



The CaFe Project: Optical Fe II and Near-infrared Ca II Triplet Emission in Active Galaxies. II. The Driver(s) of the Ca II and Fe II and Its Potential Use as a Chemical Clock

Mary Loli Martínez-Aldama¹ , Swayamtrupta Panda^{1,2} , Bożena Czerny¹ , Murilo Marinello³ , Paola Marziani⁴ , and Deborah Dultzin⁵

¹Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland; mmary@cft.edu.pl

²Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, ul. Bartycka 18, 00-716 Warsaw, Poland

³Laboratório Nacional de Astrofísica, R. dos Estados Unidos, 154—Nações, Itajubá—MG, 37504-364, Brazil

⁴INAF-Astronomical Observatory of Padova, Vicolo dell'Osservatorio, 5, I-35122 Padova PD, Italy

⁵Universidad Nacional Autónoma de México Instituto de Astronomía: Ciudad de México, Distrito Federal, MX 04510, Mexico

Received 2021 January 15; revised 2021 May 12; accepted 2021 May 19; published 2021 September 1

Abstract

In this second paper in the series, we carefully analyze the observational properties of the optical Fe II and near-IR Ca II triplet in active galactic nuclei (AGNs), as well as the luminosity, black hole mass, and Eddington ratio in order to define the driving mechanism behind the properties of our sample. The Ca II shows an inverse Baldwin effect, bringing out the particular behavior of this ion with respect to the other low-ionization lines such as H β . We performed a principal component analysis, where 81.2% of the variance can be explained by the first three principal components drawn from the FWHMs, luminosity, and equivalent widths. The first principal component (PC1) is primarily driven by the combination of black hole mass and luminosity with a significance over 99.9%, which in turn is reflected in the strong correlation of the PC1 with the Eddington ratio. The observational correlations are better represented by the Eddington ratio; thus, it could be the primary mechanism behind the strong correlations observed in the Ca II–Fe II sample. Since calcium belongs to the α -elements, the Fe II/Ca II flux ratio can be used as a chemical clock for determining the metal content in AGNs and trace the evolution of the host galaxies. We confirm the de-enhancement of the ratio Fe II/Ca II by the Eddington ratio, suggesting a metal enrichment of the BLR in intermediate- z with respect to low- z objects. A larger sample, particularly at $z > 2$, is needed to confirm the present results.

Unified Astronomy Thesaurus concepts: Quasars (1319); Active galactic nuclei (16); Spectroscopy (1558); Supermassive black holes (1663)

1. Introduction

The large diversity of the emission lines observed in the spectrum of the active galactic nuclei (AGNs) reveals a complex structure of the broad-line region (BLR). The physical conditions of the BLR such as density, ionization parameter, and metallicity can be estimated by the flux ratios of the emission lines, and their profiles supply information of the dynamics in the BLR cloud (Wandel 1999; Negrete et al. 2014; Schnorr-Müller et al. 2016; Devereux 2018). Emission lines can be divided considering their ionization potential (IP). Typically, high-ionization lines (HILs) show IP > 40 eV, while low-ionization lines (LILs) have IP < 20 eV (Collin-Souffrin et al. 1988; Marziani et al. 2019). Reverberation mapping studies have confirmed the stratification of the BLR (e.g., Horne et al. 2021), where HILs such as C IV $\lambda 1549$ or He II $\lambda 1640$ are emitted closer to the central continuum source, and LILs such as H β or Mg II $\lambda 2800$ are emitted at least three times farther. The presence of emission lines with very low IPs (IP ~ 10 eV) such as the multiple permitted Fe II transitions or the Ca II $\lambda\lambda 8498, 8542, 8662$ triplet (hereafter CaT) suggests the existence of a zone shielded from the high-energy photons emanated by the central source and likely located in the outermost portion of the BLR (Joly 1987; Dultzin-Hacyan et al. 1999; Rodríguez-Ardila et al. 2002; García-Rissmann et al. 2012; Rodríguez-Ardila et al. 2012; Marinello et al. 2016).

The physical conditions of the Fe II have been widely explored in a broad wavelength range since it provides useful

information about the energy budget of the BLR (Vestergaard & Wilkes 2001; Osterbrock & Ferland 2006). However, its complex electronic structure owing to the various ionization and excitation mechanisms complicates the model of the Fe II (Collin & Joly 2000; Baldwin et al. 2004). This ionic species manifests as a pseudo-continuum owing to the numerous blended multiplets ranging from the UV to the near-IR (NIR). In our studies (see e.g., Panda et al. 2020a, hereafter Paper I), we incorporate the Fe II data set from Verner et al. (1999), which includes a 371 level with IP up to ~ 11.6 eV, available in CLOUDY (Ferland et al. 2017). Newer Fe II models are now available that have calculated more energy levels for this species, reaching up to 26.4 eV (see Sarkar et al. 2021, for a recent compilation). This model reproduces well the UV and optical Fe II contribution observed in I Zw 1, constraining in a better way the physical conditions of the Fe II emitting clouds. For more details on the progress in understanding the Fe II emission in AGNs and its modeling, we refer the readers to Paper I.

The singly ionized calcium emission can be approximately modeled by a five-level atom: (1) the optical H and K lines ($\lambda\lambda 3933, 3968$) are emitted from the 4p level to the 4s ground level, (2) the infrared multiplet ($\lambda\lambda 8498, 8542, 8662$, CaT) arises from the 4p level to the 3d metastable level, and (3) the forbidden multiplet ($\lambda\lambda 7291, 7324$) arises from the 3d metastable level to the ground level (Ferland & Persson 1989; Marziani et al. 2014). Due to similarity between the IPs of Ly α (10.2 eV) and the singly ionized Ca II (11.8 eV), the 3d

metastable level is highly populated and the collisional excitation process leading to the infrared CaII triplet emission is efficient. Thus, the NIR CaT offers the possibility to study the properties of the very LILs in the BLR. CaT is prominent in narrow-line Seyfert 1 (NLS1) galaxies (Persson 1988; Marinello et al. 2016) and quasars (Martínez-Aldama et al. 2015a). However, when the stellar continuum has a significant contribution, the emission profile shows a central dip or, in extreme cases, only an absorption profile is observed. Therefore, a correct subtraction of the stellar component is needed, particularly in low-luminosity sources. The CaT absorption is mainly observed in Seyfert 1 and Seyfert 2 galaxies, where it may be enhanced by a population of red supergiant stars associated with a starburst (Terlevich et al. 1990). The velocity dispersion provided by the stellar CaT has been used to infer the stellar populations and determine the black hole mass throughout the relation $M_{\text{BH}} - \sigma_*$ (García-Rissmann et al. 2005).

Some theoretical and observational studies have been devoted to looking for the connections between the optical Fe II and CaT. Both ions show a strong linear relation and similar widths, narrower than $\text{H}\beta$ or $\text{Pa}\beta$ (Persson 1988; Martínez-Aldama et al. 2015a, 2015b; Marinello et al. 2016; Panda et al. 2020a), suggesting that both emission lines are emitted in the outer parts of the BLR. According to the photoionization models, both emission lines share almost identical physical conditions—large clouds (column densities $\sim 10^{24} \text{ cm}^{-2}$) with high mean densities ($\sim 10^{12} \text{ cm}^{-3}$) and relatively low temperatures ($\lesssim 8000 \text{ K}$) (Joly 1987, 1989; Ferland & Persson 1989; Panda et al. 2020a; Panda 2021).

In the first paper of the presented analysis (Paper I), we updated the observational correlation between the strengths of the two species (i.e., the flux ratios $\text{Fe II}/\text{H}\beta$ and $\text{CaT}/\text{H}\beta$, hereafter R_{FeII} and R_{CaT} , respectively) given by

$$\log R_{\text{CaT}} \approx (0.974 \pm 0.119) \log \{R_{\text{FeII}} - (0.657 \pm 0.041)\}. \quad (1)$$

We also looked extensively at the optical Fe II and CaT emission from a theoretical standpoint, using the photoionization models, which are compared with an up-to-date sample of Fe II and CaT. We tested various photoionization models in terms of ionization parameter, cloud density, metallicity, and column density and found an overlapping range of physical conditions that are required to efficiently excite these two species. We also find that the strong Fe II emitters, in order to be well modeled, require a range of metallicity from solar to supersolar (Martínez-Aldama et al. 2018; Śniegowska et al. 2021). This result is obtained by comparing the observed UV flux ratios of emission lines such as C IV $\lambda 1549$, Al III $\lambda 1860$, Si IV $\lambda 1397 + \text{O IV} \lambda 1402$, or N V $\lambda 1240$ over He II $\lambda 1640$ with the ones predicted by CLOUDY simulations. The correlation between the stronger Fe II emitters, metallicity, and Eddington ratio has been confirmed by several independent studies (e.g., Hamann & Ferland 1992; Shin et al. 2013; Panda et al. 2019).

In a subsequent paper, Panda (2021), we furthered the photoionization modeling to recover the EWs in the low-ionization emitting region in the BLR and realize the anisotropy in the accretion disk emission, leading to a better understanding of the photoionization of the low-ionization emitting regions of the BLR.

In this part of the series, we look at the observational properties and correlations from the up-to-date optical and NIR

measurements centered around Fe II and CaT emission, respectively. Usually, the stronger Fe II and CaT emitters are associated with the NLS1 AGNs, but also AGNs with higher luminosities and broader profiles show a strong emission for these two species (Martínez-Aldama et al. 2015a). Since Fe II strength (or R_{FeII}) is apparently driven by the Eddington ratio (Boroson & Green 1992; Marziani et al. 2003; Dong et al. 2011; Zamfir et al. 2010; Panda et al. 2018, 2019), it motivates us to explore the role of the Eddington ratio, black hole mass, and luminosity in the CaT and Fe II properties to decipher the primary driver leading to this observed correlation between the two species.

Additionally, since calcium belongs to the α -elements and iron is mainly produced by Type I supernovae (SNe) on relatively longer timescales, the flux ratio $\text{Fe II}/\text{Ca II}$ can be used as a proxy for estimating the chemical enrichment (Martínez-Aldama et al. 2015a), as has been tested with the UV Fe II and Mg II $\lambda 2800$ (Verner et al. 2009; Dong et al. 2011; Shin et al. 2019; Onoue et al. 2020, and references therein). Therefore, a deep observational analysis is required.

The paper is organized as follows: In Section 2, we include a short review of the sample. Section 3 describes the methods employed to estimate the black hole mass and Eddington ratio. In Section 4, we report the observational correlations of our sample, including the Eigenvector 1 sequence and the Baldwin effect. In order to confirm the correlations found, we performed a principal component analysis (PCA), the results of which are shown in Section 5. In Section 6 we discuss the potential drivers of the CaT–Fe II properties, the Baldwin effect, and the $\text{Fe II}/\text{CaT}$ ratio as a possible metal indicator. Conclusions are summarized in Section 7. Throughout this work, we assume a standard cosmological model with $\Omega_{\Lambda} = 0.7$, $\Omega_m = 0.3$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2. Observational Data

Our analysis is based on the observational properties of $\text{H}\beta$, optical FeII (4434–4684 Å), and the NIR CaII triplet collected from Persson (1988), Martínez-Aldama et al. (2015a, 2015b), Marinello et al. (2016), and Marinello et al. (2020). A detailed description of the full sample is discussed in Paper I. The full sample includes 58 objects with $42.5 < \log L_{\text{opt}} (5100 \text{ Å}) < 47.7$ at $0.01 < z < 1.68$. Due to the different selection criteria of the subsamples, the full sample shows a bimodal distribution in redshift and luminosity, where 58% of the sample shows $z < 0.1$ and $\log L_{\text{opt}} \sim 44$, while the rest of the objects are located at $z \sim 1.6$ with $\log L_{\text{opt}} \sim 47.4$ (see Figure 1). Therefore, our sample is affected by such biases, which could influence our results. These aspects are discussed in Section 4 and Section 6.

The optical measurements from Persson (1988) are originally reported by Oke & Shields (1976), Osterbrock (1976), Osterbrock & Phillips (1977), Koski (1978), Kunth & Sargent (1979), and de Bruyn & Sargent (1978). However, the quality of the data was not so high like in recent times; therefore, this sample should be treated with caution. There are five sources in common in Marinello’s and Persson’s samples. The variations in the different observational parameters are significant in three of them (Mrk 335, Mrk 493, and IZw 1; see Table A1). This could be an indication of the quality of the measurements. However, Marinello’s and Persson’s samples include typical NLS1y objects; thus, a similar behavior is expected, such as Figure 2 shows. In order to disentangle this point, new observations of the Persson sample are needed.

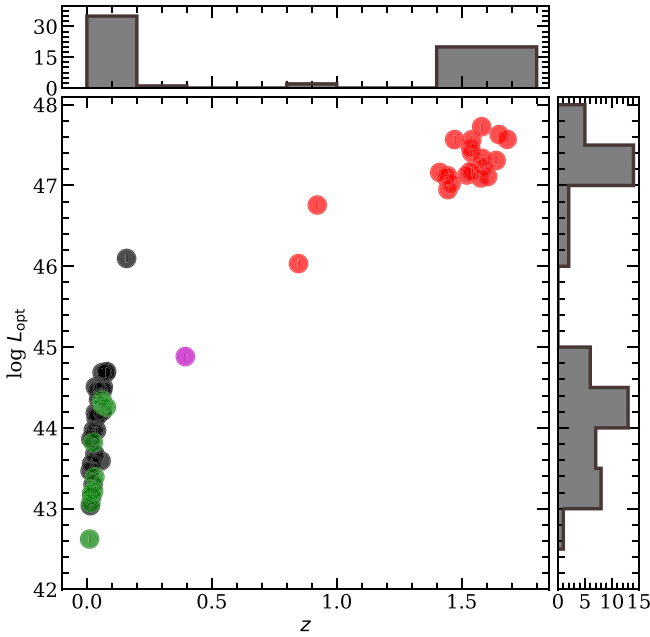


Figure 1. Redshift distribution of the sample as a function of optical luminosity at 5100 Å in units of erg s^{-1} . Black, red, green, and magenta symbols correspond to Persson (1988), Martínez-Aldama et al. (2015a), Martínez-Aldama et al. (2015b), Marinello et al. (2016), and Marinello et al. (2020) samples, respectively.

Table A1 reports the properties of the each source in the sample such as redshift, optical (at 5100 Å) and NIR (at 8542 Å) luminosity, the flux ratios R_{FeII} and R_{CaT} , and the equivalent width (EW) and FWHM of H β , CaT, and O I λ 8446. All the measurements were taken from the original papers (Persson 1988; Martínez-Aldama et al. 2015a, 2015b; Marinello et al. 2016, 2020). Since Persson (1988) does not report the luminosities at 5100 Å, we have estimated them from their apparent V magnitudes reported in the Veron-Cetty & Veron (2010) catalog. We have considered a zero-point flux density of $3.55 \times 10^{-9} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1}$ (Bessell 1990) to estimate the flux at 5500 Å in the observed frame. After correcting for the redshift, we assumed a slope of $\alpha_{\lambda} = -1.67$ (Vanden Berk et al. 2001) to estimate the flux at 5100 Å. Finally, the distance to the source was obtained through classical integration assuming the cosmological parameters specified at the end of Section 1.

3. Parameter Estimations

3.1. Black Hole Mass

The black hole mass (M_{BH}) is estimated using the classical relation given by

$$M_{\text{BH}} = f_{\text{BLR}} \frac{R_{\text{BLR}} v^2}{G}, \quad (2)$$

where G is the gravitational constant, f_{BLR} is the virial factor, R_{BLR} is the broad-line region size, and v is the velocity field in the BLR, which is represented by the FWHM of H β . The virial factor includes information of geometry, kinematics, and inclination angle of the BLR. Typically, it is assumed as constant (~ 1); however, some results point out that this factor should vary along the AGN populations (e.g., Collin et al. 2006; Yu et al. 2019). In this work, we assume the virial

factor proposed by Mejía-Restrepo et al. (2018), which is anticorrelated with the FWHM of the emission line: $f_{\text{BLR}} = (\text{FWHM}_{\text{H}\beta}/4550 \pm 1000)^{-1.17}$.

For single-epoch spectra, the R_{BLR} is usually estimated through the radius–luminosity (RL) relation (Bentz et al. 2013), given by

$$\log \left(\frac{R_{\text{BLR}}}{1 \text{ lt} - \text{day}} \right) = (1.527 \pm 0.31) + 0.533^{+0.035}_{-0.033} \log \left(\frac{L_{\text{opt}}}{10^{44} L_{\odot}} \right), \quad (3)$$

where L_{opt} corresponds to the luminosity at 5100 Å. Black hole mass estimations are reported in Table A2. The sample shows a clear distinction between low and high black hole masses ($\log M_{\text{BH}} \sim 7 - 10 M_{\odot}$).

3.2. Eddington Ratio

The accretion rate is estimated by the classical Eddington ratio defined by $L_{\text{bol}}/L_{\text{Edd}}$, where L_{bol} is the bolometric luminosity and L_{Edd} is the Eddington luminosity defined by $L_{\text{Edd}} = 1.5 \times 10^{38} \left(\frac{M_{\text{BH}}}{M_{\odot}} \right)$. The bolometric luminosity formally can be estimated integrating the area under the broadband spectral energy distribution (SED; e.g., Richards et al. 2006, and references therein). However, since this process requires multiwavelength data to constrain the SED fitting process, it is hard to get an estimation for individual sources. Mean SEDs have been used to estimate average values called bolometric correction factors (k_{bol}), which scale the monochromatic luminosity (λL_{λ}) to give a rough estimation of $L_{\text{bol}} = k_{\text{bol}} \cdot \lambda L_{\lambda}$. Usually, k_{bol} is taken as a constant for a monochromatic luminosity; however, results like the well-known non-linear relationship between the UV and X-ray luminosities (e.g., Lusso & Risaliti 2016, and references therein) indicate that k_{bol} should be a function of luminosity (Marconi et al. 2004; Krawczyk et al. 2013). Along the same line, Netzer (2019) proposed new bolometric correction factors as a function of the luminosity assuming an optically thick and geometrically thin accretion disk, over a large range of black hole mass ($10^7 - 10^{10} M_{\odot}$), Eddington ratios (0.007–0.5), spin (-1 to 0.998), and a disk inclination angle of 56° . For the optical range, the bolometric correction factor is given by

$$k_{\text{bol}} = 40 \left(\frac{L_{\text{opt}}}{10^{42}} \right)^{-0.2}, \quad (4)$$

where L_{opt} corresponds to the luminosity at 5100 Å. The wide option of parameters considered for the model process provides a better approximation corroborating previous results (Nemmen & Brotherton 2010; Runnoe et al. 2012a, 2012b). In addition, it provides a better accuracy than the constant bolometric factor correction, which led to errors as large as 50% for individual measurements. Therefore, we explore the use of k_{bol} for estimating the Eddington ratio. Table A2 reports the Eddington ratios utilizing the BH masses obtained using the classical RL relation (Equation (3)).

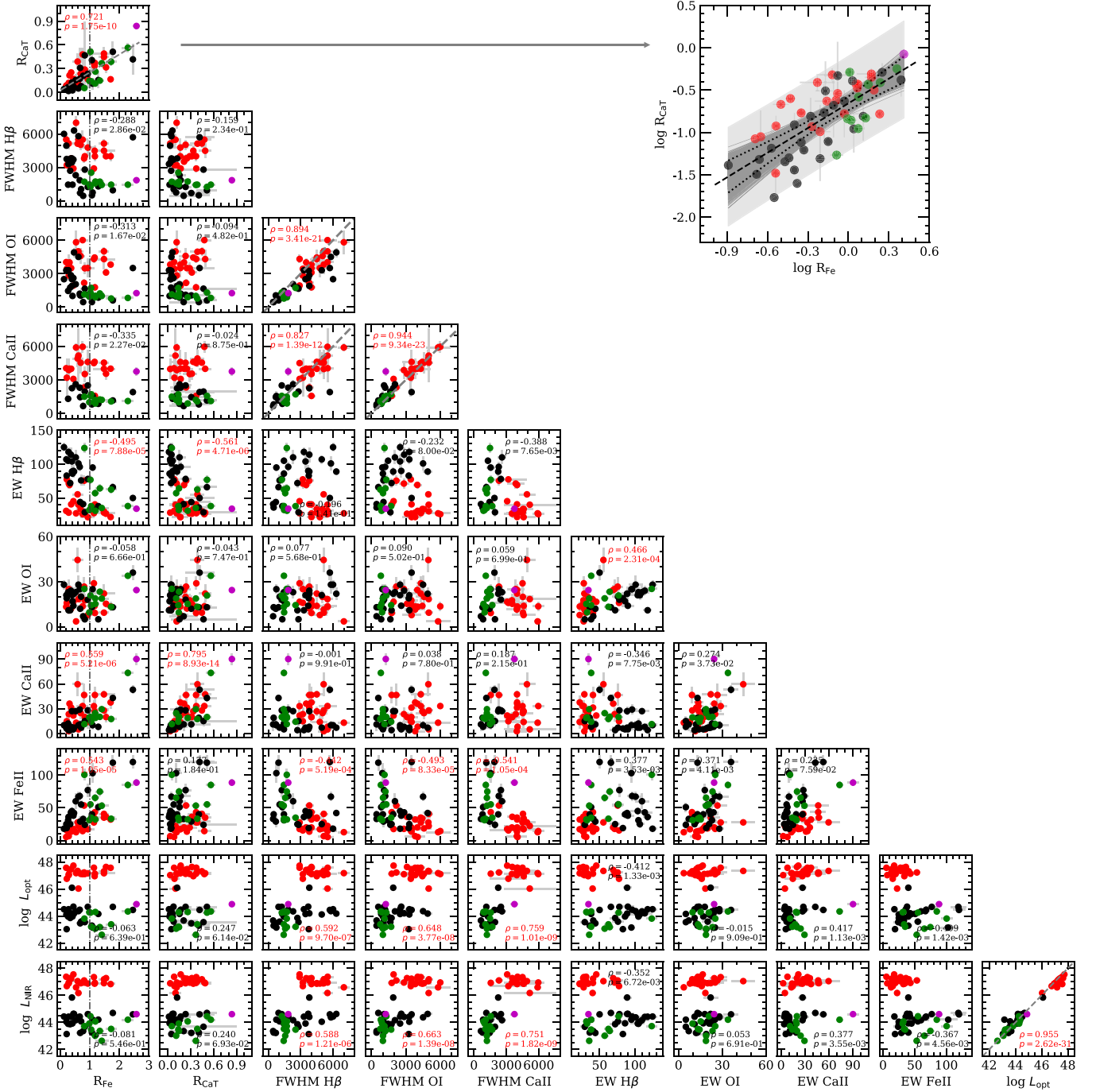


Figure 2. Correlation matrix for emission lines and continuum properties. Black, red, green, and magenta symbols correspond to Persson (1988), Martínez-Aldama et al. (2015a, 2015b), Marinello et al. (2016), and Marinello et al. (2020) samples, respectively. Each panel specifies the Spearman’s rank correlation coefficient and the p -value, where the significant correlations are colored in red. EW, FWHM, and optical and NIR luminosities are given in units of \AA , km s^{-1} , and erg s^{-1} , respectively. Gray vertical lines in the first column mark the limit for super-Eddington sources at $R_{\text{FeII}} = 1.0$. In the diagrams where the FWHMs and luminosities are correlated, the gray dashed line marks the 1:1 relation. In the correlation $R_{\text{CaT}} - R_{\text{FeII}}$ the gray dashed line corresponds to Equation (1). The inset right panel shows the relation $R_{\text{FeII}} - R_{\text{CaT}}$ in log scale with the results of the bootstrap analysis; see Section 4.5. Black dotted lines mark the confidence intervals at 95% for the 1000 realizations (dark-gray lines) of the bootstrap analysis. The light-gray patch marks the corresponding prediction interval bands for the sample.

4. The Correlation Analysis

4.1. Observational Pairwise Correlations

Figure 2 shows the correlation matrix of the observational parameters: optical (L_{opt} at 5100 \AA) and NIR (L_{NIR} at 8542 \AA) continuum luminosities, the flux ratios R_{FeII} and R_{CaT} , and the emission-line properties such as FWHM and the EW of H β , O I

$\lambda 8446$, and CaT, plus the EW of Fe II. In order to stress the difference in luminosity and FWHM values in the subsamples, they are identified by different colors. Each panel also includes the Spearman’s rank correlation coefficient (ρ) and the p -value, where significant correlations ($p < 0.001$) are colored in red (otherwise shown in black). Optical and NIR luminosities follow a linear relation (Figure 2); therefore, both luminosities

show the same behavior as the rest of the observational properties.

The top panel of Figure 2 shows the strong correlation between R_{FeII} and R_{CaT} , which is described by Equation (1) (gray dashed line; see also inset panel). The anticorrelation between R_{FeII} (or R_{CaT}) and $\text{EW}_{\text{H}\beta}$ is expected since the strength of $\text{H}\beta$ decreases as R_{FeII} (or R_{CaT}) increases. The linear correlation between R_{FeII} and EW_{FeII} is due to the fact that both parameters reflect the strength of the Fe II emission; the first one is weighted by the $\text{H}\beta$ flux, and the second is weighted by the luminosity. It is the same case for the correlation $\text{EW}_{\text{CaT}}-R_{\text{CaT}}$. Since R_{CaT} is correlated with R_{FeII} , we expect a positive linear relation between $\text{EW}_{\text{CaT}}-R_{\text{FeII}}$ and $\text{EW}_{\text{FeII}}-R_{\text{CaT}}$. On the other hand, R_{FeII} and R_{CaT} show nonlinear trends with the FWHM of the emission lines, which are further discussed in Section 4.2. The correlations between the EW and the continuum luminosity are extensively described in Sections 4.3 and 6.2.

The correlations between the FWHM of $\text{H}\beta$, O I $\lambda 8446$, and CaT are strongest according to their Spearman coefficients and their associated p -values (Figure 2). In these panels, the 1:1 line is shown for reference. $\text{H}\beta$ shows broader profiles than O I $\lambda 8446$, particularly for the sources with $\text{FWHM} > 4000 \text{ km s}^{-1}$. We obtained the trend lines by ordinary least-squares (OLS) fitting implemented in Python packages `sklearn` and `statsmodels`. The relation has a slope of 0.894 ± 0.05 and a scatter of $\sigma_{\text{rms}} \sim 0.115$ dex. The deviation at 4000 km s^{-1} could be associated with the presence of a redward asymmetry in the broadest $\text{H}\beta$ profiles, i.e., with an emitting region closer to the continuum source (Marziani et al. 2013; Punsly et al. 2020). The presence of this feature is hard to observe in the O I $\lambda 8446$ profile, since it is blended with the CaT and the NIR Fe II. On the other hand, CaT is also narrower than $\text{H}\beta$, although the scatter is larger ($\sigma_{\text{rms}} \sim 0.152$ dex) and the relation is slightly shallower than the one given by O I $\lambda 8446$ with a slope of 0.827 ± 0.08 . O I $\lambda 8446$ and Ca II show similar widths, the predicted relation gives a slope of 0.944 ± 0.05 , and the scatter is smaller ($\sigma_{\text{rms}} \sim 0.103$ dex) than in the previous cases. This general behavior corroborates that $\text{H}\beta$ is emitted closer to the continuum source than O I $\lambda 8446$ and CaT (Persson 1988; Martínez-Aldama et al. 2015a; Marinello et al. 2016).

4.2. Eigenvector 1 Sequence

The correlation between R_{FeII} and the FWHM of $\text{H}\beta$ is known as the Eigenvector 1 (EV1) sequence (Boroson & Green 1992), which is also known as the quasar main sequence (Sulentic et al. 2000). According to the EV1 scheme, the observational and physical properties of type 1 AGNs change along the sequence (Marziani et al. 2018; Panda et al. 2018, 2019). Based on the R_{FeII} strength, the accretion rate can be inferred, where the sources with $R_{\text{FeII}} > 1$ are typically associated with the highest Eddington ratios ($L_{\text{bol}}/L_{\text{Edd}} > 0.2$; Marziani et al. 2003; Panda et al. 2019, 2020b). The relation between these parameters is not linear (Wildy et al. 2019), where orientation and luminosity are also involved (Shen & Ho 2014; Negrete et al. 2018).

The EV1 sequence ($\text{FWHM}_{\text{H}\beta}-R_{\text{FeII}}$ relation) of our sample is shown in Figure 2. A displacement between the low- and high-luminosity objects (HE sample) can be appreciated; however, both kinds of sources follow the same trend. This displacement is only a luminosity effect, where the HE sample

is shifted to the larger FWHM values of the panel. An EV1-like sequence is also appreciated in the relations $R_{\text{FeII}}-\text{FWHM}_{\text{CaT}}$ and $R_{\text{FeII}}-\text{FWHM}_{\text{OI}}$, which is expected owing to the linear relation between the widths of the emission lines (Section 4.1).

The relation $\text{FWHM}_{\text{H}\beta}-R_{\text{CaT}}$ shows a kind of EV1 sequence for the low-luminosity sample, i.e. a EV1-like sequence, but it is not appreciated in the high-luminosity objects, the HE sample. It seems that in some objects the CaT increases with increasing $\text{FWHM}_{\text{H}\beta}$. The same effect is observed for the relations $\text{FWHM}_{\text{OI}}-R_{\text{CaT}}$ and $\text{FWHM}_{\text{CaT}}-R_{\text{CaT}}$. Surprisingly, the break occurs at $R_{\text{CaT}} \sim 0.2$, which corresponds to $R_{\text{FeII}} = 1$ (following Equation (1)), the limit for the highly accreting sources according to Marziani et al. (2003). A similar decoupling is also observed in the relation between the EW_{CaT} and the FWHM of the emission lines, but the scatter is quite large. Martínez-Aldama et al. (2015a) found a rough enhancement of EW_{CaT} for the HE sample at intermediate z with respect to the other objects at low z , attributing this behavior to a burst of star formation and an enrichment at intermediate-redshift sources. The new HE objects (Martínez-Aldama et al. 2015b) added to the presented analysis seem to corroborate these results; however, some selection effects could also be involved. We discuss this result in Section 6.3.

4.3. Correlations with the Equivalent Width: The Baldwin Effect

The anticorrelation between the EW and the luminosity is known as the *global* Baldwin effect (BEff; Osmer & Shields 1999; Green et al. 2001; Baskin & Laor 2004; Bachev et al. 2004; Dong et al. 2009; Zamfir et al. 2010), which was first observed between the EW of C IV $\lambda 1549$ and the continuum luminosity at 1450 \AA (Baldwin 1977). The BEff is clearly appreciated in the HILs, except NV $\lambda 1240$ owing to a second enrichment (Osmer et al. 1994). However, as the IP decreases, the slope of this anticorrelation gets shallower, and it is hard to distinguish a strong correlation for LILs (Sulentic et al. 2000; Dietrich et al. 2002). An *intrinsic* Baldwin effect (Pogge & Peterson 1992) has also been detected in the multiepoch observations for single-variable AGNs. The BEff provides information about the ionizing continuum shape (Wandel 1999), structure, and metallicity (Korista et al. 1998) of the BLR. Also, it has been used for calibrating the luminosity in cosmological studies (Baldwin et al. 1978; Korista et al. 1998).

4.3.1. Luminosity

Figure 3 shows the EWs of $\text{H}\beta$, O I $\lambda 8446$, optical Fe II, and CaT as a function of the optical and NIR luminosities. Spearman rank correlation coefficients, p -values, and the scatter of the correlations are reported in Table A3. None of the trends between the EW and the optical and NIR luminosity satisfy the criteria for a significant relation, and all of them show shallow relations with a slope of $\alpha < 0.1$. The shallow slope confirms the weak relation between the luminosity and the EW for LILs. This result is in agreement with larger samples at high redshift (Dietrich et al. 2002). The negative correlation between $\text{EW}_{\text{H}\beta}$ and L_{opt} ($\alpha = -0.064 \pm 0.032$, $\rho = -0.412$, $p = 0.001$) is expected owing to the behavior of individual variable sources (e.g., Rakić et al. 2017). This behavior is different from the one of the relation $\text{EW}_{\text{CaT}}-L_{\text{opt}}$ ($\alpha = 0.068 \pm 0.053$, $\rho = 0.417$, $p = 0.001$), which suggests the

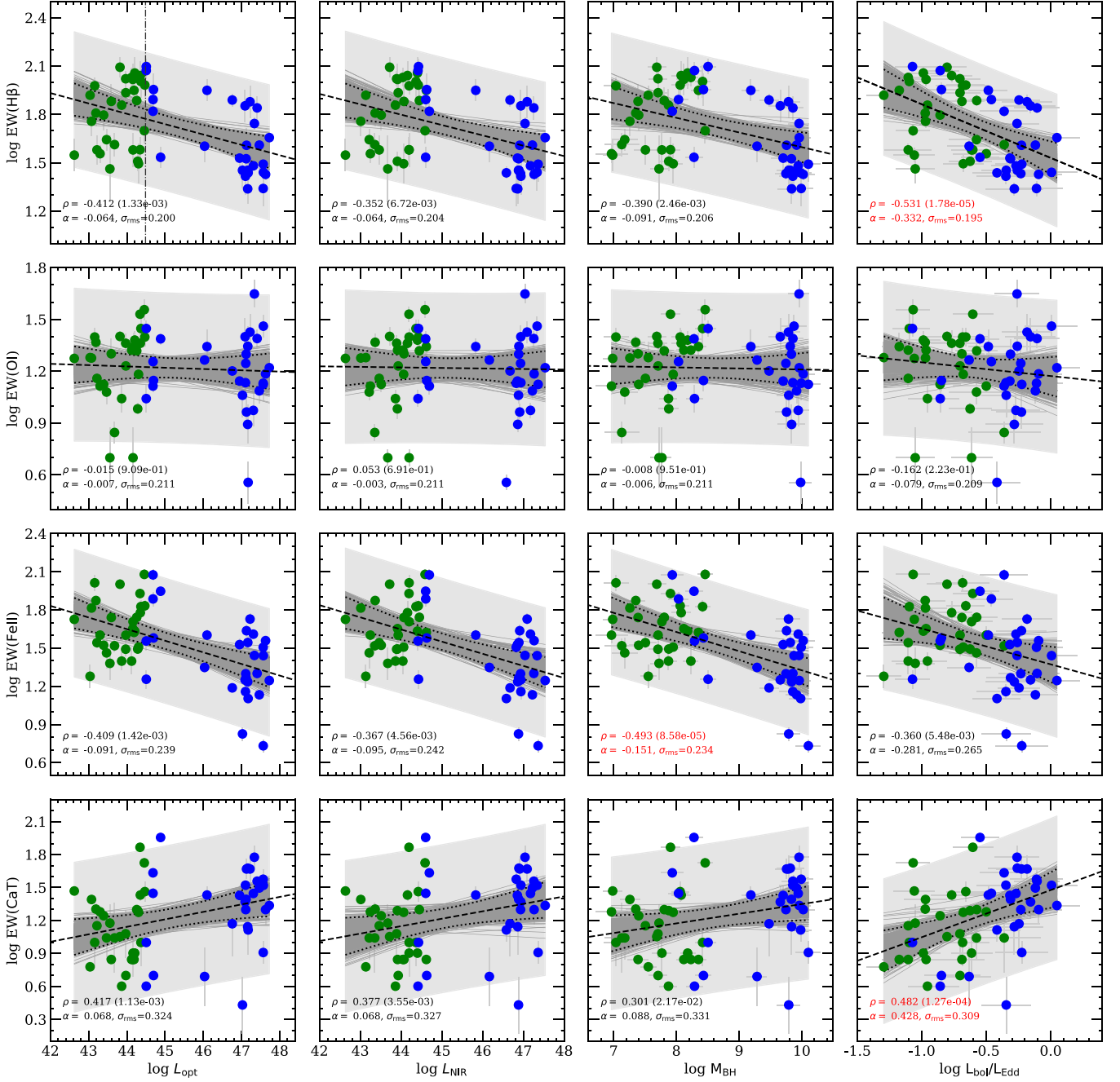


Figure 3. Correlation matrix for the optical and NIR luminosities, black hole, and the EW of H β , optical Fe II, O I λ 8446, and CaT. Green and blue symbols indicate the low- and high-luminosity subsamples, respectively (Section 4.3.3). The black dashed line represents the best OLS fitting for the full sample. The Spearman's rank correlation coefficient, p -values, slope (α), and scatter (σ_{rms}) are also shown, where the significant correlations are colored in red. Black dotted lines mark the confidence intervals at 95% for the 1000 realizations (dark-gray lines) of the bootstrap analysis. The light-gray patch marks the corresponding prediction intervals bands for the sample.

presence of an inverse Baldwin effect. A positive correlation has also been observed between the continuum at 5100 \AA and the optical Fe II in the monitoring of the variable NLS1 NGC 4051 (Wang et al. 2005). The strong correlation between the Fe II and CaT explains this behavior. However, in our sample the relation $\text{EW}_{\text{FeII}}-L_{\text{opt}}$ is negative ($\alpha = -0.091 \pm 0.039$, $\rho = -0.409$, $p = 0.001$) and is just below the criteria assumed to consider a significant correlation. Other studies reported neither a BEff for optical nor one for UV Fe II (Dong et al. 2011). Finally, the trend observed for $\text{EW}_{\text{OI}}-L_{\text{opt}}$ is

not significant and shows a slope consistent with zero ($\alpha = -0.007 \pm 0.034$), also confirmed by previous studies (Dietrich et al. 2002).

4.3.2. Black Hole Mass and Eddington Ratio

Since the black hole mass and the Eddington ratio have been considered as the main drivers of the BEff (Wandel 1999; Dong et al. 2011), we also present the correlations $\text{EW}-M_{\text{BH}}$ and $\text{EW}-L_{\text{bol}}/L_{\text{Edd}}$ in Figure 3. The parameters of the

correlations are reported in Table A3. The only significant relation involving the black hole mass is $EW_{\text{FeII}}-M_{\text{BH}}$ ($\rho = -0.493$, $\alpha = -0.151 \pm 0.060$, $\sigma_{\text{rms}} \sim 0.234$ dex).

In the correlations between the EW and the Eddington ratio, the significant relations are the ones involving $EW_{\text{H}\beta}$ and EW_{CaT} . In both cases the correlations are sharper ($\alpha_{\text{H}\beta} = -0.332 \pm 0.149$, $\alpha_{\text{CaT}} = 0.428 \pm 0.237$) and stronger ($\rho_{\text{H}\beta} = 0.531$, $\rho_{\text{CaT}} = 0.482$) than the luminosity case. Although the correlations for Fe II and O I $\lambda 8446$ are below the significance level, their slopes are steeper than the correlations with respect to the luminosities and the black hole mass. Hence, the Eddington ratio highlights the correlations with the EW, as Baskin & Laor (2004) and Dong et al. (2011) previously reported.

4.3.3. Division of the Sample

According to Dietrich et al. (2002), to avoid selection effects in the global BEff, a sample with a wide luminosity range is needed, $42 < \log L < 48$. Our sample covers this range; however, at high redshift only high-luminosity sources are available ($\log L_{\text{opt}} > 47.5$ erg s⁻¹). In order to clarify the results of Section 4.3 and the presence of possible bias, we divided the sample into two subsamples considering the median luminosity, $\log L_{\text{opt}} = 44.49$ erg s⁻¹. In Figure 3 the low- and high-luminosity subsamples are represented by green and blue symbols, respectively. The division of the sample directly affects the relations $EW-L_{\text{opt}}$, $EW-L_{\text{NIR}}$, and $EW-M_{\text{BH}}$ where no significant correlations are observed, which also reflects the bias involved in this consideration. For example, the relation $EW_{\text{H}\beta}-L_{\text{opt}}$ is positive for the low-luminosity subsample ($\alpha = 0.1 \pm 0.073$, $\rho = 0.235$, $p = 0.281$), while for the high-luminosity case the relation has a different direction ($\alpha = -0.13 \pm 0.031$, $\rho = -0.53$, $p = 0.003$), similar to the behavior of the full sample. The difference in the subsamples is also pointed out by the PCA (Appendix D). Therefore, the correlations have some relevance only when the full sample is considered.

However, the correlations $EW-L_{\text{bol}}/L_{\text{Edd}}$ are less affected by the division of the sample, at least in the significant correlations provided by the full sample analysis. In the relation $EW_{\text{H}\beta}-L_{\text{bol}}/L_{\text{Edd}}$, the direction of the best fits in the subsamples is still negative ($\alpha_{\text{low}} = -0.215 \pm 0.162$, $\alpha_{\text{high}} = -0.472 \pm 0.144$), although none of the relations are significant ($\rho_{\text{low}} = -0.26$, $p_{\text{low}} = 0.181$, $\rho_{\text{high}} = -0.35$, $p_{\text{high}} = 0.07$), whereas in the $EW_{\text{CaT}}-L_{\text{bol}}/L_{\text{Edd}}$ relation, the slope for the subsamples is positive ($\alpha_{\text{low}} = 0.240 \pm 0.158$, $\alpha_{\text{high}} = 0.694 \pm 0.242$), such as in the full sample, but without any significance ($\rho_{\text{low}} = 0.222$, $p_{\text{low}} = 0.245$, $\rho_{\text{high}} = 0.313$, $p_{\text{high}} = 0.098$). This result suggests that $L_{\text{bol}}/L_{\text{Edd}}$ is less influenced by a bias and then regulates the correlation between the EW and the luminosity, as originally suggested by Baskin & Laor (2004) and Bachev et al. (2004).

4.4. The Behavior of R_{FeII} , R_{CaT} , and the Ratio FeII/CaT

Figure 4 shows the behavior of R_{FeII} , R_{CaT} , and the ratio Fe II/CaT as a function of optical and NIR luminosity, black hole mass, and Eddington ratio. R_{FeII} and R_{CaT} do not show any significant correlation with the luminosity and black hole mass for the full sample. Only the Fe II/CaT shows a significant anticorrelation with the optical ($\rho = -0.441$, $p = 5.3 \times 10^{-4}$) and NIR luminosity ($\rho = -0.456$, $p = 3.2 \times 10^{-4}$). If the

subsamples are considered, all the best fits are below the statistical significance limit.

In the panels where $L_{\text{bol}}/L_{\text{Edd}}$ is involved, the strongest correlation is the one corresponding to the ratio Fe II/CaT ($\rho = 0.554$, $p = 6.4 \times 10^{-6}$), followed by the one with R_{CaT} ($\rho = 0.425$, $p = 8.9 \times 10^{-4}$). In both cases the trend lines for the full samples and subsamples have the same direction as in Figure 3, although the significant correlation arises only for the full sample.

The positive correlation between R_{CaT} and $L_{\text{bol}}/L_{\text{Edd}}$ confirms that the strength of CaT is driven by the accretion rate, and it remains even after the division of the sample. Although the same behavior is expected for R_{FeII} , we cannot confirm this result in our sample. The positive correlation between the optical and UV R_{FeII} and $L_{\text{bol}}/L_{\text{Edd}}$ is robust (e.g., Zamfir et al. 2010; Dong et al. 2011; Martínez-Aldama et al. 2020). Besides, the R_{FeII} has been used as a proxy for the $L_{\text{bol}}/L_{\text{Edd}}$ to correct the time delay by the accretion effect and decrease the scatter in the optical and UV RL relation (Du & Wang 2019; Martínez-Aldama et al. 2020). This indicates that the Fe II (and R_{FeII}) in our sample is affected by several factors: the sample size, the quality of the observations, and the Fe II templates employed, which could decrease the accuracy of its estimation. For instance, 10 objects from the Persson (1988) sample have only upper limits, such as Mrk 335. It is one of the five common sources observed by Marinello et al. (2016), where the R_{FeII} value is $\sim 50\%$ higher than the value estimated by Persson (1988). It confirms that the objects with upper values are highly inaccurate, and thus it could be reflected in the loss of the correlation with other parameters. A homogeneous fitting process considering the same analysis spectral procedure could help to decrease the scatter and clarify the trends that we aim to pursue in a future work.

4.5. Bootstrap Analysis

4.5.1. Random Distributions

What is the probability that an uncorrelated data set gives a correlation with Spearman rank coefficient as high as the one we observe? In order to answer this question, we modeled the distributions of each of the parameters in Figure 3 and 4 considering a random sample of 1000 elements and the probability distributions implemented in the module `stats` in Python. To determine how good of a fit this distribution is, we used the Kolmogorov–Smirnov test, which compares a sample with a reference probability distribution, and we chose the distribution with the highest $p\text{-value}_{\text{ran}}$. Since the luminosity and the black hole mass show bimodal distributions, we used two distributions to reproduce the observational one. In the rest of the cases a good fit was obtained with only one distribution. The probability distributions considered and the $p\text{-value}$ are reported in Columns (11) and (12) of Table A3. The distribution fitting of the correlation $L_{\text{bol}}/L_{\text{Edd}}-\text{Fe II}/\text{CaT}$ is shown in Figure B1 as an example of this analysis.

Later, we randomly selected 58 realizations from the original 1000, which later we correlated following the correlations in Figures 3 and 4, and in the correlation $R_{\text{FeII}}-R_{\text{CaT}}$. Finally, we repeated the process 2000 times and estimated the Spearman rank correlation coefficient (ρ_{ran}) and the corresponding $p\text{-value}$ and estimated the fractions of significant correlations (f_{ran}). Results are reported in Column (13) of Table A3. In all the cases $f_{\text{ran}} < 1 \times 10^{-3}$, which means that is very unlikely at

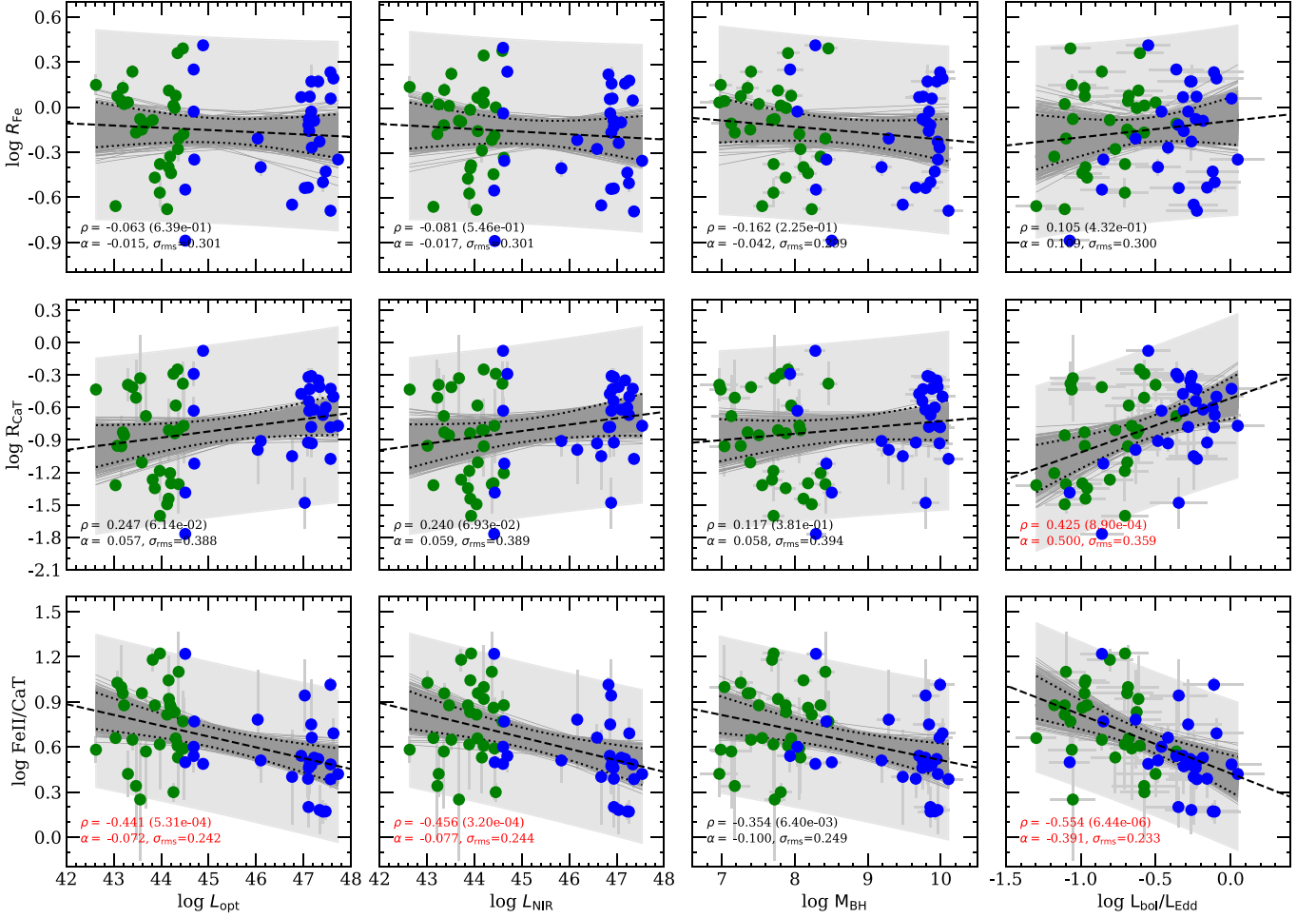


Figure 4. Correlation matrix for optical and NIR luminosity, black hole mass, Eddington ratio (L_{bol}/L_{Edd}), R_{FeII} , R_{CaT} , and Fe II/CaT ratio. Colors and symbols are the same as in Figure 3.

3σ confidence level that two independent correlations provide high correlation coefficients such as the observational sample does. Therefore, our analysis supports the reliability of the observed correlations.

4.5.2. Linear Regression Fitting

Due to the small size of our sample and the gaps in luminosity and redshift, we proved the statistical reliability of the correlations in Figures 3 and 4 via a bootstrap analysis (Efron & Tibshirani 1993). The bootstrap sample is formed by the selection of a subset of pairs from each one of the correlations by random resampling with replacement. We created 1000 realizations and then performed a linear regression fitting. The gray lines in Figures 2 and 3 correspond to the 1000 realizations, which are in agreement with the best fit of each correlation at 2σ confidence level (dotted black lines). As a reference, the figures also show the prediction interval bands (light-gray patch), which indicate the variation of the individual measurements and predict that 95% of the individual point lies within the patch. As a reference, we also analyzed the relation $R_{FeII} - R_{FeII}$ (inset panel, Figure 2) to compare the behavior of the bootstrap analysis in a very well-known correlation, obtaining an agreement within the 2σ confidence level.

In order to quantify the bootstrap results, we considered the percentiles at the 2σ confidence level and estimated the errors

of the slope (α_{BS}) and ordinate (β_{BS}) of the normal distribution drawn from the 1000 realizations for each correlation. Results are reported in Table A3. As is expected, the distributions are centered in the slope and ordinate values of each correlations, which are completely equivalent to the ones from the observational correlations. The magnitude of the errors indicates the reliability of the correlation. The larger errors are associated with the correlations below the significance criteria ($-0.4 < \rho_{BS} < 0.4$, $p > 0.001$). A clear example are the errors in the slope of the relations involving OI $\lambda 8446$ (Figure 3) or R_{FeII} (Figure 4), which indicates the inaccuracy of the results, such as the Spearman correlation coefficient shows. Meanwhile, good correlations, such as $R_{FeII} - R_{CaT}$, will show errors $< 20\%$. As the correlation coefficients indicate, the errors decrease considerably in the correlations where L_{bol}/L_{Edd} is involved. This result points out the relevance of L_{bol}/L_{Edd} in the behavior of our sample and its role in the Baldwin effect.

On the other hand, we also estimated the Spearman correlation coefficient (ρ_{BS}) for the 1000 realizations and estimated the fraction of significant realizations with respect to the total number (f_{sig}), which satisfy the significance criteria ($|\rho| > 0.4$ and $p < 0.001$). We also modeled the distribution of ρ_{BS} using a skewnorm distribution and estimated the error at the 2σ confidence level. The maximum of the ρ_{BS} distribution and f_{sig} are reported in Columns (9) and (10) of Table A3. In

the strongest correlation of the sample, $R_{\text{FeII}}-R_{\text{CaT}}$, we obtained $f_{\text{sig}} = 1$. This means that the 1000 bootstrap realizations satisfy the significant criteria and confirm the reliability of the correlation. This is also highlighted by the errors of ρ_{BS} , where the correlation remains significant within the uncertainty range. In the correlations with $|\rho_{\text{BS}}| > 0.5$, $f_{\text{sig}} > 0.75$, indicating reliable correlations. However, if the errors of ρ_{BS} are considered, there is a small possibility of dismissing the significance of the correlation. It can be expressed by the parameter $1 - f_{\text{sig}}$, which expresses the probability of failing to detect a correlation. Thus, there is probability of $< 25\%$ of detecting a false-positive correlation in $\text{EW}_{\text{H}\beta}-L_{\text{bol}}/L_{\text{Edd}}$ and $\text{Fe II/CaT}-L_{\text{bol}}/L_{\text{Edd}}$. If $|\rho_{\text{BS}}| = 0.4-0.5$, great care should be considered because the probability to detect a false-positive correlation increases to $(1 - f_{\text{sig}}) \sim 50\%$. This is the case for correlations such as $\text{EW}_{\text{CaT}}-L_{\text{bol}}/L_{\text{Edd}}$ and $\text{Fe II/CaT}-L_{\text{opt}}$. The same interpretation of false-positive probability applies in the case of no detected correlation in the observed or bootstrap samples ($\rho < 0.4$), when the probabilities are always low, particularly for the weakest correlations ($1 - f_{\text{sig}} > 80\%$).

4.6. Residual Behavior

In order to assess a possible redshift effect in our results, we estimated the residuals with respect to the best fit for the correlations in Figures 3 and 4. We divided the sample into low- and high- L subsamples, which is equivalent to a division into low and high redshift. The behavior of the distributions is shown in Figures B2 and B3. If any significant difference of the median of the distribution with respect to the zero residual level is observed, it could indicate a redshift effect. In all the correlations of Figure B2, we observed a difference within the 2σ confidence level. On the other hand, the relations of Figure 4 show the same behavior; however, the width of the distribution increases significantly, as well as the median values, particularly for the correlations involving R_{FeII} . Since this behavior is only observed in these correlations and they still show a dependency within the 2σ level, we cannot claim a redshift effect. As we mentioned previously, the R_{FeII} is not well behaved in our sample compared to previous results. Therefore, any trend involving R_{FeII} must be taken with caution.

5. Principal Component Analysis

PCA allows us to get a *better view* of the data where they can be separated in a quantitative manner, such that the relevant properties explain the maximum amount of variability in the data set. PCA works by initially finding the principal axis along which the variance in the multidimensional space (corresponding to all recorded properties) is maximized. This axis is known as eigenvector 1. Subsequent orthogonal eigenvectors, in order of decreasing variance along their respective directions, are found, until the entire parameter space is spanned (see, e.g., Boroson & Green 1992; Kuraszkiwicz et al. 2009; Wildy et al. 2019; Tripathi et al. 2020). The PCA method is particularly useful when the variables within the data set are highly correlated. Correlation indicates that there is redundancy in the data. Due to this redundancy, PCA can be used to reduce the original variables into a smaller number of new variables (principal components; PCs) explaining most of the variance in the original variables. This allows us to determine correlated parameters, and in the context of our work, we utilize this

technique to determine the physical parameter(s) that lead to the correlations illustrated in Figures 3 and 4.

Eigenvalues (or loadings) can be used to determine the numbered PCs to retain after PCA (Kaiser 1961): (a) An eigenvalue > 1 indicates that PCs account for more variance than accounted for by one of the original variables in standardized data. This is commonly used as a cutoff point for which PCs are retained. This holds true only when the data are standardized, i.e., they are scaled to have standard deviation one and mean zero. (b) One can also limit the number of components that account for a certain fraction of the total variance. Since there is no well-accepted way to decide how many PCs are enough, in this work we evaluate this using the *scree plot* (see, e.g., Figure 5), which is the plot of eigenvalues ordered from largest to smallest. A scree plot shows the variances (in percentages) against the numbered PC and allows visualizing the number of significant PCs in the data. The number of components is determined at the point beyond which the remaining eigenvalues are all relatively small and of comparable size (Peres-Neto et al. 2005; Jolliffe 2011). This helps us to realize whether the given data set is governed by a single or more dimensions, where a dimension refers to a variable. Subsequently, these PCs are investigated against the original variables of the data set to extract information of the importance of these variables in each PC.

We consider the 58 sources in our sample and collect the properties that are common among them. Due to the diversity in the studied subsamples, we only have 12 parameters that are obtained/estimated directly from the observation: z ; optical and NIR luminosity; R_{FeII} ; R_{CaT} ; FWHM and EW of $\text{H}\beta$, $\text{O I } \lambda 8446$, and CaT ; and the EW of Fe II . Among these 12 parameters, the redshift and the optical luminosities are correlated by a bias effect as shown in Figure 1, and hence we drop the redshift and only choose the optical luminosity. Thus, we have 11 parameters that are considered in the initial PCA. Later, we only adopt the NIR luminosity, which is equivalent to the optical (Figure 2). We refer the readers to Appendix C for more details on the initial PCA tests, a note on the effect redundant parameters play in this analysis, and the final set of parameters used to infer the observed correlations.

Next, we have the derived parameters—bolometric luminosity (L_{bol}), black hole mass (M_{BH}), and Eddington ratio ($L_{\text{bol}}/L_{\text{Edd}}$), which are predicted using one or more of the observed parameters that are already taken into account for the PCA run. PCA is an orthogonal linear transformation technique applied to the data and assumes that the input data contain linearly independent variables such that the eigenvectors can be represented as a sum of a linear combination of variables with associated weights (eigenvalues or loadings) corresponding to each variable. Thus, in order to remove any redundancy in the result obtained via the PCA, one needs to make sure that the parameters that go in as input are not scaled-up versions of another parameter, thereby saving us the trouble of unwanted bias that comes out of it. The effect of the inclusion of derived variables in the PCA is illustrated in Appendix C.

Similar to Wildy et al. (2019), we use the `prcomp` module in the *R* statistical programming software. In addition to `prcomp`, we use the `factoextra`⁶ package for visualizing the multivariate data at hand, especially to extract and visualize the eigenvalues/variances of dimensions.

⁶ <https://cloud.r-project.org/web/packages/factoextra/index.html>

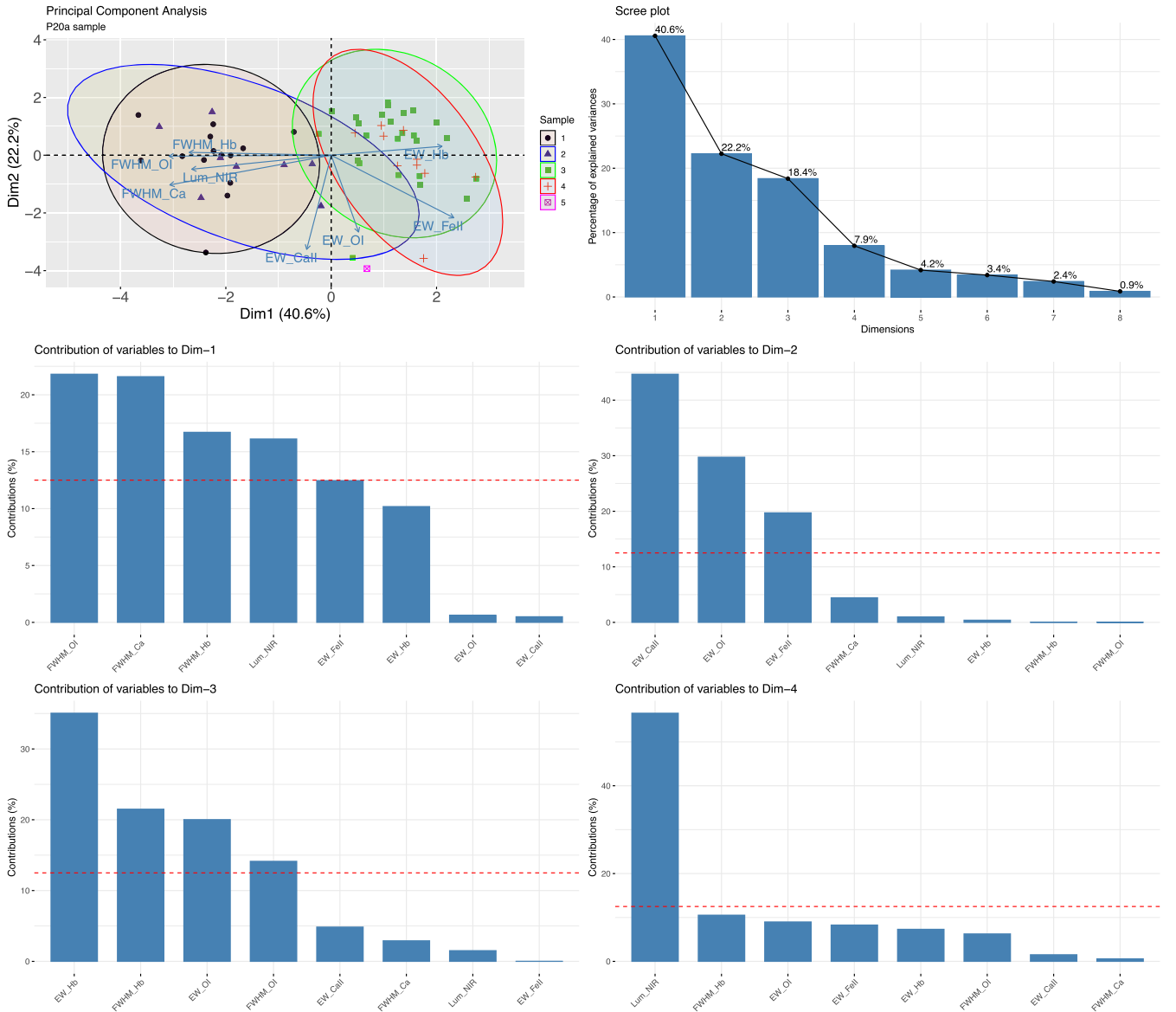


Figure 5. Graphical representation (factor map) of the PCA decomposition of our sample (58 sources) is shown in the first panel. The circles represent individual objects on the standardized PC1–PC2 plane that have variables indicated by the axis labels. The dashed lines mark the coordinate axes in the PC1–PC2 plane, and the ellipses depict the 95% occupancy of the sources in their respective subsamples. The sample is categorized based on their original source catalogs (see Panda et al. 2020a, for details on the observational sample)—(1) Martínez-Aldama et al. (2015a); (2) Martínez-Aldama et al. (2015b); (3) Persson (1988); (4) Marinello et al. (2016); and (5) PHL1092 (Marinello et al. 2020). The other panels illustrate the precedence of the first 10 PCs in the form of scree plots, followed by the contributions of the original variables to the first four PCs.

5.1. PCA on the Full Sample

The tests aimed at reducing the redundancy of the variables were applied as described in Appendix C and now allow us to perform a final PCA run with the data set that contains variables that are obtained directly from the observations and have as little redundancy as possible. The final input contains eight variables, namely, the NIR luminosity (at 8542 Å); the EWs of Fe II, H β , OI λ 8446, and CaT; and the FWHMs of H β , OI λ 8446, and CaT. The result of the PCA is illustrated in Figure 5.

In this section, we present the results of the PCA on the full sample and infer the relative importance of the *eigenvectors* as a function of the input variables. In Appendix D we describe the PCA analysis for the low- and high-luminosity samples

described in Section 4.3.3. Figure 5 shows the two-dimensional factor map between the first two PCs, the scree plot, and the contributions of the input variables to the first four PCs for the full sample. The factor map shows the distribution of the full sample categorically colored to highlight the different sources (see Section 2) in the eigenspace represented by the two PCs, Dim-1 and Dim-2 (i.e., the PC1 and PC2). The first and the second PC contribute 40.6% and 22.2%, respectively, to the overall dispersion in the data set. Combining the contributions from the two subsequent PCs (PC3 and PC4), one can explain 89.1% of the variation present in the data. We demonstrate the contributions of the original variables on these four PCs to draw conclusions on the primary driver(s) of the data set.

First PC: From the factor map we find that the vectors corresponding to the FWHMs of H β , OI λ 8446, and CaT and

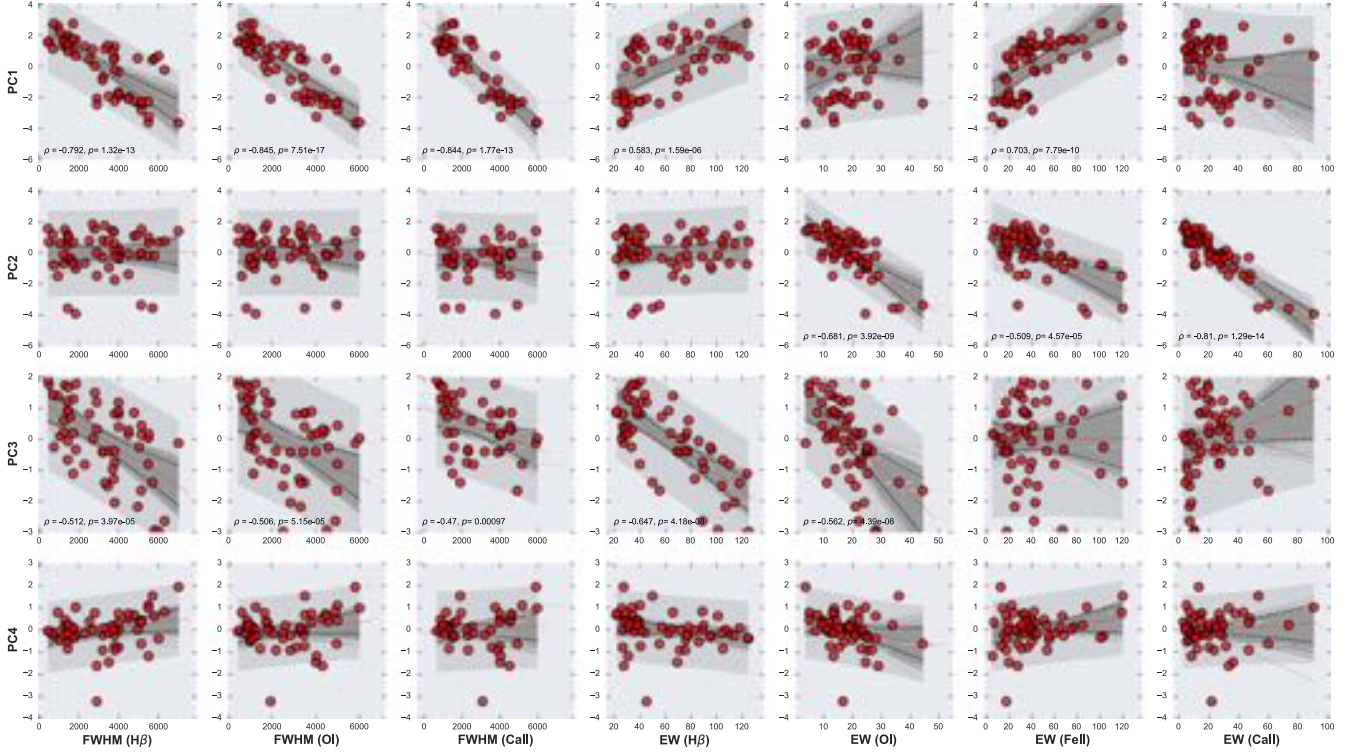


Figure 6. Correlation matrix showing dependence of the first four PCA vectors' loadings vs. the physical parameters (*observed*) for our full sample. The Spearman's rank correlation coefficient (ρ) and the p -value are reported for the correlations whenever p -value < 0.001 . The black dotted lines mark the confidence intervals at 95% for the 1000 realizations (dark-gray lines) of the bootstrap analysis. The light-gray patch marks the corresponding prediction interval bands for the sample.

the NIR luminosity are nearly co-aligned, with the FWHM vectors of H β and O I $\lambda 8446$ having almost similar orientation. These four vectors are also the major contributors to the variance along the primary PC (see third panel on the left of Figure 5). The red dashed line on the graph above indicates the expected average contribution. If the contribution of the variables were uniform, the expected value would be $1/\text{length}(\text{variables}) = 1/8 \approx 12.5\%$. For a given component, a variable with a contribution larger than this cutoff could be considered as important in contributing to the component. The EW_{FeII} is barely over this cutoff limit.

Second PC: The factor map highlights the prevalence of the EW_{CaT} , which contributes $\sim 45\%$ to this component, followed by the EW_{OI} and EW_{FeII} . The overall contribution accounts for $\sim 95\%$ from these three variables.

Third and fourth PCs: The third PC is dominated by the $\text{EW}_{\text{H}\beta}$ with a minor contribution from $\text{FWHM}_{\text{H}\beta}$, EW_{OI} , and FWHM_{OI} , whereas, the fourth PC is singularly governed by NIR luminosity. The other variables are below the expected average contribution limit.

5.2. Correlations between the Principal Eigenvectors and Observed/Derived Parameters

To quantitatively assess the relevance of the observational variables and the derived physical parameters, we show their correlation against the contributions (loadings) of the first four PCs (PC1, PC2, PC3, and PC4) for the full sample in Figures 6 and 7. The full sample is then separated into two subsamples based on the median optical luminosity of the full sample (i.e., $\log L_{\text{opt}} = 44.49 \text{ erg s}^{-1}$). A comparative analysis of the full

sample against the two subsamples is carried out in Appendix D and in Figures D2 and D3.

Figure 6 is basically the correlation matrix representation for Figure 5 that includes all the intrinsic variables (except for the NIR luminosity). We only state the values of the Spearman's correlation coefficient and the corresponding p -value when the correlation is significant ($p < 0.001$). The full information about the correlation coefficients between the four eigenvectors and the physical parameters is reported in Table A4. In the full sample, the strongest correlations with respect to PC1 (in decreasing order) are exhibited by FWHM_{OI} ($\rho = -0.845$, $p = 7.51 \times 10^{-17}$), FWHM_{CaT} ($\rho = -0.844$, $p = 1.77 \times 10^{-13}$), $\text{FWHM}_{\text{H}\beta}$ ($\rho = -0.792$, $p = 1.32 \times 10^{-13}$), EW_{FeII} ($\rho = 0.703$, $p = 7.79 \times 10^{-10}$), and $\text{EW}_{\text{H}\beta}$ ($\rho = 0.583$, $p = 1.59 \times 10^{-6}$). In the case of PC2, significant correlations are obtained only for the EWs of CaT ($\rho = -0.81$, $p = 1.29 \times 10^{-14}$), O I $\lambda 8446$ ($\rho = -0.681$, $p = 3.92 \times 10^{-9}$), and Fe II ($\rho = -0.509$, $p = 4.57 \times 10^{-5}$). For the PC3, there are negative correlations right above the significance limit for the three FWHMs and the EWs for H β and O I $\lambda 8446$. The correlations for the subsamples are described in Appendix D.

Figure D3 corresponds to the derived parameters—bolometric luminosity, black hole mass, Eddington ratio, R_{FeII} , R_{CaT} , and the ratio of the two species, Fe II/CaT. For the full sample, we see clear, strong anticorrelations with PC1 for the black hole mass ($\rho = -0.845$, $p = 6.99 \times 10^{-17}$), the bolometric luminosity ($\rho = -0.748$, $p = 1.49 \times 10^{-11}$), followed by the redshift ($\rho = -0.7$, $p = 1 \times 10^{-9}$) and Eddington ratio ($\rho = -0.519$, $p = 2.94 \times 10^{-5}$). However, in the case of the correlation with respect to the redshift, this is biased owing to the gaps in the sample distribution that is highlighted in the

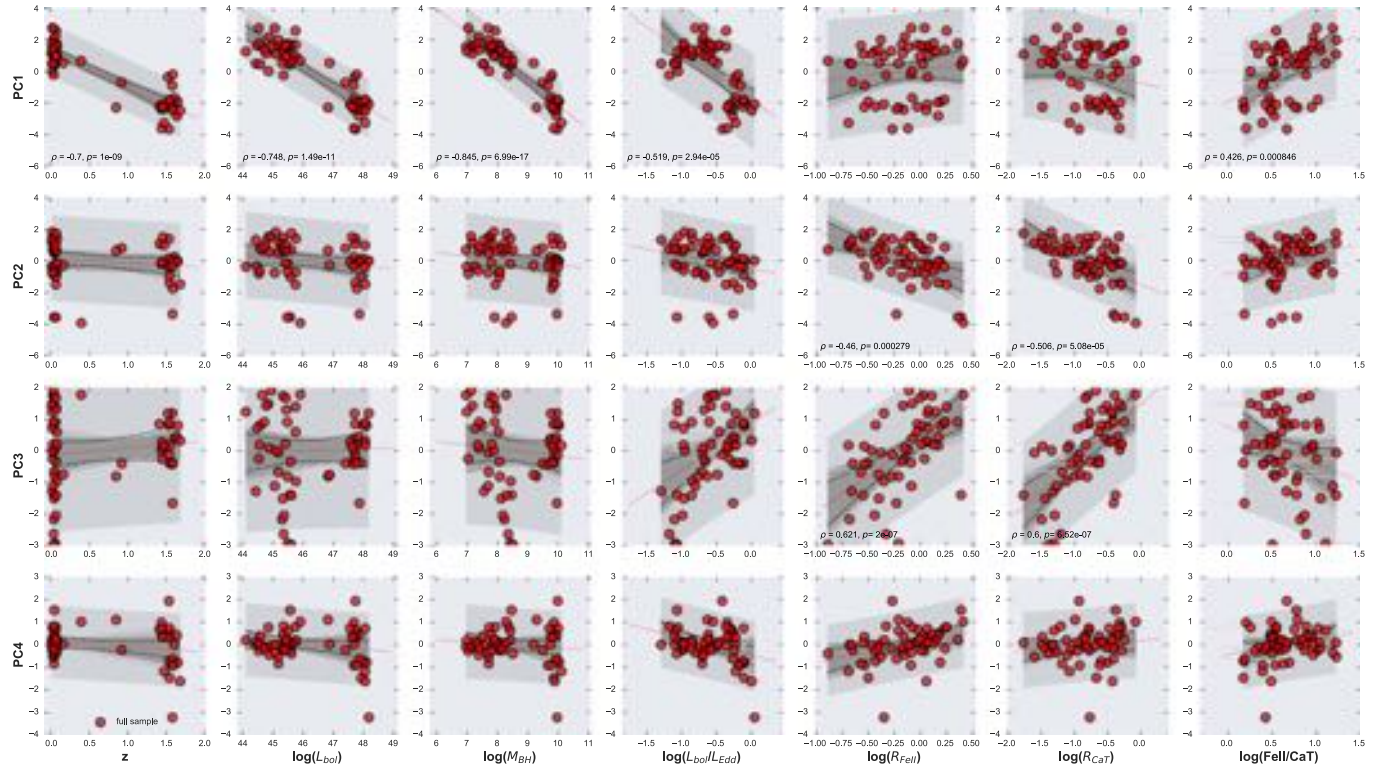


Figure 7. Correlation matrix showing dependence of the first four PCA vectors' loadings vs. the physical parameters (*derived*) for our full sample. The Spearman's rank correlation coefficients (ρ) and the p -values are reported for the correlations whenever p -value < 0.001 . The OLS fits in each panel are shown with red dashed lines. Black dotted lines mark the confidence intervals at 95% for the 1000 realizations (dark-gray lines) of the bootstrap analysis. The light-gray patch marks the corresponding prediction interval bands for the sample.

panels in the left column (this is also illustrated in Figure 1). For the remaining trends, they corroborate with the correlations that were obtained with the FWHMs in the previous figure. We also see a mild positive correlation of the PC1 with the ratio, Fe II/CaT ($\rho = 0.426$, $p = 8.46 \times 10^{-4}$). There are only two significant correlations obtained with respect to PC2, i.e., for the two line ratios— R_{CaT} ($\rho = -0.506$, $p = 5.08 \times 10^{-5}$) and R_{FeII} ($\rho = -0.46$, $p = 2.79 \times 10^{-4}$). This checks the observed correlation that is obtained and studied in this work and in our previous studies. The correlations with PC3 and the two ratios, R_{FeII} ($\rho = 0.621$, $p = 2 \times 10^{-7}$) and R_{CaT} ($\rho = 0.6$, $p = 1.52 \times 10^{-7}$), indicate that the $R_{\text{FeII}}-R_{\text{CaT}}$ correlation in the full data set is at least two-dimensional and may have multiple drivers for this observed correlation. Following the same analysis carried out in Section 4.5, we also performed a bootstrap analysis for the correlations in Figures 6 and 7 that reflects the behavior observed in Section 4.5.

6. Discussion

In this paper, we carefully look at the observational correlations from the up-to-date optical and NIR measurements centered around Fe II and CaT emission, respectively. Since our sample is affected by bias, we explored its influence in our results, showing that the correlations with the Eddington ratio and the independent observational parameters are less biased than the ones involving luminosity or black hole mass. These results are supported by a bootstrap analysis, which corroborated the statistical meaning of the correlations. We did not detect any redshift dependence in the correlations with luminosity and Eddington above the 2σ confidence level. We

also performed a PCA in order to define the main driver(s) of our sample where the black hole mass, luminosity, Eddington ratio, and Fe II/CaT show the main correlation with the PC1.

6.1. The Primary Driver(s) of Our Sample

The PCA is a powerful tool; however, the principal eigenvectors are just mathematical entities, and it not easy to connect them with a direct physical meaning. As is shown in Section 4.2 and Appendix D, the PCA gives different results if the full, low- L , or high- L samples are considered. The correlations between the PC $_i$ values and the observational parameters such as $\text{FWHM}_{\text{H}\beta}$, R_{FeII} , or $\text{EW}_{\text{H}\beta}$ for the low-luminosity subsample resemble the Boroson & Green (1992) PCA results. However, the trends are different for the high-luminosity subsample. This difference cannot be associated with luminosity or redshift effects, since PCA results are based on the space parameter considered. Hence, since the objective is to understand the general drivers of the sample, we only discuss the PCA for the full sample.

Figures 6 and 7 describe the relation between the observational and independent parameters with the principal eigenvectors, where the FWHM shows the primary correlations with a significance over 99.9%, followed by the relations with the EWs again, with a high significance (62.8% of the variation). Hence, due to the relevance of the FWHM, a high correlation between PC1 and the black hole mass is expected (Figure 7). The luminosity and black hole mass show the strongest correlations with the PC1, followed by the Eddington ratio and the ratio Fe II/CaT (see Section 6.3). The main correlations with the PC2 are with EW_{CaT} , EW_{OI} , EW_{FeII} , R_{CaT} , EW_{FeII} ,

and R_{FeII} , in decreasing order of significance. Similar to the observational trends (Figures 3 and 4), all the correlations are stronger for the CaT than for Fe II, indicating the relevance of the CaT in our sample.

On the other hand, in Figures 3 and 4 the main correlations are the ones involving the $L_{\text{bol}}/L_{\text{Edd}}$, which is also supported by the lowest errors provided by the bootstrap results (Figures B2 and B3). These facts point toward the Eddington ratio as the main driver. However, from the PCA, black hole mass and luminosity have similar relevance ($\rho_{M_{\text{BH}}} = 0.845$, $\rho_L = 0.748$), followed by the Eddington ratio ($\rho \sim -0.519$), all of them with a significance over 99.9%. Since the PCA reduces the dimensionality of the object-to-object variations, it is expected that the main correlations are associated with luminosity, black hole mass, and Eddington ratio, since the third one can be expressed by $L_{\text{bol}}/L_{\text{Edd}} \propto L_{\text{opt}}^{1/2} \text{FWHM}_{\text{H}\beta}^{-2}$. In order to test the self-dependence of the Eddington ratio and hence its role in our sample, we performed a multivariate linear regression fitting in the form $\log L_{\text{bol}}/L_{\text{Edd}} \propto \log L_{\text{opt}} + a \log \text{FWHM}_{\text{H}\beta}$. We got $a = 3.1$ ($\sigma_{\text{rms}} \sim 7.71 \times 10^{-5}$ dex), a variation of 25% with respect to the expected value ($a = -4$) from definition of $L_{\text{bol}}/L_{\text{Edd}}$. Therefore, this highlights the Eddington ratio as the driver of the correlations in the sample.

However, this result must be tested with the inclusion of more objects. In our sample, the highest $L_{\text{bol}}/L_{\text{Edd}}$ values are always associated with the highest luminosities, largest black hole masses, and highest redshifts, which is an artifact of the flux-limited sample. To verify the Eddington ratio as the main driver, one should consider samples reducing flux-limit and small-number biases, for example, including low accretors at high redshift, or enlarging the sample at low z . In addition, our sample does not include sources with $\text{FWHM}_{\text{H}\beta} > 7000 \text{ km s}^{-1}$, which usually show weak or negligible Fe II contribution. Hence, newer sources with CaT–Fe II estimates are required to confirm the current results and to certify the Eddington ratio as the driving mechanism.

6.2. Is the Eddington Ratio the Driver of the Baldwin Effect?

The driver of the Baldwin effect is still under discussion. The most accepted explanation for this effect is that high-luminosity AGNs have a soft ionizing continuum, so the number of ionizing photons available for the emission-line excitation decreases. It is supported by the fact that the spectral index between the UV (2500 Å) and X-ray continuum (2 keV) increases with luminosity (Baskin & Laor 2004; Shields 2007). Thus, the UV bump will be shifted to longer wavelengths, provoking a steeper EW– L relation as a function of the IP. Metallicity also has an important role (Korista et al. 1997), due to the correlation with the black hole mass and luminosity (Hamann & Ferland 1993, 1999). An increment in the metallicity reduces the EW of the emission lines.

In our analysis only LILs are considered; therefore, we expect a weak relation between the EW and the luminosity. The values of the slopes are around zero, $-0.1 < \alpha < 0.1$ in all the correlations, as predicted by Dietrich et al. (2002). And the correlation coefficients are below the significance level considered, except in the correlations Fe II/CaT–luminosity, although the bootstrap results predict a $\sim 50\%$ probability of detecting a false positive in this case. Therefore, the statistical significance of the EW– L relations is called into question.

In the correlations where the Eddington ratio is involved, the slope is stronger than in the luminosity case, $\alpha > 0.3$. The same

effect is found for CIV $\lambda 1549$, a high-ionization emission line. Considering the EW of CIV $\lambda 1549$, luminosities, and Eddington ratios reported by Baskin & Laor (2004), we found a stronger correlation and higher slope ($\rho = -0.5$, $\alpha = -0.3 \pm 0.06$) in the relation CIV– $L_{\text{bol}}/L_{\text{Edd}}$ than for the correlation with the luminosity ($\rho = -0.04 \pm 0.05$, $\alpha = -0.05 \pm 0.09$). Additionally, the bootstrap results predict the smallest errors and a low probability of detecting a false positive in the correlations EW– $L_{\text{bol}}/L_{\text{Edd}}$. These results suggest that the Eddington ratio has more relevance than the luminosity (Baskin & Laor 2004; Bachev et al. 2004; Dong et al. 2009), and thus the behavior of the EW of LILs and HILs is driven by the Eddington ratio.

We can probe the role of the Eddington ratio in an independent way throughout a multivariate linear regression fitting in the form $\text{EW} \propto L_{\text{opt}} + a \text{FWHM}$. For H β we obtain $a = -2.5 \pm 1.4$ ($\sigma_{\text{rms}} \sim 0.194$ dex), while for EW_{CaT} $a = -3.8 \pm 1.9$ ($\sigma_{\text{rms}} \sim 0.308$ dex). In the last case, the slope is almost similar to the expected value ($a = -4$), which again highlights the strong correlation between CaT and $L_{\text{bol}}/L_{\text{Edd}}$. At least in our sample, the CaT is a better proxy for the Eddington ratio than Fe II, although there is a 50% probability of detecting a false-positive correlation as the bootstrap results pointed out.

A novel result from the PCA is the relevance of the metallicity expressed as the ratio Fe II/CaT (Section 6.3). According to Korista et al. (1998), the metallicity has a secondary role in the Baldwin effect, so we included this parameter in the multivariate linear regression fitting: $\text{EW} \propto L_{\text{opt}} + a \text{FWHM}_{\text{H}\beta} + b \text{FeII/CaT}$. There is no improvement for the H β correlation ($a = -2.4 \pm 1.6$, $b = -0.786 \pm 1.4$, $\sigma_{\text{rms}} \sim 0.193$), while for CaT the uncertainties are high ($a = -4.8 \pm 4.1$, $b = -9.705 \pm 6.35$, $\sigma_{\text{rms}} \sim 0.279$). Therefore, we cannot confirm the secondary effect of the metallicity (expressed as the Fe II/CaT ratio) in the Baldwin effect.

6.3. Implication for the Chemical Evolution

The relative abundance of iron with respect to the α -elements has been used as a proxy for the chemical abundance in AGNs (see Hamann & Ferland 1992, for a review). The α -elements (O, Mg, Si, Ca, and Ti) are predominantly produced by Type II SNe after the explosion of massive stars ($7 M_{\odot} < M_{*} < 100 M_{\odot}$) on timescales of 10^7 yr, while Fe is mostly produced by Type Ia SNe from white dwarf progenitors on longer timescales ~ 1 Gyr (Hamann & Ferland 1993). Depending on the host galaxy type, the time delay between the manufacturing timescales varies from 0.3 to 1 Gyr for massive elliptical and Milky Way–like galaxies, respectively (Matteucci & Recchi 2001). Thus, the ratio Fe/ α can be used as a clock for constraining the star formation, the metal content, and the age of the AGN (Matteucci 1994; Hamann & Ferland 1992). The UV Fe II and Mg II $\lambda 2800$ have been widely used for this purpose since the UV spectrum is accessible in a wide redshift range, up to $z \sim 7.5$ (e.g., Dietrich et al. 2003; Verner et al. 2009; Dong et al. 2011; De Rosa et al. 2014; Sameshima et al. 2017; Shin et al. 2019; Onoue et al. 2020; Sarkar et al. 2021, and references therein). However, the Fe II/Mg II flux ratio does not show a significant redshift evolution, suggesting a large number of Type Ia SNe (Onoue et al. 2020) or AGNs being chemically mature also at high z (Shin et al. 2019).

The optical Fe II and CaT have similar IPs, and both are emitted by the same region in the BLR, although the CaT

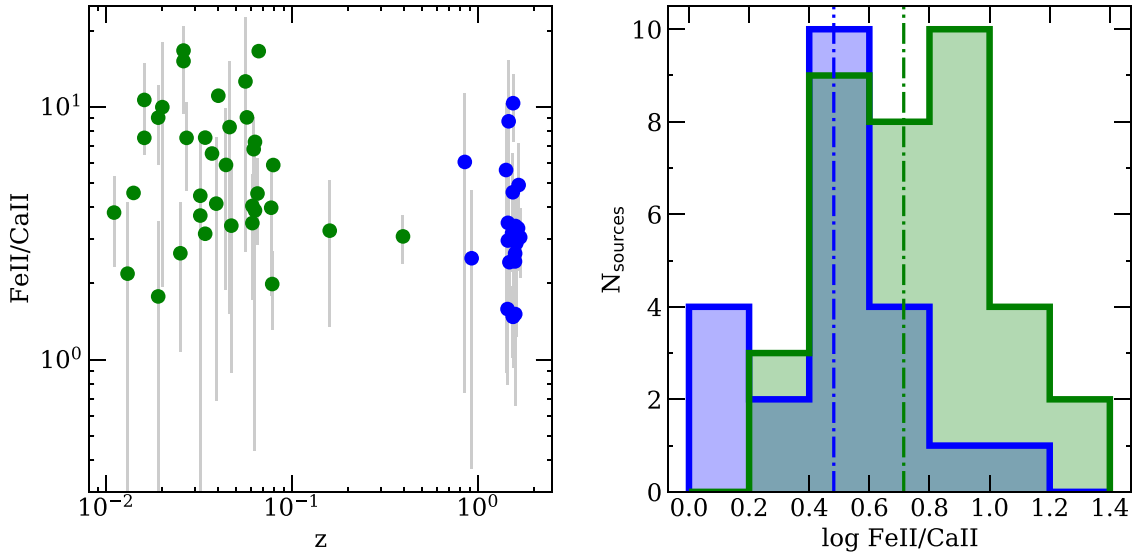


Figure 8. Left panel: Fe II/CaT distribution as a function of redshift in log scale. Green symbols correspond to sources with $z < 0.8$; the rest of the sources are marked with blue symbols. Right panel: Fe II/CaT distribution; colors are the same as in the left panel. Vertical lines mark the median redshift for the low- and high-redshift subsamples.

region is seemingly more extended (Panda 2021). Assuming that CaT scales with the rest of the α -elements and the ratio Fe II/CaT traces the abundance iron over calcium, we can use the ratio Fe II/CaT as a metal estimator. Figure 8 shows the distribution of the ratio Fe II/CaT as a function of the redshift. Dividing the sample at $z = 0.8$, which basically separates the Persson and Marinello et al. samples from the HE sample (Martínez-Aldama et al. sample), we get that the low-redshift sample has a median Fe II/CaT ratio of ~ 5.8 , while the intermediate-redshift sample shows an Fe II/CaT ratio of ~ 3.0 . A two-sample Kolmogorov–Smirnov test provides a value of 0.489 with a probability of $p_{KS} \sim 0.001$. This means that both samples are not drawn from the same distribution. The higher ratio Fe II/CaT at low redshift suggests some form of chemical evolution. Our sample reaches a maximum redshift at $z \sim 1.7$, just after the maximum star formation peak (Madau & Dickinson 2014). Thus, the Fe II/CaT ratio will be lowered by the effect of a recent starburst enhancing the α -elements with respect to iron at intermediate redshifts (Martínez-Aldama et al. 2015a). Surprisingly, the Fe II/CaT ratio also has a mild correlation with the PC1 (Figure 7), suggesting that the metal content has a relevant role in governing the properties of our sample.

Hamann & Ferland (1993) found a positive relation between the metallicity, black hole mass, and luminosity; therefore, the highest-metallicity AGN might also be the most massive, as shown in the last row of Figure 4. An exception to the Hamann & Ferland (1993) results are the NLS1 galaxies, which exhibit a high NV/CIV flux ratio (an alternative proxy of the metal content) despite their low luminosity (Shemmer et al. 2004). In our sample a clear NLS1 is the object PHL 1092 (Marinello et al. 2020), showing Fe II/CaT ~ 3.1 , which is close to the mean Fe II/CaT value of the high-redshift sample, putting this source within the regime of high metal content. Previous studies indicate that NLS1 galaxies show a deviation from the relations NV/CIV– L and NV/CIV– M_{BH} (Shemmer & Netzer 2002); in our sample we cannot confirm these results using the ratio Fe II/CaT, since the scatter for a fixed L or M_{BH} is too large.

Based on the flux ratio NV/CIV, Shemmer et al. (2004) found a positive correlation between the abundance (Z) and the Eddington ratio. Our sample shows an anticorrelation between L_{bol}/L_{Edd} and the ratio Fe II/CaT with a Spearman rank coefficient of $\rho \sim 0.554$ and a significance over 99.9%. The bootstrap results proved the reliability of this correlation, with a probability less than 11% to detect a false-positive trend. Considering that it is very likely that a recent starburst concomitantly increases Z and lowers Fe II/CaT (see, e.g., Bizyaev et al. 2019), we expect that the high Eddington ratio sources in our sample are associated with high metal content and/or low Fe/ α values. Similar to the Baldwin effect, the relation between the metal content estimator and the Eddington ratio is stronger than with luminosity, black hole mass, or redshift (Shemmer et al. 2004; Dong et al. 2011; Sameshima et al. 2017; Shin et al. 2019). Conversely, the correlation between the Eddington ratio and Fe II/Mg II remains unclear; depending on the sample considered, there is a positive (Sameshima et al. 2017; Shin et al. 2019; Onoue et al. 2020) or null correlation (Sarkar et al. 2021). Since Fe II and Mg II $\lambda 2800$ are affected by nonabundance parameters such as the density or the microturbulence, the ratio Fe II/Mg II might be affected (Sameshima et al. 2017; Shin et al. 2019). After a correction by these factors, the correlation Fe II/Mg II– L_{bol}/L_{Edd} roughly remains positive or disappears.

Under the assumption that the Fe II/CaT flux ratio is a first-order proxy of the [Fe/Ca] abundances, we found that the behavior shown by the Fe II/CaT flux ratio is in agreement with the normal chemical evolution (Hamann & Ferland 1993, 1999), where the main enrichment occurs in the early epochs. Our results also support that the main Fe II enrichment occurs 1–2 Gyr after the first starburst (Hamann & Ferland 1993, 1999) and also suggest that the strong Fe II could be associated with a second starburst. However, the over-abundance of Fe II depends on the SN Ia lifetime and the star formation epoch (Sameshima et al. 2020). In order to confirm these results, models incorporating chemical evolution and [Fe/Ca] abundances are required. In addition, we must explore the dependence on nonabundance parameters, which could