The weirdest SDSS galaxies: results from an outlier detection algorithm

arXiv: 1611.07526

Dalya Baron[1]*, Dovi Poznanski[1]†
[1] *School of Physics and Astronomy, Tel-Aviv University, Tel Aviv 69978, Israel.*

24 November 2016

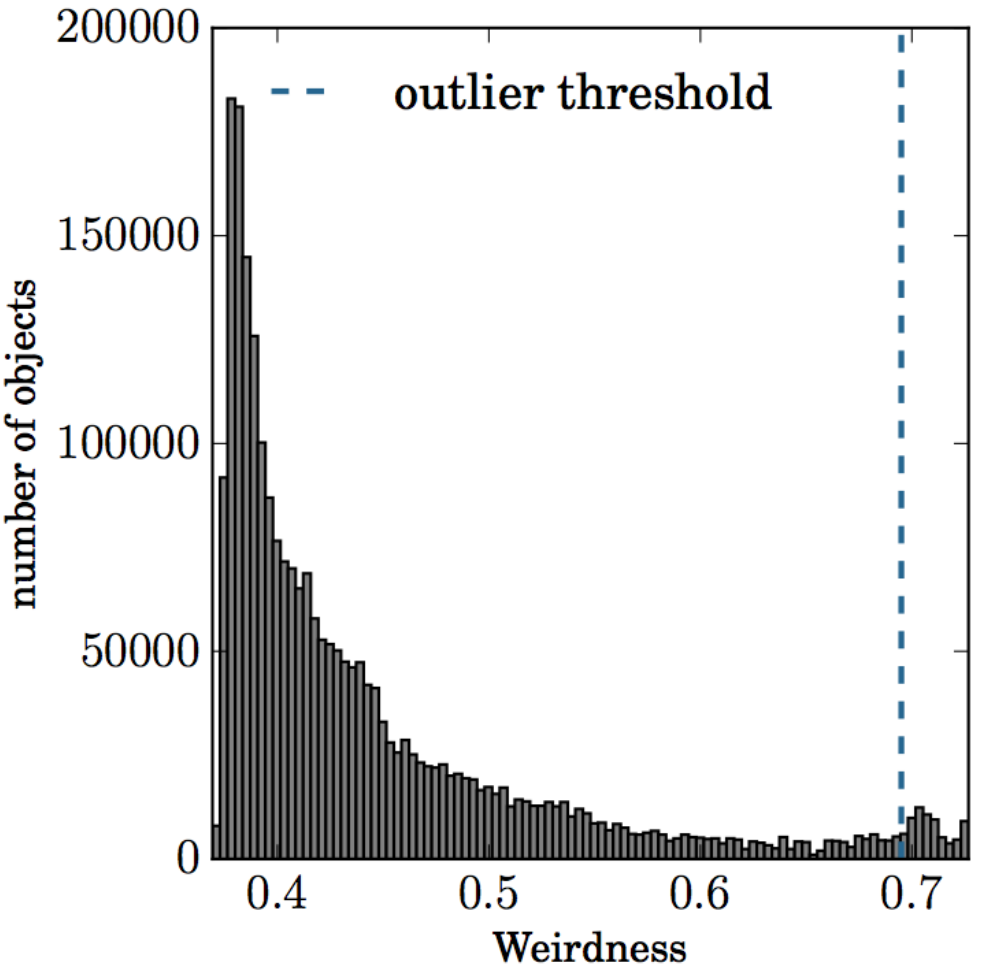
ABSTRACT

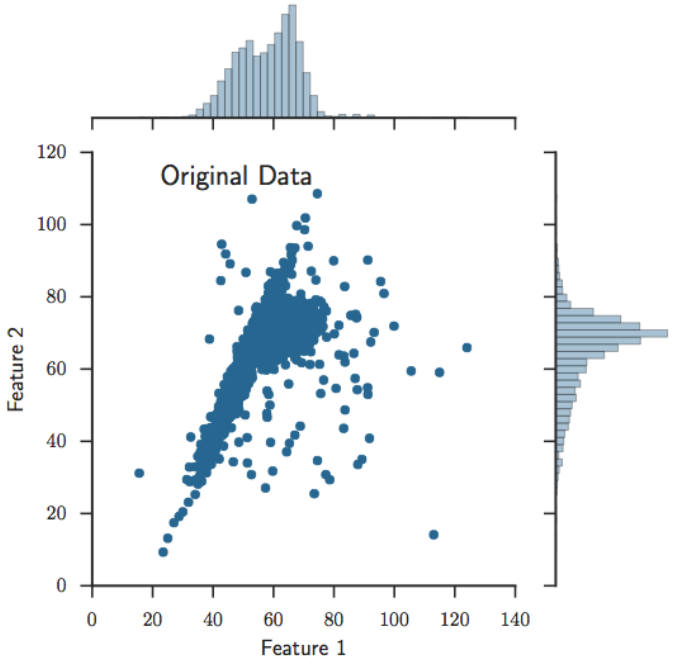
How can we discover objects we did not know existed within the large datasets that now abound in astronomy? We present an outlier detection algorithm that we developed, based on an unsupervised Random Forest. We test the algorithm on more than two million galaxy spectra from the Sloan Digital Sky Survey and examine the 400 galaxies with the highest outlier score. We find objects which have extreme emission line ratios and abnormally strong absorption lines, objects with unusual continua, including extremely reddened galaxies. We find galaxy-galaxy gravitational lenses, double-peaked emission line galaxies, and close galaxy pairs. We find galaxies with high ionisation lines, galaxies which host supernovae, and galaxies with unusual gas kinematics. Only a fraction of the outliers we find were reported by previous studies that used specific and tailored algorithms to find a single class of unusual objects. Our algorithm is general and detects all of these classes, and many more, regardless of what makes them peculiar. It can be executed on imaging, time-series, and other spectroscopic data, operates well with thousands of features, is not sensitive to missing values, and is easily parallelisable.

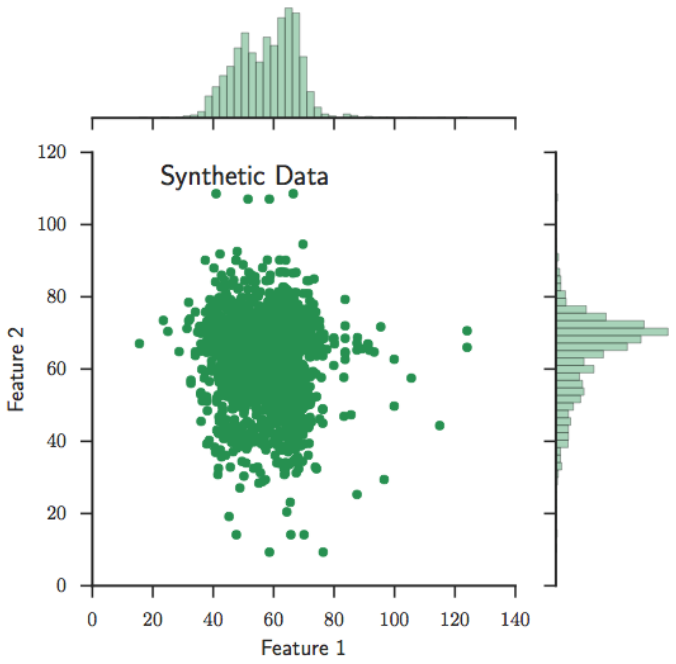
Применили алгоритмы неконтролируемого машинного обучения (Random Forest) для анализа всех спектров из SDSS DR12.
Цель – найти объекты, которые больше всех отличаются от всех остальных

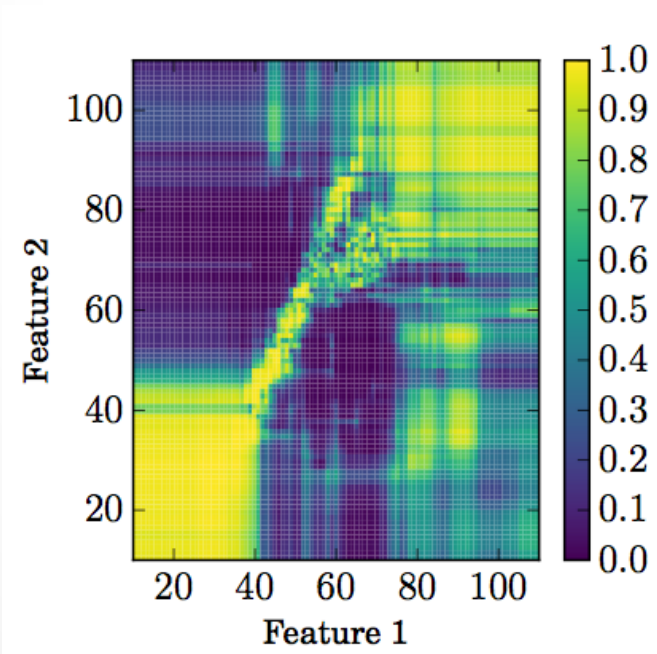
2 355 926 объектов, 15 700 параметров (потоки в линиях и континууме)
Задача – определить расстояния между объектами в 2 355 926x15 700 – мерном пространстве

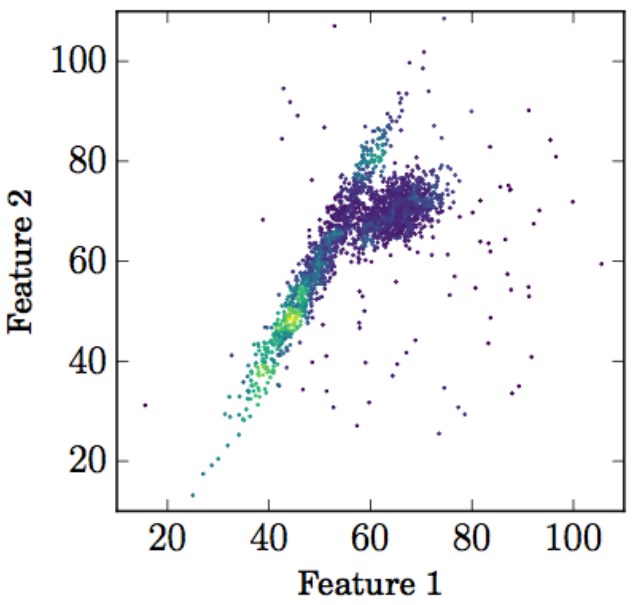
«Normal objects are the ‘known knowns’; outliers of kinds discovered before are ‘known unknowns’; but the most interesting, and challenging to find in large datasets, are the ‘unknown unknowns’, objects we did not know we should be looking for.»

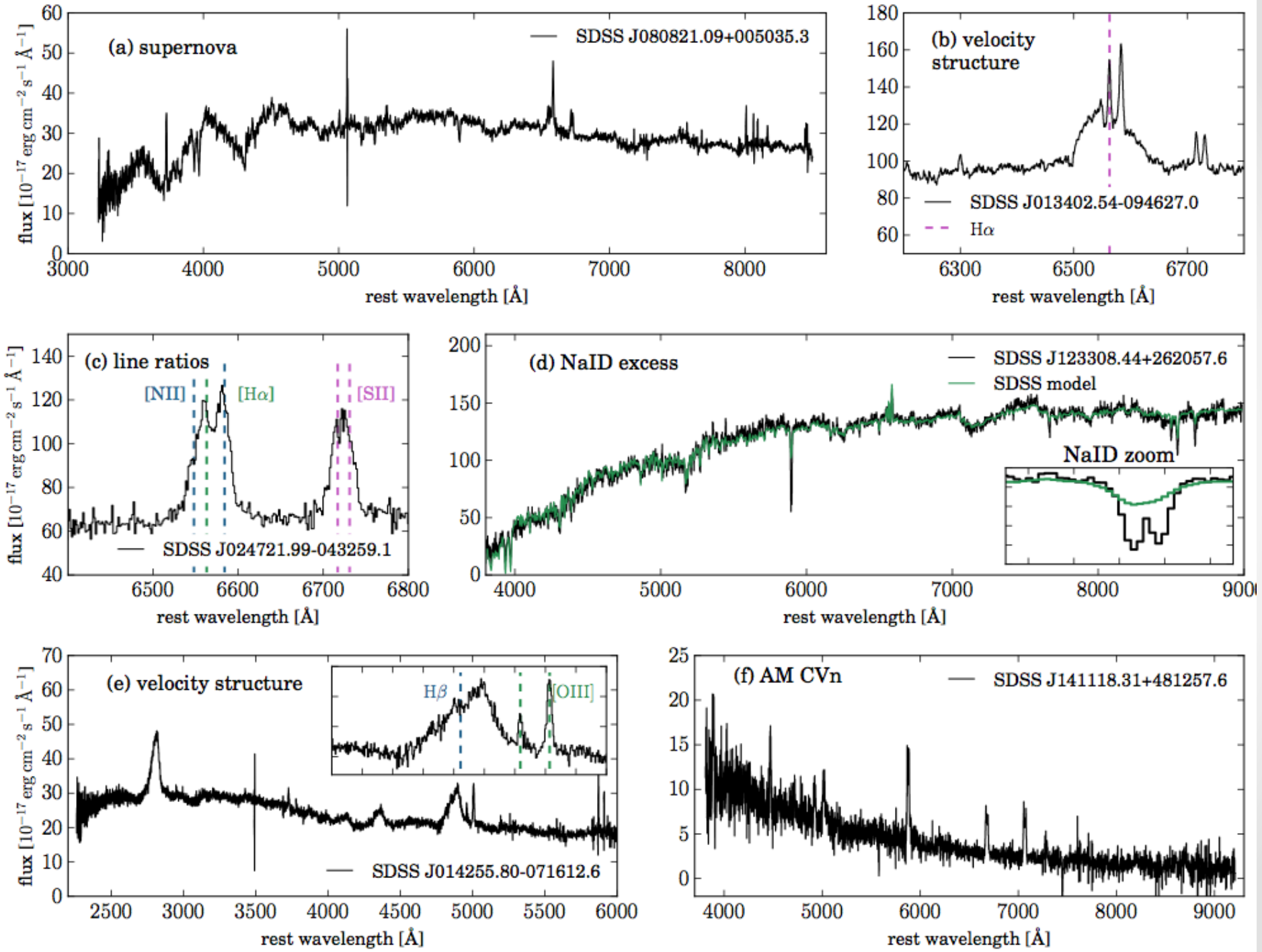


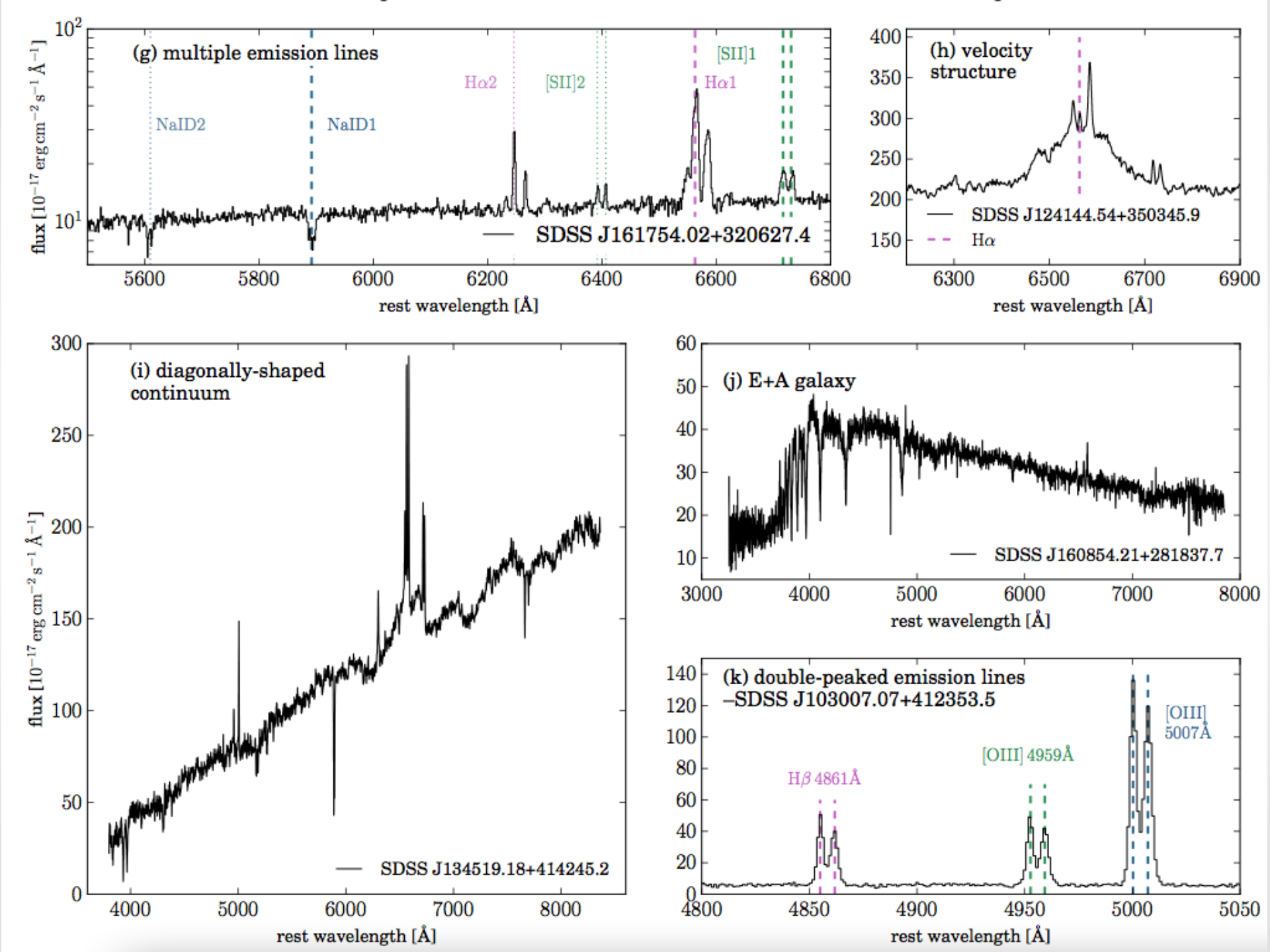

Original Data
Feature 1
Feature 2
0
20
40
60
80
100
120
140


Synthetic Data
Feature 1
Feature 2
0
20
40
60
80
100
120
140


Feature 1
Feature 2
20
40
60
80
100
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Feature 1
Feature 2
20
40
60
80
100


(a) supernova
SDSS J080821.09+005035.3
flux [10^{-17} erg cm^{-2} s^{-1} Å^{-1}]
rest wavelength [Å]
(b) velocity structure
SDSS J013402.54-094627.0
Hα
rest wavelength [Å]
(c) line ratios
[NII]
[Hα]
[SII]
SDSS J024721.99-043259.1
flux [10^{-17} erg cm^{-2} s^{-1} Å^{-1}]
rest wavelength [Å]
(d) NaID excess
SDSS J123308.44+262057.6
SDSS model
NaID zoom
rest wavelength [Å]
(e) velocity structure
Hβ
[OIII]
SDSS J014255.80-071612.6
flux [10^{-17} erg cm^{-2} s^{-1} Å^{-1}]
rest wavelength [Å]
(f) AM CVn
SDSS J141118.31+481257.6
rest wavelength [Å]


(g) multiple emission lines
NaID2
NaID1
Hα2
[SII]2
[SII]1
Hα1
SDSS J161754.02+320627.4
flux [10^-17 erg cm^-2 s^-1 Å^-1]
rest wavelength [Å]
(h) velocity structure
SDSS J124144.54+350345.9
Hα
rest wavelength [Å]
(i) diagonally-shaped continuum
SDSS J134519.18+414245.2
flux [10^-17 erg cm^-2 s^-1 Å^-1]
rest wavelength [Å]
(j) E+A galaxy
SDSS J160854.21+281837.7
rest wavelength [Å]
(k) double-peaked emission lines
–SDSS J103007.07+412353.5
Hβ 4861Å
[OIII] 4959Å
[OIII] 5007Å
rest wavelength [Å]

Выделили 400 наиболее «странных» объектов. Из них лишь около 1/5 было обнаружено ранее.

Обнаружили 2 редких класса объектов: E+A галактики с широкими эмиссионными линиями и галактики с яркими линиями [NI]

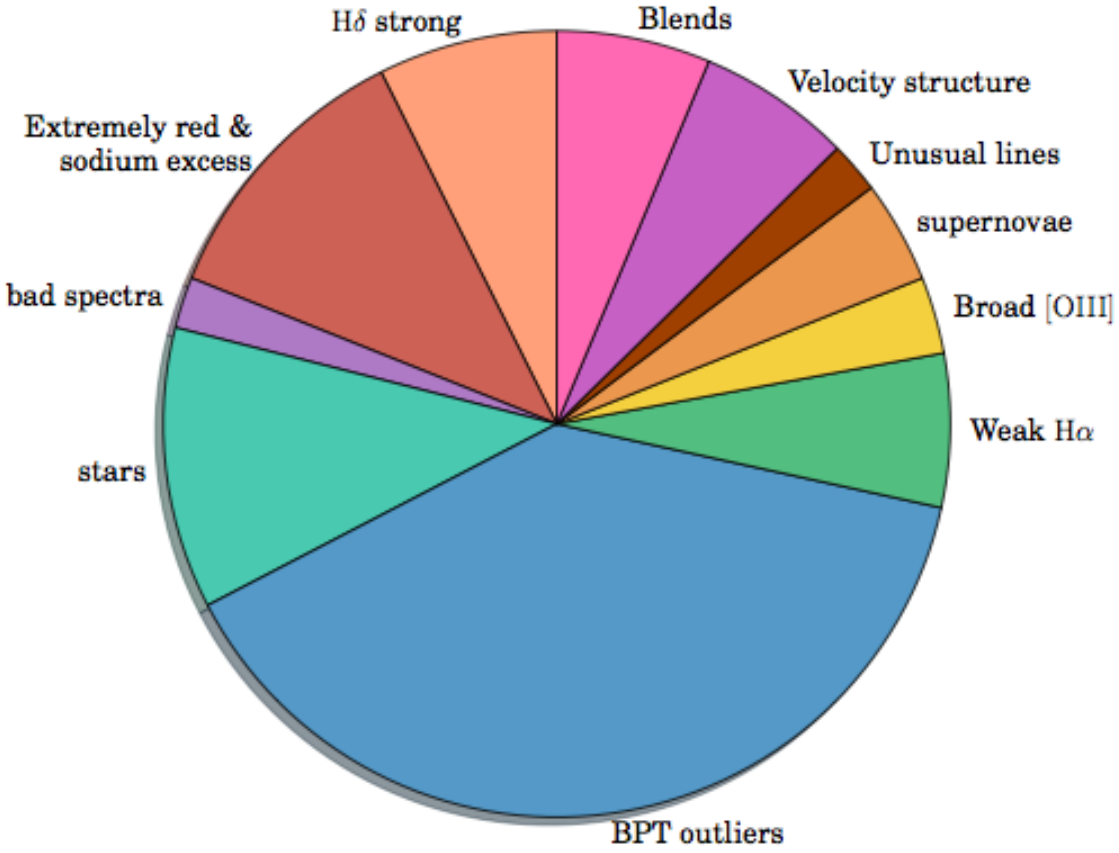


Figure 6. Distributions of the classes of outlying objects we find. Clearly, we find a wide variety of phenomena. It is interesting that only a small fraction (2.5%) are due to instrumental problems.

KINEMATICS OF EXTREMELY METAL-POOR GALAXIES: EVIDENCE FOR STELLAR FEEDBACK

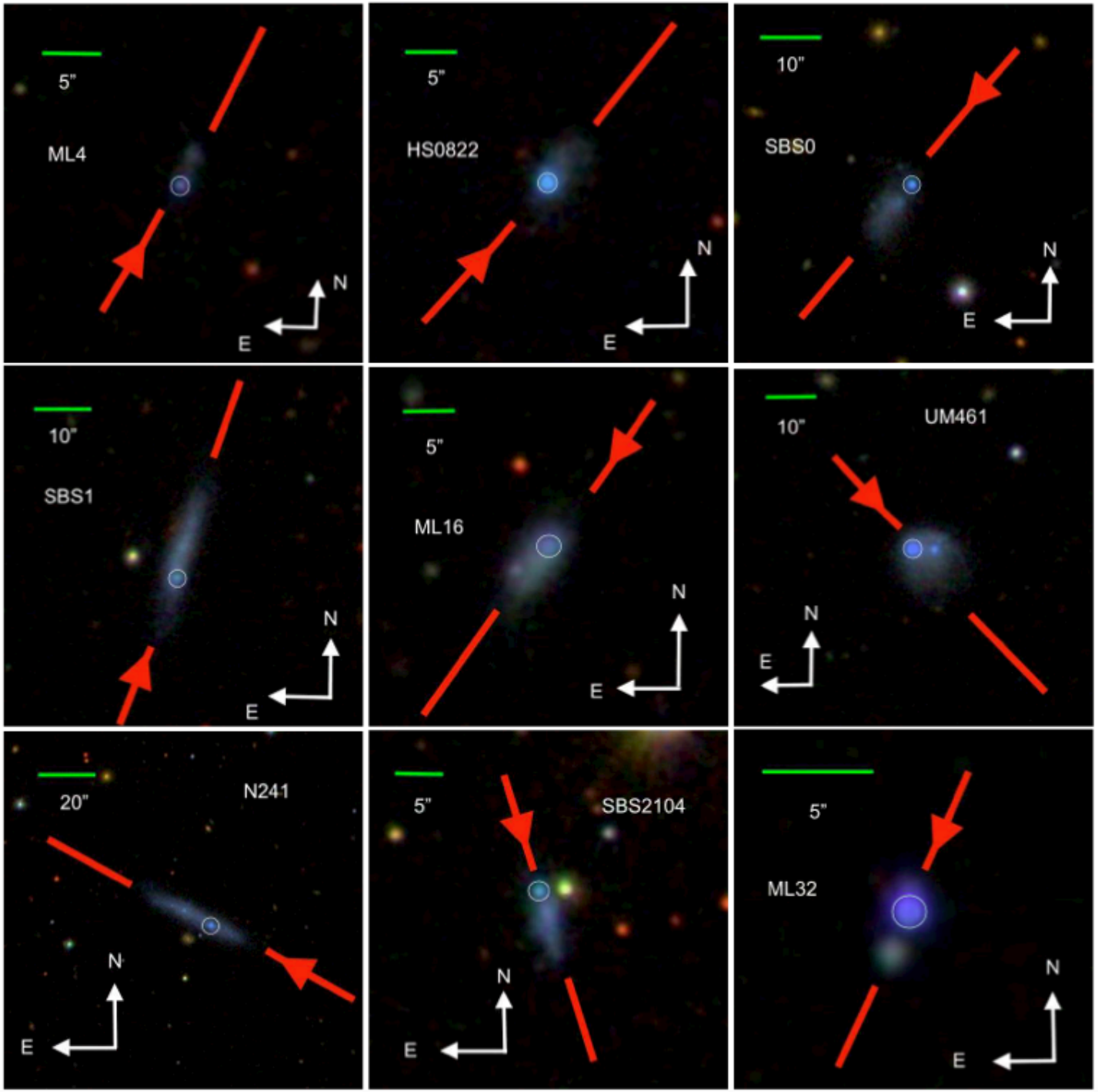
A. Olmo-García[1,2], J. Sánchez Almeida[1,2], C. Muñoz-Tuñón[1,2], M. E. Filho[1,2,3], B. G. Elmegreen[4], D. M. Elmegreen[5], E. Pérez-Montero[6], J. Méndez-Abreu[7]

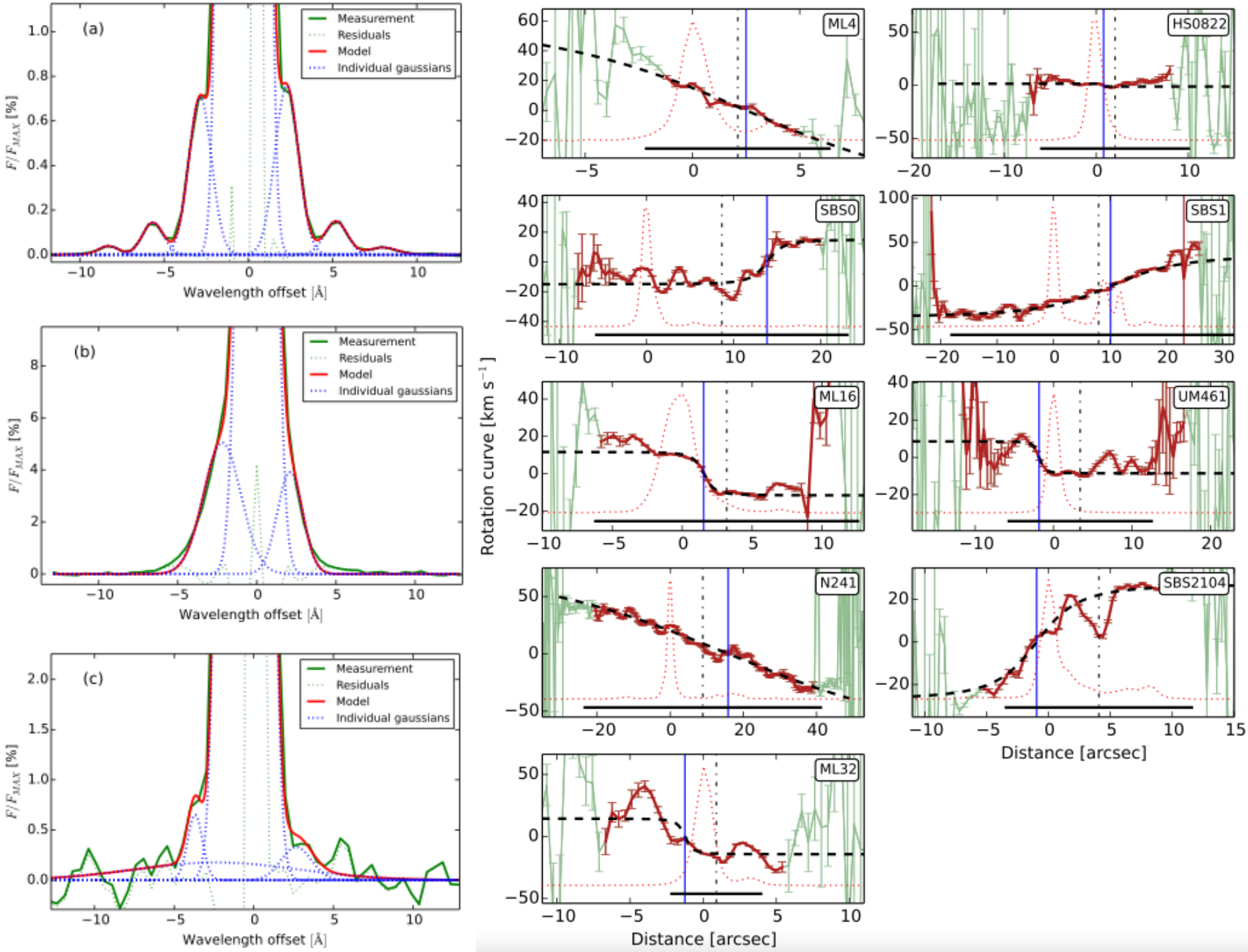
Draft version November 23, 2016

ABSTRACT

The extremely metal-poor (XMP) galaxies analyzed in a previous paper have large star-forming regions with a metallicity lower than the rest of the galaxy. Such a chemical inhomogeneity reveals the external origin of the metal-poor gas fueling star formation, possibly indicating accretion from the cosmic web. This paper studies the kinematic properties of the ionized gas in these galaxies. Most XMPs have rotation velocity around a few tens of $\mathrm{km\,s^{-1}}$. The star-forming regions appear to move coherently. The velocity is constant within each region, and the velocity dispersion sometimes increases within the star-forming clump towards the galaxy midpoint, suggesting inspiral motion toward the galaxy center. Other regions present a local maximum in velocity dispersion at their center, suggesting a moderate global expansion. The Hα line wings show a number of faint emission features with amplitudes around a few percent of the main Hα component, and wavelength shifts between 100 and 400 $\mathrm{km\,s^{-1}}$. The components are often paired, so that red and blue emission features with similar amplitudes and shifts appear simultaneously. Assuming the faint emission to be produced by expanding shell-like structures, the inferred mass loading factor (mass loss rate divided by star formation rate) exceeds 10. Since the expansion velocity exceeds by far the rotational and turbulent velocities, the gas may eventually escape from the galaxy disk. The observed motions involve energies consistent with the kinetic energy released by individual core-collapse supernovae. Alternative explanations for the faint emission have been considered and discarded.

Subject headings: galaxies: abundances – galaxies: dwarf – galaxies: evolution – galaxies: formation – galaxies: structure – intergalactic medium


5"
ML4
N
E
5"
HS0822
N
E
10"
SBS0
N
E
10"
SBS1
N
E
5"
ML16
N
E
10"
UM461
N
E
20"
N241
N
E
5"
SBS2104
N
E
5"
ML32
N
E


(a)
(b)
(c)
Measurement
Residuals
Model
Individual gaussians
F/F_{MAX} [%]
Wavelength offset [Å]
ML4
HS0822
SBS0
SBS1
ML16
UM461
N241
SBS2104
ML32
Rotation curve [km s^{-1}]
Distance [arcsec]

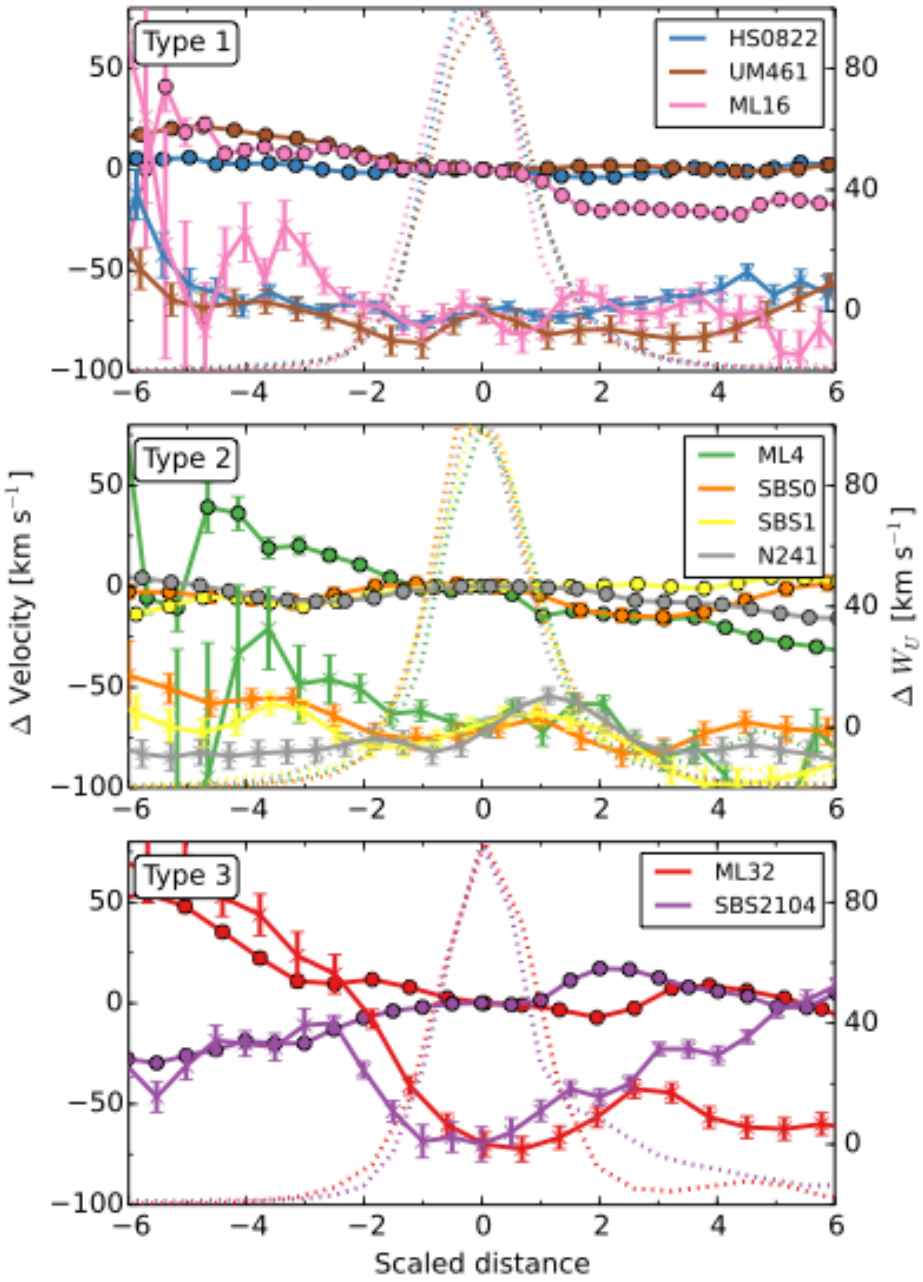
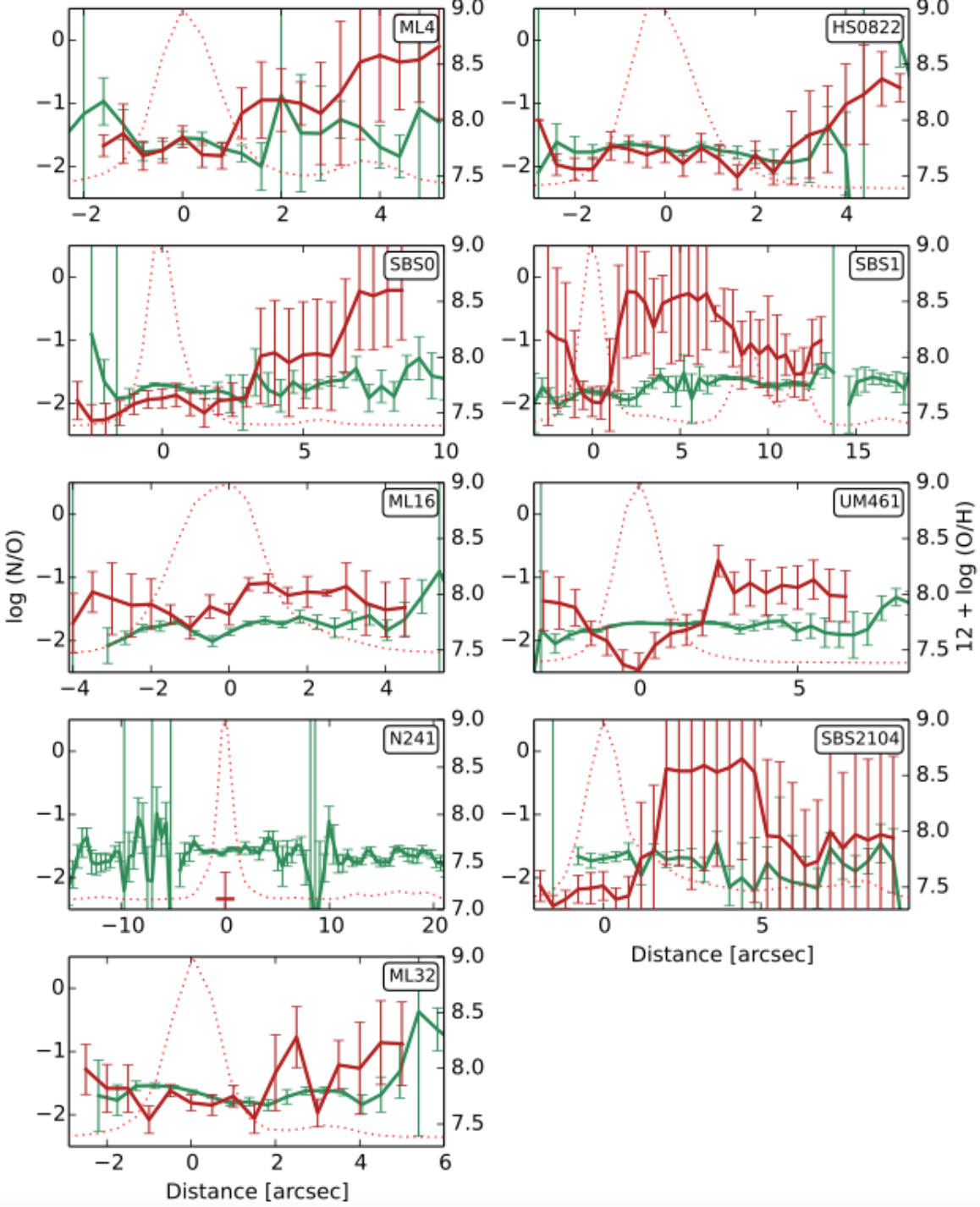


FIG. 7.— Variation of the velocity (bullet symbols) and velocity dispersion (asterisks) across the star-forming regions of the XMPs.


ML4
HS0822
SBS0
SBS1
ML16
UM461
N241
SBS2104
ML32
log (N/O)
12 + log (O/H)
Distance [arcsec]
Distance [arcsec]