalpha. Most of them, 29 out of the 32, are included in our list of candidates. Considering that the two procedures are completely independent, and the randomness inherent to the k-means procedure, the agreement is very reasonable. Recall that k-means is a powerful tool that is able to classify very large data sets in high-dimensional spaces. However, among its drawbacks, k-means does not provide a single classification for a single database. Each particular run depends on a random initialization, thus, the classification is not unique (e.g., Sánchez Almeida et al. 2010, Section 2).

Nothing ensures that the 1281 candidates are true XMPs. They have been picked up as objects where [O iii]λ4363 is particularly intense, using the rest of galaxies in the spectroscopic database of SDSS-DR7 as reference. In order to show whether or not they are XMPs, we have computed the metallicity for each to mark only those objects of truly low metallicity as an XMP. (The details are given in Section 3.) One may think of this second step as a way to clean from pollutants the systematic but otherwise rough XMP selection carried out by k-means on the SDSS-DR7 database.

### 3. METALLICITIES OF THE XMP CANDIDATES

#### 3.1. Determination of Emission Line Fluxes

We use the SDSS-DR7 spectra of the 1281 candidates to determine their metallicities. As usual, the procedure requires
Two of the observed Balmer lines are particularly broad. The spectra, with the continua removed, have been normalized to the peak intensity of the lines. The package contains a suite of routines to easily analyze individual spectra, such as those employed in our analysis.

Candidate XMP Galaxies

| Index | Name          | 12 + log(O/H)d | log(N/O)d | log τβ | CHβf |
|-------|---------------|----------------|-----------|--------|------|
| 1     | J000099.0+011423.8 | 7.67 ± 0.18 | −1.44 ± 0.15 | −2.58 ± 0.13 | 0.29 ± 0.06 |
| 2     | J001432.7+002919.9 | 7.60 ± 0.24 | −1.37 ± 0.22 | −2.63 ± 0.12 | 0.09 ± 0.03 |
| 3     | J003637.4+003826.8 | 7.66 ± 0.18 | −1.60 ± 0.14 | −2.65 ± 0.12 | 0.00 ± 0.04 |
| 4     | J001022.9−002450.5 | 7.57 ± 0.19 | −1.11 ± 0.23 | −2.54 ± 0.16 | 0.27 ± 0.03 |
| 5     | J002114.8−085809.5 | 7.56 ± 0.23 | −1.14 ± 0.20 | −2.65 ± 0.19 | 0.62 ± 0.01 |
| 6     | J002535.1+003456.1 | 7.61 ± 0.18 | −1.27 ± 0.24 | −2.68 ± 0.21 | 0.15 ± 0.04 |
| 7     | J002916.8−010212.1 | 7.61 ± 0.25 | −1.35 ± 0.15 | −2.04 ± 0.33 | 0.17 ± 0.02 |
| 8     | J003145.3−110658.8 | 7.61 ± 0.17 | −1.56 ± 0.12 | −2.24 ± 0.33 | 0.13 ± 0.18 |
| 9     | J003630.4+005234.7 | 7.59 ± 0.12 | −1.51 ± 0.14 | −1.89 ± 0.25 | 0.01 ± 0.03 |
| 10    | J003741.1+003320.1 | 7.68 ± 0.12 | −1.52 ± 0.13 | −1.89 ± 0.27 | 0.00 ± 0.03 |
| 11    | J004224.8+003315.7 | 7.63 ± 0.16 | −1.51 ± 0.14 | −2.55 ± 0.18 | 0.07 ± 0.03 |
| 12    | J004521.9−093700.1 | 7.52 ± 0.21 | −1.44 ± 0.13 | −1.92 ± 0.43 | 0.09 ± 0.03 |
| 13    | J004614.4+000635.9 | 7.55 ± 0.20 | −1.28 ± 0.25 | −2.58 ± 0.19 | 0.25 ± 0.03 |
| 14    | J005249.8−084133.9 | 7.63 ± 0.15 | −1.48 ± 0.15 | −2.66 ± 0.21 | 0.23 ± 0.03 |
| 15    | J004048.4−004010.3 | 7.60 ± 0.15 | −1.38 ± 0.16 | −2.38 ± 0.17 | 0.21 ± 0.12 |
| 16    | J010413.3−001529.0 | 7.56 ± 0.21 | −1.20 ± 0.22 | −2.65 ± 0.13 | 0.16 ± 0.04 |
| 17    | J010414.6−005040.1 | 7.31 ± 0.34 | −1.38 ± 0.22 | −1.71 ± 0.32 | 0.20 ± 0.04 |
| 18    | J011414.5−010929.8 | 7.57 ± 0.22 | −1.44 ± 0.23 | −2.41 ± 0.24 | 0.09 ± 0.23 |
| 19    | J011340.5−005239.2 | 7.04 ± 0.26 | −9.99 ± 0.00 | −1.55 ± 0.10 | 0.00 ± 0.01 |
| 20    | J011914.3−093546.2 | 7.60 ± 0.14 | −1.55 ± 0.12 | −1.92 ± 0.27 | 0.15 ± 0.02 |

Notes.

a The full table with all 332 candidates is given only in electronic format.
b Sorted according to growing R.A.
c R.A. and decl. in J2000 coordinates.
d Using HCM; see Section 3.2.
e Ionization parameter (e.g., Stasińska et al. 2012).
f Extinction coefficient in Hβ.

(This table is available in its entirety in machine-readable form.)

Figure 1. Template of four spectral classes where the emission line [O iii]λ4363 (marked with arrows) is particularly intense compared to [O ii]λ3726 and [O iii]λ5007. They were found by applying two nested k-means classifications to the 788,677 galaxies with spectra in SDSS-DR7 with a redshift < 0.25. Two of the classes (8 and 11) are AGNs, as proven by their position on the BPT diagram and the fact that the emission lines are particularly broad. The spectra, with the continua removed, have been normalized to the peak intensity of [O iii]λ4363. Wavelengths are given in Å. The label on top of each panel represents the class number and the number of galaxies in the class.

Figure 2. Using HCM; see Section 3.2.

Correcting the emission-line fluxes for reddening, which is determined from the relative fluxes of the observed Balmer lines (e.g., Hägele et al. 2008; Stasińska et al. 2012). However, Balmer emission is usually depressed due to the presence of the absorption lines of stellar origin. Therefore, prior to the measurement of the emission lines, we perform a stellar fitting to the continuum of the spectra. We use STARLIGHT (Cid Fernandes et al. 2005) to model the spectral energy distribution (SED) of the underlying stellar population. STARLIGHT fits the observed continuum as a linear combination of single stellar-population (SSP) spectra of various ages and metallicities. We chose the SSP spectra from Bruzual & Charlot (2003), including four metallicities Z = 0.0001, 0.0004, 0.004, and 0.008 (1%, 3%, 30%, and 60% Z⊙, respectively; Asplund et al. 2009). Reddening of the stellar SED was included, assuming the extinction to follow Cardelli et al. (1989) with Rv = 3.1. The stellar continuum thus obtained is subtracted out of the observed spectrum to produce a pure emission-line spectrum. The latter is employed to determine line fluxes.

We used the package SHIFU5 to obtain the flux of the emission lines. The package contains a suite of routines to easily analyze emission or absorption lines in IFU data (both in cube and RSS format). Individual spectra, such as those employed in our particular case, can be provided as a list of files (RSS). The core of the code uses CIAO’s Sherpa package (Freeman et al. 2001; Doe et al. 2007). Several custom algorithms are implemented in order to cope with general and ill-defined cases. A sigma-clipping is independently applied to the stellar continuum model, and then this is parsed to the composite line plus continuum model. Although the fitting is performed in the residual spectra, we allow the modeling of the continuum to take small deviations in the stellar continuum residuals into account.

5 Sherpa IFU line fitting package, M. García-Benito et al. (2016, in preparation).
account. A first order polynomial was chosen for the continuum, while single gaussians were selected for the lines. The continuum is evaluated in the original spectra when used to determine equivalent widths.

3.2. Method to Measure Metal Abundance

Metal abundances are computed with the code H II-CHIMisty (hereafter HCM Pérez-Montero 2014), which calculates the total oxygen abundance O/H, the nitrogen-to-oxygen ratio N/O, and the ionization parameter U. Normalized to H/β and corrected for extinction, the fluxes of the emission lines [O ii]λ3727, [O ii]λλ4363, 5007, [N ii]λ6584, and [S ii] λλ6716, 6731 are compared with the predictions of a large grid of CLOUDY (Ferland et al. 2013) photoionization models. The HCM assumes a spherically symmetric constant-density distribution of gas ionized by POPSTAR (Mollá et al. 2009) 1 Myr-old stellar clusters with the same metallicity of the gas, and covering a wide range of values in O/H, N/O, and log U. In particular, the grid includes very low values of the gas metallicity, down to 12 + log(O/H) = 6.9 equivalent to 1/60 of the oxygen abundance in the solar composition (Asplund et al. 2009). The main advantage of this method is that it leads to a derivation of the chemical abundances that is consistent with the direct method, even in the absence of one or more of the input emission lines (Pérez-Montero 2014). Because the direct method does not depend on modeling (e.g., Osterbrock 1974; Stasińska et al. 2012), it means that the metallicities provided by HCM are almost insensitive to the model assumptions. Given a set of observed line ratios, HCM assigns a $\chi^2$ to each of the models, computed as the sum of the square differences between the observed and the model line ratios. The $1/\chi^2$-weighted average of the abundance in the models yields the chemical abundance tagged to the observed line ratios. Errors are derived as the 1/$\chi^2$-weighted standard deviation of the parameters in the models. When the temperature-sensitive line [O ii]λλ4363 is available, HCM uses the whole grid of models to provide values of O/H and N/O that are consistent with the direct method, even in the absence of [O ii]λ3727. Thus using HCM results is especially appropriate when searching for XMPs in SDSS because [O ii]λ3727 is sometimes outside the observed spectral region for nearby objects. The error bars estimated by HCM quantify the internal precision of the method (i.e., they give the range of models that is consistent with the observed line fluxes). The observational errors enter into the abundance error estimate indirectly, because the observed fluxes must have enough signal to be considered for analysis.

3.3. List of XMPs

The automated procedure described in Section 3.1 provides the fluxes used by HCM (Section 3.2) to infer the oxygen abundance of the 1281 candidates (Section 2). We consider only line fluxes with a signal-to-noise ratio larger than three, including the weak [O ii]λλ4363. The scatter plot 12 + log(O/H) versus log(N/O) for the full sample is shown in Figure 2(a). Assuming that the solar composition corresponds to $12 + \log(O/H) = \approx 8.69$ (Asplund et al. 2009), XMPs are defined as those candidates having

$$12 + \log(O/H) \leq 7.69.$$  

(2)

There are 196 galaxies that fulfill this requirement, which are the bursting XMP galaxies presented in this work. Eleven of them have duplicated spectra in SDSS-DR7. We have chosen the spectrum yielding the smallest error in metallicity for analysis. The XMP galaxies are represented as red squares with error bars in Figure 2(a). The distribution of metallicities of the candidates is given in Figure 2(b). The histogram is fairly continuous at the border between XMPs and the rest, which proves the 1/10 solar metallicity threshold to be rather arbitrary. The list of XMPs is given in Table 1, which includes coordinates and the main physical properties provided by HCM. When [N ii]λ6584 is too weak for the noise level, N/O cannot be derived and it appears as $-9.99$ in the table. The distributions of N/O for the XMPs and the rest of the sample are given in Figure 2(c). They look qualitatively similar.

We carried out several sanity checks. First, we estimated the number of false positives and false negatives due to observational errors. False positives are galaxies in the XMP list that have true oxygen abundance above the threshold. False negatives are true XMP galaxies that have been discarded from the selection. Using a Bayesian approach, we work out the expected number of false positives and false negatives in Appendix. Even though the typical error of the individual abundance estimates is rather small ($\approx 0.17$ dex for the galaxies in Table 1), we expect a significant number of false positives and false negatives. Assuming the true distribution of metallicities to be given by the histogram in Figure 2(b), we expect 82 false positives and 46 false negatives (Equation (23)). If the true distribution is uniform, then the number of both false positives and false negatives is around 55 (Equation (24)). Partly to offer the possibility of recovering false negatives in the future, we include a second list with candidate XMPs in Table 2. Table 1 is based on lines with measured fluxes exceeding three times the noise. If this threshold is lowered, and lines with fluxes at two-sigma are also considered, then the number of galaxies fulfilling the condition in Equation (2) increases to 332; those are the objects included in Table 2. The reason why the number increases when lowering noise limit can be pinned down to the weakness of [O ii]λλ4363, which tends to be excluded from the metallicity estimate when the noise limit is too restrictive. When it is considered in the metallicity estimate, HCM tends to lower the oxygen abundance estimate due to arguments given in Section 2.

As a second test, we compared the metallicity provided by HCM with the metallicity obtained with the direct method (e.g., Hägele et al. 2008) when [O ii]λ3727 was available. These abundances, based on the direct method, are also included in Table 1. The scatter plot direct-method metallicity versus HCM metallicity is shown in Figure 3. As expected, the two techniques agree quite well with a dispersion of only 0.04 dex, which is smaller than the intrinsic error of both methods. In addition to this agreement, there is a tendency for HCM to overestimate the direct-method metallicity at low metallicities. Because this bias occurs below the threshold

---

8 The wavelength 3727 Å is out of the range of the original SDSS spectrograph, so that a minimum redshift around 0.03 is needed for line to appear in a SDSS spectrum.

---

$^6$ http://www.iaa.es/~epm/HII-CHI-mistry.html

$^7$ Different authors use slightly different thresholds, for example, Kunth & Ostlin (2000) and Morales-Luis et al. (2011) use 12 + log(O/H) < 7.65, Guseva et al. (2009) use ≤7.60, and we use exactly a tenth of the solar metallicity quoted in the review paper by Asplund et al. (2009).
metallicity to be XMP (Equation (2)), both methods would yield the same list of XMPs if we had access to [O II]λ3727. Moreover, the systematic differences between the abundances obtained from HCM and the direct method are small compared with their error bars, so such differences barely affect other results and conclusions of the paper. For example, adding an extra 0.06 dex error to the typical ~0.17 dex error bar increases the number of false positives in Appendix, but only from 82 to 86.

As a final test, we represented our targets in the [O III]λ5007/Hβ versus [N II]λ6583/Hα plane commonly used to separate starbursts from AGNs (the so-called BPT diagram, after Baldwin et al. 1981); see Figure 4. None are AGNs, and only two reside in the composite-spectrum region next to the divide worked out by Kauffmann et al. (2003). They may be starbursts of high N/O (Pérez-Montero & Contini 2009), or H II regions where the excitation may have contribution from shocks, evolved stars, or even AGNs. The two objects with composite-spectrum are 51 and 151 in Table 1, and their identification as XMP galaxies must be taken with caution. Most of the 196 galaxies in Table 1 seem to be new members of the XMP class. Table 2 of Paper I compiles all XMPs in the literature up to the date of publication in mid-2011; it contains 129 objects, 21 of which coincide with objects in Table 1. Paper I also carries out a systematic search for XMPs in SDSS-DR7, finding 32 such objects; 17 of the 32 targets coincide with targets selected in our study, and so are included in Table 1. In addition, there was recently a major search for XMPs by Izotov et al. (2012), which contains 41 targets; among them, 7 are in Table 1. Considering all the objects in these three lists, and discarding coincidences, only 31 among the 196 galaxies in Table 1 seem to be known. Known and unknown targets are pointed out in one of the columns of Table 1, which used the...
Table 1. This plane is commonly used to separate starbursts from AGNs (so-called BPT diagram, after Baldwin et al. 1981). The shaded region shows the area used by Izotov et al. (2012) to search for XMPs in SDSS-DR7. The colored points correspond to objects that are also found in Paper I (red) and in Izotov et al. (2012, blue). Contrarily to our targets, these previous searches pick up emitters concentrated on the leftmost region of the BPT diagram, corresponding to [NII]λ6583 ≪ Hα.

The histogram of metallicities in Figure 2 shows scatter plots of extinction versus log(12 + log(O/H)) for XMPs with [NII]λ6583/Hα > 1.3 and the corresponding expected values expected for XMPs with [NII]λ6583/Hα = 1.3. The dotted line indicates the mean value of the extinction coefficient c(Hβ) = −log(F/F0), where F and F0 stand for the reddened and original fluxes in Hβ, respectively. Therefore, the extinction coefficient is proportional to the optical depth along the line of sight in [OII]λλ3726,3729. Using the usual parametrization, the extinction coefficient at Hβ, c(Hβ), is defined to be

$$c(H\beta) = -\log(F/F_0),$$

(3)

where the symbols F and F0 stand for the reddened and original fluxes in Hβ, respectively. Therefore, the extinction coefficient is proportional to the optical depth along the line of sight in [OII]λλ3726,3729.

$$c(H\beta) = \alpha \log e,$$

with

$$\alpha = \int_0^{l_0} \kappa_D \rho_D dl.$$

(4)

The integral goes along the line of sight from the observer to the source of light at l0. The symbols \(\kappa_D\) and \(\rho_D\) stand for the dust density and the cross-section per unit mass of dust, respectively. Defining the density-weighted mean cross-section \(\langle \kappa_D \rangle\),

$$\langle \kappa_D \rangle = \int_0^{l_0} \kappa_D \rho_D dl / \int_0^{l_0} \rho_D dl,$$

(5)

then the dust-mass column density,

$$\Sigma_D = \int_0^{l_0} \rho_D dl,$$

can be inferred from the extinction coefficient as

$$\Sigma_D = c(H\beta)/[\langle \kappa_D \rangle \log e]$$

(6)

Figures 5(a) and (b) show scatter plots of extinction versus metallicity and N/O, respectively. We find no clear relationship between extinction and metallicity (Figure 5(a)). However, there is a hint of correlation between extinction and log(N/O), which means that the larger the ratio between nitrogen and oxygen the larger the reddening (Figure 5(b)). The scatter of the possible relationship is large; the extinction coefficient increases only by a factor of three over the full range of N/O, and this increase is similar to the dispersion of the individual galaxies in the scatter plot (Figure 5(b)).
There is a trend, which means that the reddening increases with increasing $N_{\rm O}/O$. The ordinate axes on the right of Figures 5(a) and (b) quantify the dust-mass column density responsible for the observed reddening. We used Equation (6) with a cross-section per unit mass $\langle \kappa_D \rangle = 10^4 \text{ cm}^2 \text{ g}^{-1}$, which is representative of this quantity at H$\beta$ (e.g., Draine 2003, Table 4). For the typical size of the giant H$\Pi$ regions in these objects, of the order of 150 pc (e.g., Sánchez Almeida et al. 2013b, 2015), the mass in dust corresponding to the observed surface density spans from $2 \times 10^2$ to $2 \times 10^4 M_\odot$. For reference, the dust mass of the prototypical XMP IZw 18, as inferred from modeling IR dust emission, is between 450 and 1800 $M_\odot$ (Fisher et al. 2014). IZw18 is included in Table 1 and its spectrum shows no extinction with an uncertainty of 0.07, which sets an upper limit of $1.5 \times 10^3 M_\odot$ that is consistent with the value inferred from dust emission by Fisher et al. (2014).

The distribution of extinction coefficients for the XMPs in Table 1 has a mean and a standard deviation given by

$$\mu_{\kappa\text{(H}\beta)} \pm \sigma_{\kappa\text{(H}\beta)} = 0.17 \pm 0.14. \tag{7}$$

### 4.2. Stellar Masses

Stellar masses are usually estimated from photometry via models that provide the light-to-mass ratio ($M/L$) given the observed colors of a galaxy (e.g., Bell & de Jong 2001). This procedure is particularly uncertain for bursting XMP galaxies, with spectra having intense emission lines that are hard to capture by the model used for $M/L$. However, stellar mass is one of the central parameters characterizing galaxies, therefore, even with the above caveat in mind, we analyzed photometry-based stellar masses of the XMP galaxies. We use those from the Max-Planck-Institute for Astrophysics—Johns Hopkins University (MPA-JHU)\(^{10}\) characterization of the SDSS-DR7 galaxy spectral data (Brinchmann et al. 2004; Salim et al. 2007). Broad-band galaxy-integrated magnitudes from the SDSS-DR7 database are compared with a grid of theoretical galaxy spectra (Bruzual & Charlot 2003) spanning a large range in star-formation histories (Salim et al. 2007). From the difference between the observed and theoretical magnitudes, a likelihood distribution for the mass of each galaxy is estimated. We then use the median and dispersion of this distribution in our analysis.

The observed stellar masses have a mean and a standard deviation of

$$\mu_{\log M_*} \pm \sigma_{\log M_*} = 8.0 \pm 1.1, \tag{9}$$

where the masses have been expressed in solar masses. The XMPs are generally dwarfs, but there is a large spread in their possible masses—log($M_*/M_\odot$) spans from 6 to 10. This can be seen in Figure 6, which shows the scatter plot metallicity versus stellar mass. We find no clear trend for the metallicity to vary with stellar mass, which is forced by the selection criteria

---

\(^{10}\) [http://www.mpa-garching.mpg.de/SDSS/DR7/](http://www.mpa-garching.mpg.de/SDSS/DR7/)
imposed on the XMPs. In a randomly chosen set of galaxies, stellar masses and metallicities are correlated (e.g., Tremonti et al. 2004). Such relationship, as updated by Andrews & Martini (2013), is included in Figure 6. By construction, XMPs are restricted to have low metallicity and therefore cannot follow the general trend dictated by the red solid line in Figure 6. However, hints of the underlying mass–metallicity relationship remain in the XMP sample, because the lowest abundances are associated with galaxies in the low-mass part of the mass distribution (Figure 6). Figure 7 shows the scatter plot N/O versus stellar mass. There is a clear trend for N/O to increase with increasing stellar mass for \( \log(M_*/M_\odot) > 8.5 \), which is consistent with the relationship found in local galaxies (e.g., Pérez-Montero & Contini 2009; Pérez-Montero et al. 2013). In addition, some low-mass XMPs do show large N/O. The combination of low metallicity and high N/O can be understood if these galaxies suffered a major metal-poor gas-accretion event that dropped O/H, keeping N/O unchanged (see Section 5 for more details). We also find a correlation between reddening and stellar mass (Figure 8), which reflects the correlation between N/O and reddening (Figure 5(b)), given the fact that N/O increases with increasing stellar mass (Figure 7). We note that none of the large N/O targets have \( 12 + \log(O/H) < 7.5 \) (Figure 2(a)).

### 4.3. Magnitudes and Colors

Figure 9 shows the color–magnitude diagram in the color filters g and r. We use Petrosian magnitudes of the integrated galaxies, as provided by SDSS. The figure includes the color threshold to be a BCD galaxy as defined by Gil de Paz et al. (2003); and transformed from U and B to g and r by Sánchez Almeida et al. (2008). Most XMPs lie below the threshold and thus look blue in broadband colors. However, a significant number are not dwarfs, with an absolute magnitude in excess of \( M_g \approx -18 \). That some are not dwarfs is also inferred from the distribution of masses in Figure 6, which contains galaxies with masses larger than \( 10^9 M_\odot \). These high-mass XMPs often correspond to the most distant objects, with redshifts above 0.05 and distances larger than 220 Mpc.

### 4.4. Number Density of XMPs in the Local Universe

The SDSS spectroscopic legacy sample was designed as a magnitude-limited survey. Therefore, it is relatively simple to correct for the luminosity bias (i.e., the Malmquist bias) and so to compute the volume density of particular objects. In Paper I we carried out an estimate of the number density of XMPs in the local universe. Because the number of XMPs has increased substantially from Paper I, such an estimate has become obsolete and is updated here.

We employ the so-called \( V_{\text{max}} \) approximation by Schmidt (1968) used to, for example, determine luminosity functions of galaxies (Takeuchi et al. 2000). The number density of galaxies with the property \( X \), \( n(X) \), is

\[
  n(X) = \sum_i \frac{1}{V_i},
\]

where the sum includes all galaxies in the sample with property \( X \), and \( V_i \) represents the maximum volume in which the \( i \)th galaxy of the sample could be observed. In a magnitude-limited
sample, so all galaxies brighter than the apparent magnitude \( m_{\text{lim}} \) are included,
\[
V_i = \frac{d_i^3}{3} \Omega,
\]
with \( \Omega \) the solid angle covered by the survey, and \( d_i \) the maximum distance at which the \( i \)th galaxy can be observed,
\[
\log (d_i/l_i) = \frac{1}{5} (m_{\text{lim}} - m_i).
\]
The maximum distance depends on the absolute magnitude of the galaxy, which can be written in terms of the distance to the galaxy, \( l_i \), and the apparent magnitude, \( m_i \), as has been done for Equation (12).

A few caveats are in order before applying Equation (10) to the data set. SDSS is not truly magnitude limited for a number of reasons. Some bright galaxies are not observed in the crowded field due to fiber collision (see Stoughton et al. 2002). Low surface brightness galaxies tend to be missed even if they have significant integrated magnitudes (e.g., Blanton et al. 2005; James et al. 2015). The actual galaxy catalog sample also contains objects coming from complementary SDSS searches, which are often fainter than the magnitude limit of the main survey. The first problem is of no relevance because most observed XMP galaxies are isolated (Filho et al. 2015). The second problem is likely more serious because very faint galaxies tend to have low surface brightness as well (Skillman 1999). Quiescent XMP galaxies like Leo P (Skillman et al. 2013) or KJ 78 (James et al. 2015) show a surface brightness around 24 mag arcsec\(^{-2}\), and objects like them would hardly appear in our selection because SDSS is almost blind at this low surface brightness (Blanton et al. 2005). As discussed in the introduction of the paper, we bypass the problem dividing XMPs into quiescent and bursting. Our search in SDSS is sensitive to the bursting XMPs only, and so are the properties derived here. This also holds for the number density to be inferred, assuming a magnitude-limited sample. The third caveat, namely the existence of XMPs from complementary searches, is also an issue because a significant part of our XMPs belong to these additional catalogs. We solve this problem by applying Equation (12) only to those XMPs with \( m_i < m_{\text{lim}} \). If we use \( m_{\text{lim}} = 17.8 \) in the filter \( r \), then Equations (10)–(12) render
\[
\eta \left( \frac{O/H}{10^{-5}} \right) = (3.4 \pm 0.9) \times 10^{-3} \text{Mpc}^{-3},
\]
with the error bar considering only the poissonian error associated with the process of galaxy counting.

The density in Equation (13) is larger than the density obtained from the sample in Paper I, which is consistent with the fact that the present work contains almost 10 times more targets. In terms of the total number of galaxies, XMPs represent some 2% of all galaxies in a given local volume. The number density of galaxies in the local universe, \( \sim 0.17 \text{ Mpc}^{-3} \), has been taken from the normalization of the total \( r \)-band luminosity function of SDSS galaxies by Blanton et al. (2005), which includes galaxies brighter than absolute magnitude \(-12\). Even with a significant increase with respect to the previous estimate in Paper I, XMP galaxies are still rare.

4.5. Low-metallicity Threshold

One of the puzzling properties of XMP galaxies is the existence of a threshold for the lowest metallicity. This threshold is also present in the XMPs revealed here. Figures 2(a) and (b) show that
\[
12 + \log(\text{O/H}) \gtrsim 7.1,
\]
which approximately corresponds to 1/40 times the oxygen abundance in the solar composition. This value is similar to, although slightly higher than, the record-breaking low-abundance galaxy SBS0335–052W (12 + log(O/H) \( \approx 7.0 \), corresponding to 1/50 of the solar composition; Izotov et al. 2009).

The reason why such a threshold exists is still a riddle. The metallicity of the gas produced during the Big Bang is virtually metal-free, and population III stars contaminate the intergalactic medium with a metallicity lower than 10\(^{-5}\) times the solar metallicity (e.g., Bromm & Larson 2004). One can envisage that some of the galaxies formed from this pristine gas have been so dull forming stars that they still retain part of the original gas. But why are none of these pristine galaxies producing stars at present? Several explanations have been offered in the literature, going from the self-enrichment of the \( \text{H} \alpha \) region used for measuring to technical difficulties for metallicity determinations below the threshold (see Kunth & Lebouteiller 2011). One additional possibility is offered by the cosmic-web gas-accretion scenario, where the starbursts in XMPs are triggered by gas recently accreted from the cosmic web (see Sánchez Almeida et al. 2014a, and references therein).

Numerical simulations predict the cosmic-web gas to accumulate metals from the metal-enriched outflows of dwarf galaxies. These metals add up along the Hubble time so that at redshift zero the cosmic-web metallicity is predicted to have a value close to the observed threshold (e.g., Oppenheimer et al. 2012). If this explanation were correct, then the observational lower limit would be tracing the metal content of the cosmic web.

4.6. Morphology

Papaderos et al. (2008), Paper I, and later Filho et al. (2013) pointed out that bursting XMPs tend to be cometary or tadpole-like. Such association between metallicity and morphology is notable because metallicity is a property dictated by the spectrum and, therefore, it implies that the spectrum conditions the morphology of the galaxy. In order to check whether the new sample of XMPs maintains this preference for the cometary shape, we repeated the eyeball classification carried out by Filho et al. (2013), which was based on the scheme presented in Paper I. The galaxies may be (1) symmetric if they present a centrally concentrated emission with no obvious structure in the outskirts, (2) cometary for a head-tail structure with an identifiable knot at the head, (3) two-knot for a structure with two knots, and (4) multi-knot for a diffuse structure with multiple star-formation knots, including irregular galaxies. Figure 10 illustrates these classes with two examples of each galaxy morphology. We use color images from SDSS for the visual inspection, and the
results are:

Cometary 57%,
Symmetric 23%,
Multi-knot 10%,
2-knot 4%,
others 6%.  

In agreement with previous estimates, cometary happens to be the dominant shape. Asymmetric shapes (i.e., cometary plus two-knot plus multi-knot) represent 71% of the sample. These results are in quantitative agreement with Paper I, where 75% of the targets are asymmetric, and with Filho et al. (2013), where cometary, multi-knot, and two-knot represented 52%, 16%, and 10% of the XMP sample, respectively. Perhaps the main difference is a slight excess of symmetric targets in the present sample, which we attribute to insufficient spatial resolution because our targets have larger redshifts than the other samples (Section 4.3), thus complicating the detection of sub-structure. In addition, we include a fifth class in the classification (others in Equation (15)) to collect objects that do not fit in the main categories easily (e.g., mergers or elongated galaxies without an identifiable knot). In addition, two of the objects are H II regions in the outskirts of large spirals (20 and 59).

We note that a significant part of the symmetric objects looks like the green peas first described by Cardamone et al. (2009). One example is shown in Figure 10—the blue symmetric galaxy J1423+22. Depending on the redshift, green-pea like galaxies may look green, pink, or blue in SDSS images (Sánchez Almeida et al. 2013a). The identification of green peas with metal-poor galaxies was carried out by Amorín et al. (2010, 2012a), and their rounded morphology seems to be associated with limitations in spatial resolution, rather than being intrinsic (e.g., Amorín et al. 2014). The finding of green peas in our sample is encouraging because it naturally connects local XMPs to those expected at higher redshifts, which will appear in surveys deeper than SDSS-DR7 (e.g., eBOSS11 within SDSS IV).

4.7. Underlying Stellar Populations

In order to decontaminate the emission-line fluxes from stellar absorption, we fitted a composite stellar spectrum to the observed continuum (Section 3.1). As byproduct of this exercise, we have information on the stellar populations that coexist with the gas responsible for the emission-line spectrum. The stellar-population decomposition is not free from uncertainties and degeneracies, however, it suffices to provide a general view of the stellar properties (e.g., Cid Fernandes et al. 2005, 2014).

We find that the underlying stellar populations are generally young, although in almost all cases there are also old stellar components contributing to the observed spectrum. Figure 11 shows the mass-weighted stellar age versus the light-weighted stellar age for the set of XMPs. The light-weighted mean is strongly biased toward young stars, and we find that most light-weighted ages are younger than 1 Gyr (i.e., they are to the left of the vertical dashed line in Figure 11). The mass-weighted mean, however, does not suffer from this bias, and most XMPs have mass-weighted ages significantly larger than 1 Gyr (i.e., they are above the horizontal dashed line in Figure 11). In other words, most XMPs seem to have been forming stars not only at present but also during the last Gyr. On top of this recent star formation, most contain old stars that

Figure 10. Examples of the various morphological types found among the XMPs; from left to right cometary, symmetric, multi-knot, and two-knot (see main text for details). The bar on top of each image gives an angular scale on the sky. The image in the top-left panel also points out north and east, which are common to all images. Galaxy images are from SDSS-DR12. The two-knot example on the upper row corresponds to I Zw18.
galaxies contains evolved stellar populations is by no means
biased toward the younger stellar populations, whereas the mass-weighted
mean is not. The vertical dashed line shows a light-weighted stellar age of
1 Gyr. Note that most XMPs are to the left of this threshold and thus present a
mass-weighted stellar age vs. light-weighted stellar age for the
Figure 11. Mass-weighted stellar age vs. light-weighted stellar age for the
XMPs, both derived using STARLIGHT (Section 3.1). The light-weighted mean
is biased toward the younger stellar populations, whereas the mass-weighted
mean is not. The vertical dashed line shows a light-weighted stellar age of
1 Gyr. Note that most XMPs are to the left of this threshold and thus present a
stellar spectrum that is typical of young stellar populations. The horizontal
dashed line shows a mass-weighted stellar age of 1 Gyr. Most XMPs are above
this line, implying that despite their seemingly young spectrum, XMPs also
contain aged stellar populations—often 10 Gyr old. The ages in the axes are
given in Gyr, and the slanted solid line corresponds to the one-to-one
relationship.

are often as old as 10 Gyr. The fact that the class of XMP
galaxies contains evolved stellar populations is by no means
unknown (see Section 1). The large difference between mass-
weighted age and light-weighted age is common to most local
star-forming galaxies (e.g., Pérez-Montero et al. 2010; Sánchez
Almeida et al. 2012).

4.8. Large Scale Environment

As it was done by Filho et al. (2015) for the XMPs in
Paper I, we explored the large-scale environment of the new list
of bursting XMPs using constrained N-body cosmological
numerical simulations of the local universe (Nuza et al. 2014a).
The cosmological parameters provide a statistical description of
the universe. The constrained simulation that we use selects a
particular realization of the current cosmological model that
reproduces the spatial distribution of galaxies in the local
universe, as observed by 2MRS (Huchra et al. 2012). The method is based on a self-consistent Bayesian machine-learning
algorithm, and provides a description of the local
cosmic-web dark matter with a resolution around 2 Mpc in a
box of 180 Mpc h⁻¹ side (Kitaura et al. 2012; Heß et al. 2013;
Kitaura 2013). This simulation, tuned to replicate the local
distribution of galaxies, tells us whether the gravitational field
at each point forces the matter to collapse in three directions
(knot), two directions (filament), or one direction (sheet), or if
the matter is expanding in all three directions (void). This
classification, based on the local gravitational potential,
correlates very well with the local dark-matter overdensity,
which increases from voids to knots. The tendency of a
particular type of galaxy to prefer one of the four environments
is quantified in terms of the excess probability ratio η(τ, ϵ),
which is defined as the ratio between the probability that a
galaxy of type τ appears in environment ϵ, and the probability
of having any type of galaxy in this environment. The excess
probability for the XMPs is shown in Figure 12. For reference,
the figure also includes the excess probability for ellipticals (E),
lenticulars (S0), spirals (Sp), and irregulars (Irr) as worked out
from 2MRS galaxies by Nuza et al. (2014a). Clearly, XMPs
have a strong tendency to appear in voids and avoid knots. This
tendency is more clear in XMPs than in any other galaxy type.
These results are very similar to those found by Filho et al.
(2015) for the XMPs in Paper I.

5. DISCUSSION AND CONCLUSIONS

Due to their astrophysical interest and rareness (Section 1),
we carried out a systematic search for XMP galaxies in
SDSS-DR7. Starting from all galaxies with spectra having
redshift smaller than 0.25, we end up with a list of only
196 XMPs (Table 1). We select galaxies with [O III]λ4363
that is particularly intense with respect to [O II]λ4959 and
[O III]λ5007. As explained in Section 2, this implies having a
high-temperature gas and, since metals are efficient coolants,
it is the fingerprint of low-metallicity gas. A first automated
search was carried out using k-means to classify all galaxy
spectra according to the shape in the region that contains
[O II]λ4363 and [O III]λ5007, 5007. The 1281 objects in the
classes of high [O II]λ4363 underwent closer scrutiny by
computing their O abundance using the method of Pérez-
Montero (2014, Section 3.2), which is a model-based approach
consistent with the direct method. Those with an abundance
smaller than a tenth of the solar value were included in the
final list (Table 1). The existence of uncertainty in the
metallicity determination, typically 0.17 dex, implies that some
of the XMPs are false positives and some of the discarded
targets are false negatives (Appendix). We offer a second
list with 332 XMP candidates (Table 2) that were selected
under less restrictive noise conditions; this list contains
objects that upon refinement of the metallicity estimate may
turn out to be XMPs. The search in the present paper is
complementary to the one carried out in Paper I, where
k-means is used in the spectral region around Hα, where the
metal-poor signpost is the weakening of [N II]λ6549 and [N II]
λ6583 with respect to Hα. From the 196 galaxies in Table 1,
only 31 are known, as inferred by comparison with the
compilation carried out in Paper I and with the new targets in
Izotov et al. (2012). The remaining 165 objects seem to be new
members of the XMP class. Among the unknown members,
18 have 12 + log(O/H) ≤ 7.50.

As part of the procedure to measure O/H, we also determine
the ratio N/O (Section 3.2). Closed-box stellar evolution over
long timescales predicts log(N/O) to be of the order −1.5 at
low metallicity, and when 12 + log(O/H) > 8 it increases
linearly with increasing metallicity, reaching ≈ −1.0 at solar
metallicity (e.g., Henry & Worthey 1999; Köppen & Hensler
2005). Most XMPs have N/O that is consistent with the value
expected at low metallicity, however, a number of them show
log(N/O) ≈ −1.0 or even larger (Figures 2(a) and 7). Metal-
poor galaxies with log(N/O) ≈ −1.0 are not very common,
but examples do exist in the literature. For instance, the so-
called green-pea galaxies have high N/O despite their low
metallicity (Amorín et al. 2010, 2012b; we have several such
green peas among our XMP galaxies—see Section 4.6).
Moreover, some of the classical XMPs also show enhanced
N/O (Morales-Luis et al. 2014b). XMPs with large N/O do not

---

12 The 2MASS Redshift Survey (2MRS) is the spectroscopic follow-up of the
two micron all sky survey (2MASS), which provides a 3D distribution of local
galaxies—R.A., decl., and redshift.
appear in Paper I because we selected galaxies with weak N lines, so our search was biased toward XMPs of low N/O. A simple interpretation for having both large N/O and low O/H is the accretion of significant amounts of metal-poor gas onto galaxies with solar metallicity ISM (e.g., Amorín et al. 2010, and references therein). The mixing of metal-poor gas with pre-existing metal-rich gas drops down the metallicity (i.e., increases H in O/H), but does not modify the original N/O, which remains as for a metal-rich ISM. Some 10% of the XMPs have log N/O $\approx -1.2$, representing objects where the accretion of metal-poor gas is arguably part of their star-formation history. Note that accretion will remain unnoticed when the pre-existing ISM presents low N/O. Interpreting N/O as a tracer of previous star-formation episodes is consistent with the trend we observe for the reddening to increase with increasing N/O (Figure 5(b)). The correlation naturally arises if part of the dust responsible for extinction was produced with N and O in previous star-formation episodes.

We quantify the global properties of the XMPs. They have a mean stellar mass around $M_{10^8}$, but with very large scatter going from $10^6 M_\odot$ to $10^{10} M_\odot$, so some are fairly massive (Section 4.2). We find that both N/O and extinction increase with increasing stellar mass, so that the contribution of evolved stellar populations to the metal content and dust is more important in massive galaxies. From the reddening of the emission-line spectrum, we infer the dust-mass column density to be around $0.2 M_\odot pc^{-2}$, which renders a very moderate extinction coefficient (Section 4.1). For a typical star-forming region size, the dust mass is between 200 and 20,000 $M_\odot$. We analyze the morphology of the XMPs, finding the results to be in line with previous studies, showing a strong tendency for the XMPs to be tadpole-like or cometary (Papaderos et al. 2008; Paper I; Filho et al. 2013). Explicitly, 71% are cometary or knotted, and only 23% are symmetric (see Section 4.6).

Since the parent sample used to select the XMPs is limited in apparent magnitude, it is relatively simple to estimate the number density of XMPs in the local universe. We worked it out in Section 4.4, finding some $3.4 \times 10^{-3}$ XMPs per cubic Mpc (Equation (13)). This density is significantly larger than the density obtained from the sample in Paper I, and it represents some 2% of all galaxies in a volume. We also work

Figure 12. Excess probability ratio for the XMPs; E, S0, Sp, and Irr galaxies are also included for reference (see the main text for details). The four panels correspond to the four types of environments: voids, sheets, filaments, and knots. XMPs have a strong tendency to appear in voids and avoid knots.
out the large-scale environment of the XMPs using constrained cosmological numerical simulations (Section 4.8). XMPs have a strong tendency to appear in voids and avoid cluster environments.

One of the most intriguing properties of XMPs is the existence of a lower limit metallicity, of the order of 2% the solar metallicity. The limit remains in our list (Section 4.5), despite the fact that the sample is significantly larger than the previous ones, and that we have systematically searched the full SDSS spectroscopic sample. One of the ways to explain the existence of a threshold is once again, the infall of intergalactic gas feeding the star-formation process. The intergalactic gas collects metals ejected from galaxies along the Hubble time, which yields a small but non-zero metallicity that is expected to be around the value of the observed threshold (Sánchez Almeida et al. 2014a and references therein).

A caveat to keep in mind is that any search for XMPs based on optical surveys, such as the one we present, overlooks faint low surface brightness galaxies. We are sensitive to star-bursting XMPs (see Section 1). As pointed out by Skillman et al. (2013; see also Section 1), for example, most XMPs may be low surface brightness quiescent XMPs, so identifying them would require resorting to blind HI surveys like the Arecibo Legacy Fast ALFA Survey (Haynes et al. 2011), which led to the discovery of Leo P (Giovanelli et al. 2013).

Thanks are due to Amanda del Olmo for pinpointing repetitions in the original list of objects, Andrés Asensio-Ramos for assistance with the estimate of false positives and negatives, Mercedes Filho for enlightening discussions and their errors derived in Section 3.3, Equations 18.

APPENDIX
NUMBER OF FALSE POSITIVES AND NEGATIVES

For the sake of compactness, we use the symbol $X_i$ to denote the value of $12 + \log(O/H)$ in the $i$th XMP galaxy. By definition, $X_i < X_0$, with $X_0$ being the threshold to be XMP. Bayes’s theorem (e.g., Martin 1971) provides the probability that the true metallicity is $X$, given that we have observed $X_i$,

$$P(X|X_i) = P(X|X)H(X)/P(X_i),$$

where $P(X|X)$ stands for the probability of measuring $X_i$ when the true metallicity is $X$, $H(X)$ represents the probability that the galaxy has a true metallicity $X$, and $P(X_i)$ corresponds to the probability that the measured metallicity of the galaxy is $X_i$ considering all possible values of the true metallicity, i.e.,

$$P(X_i) = \int_{-\infty}^{\infty} P(X|X)H(X)dX.$$  

If $X$ is larger than the threshold metallicity $X_0$, Equation (16) provides the probability that the galaxy gives a false positive if its true metallicity is $X$. Because the true metallicity yielding false positives can be any value larger than the threshold, the total probability of a false positive $P(X \geq X_0|X_i)$ is simply the sum over all possibilities, that is, an integral of Equation (16) for all $X \geq X_0$.

$$P(X \geq X_0|X_i) = \int_{X_0}^{\infty} P(X>X_0|X)H(X)dX/\int_{-\infty}^{\infty} P(X|X)H(X)dX.$$  

Assuming the distribution of observational errors to be Gaussian, the probability of measuring $X_i$ when the true metallicity is $X$ turns out to be,

$$P(X_i|X) = \exp\left\{-\frac{1}{2}\left(\frac{X_i - X}{\sigma_1}\right)^2\right\}/\sqrt{2\pi}\sigma_1,$$

with $\sigma_1$ the error (i.e., the square root of the variance of the distribution of errors). Equation (18) quantifies the probability that a single galaxy gives a false positive (i.e., it gives the number of false positives per galaxy). When we consider the full set, the total number of false positives $N_{\text{false+}}$ is the sum of the number of false positives for all XMP galaxies, or

$$N_{\text{false+}} = \sum_{i} P(X \geq X_0|X_i).$$

By definition, false negatives are true XMPs that end up having $X_i < X_0$. The number of false negatives, $N_{\text{false-}}$, can be worked out in a way that is similar to the derivation of $N_{\text{false+}}$. This time the sum is over all non-XMP galaxies,

$$N_{\text{false-}} = \sum_{j} P(X < X_0|X_j),$$

with

$$P(X < X_0|X_j) = \int_{-\infty}^{X_0} P(X|X)H(X)dX/\int_{-\infty}^{\infty} P(X|X)H(X)dX,$$

giving the probability of being XMP even though the measured metallicity $X_j$ is larger than the threshold $X_0$.

Assuming that the probability density function of possible oxygen abundances $H(X)$ is given by the observed histogram of abundances in Figure 2(b), and using the metallicities and their errors derived in Section 3.3, Equations (18)–(22) yield,

$$N_{\text{false+}} \simeq 82,$$

$$N_{\text{false-}} \simeq 46.$$  

If we assume a uniform distribution of abundances in the interval $7 \leq 12 + \log(O/H) \leq 8.5$, then

$$N_{\text{false+}} \simeq N_{\text{false-}} \simeq 55.$$
