

Galaxispopulációk fizikai paramétereinek meghatározása

és

Virtuális Obszervatóriumok

doktori értekezés

Dobos László

okleveles fizikus

Eötvös Loránd Tudományegyetem – Természettudományi Kar
Fizika Doktori Iskola – Részecskefizika és csillagászat program
iskola- és programvezető: Dr. Csikor Ferenc egyetemi tanár

témavezető: Dr. Csabai István egyetemi tanár

Eötvös Loránd Tudományegyetem – Természettudományi Kar
Komplex Rendszerek Fizikája Tanszék



Budapest, 2011.

Tartalomjegyzék

Bevezető	1
Köszönetnyilvánítás	2
I. Galaxispopulációk spektroszkópai vizsgálata	3
Bevezető az I. részhez	4
1. Extragalaktikus spektroszkópia	5
1.1 Spektroszkópai megfigyelések	5
1.2 A spektrográfok főbb típusai	9
1.3 Spektroszkópai észlelések redukciója	12
1.4 A Sloan Digitális Égtérképezés (SDSS)	17
2. Galaxispektrumok analízise	21
2.1 A galaxispektrumok eredete	21
2.2 Populációsztézis-modellek	25
2.3 Galaxispektrumok redukciója	29
2.4 Galaxisok spektroszkópai osztályozása	36
2.5 Magnitúdók K-korrekciója és evolúciókorrekciója	42
2.6 Kompozit spektrumok	44
3. Galaxispektrumok főkomponens-analízise	45
3.1 Galaxispektrumok dimenzióredukciója	45
3.2 A főkomponens-analízis korábbi alkalmazásai	50
3.3 Egy robusztus, iteratív PCA algoritmus *	51
4. SDSS galaxisok kompozitspektrum-atlasza *	53
4.1 Bevezető	53
4.2 Felhasznált adatok	56
4.3 Az osztályozási séma	56
4.4 A kompozit spektrumok	67
4.5 Modellillesztés	72
4.6 Fizikai paraméterek meghatározása	74

4.7	Kompozitok a „térképen”	74
4.8	Empirikus alapú szimulált katalógusok generálása	78
4.9	Összefoglalás	78
Irodalomjegyzék		81
II. Vörös óriásgalaxisok statisztikája *		83
Bevezető a II. részhez		84
5. A vörös óriásgalaxisok luminozitásfüggvényének extrémérték- és sorrendi statisztikája		85
5.1	A felhasznált adatok	87
5.2	A vörös óriásgalaxisok luminozitásfüggvénye	92
5.3	Extrémérték- és sorrendi statisztika	96
5.4	Galaxissűrűség meghatározása a legfényesebb óriásgalaxisok alapján . . .	102
5.5	Klaszterek legfényesebb galaxisainak statisztikája	103
5.6	Összefoglalás	109
5.7	Az extrémérték- és sorrendi statisztika lehetséges felhasználásai a jövőben	110
Irodalomjegyzék		112
III. Virtuális Obszervatóriumok		113
Bevezető a III. részhez		114
6. Adattárházak a tudományban		116
6.1	A relációs adatbázis-kezelők áttekintése	117
6.2	Nem relációs adattárházak	120
6.3	A jövő adatbázis-technológiái	121
6.4	Adattárházak optimális hardverei	122
6.5	Számítási felhők	123
6.6	Környezetvédelmi szempontok	124
7. Sokdimenziós adatbázisok építése relációs adatbázis-kezelőkkel *		125
7.1	Tömbök relációs adatbázisokban	125
7.2	Keresések térbeli indexekkel ellátott adatokon	126
7.3	Többdimenziós adatok lokalitása	127
7.4	Paraméterterek cellázása és indexelése	128

8. Az Internet felhasználása tudományos célú adatok és szolgáltatások közzétételére	131
8.1 Adatbázisok a weben	131
8.2 Webszolgáltatások	132
8.3 Metainformációk, ontológiák, szemantikus rendszerek és az adatok származási története	132
8.4 Az asztrofizikai Virtuális Observatórium	133
9. Fotometriai katalógusok a Virtuális Observatóriumban	134
9.1 Fotometriai katalógusok keresztazonosítása	134
9.2 Katalógusok keresztazonosításának bayesi modellje	134
9.3 Keresztazonosítási algoritmus relációs adatmodellekhez	136
9.4 SkyQuery: Egy robusztus, elosztott keresztazonosító rendszer implementálása relációs adatbázis-kezelőkkel *	137
9.5 Csillagászati katalógusok lábnyoma *	140
10. Csillagászati spektroszkópia a Virtuális Observatóriumban *	141
10.1 A spektroszkópiai adatok sajátosságai	141
10.2 Spectrum Services: Egy spektroszkópiai adattárház a Virtuális Observatóriumhoz	141
10.3 Gyors keresés spektroszkópiai adatbázisokban	142
10.4 Spektrumok előfeldolgozása és egyéb alapfunkciók	143
10.5 Spektrumok analízise	143
11. Összefoglalás	145
Irodalomjegyzék	145

Bevezető

Kedves Olvasó,

Ez a disszertáció három részből áll, melyből kettő szigorú értelemben véve is csillagászati témájú, míg a harmadik a csillagászat (és egyéb alaptudományok), valamint a számítástechnika határterületét érinti.

Az I. részben áttekintem az extragalaktikus spektroszkópia témakörét a megfigyelési módszerektől az adatok redukálásán át a mérések elméleti értelmezéséig. Azért döntöttem a téma részletesebb kidolgozása mellett, mert az extragalaktikus optikai spektroszkópia témaköre magyar nyelven nem hozzáférhető, és az ELTE csillagászati tantárgykínálatából is hiányzik. A rész végén részletesen bemutatom az általam és munkatársaim által összeállított kompozitspektrum-atlaszt.

A II. részben vörös óriásgalaxisok extrémérték-statisztikájáról és sorrendi statisztikájáról írt publikációknak magyar fordítását közlöm.

A dolgozat III. részében az első két résztől eltérően nem konkrét csillagászati tudományos eredményeket mutatok be, sőt szigorú értelemben véve nem is csillagászatról lesz szó; bár az általunk kidolgozott adatfeldolgozási módszerek alkalmazhatóságának illusztrációiként bemutatott lehetőségek a csillagászathoz, asztrofizikához szorosan kapcsolódnak. A tudományos megismerés módszerei az adatintenzív technológiák mindennaposá válásával forradalmi változáson mennek keresztül. Míg száz évvel ezelőtt az új matematika feltalálása vitte előre a fizikát, és matematikai tételek bizonyítása számított komoly elméleti fizikai eredménynek, addig ma az új adatkezelési technikák fejlesztése tölt be hasonló szerepet. Ugyan ma még a világ akadémiai intézményei hivatalosan nem tekintik külön tudományágnak az „e-tudományt”, témavezetőmmel és munkatársaimmal hiszünk abban, hogy egy-két évtizeden belül elnyeri az öt megillető státuszt, és egy olyan saját jogon is elismert multidiszciplináris területté válik, mely nélkülözhetetlen lesz a fizikusok, csillagászok, biológusok vagy akár óceonográfusok mindennapi munkájához.

A dolgozatban *-gal jelöltem meg azokat a fejezetcímeket, melyek (és melyek alfejezetei) teljes egészében, vagy nagyrészt saját kutatási eredményeimen, munkámon alapszanak.

Köszönetnyilvánítás

Legelsősorban köszönettel tartozom témavezetőmnek, Csabai Istvánnak, aki lassan egy évtizede segít eligazodni a tudományos élet útvesztőjében. Témavezetőmön kívül megkülönböztetett köszönet illeti Szalay Sándort és Budavári Tamást, akikkel valódi megtiszteltetés egy csapatban dolgozni. Köszönettel tartozom még Yip Ching-Wának, Frei Zsoltnak, José Blakeley-nek, Stéphane Charlot-nak, Rácz Zoltánnak, Vivienne Wildnak, és Makó Máriának akik sokat segítettek, valamint szüleimnek, akik lehetővé tették, hogy az élet problémái helyett tudománnyal foglalkozzam.

A disszertáció egyes részeinek létrejöttét az alábbi pályázatok finanszírozták: NKTH: Polányi, OTKA-80177 és KCKHA005.

The Project is supported by the European Union and co-financed by the European Social Fund (grant agreement no. TÁMOP 4.2.1./B-09/1/KMR-2010-0003).

I. rész

**Galaxispopulációk spektroszkópai
vizsgálata**

Bevezető az I. részhez

Ahhoz, hogy a galaxisokról, valamint az őket alkotó sötét anyagról, csillagokról és gázzról megfelelően részletes és átfogó képet kapjunk, spektroszkópiai méréseket kell végezni. Galaxisok vöröseltolódásának pontos megállapításához – a pontos vöröseltolódás adatok az univerzum nagyskálás szerkezetének kutatásához alapvető fontosságúak – elengedhetetlen spektroszkópia alkalmazása. Mivel a nagyskálás szerkezet kutatása eleve sok spektrum felvételét igényli, a modern égtérképezési projektek keretében – mint például a Sloan Digitális Égtérképezés¹ (SDSS, Stoughton et al. 2002, Abazajian et al. 2009), valamint az építés alatt álló kínai LAMOST² projekt (LAMOST website 2011) – olyan minőségű spektroszkópiai adatok gyűjtését tűzték ki célul, melyek a vöröseltolódás megállapításán túl lehetővé teszik a galaxisok egyéb fizikai tulajdonságainak vizsgálatát is. Az SDSS által rögzített mintegy félmillió galaxis spektrum egyedi galaxisok vizsgálatán túl teljes galaxispopulációk tulajdonságainak statisztikai vizsgálatára is lehetőséget biztosít. Az SDSS projekt a 2000-es évek eleje óta lényegesen bővítette a galaxisok csillagpopulációiról való ismereteinket.

Mivel az optikai tartományban végzett modern spektroszkópiai megfigyelések ismertetése magyar nyelven kevésbé hozzáférhető, ezért a következő fejezetekben a spektroszkópia alapvető elveit, a csillagászati spektrográfok felépítését és az adatredukciós eljárás alapjait is röviden áttekintjük. Másodikként röviden bemutatjuk a Sloan Digitális Égtérképezés spektroszkópiai részét, majd kitérünk a galaxisok különböző komponenseinek (sötét anyag, csillagok, gáz) spektroszkópián alapuló analízisére. Ezután áttekintünk néhány, a szakirodalomban leginkább elterjedt módszert, amelyekkel az egyes galaxisok fizikai paramétereiről részletes információ kapható. Ezt követően olyan statisztikai módszereket mutatunk be, melyekkel a spektrumok sokezer dimenziós sokasága hatékonyan tanulmányozható. Ezek közül a módszerek közül kiemelten foglalkozunk majd a főkomponens-analízissel, illetve annak speciális változataival, továbbá bemutatjuk a témában végzett saját kutatásaink eredményét is. Végezetül ismertetjük galaxisok kompozit spektrumainak egy általunk összeállított atlaszát, mely a különböző típusú SDSS-galaxisok spektrumainak speciális eljárással történő átlagolásával készült. Az atlasz spektrumaira alkalmazzuk a rész elején ismertetésre kerülő analitikai módszereket is, így meghatározzuk a különböző típusú galaxisok átlagos fizikai tulajdonságait is.

¹SDSS = Sloan Digital Sky Survey

²LAMOST = Large Sky Area Multi-Object Fiber Spectroscopic Telescope = nagy égtérületet vizsgáló, sok objektum megfigyelésére alkalmas spektroszkópiai teleszkóp

1. fejezet

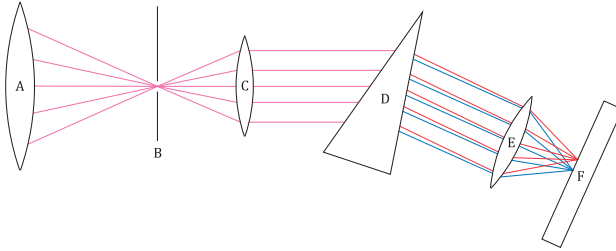
Extragalaktikus spektroszkópia

Extragalaxisok bizonyos fizikai paramétereit nagyszámú megfigyeléssel, megfelelő elméleti háttérrel és matematikai eszköztárral pusztán szélessávú fotometriai adatokból is jó közelítéssel meghatározhatóak. Ebbe a körbe tartozik például galaxisok vöröseltolódásának fotometrián és interpolációs algoritmusokon alapuló meghatározása (Connolly et al. 1995a, Csabai et al. 2000; 2003), vagy a K-korrekció fotometrikus meghatározása (Blanton et al. 2003). Mindezen módszerek közös tulajdonsága, hogy a vizsgált objektumok spektrumairól előzetes feltevéseket teszünk. Ezen feltevések természetesen konkrét spektroszkópiai megfigyeléseken kell, hogy alapuljanak.

Más fizikai paraméterek viszont kizárólag közepes vagy nagy felbontású spektroszkópiai észlelésekből állapíthatók meg. Amellett, hogy a vöröseltolódás csak így mérhető meg pontosan, ebbe a körbe tartoznak még a galaxisok belső dinamikai paramétereit (sebességdiszperzió, sötét anyag mennyisége), a galaxist alkotó csillagpopuláció tulajdonságai (a domináns csillagpopuláció kora, csillagkeletkezési ráta, fémesség), a galaxist alkotó por és gáz tulajdonságai (por mennyisége, gáz ionizációs szintje), illetve a galaxis aktivitását jellemző paraméterek (csillagkeletkezési ráta, aktív mag jelenléte).

1.1. Spektroszkópiai megfigyelések

A spektroszkópia megfigyelések célja, hogy a képpalkotó csillagászati megfigyelések esetében alkalmazott szélessávú színszűrők alkalmazásával szemben olyan nagy hullámhossztartománybeli felbontást érjünk el, mely lehetővé teszi a spektrális energiaeloszlás finom részleteinek tanulmányozását is. Az évtizedek során nagyszámú spektroszkópiai eljárás került kidolgozásra, melyek első sorban abban különböznek egymástól, hogy a különböző megfigyelési paraméterek optimalizálásakor mely szempontok élveznek előnyt. A legfontosabb megfigyelési paraméterek a következők: a megfigyelendő hullámhossztartomány, a hullámhosszbeli felbontás, a távcső mérete, valamint a megkövetelt jel-zaj viszony eléréséhez szükséges integrálási idő.



1.1. ábra. Spektrográfok elvi felépítése: A – távcső (nem a spektrográf része), B – maszk, C – kollimátor elem (a rajzzal ellentétben általában tükör), D – diszperziós elem, E – kamera, F – detektor. A különböző hullámhosszú fénysugarak a detektor más-más pontján fókuszálódnak.

1.1.1. A spektrográfok elvi felépítése

Bár az egyes spektrográfok konkrét felépítése nagyban különbözhet, egy általános, szemantikus kép mégis adható, melyen azonosíthatjuk a fő alkotóelemeket. Az 1.1. ábra a spektrográfok alapvető optikai elemeit szemlélteti. Az égi objektumokból jövő fénysugarakat a távcső úgy képezi le, hogy a fénysugarak a fókusz síkban messék egymást. Megfelelően korrigált távcsövek esetében ez a hullámhossztól függetlenül ugyanabban a pontban történik. Ez utóbbi feltétel elengedhetetlen a spektroszkópiai megfigyelésekhez. Spektrumok rögzítése esetén a távcső fókusz síkjába a képalkotó eszköz (manapság CCD kamera, régebben fotólemez) helyett egy maszk kerül, mely kiválasztja, hogy mely objektumról (vagy objektumokról) készüljön spektrum. A maszk kivágása általában keskeny rés vagy kör alakú furat, attól függően, hogy milyen az alkalmazott műszer konkrét felépítése. A maszkok szükségességére és típusaira a későbbiekben még visszatérünk.

A maszkon áthaladó fénysugarak széttartóak lesznek, így azokat egy kollimátor tükörrel párhuzamosítani kell. A párhuzamos sugarak ezután rávetülnek a diszperziós elemre (vagy elemekre), és a hullámhosszuktól függően az optikai tengelyhez képest más és más szögben térülnek el. Az azonos hullámhosszú fénysugarak a diszperziós elemnél történő áthaladás, vagy arról való visszaverődés után is párhuzamosak maradnak egymással.

A hullámhosszanként párhuzamos sugarak ezután egy fókuszáló tükörről visszaverődve, vagy lencserendszeren áthaladva képződnek le a detektorra. Amennyiben a spektrográf diszperziója nagy, és a különböző hullámhosszúságú sugarak nagy szögeket zárnak be az optikai tengellyel, valamint egymással, akkor a detektorok kis mérete miatt a teljes hullámhossztartomány egyidejű leképezése nem lehetséges. Ilyenkor szükség van a diszperziós elem, vagy a detektor mozgatására, mely a spektrográf mechanikai felépítését jelentősen megbonyolítja, méretét megnöveli.

1.1.2. A maszk változatai

Az előző fejezetben említésre került a spektrográfok egyik fő alkatrésze, a *maszk*. Alapvető jelentőségű, hogy egy adott spektrográf milyen maszkot alkalmaz, illetve egyáltalán rendelkezik-e maszkkal.

Kiterjedt objektumok vizsgálatának esetében feltétlenül szükség van valamilyenféle maszkra, hiszen ha csak a diszperziót okozó elem kerülne beépítésre, úgy a detektoron az objektum különböző területeiről származó, de egymáshoz képest hullámhossz szerint is eltolt spektrumok szuperpozíciója jelenne meg, ami a mérést értelmezhetetlenné tenné. Pontforrások esetében azonban – ha a pontterületi függvény megfelelően kicsi – kis diszperziójú spektrumok sikeresen rögzíthetők maszk alkalmazása nélkül is.

A leggyakrabban alkalmazott maszk az egyszerű optikai rés. A diszperziós elem részspektrográfok esetében úgy kerül beépítésre, hogy az pontosan a résre merőlegesen térítse el a fénysugarakat. Ezáltal a detektoron kétdimenziós kép jön létre: az egyik irányban a hullámhossz függvényében, a másik irányban a rés menti pozíció függvényében rögzíthető az intenzitáseloszlás. A résspektrográfokra az 1.2.2. fejezetben részletesebben kitérünk.

A nagy felbontású spektrográfia alapvető eszköze az échelle-spektrográf, mely két, egymásra merőlegesen beépített diszperziós tagot tartalmaz. Ilyen esetben a maszk értelemszerűen csak kis átmérőjű lyuk lehet. Az échelle-spektrográfokra a 1.2.3. fejezetben külön kitérünk. A több objektum egyidejű megfigyelésére kiválóan alkalmas, optikai fényvezető szálakat alkalmazó spektrográfok ismertetésére később, a 1.2.4. fejezetben kerítünk sort.

1.1.3. A diszperziós elemek típusai

A diszperziós elem feladata, hogy az azon áthaladó, vagy arról visszaverődő fénysugarakat a hullámhossztól függően különböző szögben térítse el. A három alapvető optikai eszköz, ami erre a célra alkalmas a *prizma*, az *optikai rács*, valamint e kettő kombinációja, melyet a nemzetközi szakirodalomban *grism* néven említenek.

A jól ismert prizma valamilyen nagy törésmutatójú üvegből készülő elem, melynek a szükséges tartományban lehetőség szerint lineáris, vagy legalább is lassan változó diszperziós tulajdonságai vannak. A prizmák spektroszkópiai alkalmazásának nagy hátránya az elérhető kis diszperzió, valamint a törésmutató bonyolult hullámhossz függése. Ezen kívül gyártásuk a felhasznált nagy mennyiségű és jó minőségű anyagok miatt drága.

A leggyakrabban használt diszperziós elem az optikai rács, hiszen ezek anyagigénye a prizmákéval szemben minimális, és csak a felületüket kell megfelelően megmunkálni. Alkalmazásukkal szinte tetszőleges nagyságú lineáris diszperzió elérhető. Az optikai rácsok – szintén szemben a prizmákkal – olyan ultraibolya tartományokban is használhatók, ahol az üveg már nem ereszti át a fényt. Ugyanakkor optikai rács használatakor elkerülhetetlen

a spektrográfok nagy mérete. További nehézség, hogy a rácsokról visszaverődő fénysugarak több rendben is elhajlanak, és ezek a rendek egymásra szuperponálódhatnak. A rendeket vagy további optikai elemekkel, vagy speciális adatfeldolgozó eljárásokkal lehet szétválogatni. A rácsosaz kialakítható úgy is, hogy a rovátkák oldalai adott szöget zárjanak be a rács felületének normálisával¹. Ezzel a módszerrel elérhető, hogy a visszaverődés határfoka egy adott hullámhossz környékén jóval nagyobb legyen, mint más hullámhossztartományokban; ez segít a zavaró szuperponálódott magasabb rendek intenzitásának lecsökkentésében is. Échelle-rendszerű spektrográfok esetében két optikai rácsot alkalmaznak. Az első rácsról visszaverődő fénysugarakat egy második, 90°-kal elforgatott rács tovább bontja, így a detektor mátrixon nagy felbontású, két dimenziós kép jelenik meg.

A grism az optikai rácsot és a prizmat úgy kombinálja, hogy a rácsot a prizma kilépő felületén alakítják ki. Az eszköz általában úgy van méretezve, hogy egy bizonyos hullámhosszúságú fény egyenesen haladjon át rajta. Ez a tervezés lehetővé teszi a grism objektívprizmaként történő alkalmazását is (ld. 1.2.1. pont). Az eszközzel jelentősen nagyobb diszperzió érhető el, mint az egyszerű prizmával, és a spektrográf egészének mérete is csökkenthető. Sokszálas spektrográfok esetében általában ilyen diszperziós tagot használnak.

1.1.4. Detektortípusok

Bár a vöröseltolódás meghatározására a spektrofotometriailag igen nehezen kalibrálható *fotólemezek* is alkalmasak voltak, galaxisok részletes vizsgálatának céljára mindenképp jól kalibrálható, közvetlenül digitalizálható jelet előállító detektorok alkalmazása szükséges. Korábban erre a célra *fotométerekeket* alkalmaztak, azonban a fotométerekekből nem alakítható ki nagy felbontású lineáris vagy mátrix elrendezés. Emiatt a fotométert, vagy a spektrográf diszperziós elemét mozgatni kellett, hogy a hullámhosszfüggő intenzitást fel lehessen térképezni.

Manapság spektrográfok esetében is jóval elterjedtebb a *CCD mátrixok* alkalmazása. Ebben az esetben a spektrográfot megfelelő lencserendszerrel ellátott kamerával kell kiegészíteni, ami a spektrográf diffrakciós képét valódi képpé képezi le a detektormátrixon. Échelle-elrendezés esetén, CCD mátrix alkalmazásával, mozgó tag nélkül is nagy felbontású spektrográf készíthető.

1.1.5. Spektrális felbontás

Értelemszerűen a spektrográfok egyik fő paramétere, hogy milyen felbontóképességgel rendelkeznek a hullámhossztartományban. Mivel a spektrumok hullámhossztartománybeli mintavételezése gyakran nem lineáris, hanem logaritmikus skálán történik, ezért a

¹blazing

felbontás jellemzésére az $R = \lambda/\Delta\lambda$ mennyiség használatos. A Doppler-eltolódás formulája szerint definiálható továbbá az ú.n. sebességszórás: $\sigma_v = c\Delta\lambda/\lambda$.

A spektrális felbontóképesség megtervezésekor számos szempont alapján kell döntést hozni. A nagyobb felbontóképesség nem csak a kívánt jel-zaj arányú mérés eléréséhez szükséges integrálási időt emeli meg, de a spektrográf geometriai méreteit is jelentősen megnöveli. Legtöbb esetben ez utóbbi a fő korlátozó tényező. Ugyanakkor a vizsgálandó objektumok bizonyos spektrális tulajdonságainak meghatározásához (vonalprofilok, vöröseltolódás) szükséges minimális felbontóképességet a spektrográfnak biztosítania kell.

További jelentős szempont, hogy közepes felbontású ($R \approx 2000$) spektrográfok esetében fixen beépített optikai tagok és detektorok alkalmazhatók. Nagy spektrális felbontás esetén a műszer méretei szükségessé tehetik a diszperziós tag, vagy a detektor mozgatását, mely a mechanikai megoldások bonyolultságán túl az észleléshez szükséges integrálási időt is sokszorosára növeli. Ugyancsak problémát okoz, hogy nagy felbontás esetén a különböző diffrakciós rendek szuperponálódnak, mely az adatfeldolgozási eljárást jelentősen bonyolítja.

Mivel a nagy felbontású megfigyelésekhez szükséges integrálási idő a műszerek áránál is jelentősebb korlátozó tényező, ezért nagy felbontású spektrumok extragalaktikus objektumokról ritkán készülnek. Ha mégis, akkor az ilyen spektroszkópiai minták statisztikai elemzésre nem alkalmasak. A sok megfigyelést igénylő extragalaktikus spektroszkópiai égfelmérések közepes felbontással készülnek.

1.2. A spektrográfok főbb típusai

1.2.1. Az objektívprizma

E spektrográfia módszer esetében ú.n. *objektívprizmát*, egy kis diszperziójú prizmat alkalmaznak, melyet közvetlenül a távcső optikai rendszerébe építenek be. Az objektumok képei ezáltal hullámhossz szerint elnyúlt foltokként jelennek meg az egész fókusz síkban, és a spektrumok a képalkotásra alkalmas detektorral rögzíthetők. Az eljárás leginkább Schmidt-távcsövek esetében kerül alkalmazásra, és kiválóan alkalmas nagyszámú objektum egyidejű megfigyelésére. Az igen kis spektrális felbontás miatt azonban – nagy diszperzió esetén ugyanis az objektumok elnyúlt képei egymásra kerülnének – a módszer csak korlátozottan alkalmazható. Ezt az eljárást régebben szélesebb körben alkalmazták bizonyos típusú pontszerű források kiválasztására, ugyanis a különböző típusú csillagok, valamint a különböző vöröseltolódású kvazárok igen jellegzetes alakú foltokat produkálnak a hagyományos fotólemezeken. Az objektumok a kis diszperziós felvételek alapján szemmel könnyen katalogizálhatóak, és későbbi nagy felbontású spektroszkópiai analízis céljára kijelölhetőek.

1.2.2. Résspektrográf

Optikai rés alkalmazása esetén lehetőség nyílik arra is, hogy kiterjedt objektumokról ún. *kétdimenziós spektrumot* készítsünk. Ezt a módszert *hosszú réses spektroszkópiának* nevezik. Ebben az esetben az optikai rést a galaxis képének valamelyik tengelye mentén pozícionáljuk, a diszperziós elemet pedig úgy állítjuk be, hogy a rés irányára pontosan merőleges irányban okozza a különböző hullámhosszú fénysugarak eltérülését. Eredményül egy hullámhossztól és radiális pozíciótól egyaránt függő intenzitás képet kapunk, mely képalkotó eszközzel (például egy kamerával és egy CCD detektorral) rögzíthető. A módszer igen hatékonyan használható közeli, nagy látszólagos átmérőjű galaxisok tanulmányozására.

Amennyiben egyidejűleg több keskeny rést alkalmazunk, akkor egyszerre több objektum spektruma is rögzíthetővé válik – hasonlóan a kis diszperziós, maszk nélküli eljárásához. Ebben az esetben a rések elhelyezését (azaz a látómezőben észlelendő objektumok listáját) úgy kell optimalizálni, hogy a képalkotó detektoron megjelenő spektrumok ne essenek egymásra. A módszer segítségével kis és közepes felbontású spektrumok rögzíthetők nagy számban, így az eljárás ma is széles körben elterjedt. Az eljárás neve *többréses spektroszkópia*. Ilyen rendszerű spektrográf például az Európai Déli Observatórium VLT távcsöveinek VIMOS műszere, valamint a Keck-távcsövek DEIMOS műszere. A többréses spektroszkópia hátránya, hogy az elérhető felbontás korlátos, valamint a teljes detektorfelületet nehéz kihasználni.

1.2.3. Échelle-rendszerű spektrográfok

Egyetlen diszperziós elem alkalmazása esetén nagy spektrális felbontás csak sűrű rácsmintázattal érhető el. Ez azonban együtt jár azzal a kellemetlenséggel, hogy a magasabb diszperziós rendek egymásra vetülve jelennek meg. Ennek kiküszöbölésére fejlesztették ki az échelle-rendszerű spektrográfokat, melyekben a nagy spektrális felbontást két, külön-külön lényegesen kisebb diszperziójú elem biztosítja, mint amilyen diszperzióra egyetlen rács esetében szükség lenne. Az első diszperziós tag egy közepes diszperziójú rács vagy *grism*. A második pedig egy erre pontosan merőleges irányú diszperziót okozó rács. Ez az optikai elrendezés egyetlen pontból kiinduló fénysugarak diffrakciós képét állítja elő egy kétdimenziós detektoron, így egy spektrográffal egyszerre csak egy objektum megfigyelésére van lehetőség.

1.2.4. Üvegszálás spektroszkópia

Spektroszkópiai megfigyelések esetén az expozíciós idő a leginkább szorító tényező. Ha a cél minél több spektrum rögzítése, olyan távcsövet és spektrográfot érdemes tervezni, melyekkel egyszerre minél több objektumról lehet spektrumot készíteni. Ehhez nagy fényerő mellett széles látómezőre van szükség, valamint olyan többnyílásos maszkra, ami lehető-

ség szerint egyszerűen kicserélhető, vagy akár automatikusan átkonfigurálható aszerint, hogy a látómező mely objektumairól kívánunk spektrumot készíteni. A fő problémát az jelenti, hogy egy nagylátószögű távcső esetében a fókuszszík átmérője a 60-100 cm-t is elérheti, ami hatalmas spektrográf méreteket tesz szükségessé.

A spektrográfok felépítése azonban módosítható úgy, hogy a maszkon áthaladva a fény nem közvetlenül a diszperziós elemre vetül, hanem egy fényvezető üvegszállal a távcsőből kivezetésre kerül. Mivel az optikai szálak tetszőleges geometriai elrendezésben végződthetők, az objektumok fókuszszíkbeli pozíciója csak gyenge feltételeket szab egyidejű spektroszkópiai vizsgálhatóságukra (a fényvezető szálak nem kerülhetnek egymáshoz túl közel). Az üvegszálás technológia lehetővé teszi, hogy egy széles látószögű távcső és több száz optikai szál alkalmazásával egyszerre igen nagy számú spektrumot rögzítsünk.

A gyakorlatban az üvegszálak fókuszszík felőli végét egy előre legyártott maszk fura-taiban rögzítik, vagy különböző más megoldásokkal pozicionálják. Létezik spektrográf, melynél erős mágnesek rögzítik a szálakat, melyeket a megfigyelés előtt egy robot helyez el a megfelelő pozíciókban egy acéllemez felületén.

A fényvezető szálak spektrográf felőli végét egy egyenes mentén kell végződtetni, így a konfiguráció pontosan megfelel a hosszúréses spektrográf konfigurációjának.

A legjelentősebb optikai szálak spektrográfok a Sloan Digitális Égtérképezés (Stoughton et al. 2002) távcsővének spektrográfjai, a 2dF (Colless et al. 2001) és 6dF (Jones et al. 2004) spektroszkópiái égfelmérésekhez épített spektrográfok, valamint a kínaiak új LA-MOST távcsőve és spektrográfja.

1.2.5. Többdimenziós spektroszkópia

Ha egy megfigyelt objektumról készített spektrum esetében a fluxussűrűséget csak a hullámhossz függvényében adjuk meg, úgy egydimenziós spektrumról beszélünk. A kétdimenziós spektroszkópiáról már ejtettünk szót a hosszúréses spektroszkópia kapcsán. Ebben az esetben a hullámhosszon kívül egy lineáris koordinátától is függ a mért fluxus értéke. Ez a módszer hatékony eszközt jelen galaxisok rádiálisan változó paramétereinek meghatározásakor, mint amilyen például a rotációs görbe.

Az optikai szálakat alkalmazó spektrográfok egy speciális típusa esetében a szálakat egyik végükön szoros kötegbe rendezik. Ezzel a megoldással a távcső fókuszszíkjának egy jelentős területe mozaik szerűen lefedhető. Ezt a módszert *teljesmező-spektroszkópiának*² vagy *háromdimenziós spektroszkópiának* nevezzük, hiszen a hullámhossz dimenzión túl a két égi koordináta függvényében is készül mérés.

²integral field spectroscopy

1.2.6. Spektroszkópia az időtartományban

A megfigyelhető időskálán változó objektumok esetén beszélhetünk idődimenzióról is. Változócsillagok vizsgálatának egyik alapvető fontosságú módszere a rövid időskálán, akár napi rendszerességgel végzett spektroszkópiai megfigyelés. Az SDSS adatbázisa tízezer számra tartalmaz ismételt megfigyeléseket, ahol két megfigyelés között akár több év is eltelhetett. Kvazárok, aktív magvú galaxisok esetében ez kivételes lehetőséget jelent a változó spektrális komponensek azonosítására és vizsgálatára (Vanden Berk et al. 2004, Wilhite et al. 2005; 2006, Yip et al. 2009).

1.3. Spektroszkópiai észlelések redukciója

CCD detektorral készített spektroszkópiai észlelések esetén a redukciós eljárás a képalakító észlelések esetében ismert eljárásnál lényegesen összetettebb. A detektor sötét áramának és kiolvasási nulla szintjének meghatározása ugyanúgy történik, mint képpalkotó rendszerek esetén, de a további lépéseknél már figyelembe kell venni a detektor érzékenységének hullámhossz függését.

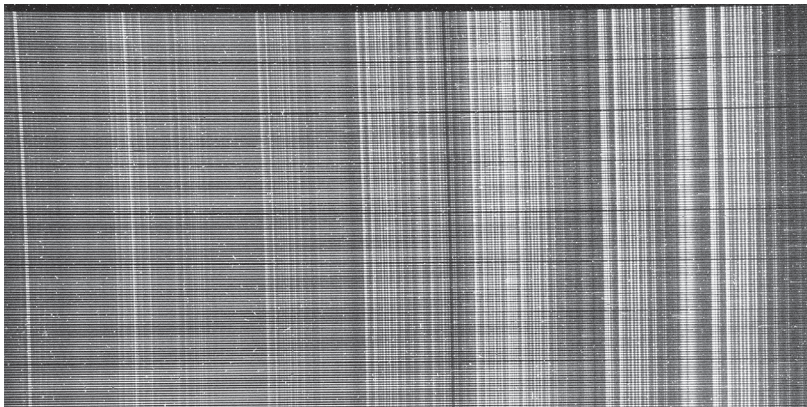
A hullámhossz kalibrációt magát viszonylag könnyű elvégezni, hiszen elegendő egy ismert vonalakat kibocsátó ívlámpa fényét áttereszteni a spektrográfon. A képpontonkénti érzékenység kalibrációjához spektrográfok esetében is egy egyenletes megvilágítású felületről (vagy néha az alkonyati égről) kell képet készíteni, ám itt a hullámhosszfüggés miatt a fényforrást körültekintően kell megválasztani. Célyszerű minél laposabb spektrumú, lehetőleg magas hőmérsékletű izzólámpát használni.

A CCD mátrixot alkalmazó spektrográfok esetében a detektoron létrejövő vonalas mintázat bonyolult, feldolgozásához a spektrográf geometriáját pontosan ismerni kell. Az 1.2. ábra egy üvegszálás spektrográf CCD detektorán létrejövő nyers mintázatot szemlélteti.

Hosszú idejű megfigyelések esetén a spektrográf geometriája különböző környezeti tényezők hatására megváltozhat (pl. a hódilatáció, vagy a távcső mozgatásából eredő deformációk miatt). Emiatt a mérési paramétereket az éjszaka során rendszeresen újra kell kalibrálni.

1.3.1. Fluxuskalibráció

A szélessávú megfigyelésekhez hasonlóan a spektroszkópiai megfigyeléseket is standard csillagokhoz kalibráljuk, ám a kalibráció művelete annál jóval összetettebb, hiszen a kalibrációt a hullámhossz függvényében kell elvégezni. Ehhez pontosan kell ismerni a standard csillag elméleti spektrumát és a köztünk és a csillag között levő intersztelláris gáz – szintén hullámhossz függő – abszorpcióját.



1.2. ábra. Űvegszálás spektrográf CCD-je által rögzített nyers kép. A vízszintes csíkok egy-egy objektum űvegszál által továbbított spektrumai. Jól megfigyelhetők az emissziós vonalak, melyek világos, függőleges sávokként látszanak. Az ilyen nyers spektrumokból összetett képfeldolgozási módszerekkel nyerhetők ki az egyedi objektumok kalibrált spektrumai. Forrás: SDSS.

A kalibráció elvi folyamata a következő: az ismert spektrumú csillagról megfigyelést készítenk, a megfigyelést kiredukáljuk (elvégezve az ég szintjének levonását, valamint az extinkciókorrekciót, ld. később), majd az eredményt elosztjuk a csillag elméleti spektrumával. Így egy korrekciós függvényhez jutunk, amit minden későbbi megfigyelés eredményére alkalmazni kell. Ez a korrekció elvileg automatikusan tartalmazza a légkör hullámhosszfüggő fényszórásának hatását az adott légtömeg mellett.

Természetesen a gyakorlatban a fent felvázolt, leegyszerűsített eljárás számos nehezen megoldható problémát vet fel. Ezek közül leginkább az ég szintjének megbecslése problémás.

1.3.2. Az ég spektrumának meghatározása és levonása

Az éjszakai ég spektruma meglehetősen összetett, melyet sok tényező befolyásol. Az optikai tartományban a spektrum egy viszonylag lapos kontinuumból és jellegzetes vonalakból áll. Az ég éjszakai sugárzását a természetes jelenségeken túl az obszervatórium közelében fekvő települések fényszennyezése is erősen befolyásolja. Külön gondot okoz, hogy az éjszakai spektrum az éjszaka folyamán jelentősen megváltozhat.

Az ég spektrumáról összefoglaló dolgozatot készített Benn & Ellison (1998). Az általuk rögzített spektrumot az 1.3. ábra mutatja. A kontinuum jelentős része az éjszakai égboltfényből³, az állatövi fényből, a háttérobjektumok diffúz sugárzásából, valamint a légköri

³airglow

O₂ molekulák rezgési és forgási módusaiból származik.

Az 7000 Å-nél rövidebb hullámhosszúságú vonalak közül az O atom vonalai, és a Na atom dublettje a legjelentősebbek. Ez utóbbi vonalak egyrészt a természetes eredetű, mezoszférában található több kilométer vastag nátriumrétegből származnak, másrészt az utcai világítás nátriumgőz-lámpáinak légkör által visszavert fényéből. Megfigyelhető továbbá egy határozott Hg vonal is, mely egyértelműen a fluoreszcens világítótestek által okozott fényszennyezésből származik. A Hg és Na vonalak jól használhatók a spektrumok hullámhossz-kalibrációjára is.

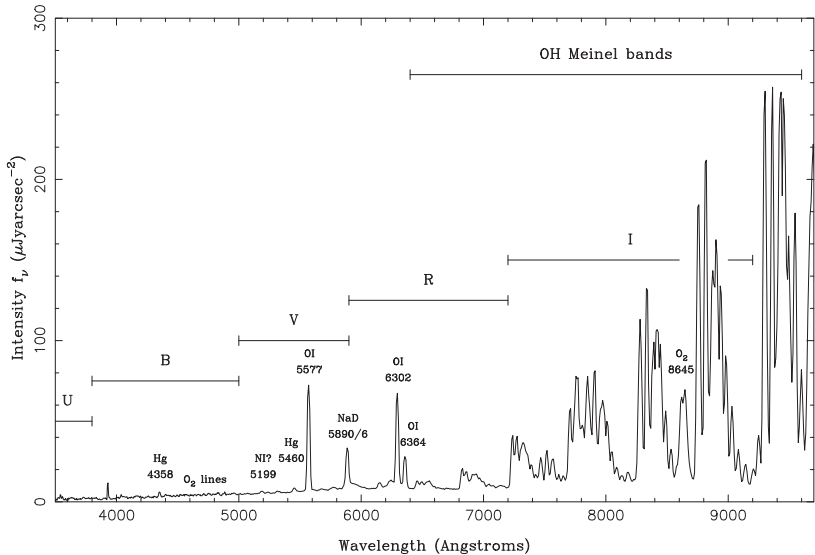
7000 Å fölött a karakterisztikus vonalak helyett az OH molekulagyök különböző gerjesztési módusaiból származó igen erős sávok, az ún. Meinel-sávok jelennek meg. Az ég levonásakor ezek okozzák a legjelentősebb problémát.

Az ég spektrumának levonása a megfigyelt objektumok spektrumából a gyakorlatban a következőképp történik: Előzőleg külön rögzítjük az ég több különböző, a vizsgálandó objektumhoz közeli területének spektrumát, majd ezeket átlagoljuk, és levonjuk az objektum spektrumából. Ez általában jó eredményt szolgáltat a közeli ultraibolya tartományban és a látható fény tartományában is, ám közeli infravörösben a Meinel-sávok eltávolítása meglehetősen nehéz. A problémára Wild & Hewett (2005) talált hatékony megoldást. Ők az SDSS megfigyeléseiből vett ég spektrumok főkomponenseit határozták meg, és az így nyert bázison kifejtve az égzeviduumokat az OH sávokat sikeresen eltávolították. A főkomponens-analízis módszeréről bővebben ld. a 4.4.5. fejezetet.

Az OH sávok szennyezésének kitett közeli infravörös tartomány spektroszkópiai vizsgálata különösen fontos a nagy vöröseltolódású kvazárok tanulmányozása szempontjából, hiszen ezek látható és közeli ultraibolya vonalai a vöröseltolódás miatt a közeli infravörös tartományban figyelhetők meg. Galaxisok spektroszkópiája szempontjából is fontos a 8500 Å körüli tartomány, ahol a CaII ion jellegzetes triplette található. A CaII triplett a csillagpopuláció fémességének fontos indikátora (Bica & Alloin 1987, Armandroff & Zinn 1988, Carrera et al. 2007, Foster et al. 2010).

1.3.3. A galaktikus extinkció korrigálása

Csillagászati spektrumok redukciójakor – miután a földi légkör hatását kiküszöböltük – figyelembe kell vennünk azt is, hogy a hozzánk eljutó fény a Tejútban található jelentős mennyiségű poron és gázon is áthalad. A különböző összetételű por- és gázfelhők abszorpciójának elméleti leírása körülményes, ezért a gyakorlatban alkalmazott korrrekciós eljárások empirikusak. A felállítható törvények a fényelnyelés nagyságát adják meg a hullámhossz függvényében. A Tejút anyagának extinkciós törvényei ismert típusú csillagok megfigyeléséből határozhatók meg. Ezt a munkát többek között Cardelli et al. (1989) végezték el, mely eredményeket O'Donnell (1994) pontosított. Előre utalunk a 2.1. ábrára, ahol folytonos vörös görbe mutatja a Tejútban levő por hullámhosszfüggő extinkcióját.



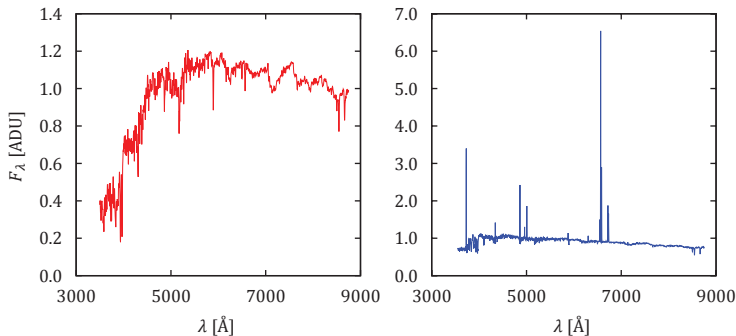
1.3. ábra. A La Palma-i Hold nélküli éjszakai ég spektruma. A betűkkel jelzett intervallumok a szabványos Johnson-színszűrők tartományait jelölik. Kép forrása: Benn & Ellison (1998).

Az extinkció értékének meghatározásához nem elegendő maga az elnyelési törvény, hanem szükség van az elnyelő anyag optikai mélységének ismeretére is. Galaktikus objektumok spektrumának extinkcióra történő korrekciójakor valamilyen módon meg kell becsülnünk a köztünk és a megfigyelt csillag között levő por optikai mélységét. Extragalaktikus objektumok esetében szerencsésebb a helyzet, mert ekkor a por mennyisége nem függ az objektum távolságától, csak a megfigyelés irányától. Infravörös tartományú megfigyelésekkel a Tejútban található por feltérképezhető. A legelterjedtebb ilyen portérképét Schlegel et al. (1998) készítették.

1.3.4. A vöröseltolódás és a spektroszkópiai típus meghatározása

A vöröseltolódás meghatározására több lehetőség is adódik. Ha valamilyen módon meghatározzuk az emissziós vonalak pozícióját, és sikerül a jól ismert vonalsorozatokat azonosítani (például a hidrogén Balmer-sorozatát), akkor a vonalak hullámhosszaiból kiszámítható a vöröseltolódás. A módszer gyengéje, hogy csak erős vonalakkal rendelkező, jó jel-zaj arányú spektrumok esetében alkalmazható.

A legelterjedtebb módszer esetében a spektrumokat sablonokkal keresztkorreláltatjuk. Az 1.4. ábra két ilyen sablont mutat, egy tipikus vörös és egy tipikus kék galaxisét. Az



1.4. ábra. Tipikus galaxis keresztkorrelációs sablonok: vörös – korai típusú (elliptikus) galaxis, kék – késői típusú (spirál) galaxis.

eljárás igen egyszerű: a sablonspektrumot különböző vöröseltolódások mellett hozzáillesztjük a méréshez (ez egyetlen normálási faktor meghatározását jelenti), és kiszámoljuk az eltérést jellemző χ^2 értékét. Az eljárást minden egyes sablonnal megismételjük. A legkisebb χ^2 -et adó sablon-vöröseltolódás párt tekintjük a vizsgált objektum spektroszkópiai típusának, illetve vöröseltolódásának.

A keresztkorreláltatást érdemes logaritmikus hullámhosszrácson megadott spektrumokkal végezni, hiszen ekkor a vöröseltolódás változtatásakor nincsen szükség a spektrumok újra-mintavételezésére, csak a hullámhosszakat kell egy konstanssal eltolni. Az SDSS spektrumai ezért (továbbá megfigyelési okokból is) eleve logaritmikus rácson vannak megadva.

1.3.5. Spektrofotometria

Bár a CCD detektorok csak a beeső fotonok számát tudják meghatározni (hullámhossz függő hatásfokkal), a hullámhossz-tartománybeli felbontottság miatt pontosan meghatározható a beeső energiafluxus.

Jelölje $F_\lambda(\lambda)$ a spektrális fluxussűrűséget, ahol a λ alsó index arra utal, hogy a fluxus hullámhosszegységre vonatkoztatva adjuk meg. A spektrális fluxussűrűség egysége tehát $[F_\lambda(\lambda)] = \text{erg s}^{-1} \text{cm}^{-2} \text{\AA}^{-1}$. Figyelembe véve, hogy képalkotásra használt, szélessávú színszűrőkkel ellátott detektorok a szűrő teljes hullámhossztartományában összegzik a fotonokat, de energia szerint nem képesek differenciálni, az energiafüggetlen fotonfluxus definíciójára a következő adódik:

$$F = \frac{\int F_\lambda(\lambda) r(\lambda) \lambda d\lambda}{c \int r(\lambda) \frac{1}{\lambda} d\lambda}, \quad (1.1)$$

ahol $r(\lambda)$ a szélessávú színszűrő és a CCD detektor kvantumhatásfokának együtteséből meghatározható átviteli függvény, c pedig a fénysebesség. Ebből az ún. szintetikus AB magnitúdó definíció szerint az $m_{AB} = -2,5 \log F - 48,6$ formula szerint számolandó (Fukugita et al. 1996).

A spektrofotometriai kalibráció során az a cél, hogy a spektrum megfelelő skálázásával elérjük, hogy a szintetikus magnitúdók azonosak legyenek a képalkotó detektor által rögzített felvételekből számolható magnitúdókkal. Csillagok esetében elvileg megoldható, hogy a pontterületi függvény alapján számított modellmagnitúdóhoz skálázzuk a spektrumokat. Kiterjedt objektumok esetén valamilyen rögzített apertúrára számított magnitúdót kell nézni, ahol az apertúra megegyezik a spektrum felvételekor használt maszkon levő rés vagy furat méretével.

A spektrofotometriai kalibráció általában nem végezhető el tetszőleges típusú objektum esetére. Ha a csillagokhoz kalibrálunk, az eltéréseket okoz a galaxisok esetében, és fordítva. Gyakran problémát okoz az is, hogy a spektroszkópiai megfigyelések nem „fotometriai minőségű” éjszakákon történnek; a spektrofotometriai kalibráció a légkör változó nyugodtságától⁴ való jelentős függése, valamint a túl fényes ég további problémákat okoz (Abazajian et al. 2009).

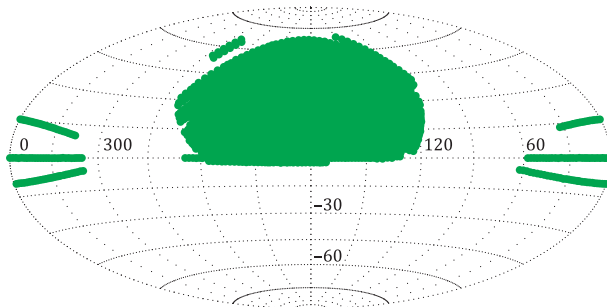
1.4. A Sloan Digitális Égtérképezés (SDSS)

A közelmúlt legjelentősebb égtérképezési projektjét a Sloan Digitális Égtérképezés keretében hajtották végre (Stoughton et al. 2002, Abazajian et al. 2009). A 90-es évek közepén indult projekthez egy speciális, nagy látószögű dedikált távcsövet (Gunn et al. 2006) építettek egy, a kor szintjén csúcstechnológiának számító, 5 különböző színszűrővel (Fukugita et al. 1996) ellátott 120 megapixeles CCD képalkotó kamerával (Gunn et al. 1998) és két üvegszálás spektrográffal (SDSS website 2011). Az égtérkép adataiból Szalay et al. (2000) úttörő jelleggel relációs adatbázis-kezelőn alapuló adattárházat építettek. Az SDSS adatbázisa máig a legátfogóbb csillagászati adatbázis, mely közel 200 millió galaxis fotometriai, és egymillió galaxis kalibrált spektroszkópiai adatait tartalmazza.

Az SDSS távcsöve és kamerája az égbolt csaknem egynegyedét öt év alatt tudta lefényképezni, így megismételt megfigyelésekre csak kevés lehetőség volt. Az égtérképezés egyik részprojektje keretében az ún. *egyenlítői sávot* viszont 20-40 alkalommal figyelték meg. Ez egyrészt időtartománybeli megfigyeléseket jelent, melyen kitűnően azonosíthatók a néhány éves időskálán változó objektumok, másrészt a képek kiátlagolásával közel 24 magnitúdós detektálási mélység érhető el.

Az SDSS projekt folytatásaként a 2000-es évek végén megvalósult a csillagok spektroszkópiájára koncentráló SDSS-II projekt is, valamint jelenleg is folyamatban van az SDSS-

⁴seeing



1.5. ábra. Az égbolt az SDSS égfelmérés által spektroszkópiailag feltérképezett része ekvatoriális koordinátákban. A *lábnym* területe 8378 négyzetfok, ami a teljes égbolt csaknem egynegyede. A nagy, egybefüggő terület az északi galaktikus pólus körüli rész A Tejút síkját a megfigyelésekből teljesen kihagyták. Kép forrása: SDSS.

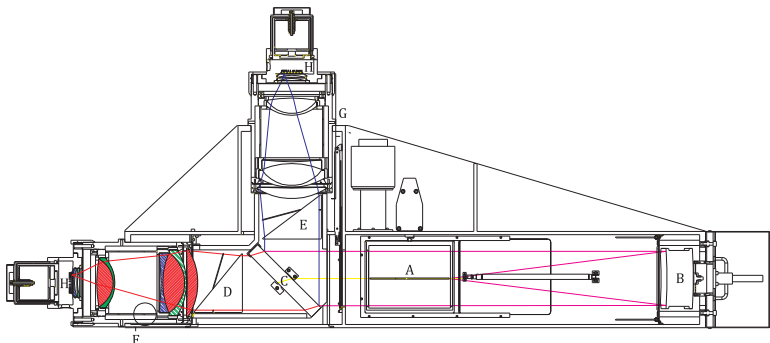
III projekt, mely többek között távoli vörös óriásgalaxisok vöröseltolódásának meghatározására koncentrálna. Ez utóbbi galaxistípus kiváló nyomjelzője az univerzumot alkotó egyik anyagfajta, a sugárzó, látható anyag kozmikus skálán való eloszlásának. Ezekkel a galaxisokkal a dolgozat II. részében részletesen is foglalkozunk.

1.4.1. Az SDSS spektrográfjai

Az SDSS távcsövéhez két azonos felépítésű üvegszálal spektrográf épült, melyek együttesen 640 spektrum rögzítését teszik lehetővé a 3800-9100 Å közötti hullámhossztartományban. A spektrográf felbontása 69 km s^{-1} , ami $R = 1850\text{-}2200$ -nak felel meg. A hullámhossz-kalibráció pontossága kb. 10 km s^{-1} . Az átviteli görbék maximuma 0,2-0,25 körül van.

A spektrumok rögzítését spektrográfonként két-két, a fényképező kameránál használtal azonos típusú, 4 megapixeles CCD végzi el. A spektrográfok egy-egy dichroikus sugárzó elemmel rendelkeznek, melyek a kollimált fénynyalábok útjában helyezkednek el, és a 6000 Å-nál hosszabb hullámhosszakat áteresztik, az ennél rövidebbeket visszatükrözik. A sugárzókkal ketté választott fénynyalábokat spektrográfonként két-két, a kék, illetve a vörös tartományra optimalizált diszperziós elem bontja fel. A felbontott fényt két-két nagy átmérőjű, egyedileg tervezett és épített kamera képzi le a detektorokra. A spektrográfok felépítését az 1.6. ábra szemlélteti.

Spektroszkópiai megfigyelések során a távcső fókuszsíkjába a fényképező kamera helyett egy furatokkal ellátott alumínium lemez (a spektroszkópiai maszk) kerül. A furatok a spektroszkópiára kijelölt objektumok pozíciójának megfelelő helyeken vannak. A maszk behelyezése előtt a spektrográfok üvegszálait manuálisan kell a furatokba helyezni.



1.6. ábra. Az SDSS spektrográfok felépítése. A főbb szerkezeti elemek a következők: A – üvegszálak végződésai, B – kollimátor tükör, C – sugáróztó dichroikus elem, D – vörös grism, E – kék grism, F – vörös kamera, G – kék kamera, H – CCD detektor. Kép forrása: SDSS.

A 640 furat mellett maszkonként további 10 furat is készül vezetécsillagok megfigyelése számára, melyek a távcső pontos pozicionálást biztosítják. Az SDSS spektrográfokhoz tíz teljes üvegszálkészlet készült, azért, hogy az üvegszálakat a maszkokba még a megfigyelések előtti este be lehessen dugdosni, ezzel is megspórolva az értékes megfigyelési időt.

1.4.2. Spektroszkópiai kiválasztási algoritmusok

Mivel a spektroszkópiai megfigyelés nagyon időigényes, a megfigyelendő objektumokat gondosan kell kiválasztani. Az SDSS megfigyelések esetében egy fotometriai megfigyelés 54 másodpercet vesz igénybe, míg a 640 spektrum egyidejű rögzítése minimum háromszor 15 perc expozíciót igényel. Ez az idő a megkívánt jel-zaj aránytól függően hosszabb is lehet.

A kiválasztási algoritmusok⁵ az előzőleg elvégzett, kiredukált fotometriai megfigyelések adatai alapján jelölik ki azokat az objektumokat, melyek spektroszkópiai megfigyelése szükséges és indokolt. Az algoritmusokat sok szempont együttes figyelembe vételével tervezték meg, több különböző objektumosztály lehető legszéleskörűbb megfigyelését tűzve ki célul. Az algoritmusok főként látszólagos magnitúdó, felületi fényesség és színindexek alapján osztályozzák az objektumokat. Az SDSS által spektroszkópiára kijelölt főbb objektumosztályok a következők:

- Fő galaxis minta: $m_r < 17,77$ mag-nál fényesebb, legalább $\mu_{r,50} < 23$ mag arcsec⁻² felületi fényességű extragalaxisokat tartalmazza (Strauss et al. 2002). A dolgozat I. részében erre a mintára koncentrálnunk.

⁵targeting

- Vörös óriásgalaxis⁶ (LRG) minta: összetett kiválasztási feltételek alapján jelöli ki a passzívan fejlődő, igen fényes, vörös galaxisokat (Eisenstein et al. 2001). A minta specialitása, hogy a fő galaxis mintával szemben igen magas, $z \approx 0,38$ vöröseltolódásig teljes, és $z \approx 0,5$ környékén is tartalmaz objektumokat. A vörös óriásgalaxisokkal a dolgozat II. részében külön foglalkozunk.
- Kvazár minta (Richards et al. 2002).
- Csillag minta: galaktikus csillagokból áll.
- SEGUE⁷ minta: A Tejút csillagpopulációjának megismerése szempontjából összeállított minta, mely a ritka csillagtípusokra külön figyelmet fordít (Yanny et al. 2009).

A kiválasztási algoritmusok paramétereinek körültekintő meghatározása különösen fontos, hiszen a túl szűk tartományok hiányos mintát, míg a túl bő tartományok sok feleslegesen megfigyelt objektumot eredményeznek. A nagyobb megfigyelendő minta nyilvánvalóan hosszabb megfigyelési időt igényel. Természetesen ezeket a kiválasztási szabályokat az adatok statisztikai analízisének a legmesszebbmenőkig figyelembe kell venni. Erre jó példát fogunk mutatni a dolgozat II. részében, ahol többek között meghatározzuk a vörös óriásgalaxisok luminozitásfüggvényét is.

1.4.3. Az SDSS spektroszkópiai adatai

Az eredeti SDSS térképezés keretében összesen 1,6 millió objektum spektrumát rögzítették, ebből 929 ezer galaxis. A spektrumok medián jel-zaj aránya $\overline{SN} = 16,9$, a medián vöröseltolódás (a vörös óriásgalaxis mintát is figyelembe véve) $\bar{z} = 0,113$. Közel 50 ezer objektumról megismételt spektroszkópiai megfigyelés is rendelkezésre áll. A vörös óriásgalaxis minta külön figyelmet érdemel, hiszen ez $z \leq 0.38$ vöröseltolódásig teljes (Loh & Strauss 2006), így a nagyskálás szerkezet tanulmányozására kiválóan alkalmas.

A spektrumok jel-zaj viszonya természetesen nem éri el a nagy távcsövekkel készített spektrumok minőségét, de azok a vöröseltolódás meghatározásán túli részletes analízisre mégis alkalmasak. A minta rendkívüli mérete (máig a legnagyobb) a galaxispopulációk paramétereinek igen széleskörű statisztikai elemzését teszi lehetővé. A spektrumok többi fontos paraméterét a 1.4.1. fejezetben már felsoroltuk.

Az SDSS spektrográfiai adatai elérhetőek a *Virtuális Observatóriumon* (VO) keresztül is. A Virtuális Observatórium szabványosított adatformátumok és keresési protokollok segítségével egyesíti az Interneten fellelhető csillagászati adattárházakat, archívumokat. Az SDSS spektrumait publikáló webszolgáltatás a dolgozat szerzőjének munkája, és a III. rész 10. fejezetében kerül bemutatásra.

⁶LRG = Luminous Red Galaxy

⁷SEGUE = Sloan Extension for Galactic Understanding and Exploration

2. fejezet

Galaxisspektrumok analízise

Extragalaxisok esetében három alapvető komponens határozza meg a megfigyelhető optikai spektrumot: a csillagok fényéből származó, főként abszorpciós vonalakat mutató, de alapvetően termikus jellegű *kontinuum*; a környező csillagokból és esetleges aktív magból származó sugárzás által gerjesztett csillagközi gáz diszkrét *vonalas emissziója*, illetve a néhány száz kelvin hőmérsékletű csillagközi *por UV és optikai elnyelése*, közeli infravörös sugárzása.

Amint azt ebben a fejezetben bemutatjuk, a kontinuumból közvetlen információt kaphatunk a galaxist alkotó csillagok sokaságáról, a csillagok kinematikájáról (azaz közvetve a sötétanyag-haló térbeli eloszlásáról is), és a csillagközi por mennyiségéről. Az emissziós vonalak tanulmányozásával a csillagközi porról, a galaxisban jelen levő fiatal csillagokról és az aktív mag jelenlétéről nyerhetünk információt.

A galaxisspektrumok analíziséhez elméleti vagy félempirikus modellekre támaszkodunk. Ezek a modellek a 2000-es évek közepére jutottak el olyan szintre, hogy a megfigyeléseket kellő pontossággal megmagyarázzák. Főként a Sloan Digitális Égtérképezésnek köszönhetően mára annyi galaxisspektrum került rögzítésre, amiből a közeli univerzum galaxispopulációinak alapos statisztikai vizsgálata is megvalósítható.

Mivel elsődleges célunk az, hogy a galaxisok evolúcióját kozmológiai kontextusba téve derítsük fel, a fizikai paramétereket általában a vöröseltolódás függvényeként igyekszünk meghatározni, és elsősorban a galaxispopulációk jellemzőinek időbeli változására koncentrálnak. A másik lényeges szempont, amit az eredmények értelmezésénél figyelembe kell venni, hogy a galaxisok közötti kölcsönhatásokat – elsősorban azok gyakoriságát – egyértelműen a galaxisok közvetlen környezete, a lokális galaxissűrűség határozza meg.

2.1. A galaxisspektrumok eredete

Galaxisspektrumok analíziséhez meg kell ismernünk a spektrumok létrejöttében szerepet játszó folyamatokat, illetve azt, hogy a spektrumok egyes jellegzetességei a galaxist alkotó

komponensek közül melyektől származnak.

2.1.1. A kontinuum emisszió

Mivel az extragalaxisokat alkotó egyedi csillagokat távcsöveink nem képesek felbontani, megfigyeléskor a spektroszkópiai apertúrának megfelelő területéről érkező fényt integrálva detektáljuk, azaz a sok csillag együttes fényét látjuk. Ha a galaxisokban nem lenne por és gáz, akkor tehát pontosan olyan spektrumot várnánk, mint sok csillag spektrumának az összege. Ez egy termikus jellegű, abszorpciós vonalakkal szabdaltspektrum, mely általában nagyon hasonlít valamilyen konkrét spektrális osztályba tartozó csillagspektrumra: olyan csillagokéra, melyek éppen dominálják a galaxis csillagpopulációját. A galaxisok spektrumában megfigyelhető abszorpciós vonalak fő jellemzője, hogy azok a csillagok nagy sebességű (jellemzően $150\text{-}250\text{ km s}^{-1}$) mozgása miatt jelentő Dopplerkiszéledést mutatnak.

Galaxisok spektrumának sztelláris kontinuum elméleti alapokon is jól modellezhető. Ennek bemutatására a 2.2. fejezetben még visszatérünk.

2.1.2. Nebuláris emisszió

Galaxispektrumok szembeötlő jellegzetességei a sokszor rendkívül erős emissziós vonalak. Ezeket a galaxisokban levő gázfelhők bocsájtják ki a galaxis csillagaiból, vagy aktív magjából származó gerjesztő sugárzás hatására. A galaxisok között ezért alapvető különbséget lehet tenni az emissziós vonalak erőssége és a különböző elemek gyakorisága, illetve ionizációs állapota alapján. A különböző elemek ionizációs szintjeinek arányaiból megbecsülhető az ionizáló sugárzás spektruma. Ez utóbbi alapján eldönthető, hogy az elsődleges sugárzás forrása a galaxis fiatal csillagpopulációja, vagy az aktív galaxismag.

Irreguláris és spirális galaxisokban általában aktív csillagkeletkezés zajlik, mely szám szerint viszonylag kevés nagy tömegű csillag megjelenésével jár, ám ezek a sugárzási mérleg szempontjából mégis igen jelentősek. Az ilyen nagy tömegű, de rövid életű csillagok által kibocsátott *termikus ultraibolya sugárzás* ionizálja, illetve gerjeszti a galaxis gázfelhőit. Ionizáció elsősorban a csillagokhoz közeli tartományokban történik, így jönnek létre az ún. *HII régiók*, melyeket főként ionizált hidrogén alkot. Fontos kiemelni, hogy a *HII régiók* többnyire sűrűbbek a galaxis egyéb részein található gázfelhőknél, hiszen ezek a tartományok közvetlenül közrefogják a csillagkeletkezési zónákat, melyek eleve összesűrűsödött gázból alakulhattak csak ki. Összefoglalva: a csillagkeletkezési régiókat mindig *HII régiók* veszik körül. Egy galaxis legfiatalabb csillagai mindig sűrű gáz- és porfelhőbe burkolóznak.

Az aktív galaxismagok a galaxisban található központi fekete lyuk által lassan elnyelt anyag gravitációs energiáját (feltehetően a fekete lyuk forgási energiáját is felhasználva)

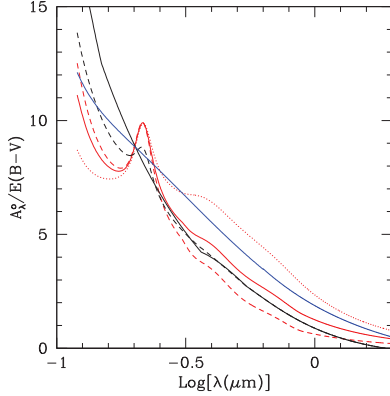
alakítják meglehetősen jó hatásfokkal sugárzássá. A magba hulló anyag a fekete lyuk körül pályára áll, és az akkréciós korong mentén befelé spirálozik. A lassú bespirálozás annak köszönhető, hogy az akkréciós korongban jelen levő dinamikai (feltehetően magneto-hidrodinamikai eredetű) instabilitások a gázt viszkózussá teszik. A viszkózus gáz így a bespirálozás folyamata során felforrósodik, és az ultraibolya tartományban domináns, nagyjából termikus sugárzással sugároz. Az aktív mag környezetében levő gázfelhők gerjesztését és ionizációját mégsem ez a sugárzás dominálja, hanem a forgó fekete lyuk pólusai irányába kiinduló relativisztikus anyagáramlások, az ún. *jetek* sugárzása. A modellek szerint a jetek pozitronokból és elektronokból álló anyagáramok. Az elektron-pozitron párokat a fekete lyuk eseményhorizontja közelében az akkréciós korong által táplált extrém mágneses tér gyorsítja relativisztikus sebességre. A felgyorsított részecskék a mágneses tér erővonalai mentén kifelé spiráloznak, így szinkrotron sugárzást bocsátanak ki. Az alap szinkrotronfrekvencián túl rengeteg felharmonikus is megjelenik, így az eredő sugárzás gyakorlatilag egy folytonos, hatványfüggvény jellegű spektrummal rendelkező elektromágneses sugárzás lesz. Ez a sugárzás ionizálja és gerjeszti a környező gázfelhőket, melyek így karakterisztikus emissziós spektrumot mutatnak. Az emissziós vonalak szélessége attól függ, hogy az azokat kibocsátó gázfelhő milyen sebességgel kering a központi mag körül.

Baldwin et al. (1981) megmutatták, hogy ún. *diagnosztikai diagramok* (a cikk szerzőiről BPT-diagramnak nevezett ábrák) segítségével hogyan lehet meghatározni a gázfelhőket gerjesztő elsődleges sugárzás jellegét, ezáltal megmondani annak eredetét. Elméleti úton kiszámítható, hogy egy adott spektrumú ionizáló sugárzás hatására egy gázfelhő milyen egyensúlyi ionizációs állapotba kerül. Ebből közvetlenül megadható a különböző karakterisztikus vonalak erősségének aránya. Az eljárást megfordítva a vonalak arányaiból meghatározható az elsődleges ionizáló sugárzás jellege, azaz eldönthető, hogy a galaxisban a csillagkeletkezés, vagy az aktív fekete lyuk sugárzása-e a domináns. Előre utalunk a 4.2 ábrára, ahol SDSS galaxisok BPT-diagramját ábráztuk.

2.1.3. Az intersztelláris por hatása

A galaxisok megfigyelhető spektrumát nagyban befolyásolja a csillagközi por. Extragalaxisok esetében meg kell különböztetni a távoli galaxisban található port, illetve azt a port, amin a távoli galaxisból érkező fény a Tejútba belépve halad át. A Tejút által okozott vörössedésre jól alkalmazható modellekkel rendelkezünk, hiszen infravörös megfigyelésekkel a különböző galaktikus irányokba a por mennyisége jól megbecsülhető (Schlegel et al. 1998, O'Donnell 1994).

Extragalaxisok esetében a fő problémát az jelenti, hogy a bennük levő por összetétele nem ismert. Calzetti et al. (1994) felállítottak egy empirikus extinkciós formulát, és kimutatták, hogy az extragalaxisokra általában a Tejútétól eltérő porösszetétel jellemző. A



2.1. ábra. Néhány, a közeli Univerzum galaxisaira jellemző extinkciós görbe. A piros görbék a Tejútbeli por hullámhosszfüggő csillapítását ábrázolják három különböző irányban. A szaggatott fekete görbén a Nagy Magellán Felhő, a folytonos fekete görbén a Kis Magellán Felhő küllőjének extinkciós görbéje látható. A kék görbe a csillagkeletkezési régiókra jellemző elméleti törvény alapján lett felrajzolva (Cardelli et al. 1989, Gordon & Clayton 1998, Misselt et al. 1999). Forrás: Calzetti (2001).

2.1. ábrán tipikus extinkciós görbék láthatók. A Tejút esetében 2175 Å-nél jól megfigyelhető egy csúcs, mely eredete pontosan nem ismert, de valószínűleg grafit szemcsék és policiklikus aromás szénhidrogén molekulák¹ (PAH) keverékének tulajdonítható. A csillagközi por tulajdonságainak részletes ismertetését jelen dolgozat keretei nem teszik lehetővé, de ajánljuk Calzetti (2001) összefoglaló munkáját, illetve kiemelünk két közelmúltban készült tanulmányt, melyek az SDSS adatai alapján, nagy minta segítségével tanulmányozták a témát (Yip et al. 2010, Wild et al. 2011).

Az extinkció mértékét legtöbbször a τ_λ optikai mélység megadásával jellemezzük, ahol a λ index a hullámhosszfüggésre utal. A szokásos definíció szerint az optikai mélység az intenzitáscsökkenés exponense: $I_\lambda = I_{\lambda,0} e^{-\tau_\lambda}$. Sokszor az indexbe egy szélessávú színszűrő betűjele kerül, például τ_V a látható fény tartományban vett optikai mélységet jelenti. Az extinkció mértékét megadhatjuk magnitúdó egységekben is, ekkor jele $A_\lambda = -2,5 \tau_\lambda \log_{10} e$.

A por az észlelt optikai színindexek vörösödését (számértékük növekedését) okozza, így az extinkció jellemzésére legtöbbször a *színindextöbblet* (avagy szelektív extinkció) nevű mennyiséget használjuk, melyet $E(B-V)$ -vel jelölünk, és a B és V Johnson-magnitúdók segítségével a következőképpen definiálunk:

$$E(B-V) = (B-V)_{\text{megfigyelt}} - (B-V)_{\text{eredeti}} \quad (2.1)$$

¹PAH = polycyclic aromatic hydrocarbons

Szokás még bevezetni az $R(V) = A(V)/E(B - V)$ mennyiséget is. Ez a mennyiség korrelál az extinkciót okozó porszemcsék méretével. Ahhoz, hogy $E(B - V)$ -ből τ_λ -t megkapjuk, a hullámhosszfüggő extinkciós törvény ismeretére van szükség.

2.2. Populációsintézis-modellek

Amennyiben galaxisok spektrumát elméleti úton kívánjuk modellezni, a modellnek a korábban ismertetett valamennyi galaxis-összetevőről számot kell adnia. A sztellárkontinuum leírására az ún. *populációsintézis-modellek* vállalkoznak. A legújabb modelleket megfelelő pormodellekkel kiegészítve a megfigyelt galaxispektrumok kontinuumát jól reprodukálható.

A populációsintézis-modellek galaxisok spektrumát egyedi csillagok spektrumaiból építik fel. Ehhez feltételezik, hogy a csillagok egyszerre nagy számban, nagyjából homogén összetételű gázfelhőkből születnek, és a fiatal csillagok tömegeloszlása jól leírható egy univerzális *kezdeti tömegeloszlás-függvénnyel*. Ha a különböző tömegű csillagok spektrumának időbeli fejlődését ismerjük, akkor ez alapján az egy időben keletkezett teljes csillagpopuláció együttes spektrumát és annak időfejlődését is modellezni tudjuk. Galaxisok esetében beszélhetünk a *csillagkeletkezési történetről*, mely mennyiség a kozmikus idő függvényében adja meg a múltbeli csillagkeletkezési rátát. Ez utóbbit konvolválva az egykorú csillagpopulációk időfüggő spektrumával komplett galaxisok spektrumát is modellezni tudjuk. A továbbiakban részletesen bemutatjuk ezt az eljárást, mert ennek ismerete elengedhetetlen, hogy a galaxisok fizikai paramétereinek meghatározására kidolgozott eljárásokat megértsük.

A konkrét jellemzőik részletes tárgyalása nélkül felsorolunk néhány széles körben elterjedt modellt: PÉGASE (Fioc & Rocca-Volmerange 1997), GALAXEV (Bruzual A. & Charlot 1993, Bruzual & Charlot 2003), Maraston (Maraston 2005, Maraston et al. 2009), MILES (Vazdekis et al. 2010).

2.2.1. Csillagspektrumok modellezése

Egyedi csillagok spektrumának modellezése az asztrofizika egyik sikeres területe. A modellek általában több részből tevődnek össze: Egyfelől szükséges modellezni az adott tömegű és fémességű csillag energiatermelő folyamatainak, másfelől a felszín felé irányuló energiátranzportnak az időbeli alakulását. E kettőből meghatározható a fotoszféra hőmérséklete és sugara, így a csillag minden időpontban elhelyezhető a Hertzsprung–Russell-diagramon. Egy adott csillag a tömegétől és fémességétől függően fejlődése során egy igen jól definiált utat jár be a HRD-n, ezt az utat *csillagfejlődési trajektóriának* nevezzük. A csillag HRD-n levő pozíciója, valamint a fotoszféra összetételének pontos ismeretében a csillag

spektruma a Fowler–Milne-elmélet szerint életének bármely időpontjában meghatározható.

A csillagok evolúciós trajektóriáit leíró modellek közül elterjedtebbek: Padova (Alongi et al. 1993, Bressan et al. 1993, Fagotto et al. 1994a;b, Girardi et al. 1996), Geneva (Schaller et al. 1992, Charbonnel et al. 1996; 1999). A különböző csillagtypusok elméleti vagy empirikus spektrumait populációsztézis céljára összefoglaló atlaszok közül kiemelendők: Pickles (Pickles 1985), Kurucz (Kurucz 1992), BaSeL (Lejeune et al. 1997), STELIB (Le Borgne et al. 2003).

A továbbiakban az \mathcal{M} tömegű, Z fémességű és t korú csillag által kibocsátott fény spektrális energiasűrűségét $S_\lambda^*(\mathcal{M}, Z, t)$ -vel jelöljük. Az alsó indexbe írt λ azt jelöli, hogy a spektrális energiasűrűséget hullámhosszegységre vonatkoztatjuk. Ezt a mennyiséget $1\mathcal{M}_\odot$ egységre vonatkoztatjuk, így dimenziója $[S_\lambda^*] = \text{erg s}^{-1} \text{Å}^{-1} \mathcal{M}_\odot^{-1}$.

Nagy tömegű csillagok spektrumát általában a termikus pulzáció szakaszáig érdemes követni, a szupernóva robbanást a bemutatandó modellbe nem építjük be. Általában is igaz, hogy a csillagok nagyon rövid ideig tartó tranzien viselkedésével galaxispektrumok modellezésekor nem érdemes foglalkozni. A modell egyszerűsége érdekében úgy tekintjük, hogy egy nagy tömegű csillag spektruma az életkorát meghaladó t értékekre zérus. Törpecsillagok modellezését általában a héliummag-villanásig szokás követni, hiszen ez az idő már gyakran meghaladja az Univerzum jelenlegi korát. A Napnál kisebb tömegű csillagok spektrumát lehetséges egy konstans fősorozati spektrummal helyettesíteni.

2.2.2. Egykorú csillagpopulációk

Ha egy galaxisban az interisztelláris gáz valamilyen perturbáció (mely lehet a spirálkarokat létrehozó sűrűség hullám, vagy akár közeli elhaladás egy másik galaxis mellett) hatására gravitációs kollapszust szenved, és a gáz az összehúzódás során kellően hatékonyan tud hűlni ahhoz, hogy az összehúzódás ne álljon meg a felmelegedés hatására, akkor az összehúzódó gázfelhőből csillagok jönnek létre.

A Jeans-elmélet alapján az így összeomló gázfelhők minimális mérete néhány ezer nap-tömeg. Ez azt jelenti, hogy a csillagok nem egyenként születnek, hanem a jó közelítéssel homogén kémiai összetételű felhőből egyszerre sok csillag keletkezik. A megfigyelések szerint az egy gázfelhőből létre jövő csillagok tömege egy univerzális tömegeloszlást követ. Ezt az eloszlást nevezzük *kezdeti tömegeloszlás-függvénynek*. Ez egy valószínűségi eloszlás, jele $\xi(M)$. A kezdeti tömegeloszlás-függvényre az első közelítést Salpeter (1955) adta, mára leginkább a Chabrier (2003) által kidolgozott formula terjedt el a hétköznapi használatban.

A kezdeti tömegeloszlás-függvény univerzalitása vitatéma a szakirodalomban. A megfigyelések alapján közeli (tehát késői kozmikus időben létrejött) csillagpopulációkra a kezdeti tömegeloszlás-függvény nagyon hasonló, de vannak arra utaló jelek, hogy a korai Uni-

verzumban a nagy tömegű csillagok részaránya a mainál jóval nagyobb volt. Amikor a galaxisok spektrumának modellezését tárgyaljuk, akkor kis vöröseltolódásokra korlátozódunk, és a kezdeti tömegeloszlás-függvényt az egyszerűség kedvéért univerzálisnak tekintjük.

Az egyedi csillagok spektrális fejlődését leíró S_λ^* és a $\xi(\mathcal{M})$ eloszlásfüggvény segítségével felírható egy adott Z fémességű gázfelhőből, azonos időpontban keletkezett csillagok $1\mathcal{M}_\odot$ -re vonatkoztatott együttes spektruma:

$$S_\lambda(Z, t') = \int_0^\infty S_\lambda^*(\mathcal{M}, Z, t') \xi(\mathcal{M}) d\mathcal{M}, \quad (2.2)$$

ahol t' a csillagkeletkezés időpontja óta eltelt idő. Az egyedi csillagok spektrumai helyett az *egykorú csillagpopulációk spektrumai* képezik a populációsintézis-modellek alapját.

Az egykorú csillagpopulációk nem csupán elméleti jelentőségűek, hiszen a nyílt- és gömbhalmazok pontosan egy időben keletkezett csillagokból állnak. Ugyanakkor megjegyezzük, hogy a legtöbb ilyen modell szigorúan véve mégis csak elméleti jellegű, hiszen protocsillagokkal nem számol, és úgy tekinti, hogy minden csillag a $t = 0$ időben kezdi meg életét a fősorozaton. Tudjuk, hogy a legnagyobb tömegű csillagok már szupernóvaként be is fejezik életüket, mire a törpecsillagok egyáltalán elérik a fősorozatot.

2.2.3. A csillagkeletkezési ráta és a kémiai evolúció

Az előzőek mintájára összeállíthatjuk galaxisok spektrumát is. Ehhez két új mennyiséget kell bevezetnünk.

Az egyik a *csillagkeletkezési ráta*, mely megadja, hogy a galaxisban mennyi csillag keletkezett a kozmikus idő függvényében. A mennyiség jele $\psi(t)$, és leggyakrabban $[\psi(t)] = \mathcal{M}_\odot \text{ év}^{-1}$ egységekben szokás megadni. A csillagkeletkezési ráta jellemző értéke Tejút típusú spirálgalaxisokban $\psi \simeq 0,1 - 1,0\mathcal{M}_\odot \text{ év}^{-1}$, csillagotó galaxisokban azonban elérheti a $\psi \simeq 10^4 \mathcal{M}_\odot \text{ év}^{-1}$ értéket is. Ezzel szemben vörös galaxisokban a csillagkeletkezés gyakorlatilag zérus (Kauffmann et al. 2003a).

A másik bevezetendő mennyiség a *kémiai evolúció* paramétere, jele $\zeta(t)$. Ez azt adja meg, hogy az idő előrehaladtával hogyan változik a galaxis gázfelhőiben a héliumnál nehezebb elemek mennyisége, azaz a csillagképzésre alkalmas csillagközi gáz fémessége. Ősi galaxisoknál ez az érték nyilvánvalóan csaknem zérus volt, hiszen a legelső csillagok csak primordiális gázból jöhettek létre. A fémesség időbeli növekedését a nagy tömegű csillagok szupernóvaként történő felrobbanásakor szétszóródó nehéz elemek feldúsulása okozza. Ez alapján a fémességnek időben monoton nőnie kell. Egyes elméletek szerint előfordulhat, hogy a galaxis gázanyagot akkretál a környező galaxisközi térből. Ez a primordiális gáz ellene hat a szupernóvak okozta nehézelem-feldúsulásnak, így a feldúsulás ütemét lassíthatja.

Fontos kiemelni, hogy a kémiai evolúció hatása nem jelenik meg azonnal a csillagok fényében. A fősorozati csillagok konvektív zónája nem elég mély, hogy a magban létrejött elemeket a felszínre szállítsa. A megnövekedett fémesség ezért csak az újonnan született csillagok spektrumában figyelhető meg.

2.2.4. Galaxisok spektruma

A csillagkeletkezési rátát és a kémiai evolúciót is figyelembe véve a galaxis spektruma az egykorú csillagpopulációk spektrumából kiindulva a következő konvolúciós integrállal írható fel:

$$F_{\lambda}(t) = \int_0^t \psi(t-t') S_{\lambda}(\zeta(t-t'), t') dt'. \quad (2.3)$$

A formulában t a kozmikus idő, a csillagkeletkezési rátát ennek függvényében érdemes megadni.

Mivel a galaxisok kémiai evolúciójának részletei egyelőre nem pontosan ismertek, továbbá az effektus kisebb, mint a rendelkezésünkre álló modellek jelenlegi pontossága, gyakran a kémiai evolúciót külön nem modellezzük, hanem egy átlagos $\zeta(t) = \bar{\zeta}$ értéket használunk. Ezzel a 2.3 integrál a következőre egyszerűsödik:

$$F_{\lambda}(t) = \int_0^t \psi(t-t') S_{\lambda}(\bar{\zeta}, t') dt'. \quad (2.4)$$

2.2.5. A por hatásának modellezése

Ha a galaxisok spektrumának sztellárkontinuumát kívánjuk megérteni, a csillagok hozzájárulásán kívül mindenképp figyelembe kell venni a galaxisokban levő por extinkciós hatását is. Ebbe a modellbe nem értjük bele a Tejútban levő por által okozott vörösödést, mert ez utóbbira könnyen tudunk korrigálni.

Az interszstelláris por főleg szénből, és annak vegyületeiből összeállt, néhány száz atomot tartalmazó molekulákból áll. Legfontosabb tulajdonságuk, hogy az elnyelt rövid hullámhosszú (főleg ultraviolet) sugárzást a molekulák rezgési és vibrációs módusaiknak megfelelő infravörös hullámhosszakon sugározzák ki. Mivel az infravörös sugárzás nem tudja ionizálni a környezetében levő gázfelhőket, a sugárzás könnyen kijut a porfelhőből. Ez a mechanizmus teszi lehetővé, hogy a port tartalmazó gázfelhők gyorsan hűljenek, és csillagkeletkezési régiókat alakítsanak ki. Amikor tehát galaxisok portartalmát próbáljuk modellezni, külön figyelmet kell fordítani arra, hogy a fiatal csillagok poros környezetben jönnek létre (Calzetti et al. 1994, Calzetti 2001, Silva et al. 1998, Charlot & Fall 2000).

Amennyiben a célunk csupán a közeli ultravioleta és az optikai hullámhossztartomány spektrumának modellezése, úgy elegendő a por extinkciós hatására koncentrálni, és nem kell külön figyelembe venni az extinkció mellett megjelenő infravörös emissziót. A por

infravörös emissziójának modellezésével számos publikáció foglalkozik (Silva et al. 1998, da Cunha et al. 2008).

Megállapítást nyert (Calzetti 2001, Wild et al. 2011), hogy univerzális pormodellről nem lehet beszélni, hiszen a por összetétele galaxisról galaxisra változhat, ami maga után vonja az extinkció hullámhosszfüggőségének változékonyságát is. Ennek ellenére közeli ultrabolya és optikai hullámhosszakra megfelelően használható egy általános formula.

Az extinkció hullámhosszfüggésének megállapításán túl fontos kérdés a por eloszlása a galaxisok belül. Nyilvánvaló, hogy homogén eloszlású porral a galaxisok nem modellezhetők. A modell kezelhetősége szempontjából viszont fontos, hogy a por geometriájának leírása egyszerű legyen. Charlot & Fall (2000) abból a megfontolásból, hogy a por hatása mindig jóval jelentősebb a fiatal csillagok környezetében, mint a galaxis többi részén, a por pontos geometriáját külön nem modellezi, hanem a fiatalabb csillagok esetében jelentősebb extinkciót tételez fel. Ezzel szemben pl. Silva et al. (1998) bonyolult geometriájú porral dolgoznak.

Bruzual & Charlot (2003) az alábbi egyszerű pormodellt alkalmazta nagy sikerrel.

$$\tau_{\lambda}(t') = \begin{cases} \tau_V(\lambda/5500 \text{ \AA})^{-0.7}, & \text{ha } t' \leq 10^7 \text{ év} \\ \mu\tau_V(\lambda/5500 \text{ \AA})^{-0.7}, & \text{ha } t' > 10^7 \text{ év} \end{cases} \quad (2.5)$$

A formula egykorú csillagpopulációra vonatkozik, τ_V a látható fényre vonatkozó optikai mélység, t' a csillagkeletkezés óta eltelt idő, μ pedig egy paraméter, ami azt szabályozza, hogy a fiatal csillagok körül mennyivel több a por, mint általában a galaxisban. A modell nagy előnye – főként a bonyolult geometriával operáló modellekkel szemben –, hogy ha a modellt mért spektrumokhoz illesztjük, csupán két szabad, illesztendő paraméterrel rendelkezik (τ_V és μ).

2.2.6. A sebességszperzió

Mivel a csillagok a galaxisokon belül jelentős sebességgel mozognak, az abszorpciós vonalak kiszélesedésében a látóiránnyal párhuzamos sebességkomponensből eredő Dopplereffektus dominál. Ezt az effektust mindenképpen figyelembe kell venni a spektrum értelmezésekor, és a kontinuum illesztésekor. A σ_v sebességszperzió tipikus értéke $\sigma_v \approx 150\text{-}250 \text{ km s}^{-1}$ között van.

2.3. Galaxispektrumok redukciója

Amikor galaxisok fizikai paramétereit kívánjuk meghatározni, úgy járunk el, hogy csillagpopulációsintézis-modelleket illesztünk a mérésekre, illetve meghatározzuk a spektrumvonalak erősségét. Az illesztés előtt feltételezzük, hogy a mért spektrum spektrofotomet-

riai kalibrációja helyes, és a Tejút által okozott vörösödéésre a spektrum már korrigálva van.

Első dolgunk az, hogy a spektrumot szétválasszuk a csillagokból és a porból, valamint a nebuláris emisszióból származó részekre. Ezt az első lépést *kontinuumillesztésnek* nevezzük. Ekkor a jól ismert emissziós vonalak tartományát figyelmen kívül hagyva keressük meg a mérésekre legjobban illeszkedő sztellar-kontinuum-modellt. Ez az illesztés többféle matematikai módszerrel is elvégezhető, ezek közül két alapvető eljárás kerül majd ismertetésre. Az emissziós vonalak paramétereit a kontinuum eredeti spektrumából való levonása után tudjuk meghatározni.

Látni fogjuk, hogy a csillagokból származó kontinuum pontos megbecslése nagyon fontos az emissziós vonalak erősségének pontos megméréséhez is, hiszen általában minden fontos emissziós vonal helyén a kontinuum valamilyen mértékű abszorpciót mutat.

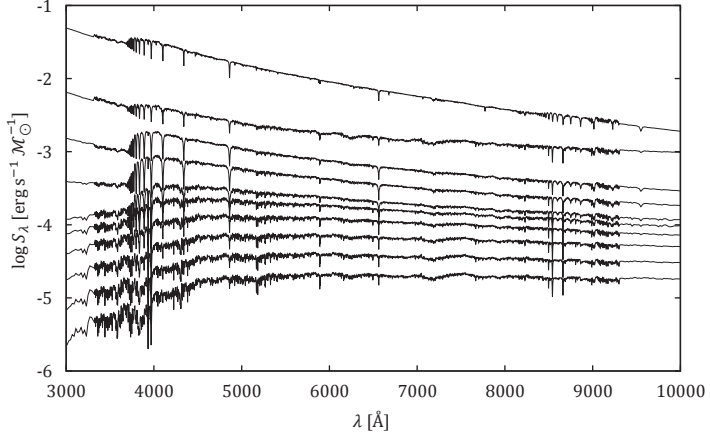
2.3.1. Kontinuumillesztés

A kontinuum illesztésekor az emissziós vonalak, esetleges ég vonalak vagy hibás mérési pontok környezetét ki kell maszkolni, és figyelmen kívül kell hagyni. A kontinuumillesztés célja, hogy olyan populációsintézis-modellből származó elméleti spektrumot találjunk, amely legjobban illeszkedik a méréshez.

Korábban láttuk, hogy a populációsintézis-modelleknek számos szabad paraméterük van, melyek együttes illesztése túl sok szabad paramétert jelentene. A $\psi(t)$ csillagkeletkezési ráta rögtön egy teljes szabad függvényt jelent, érthető tehát, hogy ennek pontos meghatározására egyetlen spektrumból nincsen lehetőség. A kontinuumillesztés mikéntjére a szakirodalomban csupán néhány módszer terjedt el, ezek közül kettőt mutatunk be, melyek a legtöbb esetben sikeresen alkalmazhatók. Az első módszer az illesztendő paraméterek számát úgy redukálja, hogy néhány előre generált sablon lineárkombinációjaként állítja elő a kontinuumot, míg a másik Bayes-statisztikai módszerrel igyekszik megbecsülni a fizikai paraméterek valószínűségi eloszlását.

A nem-negatív együtthatós lineárkombinációk módszere. Ez a módszer azon múlik, hogy az egykorú csillagpopulációk spektrumának megváltozási üteme a populáció öregedésével egyre lassul. A kezdeti néhány millió év alatt a spektrum annyit változik, amennyit egy idős populáció néhány milliárd év alatt. A változás sebessége nagyjából exponenciálisan csökken az idővel, ahogy azt a 2.2. ábra is szemlélteti.

A spektrumok változásának ezt a tulajdonságát kihasználhatjuk úgy, hogy ahelyett, hogy az egész $\psi(t)$ csillagkeletkezési történetet szabad paraméterként illesztjük, csak néhány előre megadott korú populációból kombináljuk ki a megfigyelt galaxis kontinuumát.



2.2. ábra. Egy $Z = Z_{\odot}$ fémességű, egy időben keletkezett csillagokból álló populáció spektrumának változása az időben. A görbéknek megfelelő populációk kora (felülről lefelél): 5×10^6 , $2,5 \times 10^7$, 1×10^8 , $2,9 \times 10^8$, $6,4 \times 10^8$, 9×10^8 , $1,4 \times 10^9$, $2,5 \times 10^9$, 5×10^9 és $1,1 \times 10^{10}$ év. Fiatal populációk esetén jól megfigyelhető a fényes csillagok dominanciája. Látható továbbá a fényesség-tömeg arány drasztikus csökkenése, és a 4000 Å-ös törés erősödése a korrall.

Legyen $\psi(t)$ a következő alakú:

$$\psi(t) = \sum_{i=1}^k a_i \delta(t - t_i), \quad (2.6)$$

ahol $\delta(t)$ a Dirac-delta függvény, az a_i együtthatók pedig mind pozitívak. A konvolúció linearitása miatt ilyen alakú csillagkeletkezési történetet választva a galaxis spektruma (2.4) alapján szintén lineárkombinációként írható fel:

$$F_{\lambda}(t_0) = \sum_{i=1}^k a_i S_{\lambda,i}(\bar{Z}), \quad (2.7)$$

ahol az $S_{\lambda,i}$ -k különböző korú csillagpopulációk naptömegre normált spektrumai, így az a_i együtthatók naptömeg dimenziójúak. F_{λ} argumentumában a rögzített t_0 időpont arra utal, hogy a galaxist csak egyetlen, fix időpontban tudjuk megfigyelni. A fentiekben elmondottak szerint az $S_{\lambda,i}$ populációk korát úgy érdemes megválasztani, hogy az időlépések között a spektrum megváltozásának nagysága körülbelül azonos legyen. Tremonti et al. (2004) ilyen megfontolások alapján a következő korú populációkat választotta: $t_i = (0.005, 0.025, 0.100, 0.290, 0.640, 0.900, 1.4, 2.5, 5.0, \text{ valamint } 11.0) \cdot 10^9$ év.

Felmerül még a kérdés, hogy miként érdemes a fémesség \bar{Z} értékét meghatározni. Bár ez egy folytonosan változó paraméter, általában csak néhány diszkrét értékkel szokás dolgozni. Ennek az az oka, hogy a modelleket nagyon nehéz a Napétól ($Z = Z_{\odot}$) eltérő fémességű csillagokhoz kalibrálni, hiszen ezekből a Tejútban igen kevés van. A legtöbb csillagpopulációsintézis-modell néhány rögzített fémességű alapkönyvtárat tartalmaz, melyek között bizonyos megkötésekkel lehet interpolálni.

Megjegyezzük még, hogy (2.6)-ben a delta függvény helyett használható lenne például exponenciálisan lecsengő függvény is. A delta függvény használata mellett szól az egyszerűsége, és az a fizikai tény, hogy a galaxisokban a jelentős csillagkeletkezés ún. *csillagontó periódusok* során történik. Ilyenkor a csillagkeletkezési ráta az átlagos spirálgalaxisoknál tapasztalható rátának $10^2 - 10^4$ -szerese, viszont ezek a periódusok kozmikus skálán nézve igen rövid ideig tartanak.

Az elmondottak alapján tehát a galaxis kontinuumának illesztése a következőre illesztési problémára vezet: Vesszük a megmért és kalibrált, az $S_{\lambda,i}$ spektrumsablonokkal azonos hullámhossz rácsra konvertált spektrumot², majd kimaszkoljuk az emissziós vonalakat. Ezek után valamilyen kezdeti értékre állítjuk a por modelljének paramétereit, valamint a csillagok sebességszórását. Az előre legenerált $S_{\lambda,i}$ spektrum sablonokra alkalmazzuk a sebességszórás okozta kiszélesedést, majd a por okozta extinkciót. Így az $S'_{\lambda,i}$ módosított sablonokhoz jutunk. Ezek nem lineáris kombinációjából kell előállítani a megfigyelt spektrumot, melyek a következő optimalizációs problémával írhatunk le:

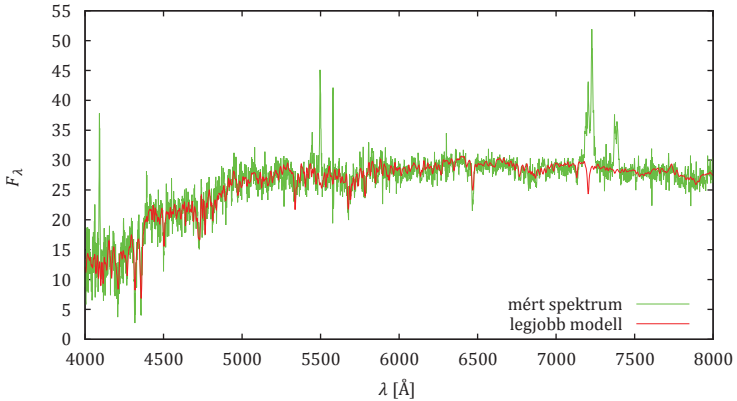
$$\arg \min_{a_i \geq 0} \sum_{\lambda} w_{\lambda} \left(F_{\lambda} - \sum_i a_i S_{\lambda,i} \right)^2 \quad (2.8)$$

Itt nagyon lényeges az $a_i \geq 0$ feltétel. Erre azért van szükség, mert csak csillagpopulációk összegének van fizikai értelme, különbségüknek nincsen. A λ -ra való összegzést diszkrét szummaként írtuk, hiszen itt egy valódi mérésről van szó, ahol a fluxust egy hullámhossz-rácson adjuk meg. A w_{λ} a hullámhosszrácspontjaihoz rendelt súly, értéke zérus, ha az adott hullámhossz valamiért ki lett maszkolva (például azért, mert ott egy emissziós vonal található). Egyéb esetben a súlyok értéke számítható például a mérési hibából is.

A felvázolt módszert a nem negatív legkisebb négyzetek módszerének nevezik, és a megoldására létezik bizonyítottan konvergens algoritmus (Lawson & Hanson 1974). Sajnos ettől függetlenül a por paramétereit és a sebességszórásértéket egy külső ciklusban, külön kell optimalizálni.

A 2.3. ábra egy SDSS spektrumot, és az illesztett kontinuumot mutatja. Megfigyelendő,

²Itt feltesszük, hogy a galaxis vöröseltolódását már korábban meghatároztuk, ezt a műveletet nem a kontinuum illesztések céljára elvégezzük. A vöröseltolódás általában egyszerűen meghatározható különböző sablonokkal vett keresztkorrelációval.



2.3. ábra. Egy SDSS galaxis mért spektruma, és a nem negatív legkisebb négyzetek módszerével illesztett modellspektrum.

hogy az emissziós vonalak alatt majdnem mindenütt megjelenik a kontinuumhoz tartozó abszorpció. Ha a vonal erősségét naiv megoldással úgy próbálnánk meghatározni, hogy lineáris háttérrel (kontinuumot) tekintünk, és egy Gauss-függvényt illesztünk, akkor a vonalban levő fluxust alulbecsülünk.

A kontinuumillesztés itt megadott módszere elterjedt, és az emissziós vonalak ekvivalens szélességének meghatározásához könnyen használható. Felhívjuk azonban a figyelmet arra, hogy az illesztett modell paramétereiből a galaxis paramétereit (kor, fémesség) csak nagy hibával lehet megbecsülni. Sokkal pontosabb mérésre adnak módot a későbbiekben ismertetésre kerülő abszorpciós vonalindexek.

A sztochasztikus csillagképződési modelleken alapuló bayesi módszer. A korábban ismertetett, nem negatív együtthatós lineáris kombináció alapuló kontinuumillesztési módszer ugyan megfelelően és gyorsan használható arra, hogy az emissziós vonalak megmérése céljából a kontinuumot a megfigyelt spektrumból levonjuk, de a galaxist alkotó csillagok fizikai paramétereiről (fémesség, átlagos kor, csillagkeletkezési történet) nem kapunk elegendő információt. Ennek legfőbb oka az, hogy optikai mérések esetében az említett paraméterek tere erősen degenerált: nagyon hasonló spektruma van egy öreg csillagokat tartalmazó galaxisnak, és egy nem annyira öreg, de nagy fémességű csillagokat tartalmazó galaxisnak.

Amennyiben csak optikai hullámhossztartományban készült mérések állnak rendelkezésünkre, kénytelenek vagyunk ezekkel a degenerációkkal együtt élni. Emiatt olyan statisztikai módszert kell találnunk, mely – ha a fizikai paraméterek pontos értékét nem is –

legalább lehetséges eloszlásukat megadja. Az eloszlások ismeretében könnyen kiválaszthatjuk a paraméterek néhány legvalószínűbb kombinációját, illetve megadhatjuk azok bizonytalanságát. A módszer lényege röviden abban áll, hogy feltérképezzük a modell paramétereinek terét, és minden modellt hozzápróbálunk a méréshez. Egy paraméterkombinációhoz az illeszkedés jószágától függően rendelünk valamilyen valószínűséget. Természetesen a problémát az jelenti, hogy a modellek paramétertere óriási, így a pontos feltérképezés nem hajtható végre. Pontosan ezen a ponton segít a *Bayes-statisztika* módszere, melyet követve úgy járhatunk el, hogy a paraméterek eloszlására *a priori* feltevéseket fogalmazzunk meg. Az *a priori* feltevések a fizikai paraméterekre vonatkozó előzetes tudásunkat kell, hogy kódolják. Egyszerűbb esetben az *a priori* feltevés lehet annyi, hogy az adott paramétert szoros határok közé szorítjuk (erre lehet példa a galaxist alkotó csillagok korának maximalása az Univerzum életkorában). Ha egy paraméter nagy határok között változhat, akkor a logaritmikus *a priori* feltevés lehet hasznos.

Amennyiben az *a priori* eloszlások tehát rendelkezésünkre állnak, úgy véletlenszerűen választott, de ezeknek az eloszlásoknak megfelelő paraméterekkel számítjuk ki a modellt, és ezt ismételve térképezzük fel a paraméterek terét. Természetesen annak nem sok értelme van, hogy minden egyes illesztendő spektrum esetében teljesen új modelleket generáljunk, ezért a gyakorlatban úgy érdemes eljárni, hogy legenerálunk egy modelsokaságot, melyek paramétereit az *a priori* eloszlások alapján határoztuk meg. Ezután már csak a legenerált modelleket kell a méréshez hozzápróbálni. A fő problémát a csillagkeletkezési történetet megadó függvény jó paraméterezése jelenti. Kauffmann et al. (2003a), illetve Gallazzi et al. (2005) SDSS galaxisok analízise kapcsán részletesen bemutatják a sztochasztikus csillagképződési modelleken alapuló módszert.

Felmerül a kérdés, hogy milyen módon lehet a megfelelő *a priori* eloszlásokat meghatározni. Ebben egyfelől az elmélet segíthet, másfelől pedig az, ha az optikai észleléseket más hullámhosszúságú megfigyelésekkel is kiegészítjük (Salim et al. 2005).

2.3.2. Az emissziós és abszorpciós vonalak jellemzőinek meghatározása

Miután a kontinuum meghatározásra került, az levonható az eredetileg megmért spektrumból, így csak a zaj és az emissziós vonalak maradnak. Az emissziós vonalak illesztésekor a nagy sebességszperzió miatt általában elegendő Gauss-függvények lineárkombinációját használni. Illesztéskor kihasználható, hogy a vonalak szélessége általában nagyjából azonos sebességszperziót mutat (ez alól fontos kivételek az aktív maggal rendelkező galaxisok), valamint hogy a multiplettben megjelenő vonalak arányai jól ismertek.

Az ekvivalens szélesség. Bár az emissziós vonalak erőssége megadható fluxus dimenziójú mennyiségként is, általában egy olyan mennyiséget használunk, amely független a

megfigyelt objektum luminozitásától, illetve a spektrum abszolút kalibrációjától. Ilyen mennyiség az ún. *ekvivalens szélesség*, melyet az alábbi integrál definiál:

$$W_\lambda = \int 1 - \frac{F_{\lambda, \text{vonal}}}{F_{\lambda, \text{kontinuum}}} d\lambda. \quad (2.9)$$

Itt F_{vonal} az emissziós vonal spektrális fluxussűrűsége, amiről feltesszük, hogy a vonal szélei felé haladva zérushoz tart. $F_{\text{kontinuum}}$ a korábban definiált sztellárkontinuum. Az ekvivalens szélesség egysége hullámhossz dimenziójú, és valójában annak az elméleti kontinuum tartománynak a szélességét adja meg, melyből ugyanannyi integrált fluxus érkezne, mint amennyi az emissziós vonalból érkezik.

A Lick-indexek rendszere. Értelemszerűen abszorpciós vonalakra nem definiálható a korábbiak szerinti ekvivalens szélesség, lévén az abszorpciós vonalak magából a kontinumból származnak, vagyis az ekvivalens szélesség eredeti definíciója szerinti érték mindig zérus lenne.

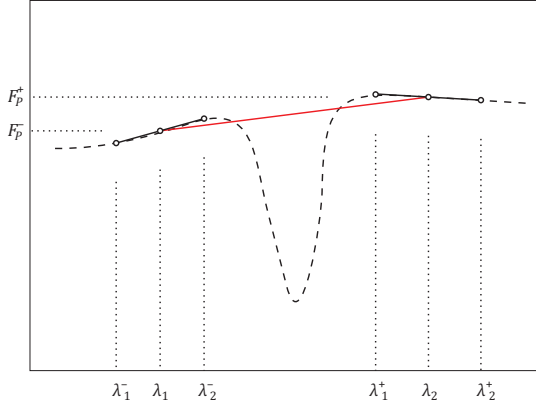
Az abszorpciós vonalak jellemzésére Worthey et al. (1994) bevezették az ún. *Lick-indexek* rendszerét. Méréseiket a kaliforniai Lick-obszervatórium távcsöveivel végezték, a rendszer neve innen ered. A Lick-indexeket galaktikus csillagok spektrumain kalibrálták. Az indexek úgy lettek meghatározva, hogy azok a spektrum alapján könnyen megmérhetőek legyenek, ugyanakkor erősen korreláljanak a csillagok fizikai paramétereivel. Az indexek rendszerének elsődleges célja eredetileg is az volt, hogy egyedi csillagok helyett csillagpopulációk paramétereinek (elsősorban kor és fémség) becslésére lehessen őket felhasználni. Később mások a rendszert további indexekkel egészítették ki (Bruzual A. 1983, Diaz et al. 1989, Huchra et al. 1996, Worthey & Ottaviani 1997, Balogh et al. 1999).

A rendszer három különböző fajta indexet definiál. Néhány indexet az ekvivalens szélesség módosított formulája szerint definiáltak, míg más indexeket integrált fluxus (tehát magnitúdó) egységekben szokás megadni. A harmadik típusú index fluxus arányokat definiál.

Az ekvivalens szélesség általánosításához be kell vezetni egy ún. *pszeudokontinuumot*, melyet az következő integrál definiál:

$$F_p^\pm = \int_{\lambda_1^\pm}^{\lambda_2^\pm} \frac{F_\lambda}{\lambda_2^\pm - \lambda_1^\pm} d\lambda, \quad (2.10)$$

ahol λ_1^\pm és λ_2^\pm indexenként két-két hullámhossz intervallum két vége. Az F_{ca} *lokális kontinuum* ezek után az a szakasz, mely két végpontját a $[\lambda_1^-, \lambda_2^-]$ és a $[\lambda_1^+, \lambda_2^+]$ intervallumok felezőpontjai, illetve az F_p^\pm pszeudokontinuum fluxusok definiálnak, ld. 2.4. ábra. Ezekkel



2.4. ábra. A Lick-indexek magyarázatához.

a definíciókkal a módosított ekvivalens szélesség a következő:

$$W_{\lambda}^{\text{abs}} = \int_{\lambda_1}^{\lambda_2} 1 - \frac{F_{\lambda}}{F_{c\lambda}} d\lambda. \quad (2.11)$$

Más indexeket inkább magnitúdó egységekben érdemes megadni, ekkor a definíció az alábbi:

$$M = -2,5 \log_{10} \left[\frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \frac{F_{\lambda}}{F_{c\lambda}} d\lambda \right]. \quad (2.12)$$

A harmadik esetben egyszerűen hullámhossz intervallumok fluxusainak arányát definiáljuk:

$$R = \frac{\int_{\lambda_1^-}^{\lambda_2^-} F_{\lambda} d\lambda}{\int_{\lambda_1^+}^{\lambda_2^+} F_{\lambda} d\lambda} \quad (2.13)$$

Az indexek konkrét $\lambda_{1,2}^{\pm}$ paraméterei a korábban idézett munkákban találhatóak. Az galaxisok abszorpciós indexek szerinti osztályozására még visszatérünk.

2.4. Galaxisok spektroszkópiai osztályozása

Egyedi csillagok spektruma néhány alapvető fizikai paramétértől függ (tömeg, fémesség, kor). Mint a 2.2.4. fejezetben láttuk, galaxisok teljes csillagpopulációinak leírásához az egyedi csillagok spektrumát megfelelő módon integrálni kell. Ugyanakkor ismert az is, hogy bár a kisebb tömegű csillagok száma jóval nagyobb, egy csillagpopuláció teljes spektrumát mégis az aktuálisan legnagyobb tömegű (következésképpen legfényesebb) csillagok

fénye dominálja. Ez azt eredményezi, hogy a fiatal populációk spektruma ultraibolya tartományban erős, A–B típusú csillagok spektrumára hasonlít, míg öreg populációk spektruma inkább törpecsillagok fénye által dominált, G vagy K típusú csillagokéra. Első ránézésre tehát a teljes csillagkeletkezési történet nem állapítható meg a spektrumból, viszont a legutolsó csillagkeletkezési periódus ideje jó biztonsággal kikövetkeztethető.

2.4.1. Színindexek szerinti osztályozás

Bár a színindexek szélessávú fotometriai felvételekből is megállapíthatók, az itt bemutatandó osztályozáshoz spektrumok felvételére is szükség van. Egyrészt ismerni kell a galaxisok vöröseltolódását, másrészt el kell végezni a K-korrekciót. (Bár a K-korrekció és a vöröseltolódás egyedül szélessávú fotometriából is megbecsülhető.)

A dolgozat későbbi fejezetére előre utalva hivatkozunk a 4.1. ábrára. Ezen az ábrán az SDSS galaxisok K-korrigált $g - r$ színindexét ábrázolja a vöröseltolódás függvényében. Az ábrát $\Delta z = 0,05$ felbontással készítettük, és ez egyes vöröseltolódás intervallumokban normáltunk a galaxissűrűsége. Jól megfigyelhető, hogy a galaxisok két csoportra oszlanak: az ún. *vörös sorozatra* és a *kék felhőre*. Előbbieket öreg, míg utóbbiakat fiatal csillagpopuláció dominálja. Megfigyelendő, hogy míg a vörös sorozat viszonylag keskeny sávot rajzol ki, azaz a galaxisok színe jól korrelál a vöröseltolódással, addig a kék felhő sokkal jobban szétterül. A jelentősen nagyobb szórás oka a galaxisonként jelentősen eltérő csillagkeletkezési ráta (Kauffmann et al. 2003a).

A vörös sorozatról, és annak evolúciójáról részletes tanulmányt készítettek Bell et al. (2004), valamint Faber et al. (2007). Győry & Bell (2010) kimutatták továbbá, hogy a vörös sorozat fiatalabb tagjainak jelentős része megfigyelhető morfológiai perturbációkat mutat, amely korábbi ár-apály hatások eredményei, és azt erősítik meg, hogy a fiatal vörös galaxisok a közelmúltban dinamikailag is heves életet éltek, azaz galaxisok összeolvadásával jöttek létre.

Strateva et al. (2001) megmutatták, hogy a két jelentős galaxis csoport között megfigyelhető egy átmeneti osztály is, melyet *zöld völgynek* neveztek el. Wyder et al. (2007), valamint Martin et al. (2007) a zöld völgy galaxisait részletesen megvizsgálták, és megállapították, hogy ezt az osztályt jelentős csillagkeletkezésen nem rég átesett galaxisok alkotják, melyek között jelentős számban fordulnak elő aktív központi maggal rendelkező galaxisok is. Számuk azért jóval alacsonyabb, mint a vörös vagy kék galaxisoké, mivel a galaxis színét domináló óriáscsillagok élete rövid, így a galaxis hamar átvándorol a kék felhőből a vörös sorozatra. Az elmélet szerint a zöld völgy galaxisai az éppen átvándorló, átmeneti galaxisokból állnak.

Röviden összefoglalva az állítható, hogy a vörös sorozat galaxisai kék galaxisok összeolvadásával jöttek létre, mely összeolvadás előbb jelentős csillagkeletkezést váltott ki, majd a csillagkeletkezés (feltehetően a központi mag aktivitásának hatására) hirtelen megszűnt,

így a galaxisok átvándoroltak a kék felhőből a vörös sorozatra (Faber et al. 2007).

A kor-fémesség-por degeneráció Bár a galaxisok akár egyetlen, jól megválasztott szín-index alapján is jól elkülöníthetők két nagy csoportra, a csillagpopulációk paramétereinek pontos becslése még igen jó minőségű optikai szélessávú fotometriai adatok alapján is nehéz. Ennek elsődleges oka az, hogy a csillagpopuláció öregedése, és a galaxisokban levő por hatása, illetve a csillagok légkörének nagyobb fémtartalma az optikai szélessávú fluxusra ugyanaz: a por nagyon hasonlóan vörösríti el a galaxis színét, mint ahogyan az öregedő csillagpopuláció vörösödik a nagy tömegű csillagok eltűnésével, illetve ahogyan a csillagok vörösebbek a nagyobb fémtartalom miatt. Ez a vörösödés az optikai tartományban nagyjából egyenletes, így a színszűrők átviteli görbéjének megváltoztatásával sem oldható fel a degeneráció. A megoldást ultraibolya és infravörös mérések elvégzése jelenti, melyek azonban technikai akadályok miatt csak más műszerrel (leginkább csak műholdon elhelyezett műszerekkel) kivitelezhetőek. Ez a jelenség számottevően korlátozza a szélessávú optikai adatokon futtatott fotometrius vöröseltolódás-becslő algoritmusok lehetőségéit is.

Mindezek miatt a galaxisok paramétereinek biztonságos becsléséhez optikai spektroszkópiai megfigyelésekre van szükség.

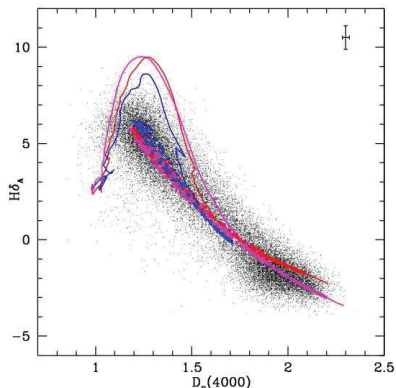
2.4.2. Abszorpciós indexek szerinti osztályozás

Mivel a 2.3.2. fejezetben elmondottak szerint az abszorpciós indexek a sztellárkontinuumot jellemzik, segítségükkel közvetlenül nyerhető információ a galaxist alkotó csillagokról. Ehhez az is hozzájárul, hogy a galaxisokban jelen levő por hatásának hullámhosszfüggése elég gyenge ahhoz, hogy az abszorpciós index értékét számottevően ne befolyásolja. Az alábbiakban bemutatunk egyféle abszorpciós indexek szerinti osztályozást, amivel meghatározható a galaxisokban végbemenő csillagkeletkezési folyamatok rátája. A tárgyalásban Kauffmann et al. (2003a) munkáját követjük.

Idősebb, fémekben gazdag csillagpopulációk spektrumában 4000 Å-nél egy jelentős törés figyelhető meg: a rövidebb hullámhosszakon rengeteg abszorpciós vonal jelenik meg, melyek a fluxus jelentős csökkenését okozzák. Ezek a vonalak alacsonyan ionizált fémektől származnak, és az általuk okozott jelenség neve *kitakarás*³. Az alacsonyan ionizált fémek tipikusan törpecsillagok spektrumában vannak jelen, a forróbb légkörökben ezek az elemek többszörösen ionizálódnak, és máshol okoznak vonalakat. Bruzual A. (1983) nyomán Balogh et al. (1999) definiálták a $D_n(4000)$ kontinuum indexet, mely a 4050-4250 és 3750-3950 Å közötti integrált fluxusok arányával azonos.

A $H\delta_A$ abszorpciós vonal megjelenése azokra a csillagpopulációkra jellemző, melyek az utóbbi 0,1-1 milliárd évben csillagkeletkezésen estek át. Az abszorpciós vonal megjelené-

³line-blanketing



2.5. ábra. Az $H\delta_A$ index a $D_n(4000)$ függvényében. A fekete pontok az SDSS galaxisok eloszlását jelölik, míg a görbék és színes pontok modellekből lettek számolva. A három folytonos görbe egy-egy hirtelen csillagkeletkezésből ($\psi(t) = \delta(t - t_0)$) létrejött három különböző fémmasszú populáció idő szerinti trajektóriáját ábrázolja. A trajektóriák az ábra bal oldala felől indulnak, és a jobb alsó sarok irányába öregednek. A színes pontok három különböző fémmasszú, folytonos lassú csillagkeletkezési rátával rendelkező galaxisok helyét jelölik. Forrás: Kauffmann et al. (2003a).

sének oka, hogy a fényes O és korai B típusú csillagok már eltűntek, de a késői B és korai F típusú csillagok még dominálják a vonal tartományában kibocsátott fluxust.

A 2.5. ábrán az előbb tárgyalt két abszorpciós index szerint látható az SDSS galaxisok eloszlása, illetve néhány populációsintézis-modell időbeli trajektóriája. A folytonos vonalak egy időben keletkezett, különböző fémmasszú csillagpopulációk trajektóriái. A $t = 0$ időpont a görbék baloldali végén van, és az idő előrehaladtával az egyidejű csillagpopulációk helye jobbra tolódik, míg át nem fed a passzív galaxisokkal. A színes pontok folytonos, lassú csillagkeletkezési rátával jellemezhető populációk helyét mutatják. Ez utóbbiak szépen átfednek az SDSS galaxisokkal, vagyis azt állíthatjuk, hogy a legtöbb alacsony vöröseltolódáson megfigyelhető galaxis a lassú, folytonos csillagkeletkezési régió, és a passzív régió által meghatározott tartományban helyezkedik el. A csillagkeletkezési ráta pontos értéke úgy határozható meg, hogy megnézzük az adott galaxis mely modellekkel fed át.

Természetesen ez csak egy példa az abszorpciós indexek felhasználására. Számos más módszer létezik, mellyel a csillagkeletkezési rátán túl a kor, fémmasszú stb. meghatározható.

2.4.3. Emissziós vonalak szerinti osztályozás

A galaxisok spektrumainak optikai tartományban megfigyelhető emissziós vonalai a gerjesztett csillagközi gáztól származnak. A ritka csillagközi gáz egyik legszembetűnőbb tulajdonsága, hogy magas ionizációs állapotú elemeket tartalmaz, és az emissziós vonalak között számos tiltott átmenet is megtalálható. Amennyiben a távoli galaxisban és a Tejútban levő csillagközi por hullámhosszfüggő elnyelésére megfelelően korrigálunk, az azonos ionhoz tartozó átmenetekből származó emissziós vonalak arányainak meg kell egyeznie a laboratóriumban mérhető (illetve tiltott átmenetek esetén elméletileg kiszámolható) vonalarányokkal. Ez lehetőséget teremt a pormodellek pontos kalibrálására (Calzetti et al. 1994, Calzetti 2001).

Számunkra ennél is érdekesebb kérdés, hogy meghatározzuk a vonalakat kibocsátó gáz különböző összetevőelemeinek ionizációs arányát. Az ionizációs arányok, illetve ionizációs energiák ismeretében a Saha-formula felhasználásával meghatározható a gázt ionizáló sugárzás körülbéli spektrális energiaeloszlása, melyből következtetni tudunk a gerjesztés eredetére, vagyis a galaxisban végbemenő energiatermelő folyamatokra.

A galaxisokban két alapvető folyamat képes ionizáló ultraibolya sugárzást kelteni: A nagy tömegű, forró, fiatal csillagok *termikus* sugárzása az UV tartományban domináns, ez felelős a HII régiók gerjesztésért. A központi fekete lyuk közreműködése által kibocsátott sugárzás spektruma viszont *hatványfüggvény* jellegű, így a különböző elemeket a termikus sugárzástól eltérő arányban ionizálja. Ez utóbbi galaxisokat *aktív magvának* (AGN) nevezzük.

Baldwin et al. (1981) ún. *diagnosztikai diagramokat* készített (BPT-diagram), melyeken különböző égi objektumok emissziós vonalainak arányát tüntették fel, és megállapították, hogy az objektumok a vonalarányok szerint egyszerűen osztályozhatók. A módszert Osterbrock & De Robertis (1985), majd Veilleux & Osterbrock (1987) fejlesztette tovább.

2.4.4. Aktív galaxismagok és csillaképző galaxisok

Az emissziós vonalak alapján a galaxisok három alapvető csoportra oszthatók. A *passzív galaxisok* (illetve kevésbé aktív galaxisok) esetében az emissziós vonalak nem, vagy csak alig detektálhatók, így ez a csoport a BPT-diagramokon nem ábrázolható. Jól elkülöníthetőek a diagramon a *csillagképző galaxisok*, ahol a csillagközi gáz ionizációnak forrásai a fiatal, ultraibolya tartományban termikusan sugárzó óriáscsillagok (Charlot & Longhetti 2001). A harmadik nagy csoportot az *aktív magvú galaxisok* jelentik, ahol a hatványfüggvény jellegű ionizáló sugárzás eredete az aktív mag szinkrotron sugárzása.

Kewley et al. (2001) populációsintézis-modellek segítségével, teljesen elméleti alapon vizsgálták meg a problémát, és két fő eredményre jutottak: Egyrészt a BPT diagra-

mon létezik egy, az extrém csillagontó galaxisokat határoló vonal, melyen túl elvileg sem helyezkednek el termikusan gerjesztett rendszerek. Másrészt létezik egy átmeneti zóna, ahol a termikus és hatványfüggvény jellegű gerjesztések szignatúrája egyszerre van jelen. Ez utóbbi arra utal, hogy számos galaxisban egyszerre történik csillagkeletkezés, és van jelen aktív mag.

Az SDSS égtérképezés által vizsgált galaxisok emissziós vonalak szerinti osztályozását Kauffmann et al. (2003c), valamint Kewley et al. (2006) végezték el. Előre utalunk egy későbbi fejezetre, ahol a 4.2. és 4.15. ábrákon reprodukáljuk az ő eredményeiket, és feltüntetjük az általuk meghatározott szegregációs görbéket is. Ezekon az ábrákon jól elkülöníthető az eddig felsorolt három tartomány: csillagképző galaxisok, aktív magvú galaxisok, valamint a kettő közti átmeneti zóna. Kewley et al. (2006) az AGN-ek osztályozását tovább vitték, és sikerült elkülöníteniük a Seyfert-galaxisokat az ún. *LINER*⁴-ektől (Heckman 1980). Ez utóbbiak neve mozaikszó, az angol rövidítés megfelelője alacsony ionizációs szintű keskeny emissziós régiók. A LINER-ek és Seyfertek közötti elkülönülést a 4.3. ábra mutatja.

Az ionizációs arányok vizsgálatán túl további megfigyelésekkel is bizonyítható, hogy a Seyfert-galaxisok emissziójáért az aktív mag felelős. Seyfert 1 típusú galaxisok esetében a széles emissziós vonalak egyértelművé teszik, hogy a sugárzó gázfelhők egy része a nagy tömegű központi maghoz közel kering. Erre szintén bizonyíték a széles vonalak időbeli változékonysága. Néhány Seyfert 2 galaxis esetében polarimetriai mérésekkel kimutatták, hogy a keskeny vonalas Seyfert 2 komponens mellett megjelenik egy gyenge, erősen polarizált széles emissziós vonalak mutató spektrum is (Antonucci & Miller 1985). Ennek a komponensnek a származási helye szintén a mag, de a látóirányba eső porfelhők miatt a központi régió közvetlenül nem figyelhető meg, hozzánk csak a magtól távolabbi porfelhőkről visszaverődő fény jut el. A széles vonalak mutató spektrum polarizációját pontosan ez a visszaverődés okozhatja. Az aktív galaxisok témakörét áttekintő aktuális munkáért ld. Urry & Padovani (1995), Antonucci (1993), Ho (2008).

Míg az aktív galaxismagokról többé-kevésbé elfogadott modellekkel rendelkezünk, addig a LINER-ek pontos mibenléte, az őket tápláló energiaforrások eredete máig nyílt kérdés. Annyi bizonyos, hogy a galaxismagok mintegy 30 %-a mutat LINER jellegzetességeket. A szóba kerülő energiaforrások között szerepelnek kis energiájú AGN-ek, lökéshullámok, vagy extrém magas hőmérsékletű csillagok is. Részletekért ld. Kewley et al. (2006) és az ott szereplő hivatkozásokat.

⁴LINER = Low Ionization Narrow Emission Region

2.5. Magnitúdók K-korrekciója és evolúciókorrekciója

Ha különböző vöröseltolódásokon levő galaxispopulációkat akarunk összehasonlítani szélessávú abszolút magnitúdóik alapján, akkor két fontos effektust kell figyelembe vennünk. Mindkettő az objektumok spektrumán, illetve spektrumaik változásán alapul.

Az első effektus abból adódik, hogy magasabb vöröseltolódáson a spektrumok jellegzetességei (leginkább a 4000 Å körüli törésnek van nagy jelentősége) nagyobb hullámhosszakra tolódnak. Ez azt eredményezi, hogy a galaxisok színindexei – még ha az eredendő sugárzásuk spektruma meg is egyezik – különböző vöröseltolódásokon különböző lesz. A 2.6. ábra egy vörös és egy kék galaxis színindexeinek változását mutatja a vöröseltolódás függvényében. A változás olyan erős, hogy amikor az abszolút magnitúdót meghatározzuk, ezt az effektust mindenképp figyelembe kell vennünk, és korrigálnunk rá. Az abszolút magnitúdó képlete a következőképpen módosul:

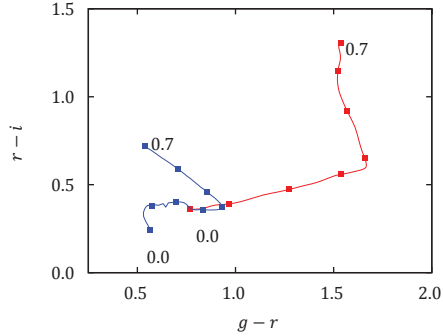
$$M_K = m - DM - K, \quad (2.14)$$

ahol az indexbe írt K betű az abszolút magnitúdó K -korrigáltságára utal; m a megfigyelt magnitúdó; DM a szokásos távolságmodulus, amit a vöröseltolódásból határozzunk meg (Hogg 1999); K pedig az ún. *K-korrekció*, ami a tárgyalta effektust figyelembe veszi. K értéke függ a vöröseltolódástól is, és a szélessávú színszűrőtől is. Könnyen belátható, hogy a K -korrekció értéke az adott színszűrőben megfigyelt fluxus, valamint annak a fluxusnak az arányából számolható, amit úgy kapunk, mintha az objektum azonos spektrummal, de $z = 0$ vöröseltolódáson lenne megfigyelve (Hogg et al. 2002). A formulában (ld. alább) megjelenik egy $\frac{1}{1+z}$ faktor is, mely abból ered, hogy a fluxus hullámhosszegységre vonatkoztatva van megadva. A K -korrekció meghatározásához tehát szükségünk van az objektum illesztett spektrumára (legalább is az illesztett kontinuumra, mivel az emissziós vonalak fluxusa ilyen szempontból nem számottevő). Természetesen a spektrum akkor is illeszthető, ha csak a szélessávú magnitúdók ismertek (Blanton et al. 2003). A spektrum alapján a fotonfluxus (1.1) alapján számítható. Ezt általánosítva a K -korrekcióra

$$K = -2,5 \lg \left[\frac{1}{1+z} \frac{F_z}{F_{z=0}} \right] = -2,5 \lg \left[\frac{1}{1+z} \frac{\int F_\lambda \left(\frac{\lambda}{1+z} \right) r(\lambda) \lambda d\lambda}{\int F_\lambda(\lambda) r(\lambda) \lambda d\lambda} \right], \quad (2.15)$$

ahol F_z a modell spektrum integrált fluxusa z vöröseltolódáson; $F_{z=0}$ ugyanez, de $z = 0$ vöröseltolódáson; F_λ a korábbi spektrális fluxussűrűség, míg r a színszűrő átviteli függvénye.

A K -korrekció mindaddig elegendő, amíg nem vesszük figyelembe a galaxisok csillagpopulációinak öregedését. Távoli vöröseltolódásokon ugyanis a galaxisok egy fiatalabb állapotát látjuk, tehát ha minden szempontból korrektül kívánunk eljárni, akkor a galaxi-



2.6. ábra. Egy tipikus vörös és kék galaxis színindex trajektóriája az SDSS g' , r' és i' színszűrőjében. A pontok a vöröseltolódást jelölik $\Delta z = 0,1$ -es felbontással. Az effektus még közeli vöröseltolódásokon is olyan erős, hogy a K-korrekció elkerülhetetlen. Az ábra az evolúciós effektusokat nem szemlélteti.

sok z és $z = 0$ közötti „öregedését” is figyelembe kell vennünk. Ha a spektrális evolúcióra valamilyen modellt alkotunk, akkor az alapján kiszámíthatjuk az *evolúciókorrekciót*. Ez egy újabb tagot az az abszolút magnitúdó képletéhez:

$$M_{K+e} = m - DM - K - E, \quad (2.16)$$

ahol az indexbe írt $K+e$ az abszolút magnitúdó K - és evolúciókorrigáltóságára utal; E pedig az evolúciókorrekció értéke. A korábbi logikát folytatva az evolúciókorrekcióra a következő adódik:

$$E = -2,5 \lg \left[\frac{F_t}{F_{t_0}} \right] = -2,5 \lg \left[\frac{\int F_{\lambda,t}(\lambda) r(\lambda) \lambda d\lambda}{\int F_{\lambda,t_0}(\lambda) r(\lambda) \lambda d\lambda} \right], \quad (2.17)$$

ahol a t_0 index arra utal, hogy a galaxist modellező spektrumot úgy vesszük, mint amilyen idős $z = 0$ -nál lenne; a sima t pedig arra, hogy a megfigyeléseknek megfelelő korú modellt veszünk. $F_{\lambda,t}$ azonos a K -korrekciónál használt F_{λ} -val.

Kék galaxisok esetében a spektrum evolúciója nem számottevő, mivel itt mindig a legfiatalabb csillagpopuláció dominálja a galaxis spektrumát. Ettől függetlenül megfigyelhető a galaxisok halványodása, ami az Univerzum csillagkeletkezési rátájának csökkenésével függ össze. Mindenféle típusú galaxisok esetében egy általánosan használt átlagos evolúció korrekciós érték az $E(z) = -1,6(z - 0,1)$ alakban adható meg (Tegmark et al. 2004).

Vörös óriásgalaxisok esetében ennél jóval pontosabb becslés adható, hiszen ezekben a galaxisokban nincsen számottevő csillagképződés, vagyis csupán a bennük levő csillagok passzív öregedésével kell számolni. Vörös galaxisokra Maraston et al. (2009) adtak jól használható, konzisztens modellt.

2.6. Kompozit spektrumok

Az SDSS égtérkép legnagyobb előnye, hogy a rögzített nagy mennyiségű galaxisspektrum a galaxispopulációk statisztikai vizsgálatát is lehetővé teszi. A nagy mennyiségű megfigyeléshez alkalmazott megfigyelési technika ugyanakkor kihívások elé állítja az adatokat feldolgozó csillagászokat. Itt azt emelnénk ki, hogy a kis távcsőméret és a rövid integrációs idő miatt a spektrumok jelentős része meglehetősen alacsony jel-zaj arány mellett került rögzítésre. A jel-zaj viszony növelésére viszont lehetőséget ad, ha a különböző, ám nagyon hasonló objektumoktól származó spektrumokat kiátlagoljuk, ún. *kompozit spektrumokat* készítünk.

A 4. fejezetben saját munkánkat mutatjuk be, melyben az SDSS égtérkép galaxisspektrumainak átlagolásával nyert kompozit spektrumok atlaszát készítettük el.

3. fejezet

Galaxisspektrumok főkomponens-analízise

Az eddig bemutatásra került analitikus módszerek (a kontinuumillesztéstől eltekintve) mindig csak a spektrum egy kis tartományát vették figyelembe (emissziós vonalak, abszorpció indexek). Felmerül a kérdés, hogy mi módon lehet a teljes spektrumban levő információ mennyiséget kinyerni és felhasználni. Egyáltalán mennyi információ van egy galaxisspektrumban?

Ezek a kérdések több szempontból is lényegesek. Egyrészt ha a teljes spektrumot felhasználjuk az analízis során, az a véletlenszerű zajjal szemben robusztusabbá teszi a paraméterbecslő eljárásainkat, hiszen lényegesen több adatból tudunk dolgozni, mint a szűk tartományra koncentrált indexek esetében. Másrészt ha tudjuk, hogy mennyi információ van kódolva egy spektrumban, megfelelő paraméterezésre áttérve minimalizálhatjuk a spektrummodelljeink paramétereinek számát. A 2. fejezetben megismertek alapján arra számítottunk, hogy a galaxisspektrumok valóban néhány paraméterre (kor, fémesség, valamilyen módon paraméterezett csillagkeletkezési történet) redukálhatók. Látni fogjuk, hogy ezen paraméterek erős korrelációja miatt valóban elegendő néhány paraméter a spektrumok 90-95 %-os rekonstruálásához.

3.1. Galaxispektrumok dimenzióredukciója

Ebben a fejezetben bemutatunk néhány lineáris eljárást, melyek segítségével a spektrumok sokezer dimenziós tere néhány dimenzióra redukálható. Ezeknek az eljárásoknak lassan két évtizedes történetük van, és mára széles körben elterjedt módszereknek számítanak. Újdonság viszont az a robusztus eljárás, mely segítségével zajos, hiányos spektrumok dimenzióredukciója iteratíván és gyorsan végezhető el. Ez utóbbi algoritmus jelenleg egyetlen implementációja a dolgozat szerzőjének munkája, és ennek felhasználásával készült a 4. fejezetben bemutatandó kompozitspektrum-atlasz is.

Mielőtt a galaxisspektrumok egy sokaságát dimenzióredukciós eljárásoknak vetnénk alá, el kell végeznünk pár fontos műveletet. Egyrészt szükséges az összes spektrum nyugalmi rendszerbe konvertálása, másrészt a fluxusok valamilyenfajta normálása, harmadrészt a hullámhosszbinék egy közös hullámhosszrácsra konvertálása. Ez utóbbi rögtön problémát vet fel, hiszen a változó vöröseltolódás miatt a spektrumok rövid, illetve hosszú hullámhosszú végei vagy lelőgnak a hullámhosszrácsról, vagy nem érik el annak végét. A lelógó részt egyszerűen elhagyjuk, viszont a hiányos részek kezelése külön körültekintést igényel. Erre a problémára a 3.3. pontban még visszatérünk. A spektrumok normálásáról szót ejtünk a 4.4.2. fejezetben, amikor a módszer egy konkrét alkalmazását mutatjuk be.

3.1.1. A dimenzióredukciós eljárások áttekintése

A lineáris dimenzióredukciós eljárásokban közös, hogy az adatvektorok (esetünkben a sokezer dimenziós spektrumok) lineáris terét valamilyen projekcióval egy lényegesen kisebb dimenziószámú lineáris térre vetítjük, és az eredeti adatvektorokat a kisebb tér dimenziószámának megfelelő számú paraméterrel helyettesítjük.

A dimenzióredukció eredményeként kapott kis számú paramétert érdemes a galaxisok más módon meghatározott fizikai paramétereivel korreláltatni, mert ily módon a pusztán matematikai eljárásból nyert új paraméterekhez fizikai tartalmat is rendelhetünk. Mivel a dimenzióredukciós eljárással származtatott paraméterek az egész spektrumot használják egy kis szakasz helyett, könnyen előfordulhat (és erre konkrét példát is adunk), hogy az új paraméterek zaja, korrelációja a direkt módon meghatározott modellparaméterek zajánál kisebb, illetve korrelációjánál nagyobb.

A főkomponens-analízis. A legelterjedtebb lineáris dimenzióredukciós eljárás a főkomponens-analízis¹ (PCA), más néven a Karlhunen–Løve-transzformáció (Pearson 1901). Fontossága miatt erre az eljárásra a 3.1.2. fejezetben külön is kitérünk.

A PCA által szolgáltatott főkomponensvektorok galaxisspektrumok esetében fizikai értelemmel is bírnak (Yip et al. 2004a). Ha a főkomponens-analízist az átlagok levonása nélkül végezzük el, akkor az első főkomponens az átlagos spektrumhoz lesz hasonló. A második és harmadik főkomponens általában a teljes spektrum meredekségét, és a 4000 Å-ös törés nagyságát határozza meg, míg a magasabb rendű főkomponensek az emissziós vonalak arányaira érzékenyek, ld. Yip et al. (2004a, 5. ábra).

A CX-dekompozíció. Általános PCA problémák esetében azonban a főkomponensek nagyon sok adatvektor lineárkombinációi, így semmi sem garantálja az eredményül kapott főkomponensek fizikai értelmezhetőségét. Ezt a problémát megcélözva dolgozta ki Mahoney & Drineas (2009) az adatmátrixok CUR-dekompozícióján alapuló CX-dekompozíciós

¹PCA = Principal Component Analysis

eljárást (Drineas et al. 2008). Röviden összefoglalva az eljárás lényegét: Az eredeti \mathbf{A} adatmátrixot $\mathbf{A}' = \mathbf{C}\mathbf{X}$ közelítő alakban faktorizáljuk, ahol a \mathbf{C} mátrix egy alacsony rangú, az eredeti adatmátrix néhány oszlopából álló mátrix. Ez nagyon fontos különbség a PCA-hoz képest, ahol az SVD faktorizáció eredményeként a bázisvektorok nagyon bonyolultan függenek az eredeti adatvektoroktól. Spektroszkópiai nyelvezetre lefordítva ez annyit tesz, hogy bármely spektrumot néhány „jól megválasztott” valódi spektrum lineárkombinációjaként írunk fel. A CX-dekompozíciós algoritmus ezeket a „jól megválasztott” spektrumokat hivatott meghatározni. Ebben az esetben a fizikai értelmezhetőséghez nem fér kétség. A CX-dekompozíció elvégezhető a transzponált problémára is, amikor is a \mathbf{C} mátrix nem az eredeti adatvektorokból, hanem az adatmátrix soraiból áll. Ebben az esetben azt várjuk, hogy az eljárás megadja azokat a hullámhossz-tartományokat, amelyek a legtöbb információt hordozzák a spektrumokról (Yip & Csabai 2011).

A MOPED tömörítés. Heavens et al. (2000) kidolgozták a MOPED nevű lineáris, veszteségmentes tömörítési eljárást. Megmutatták, hogy ha egy M paraméteres modell alapján generálunk spektrumokat, és azokra szimulált zajt teszünk, akkor a MOPED eljárással az egyes spektrumok M számú adatra tömöríthetők (a jellemzően 10^3 számú adat helyett). Amennyiben a spektrumokra tett zaj független a modell paramétereitől, úgy a tömörítési eljárás veszteségmentes. Az eljárás kidolgozói megmutatták, hogy a tömörítési eljárással kapott M adat kapcsolatba hozható az eredeti modellparaméterek értékeivel. Az eljárás nagy előnye, hogy erősen zajos spektrumok esetén is használható. A MOPED tömörítés elsősorban akkor nyújt nagy segítséget, ha egy adott spektrumhoz legközelebb eső modellspektrumok megtalálására van szükség. Ekkor a modellspektrumok egyedi illesztése helyett elegendő a tömörített adatok terében keresni, ami sokkal kisebb dimenziós a spektrumok eredeti terénél (Panter et al. 2003).

A MOPED algoritmus használatához mindenképpen szükség van ismert paraméterekkel generált modellspektrumokra, míg a PCA algoritmus kizárólag az adatokkal dolgozik. A PCA definíció szerint az adatpontok szórására érzékeny, de nem képes elkülöníteni, hogy a szórás az alapvető paraméterek szórásából, vagy a mérési hibából következik-e.

3.1.2. A Karlhunen–Loève-transzformáció

A PCA, főkomponens-analízis, avagy Karlhunen–Loève-transzformáció lényege, hogy az adatvektorokat egy sokdimenziós euklideszi tér adatpontjainak tekintjük, majd az adatpontok alapján egy olyan ortogonális bázist származtatunk, mely vektorai az eredeti vektorok legnagyobb szórású irányiba mutatnak. A eljárás szemléletesen a következő: A D dimenziós térben dolgozva megkeressük azt a $D - 1$ dimenziós hipersíkot, melytől véve az adatpontok távolsága a legnagyobb szórást mutatja. Ennek a síknak a normálisa lesz az első bázisvektor, melynek a nagyságát a szórással arányosan választjuk meg. Az összes

további bázisvektort ugyanígy választjuk, de azzal a kikötéssel, hogy a már meglevő bázisvektorokra merőlegesnek kell lenniük. Az eljárást addig folytatjuk, ahány főkomponenst szeretnénk meghatározni.

Algebrailag a megfogalmazás tömörebb, és kétféle eljárás is létezik. Az egyik szerint meghatározzuk az adatpontok kovarianciamátrixának sajátérték dekompozícióját, a másik szerint pedig magának az adatmátrixnak a szinguláris érték dekompozícióját.

Jelölje \mathbf{X}^T az adatvektorokból képzett mátrix transzponáltját (vagyis az adatvektorok a mátrix sorai, tehát a mátrix N sorból és D oszlopból áll, ahol N az adatpontok száma, D pedig a vektortér dimenziója). Ekkor az adatmátrix szinguláris értékek szerinti faktorizációja

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (3.1)$$

ahol \mathbf{U} egy $D \times D$ méretű mátrix, mely oszlopaiban a $\mathbf{C} = \mathbf{X}\mathbf{X}^T$ kovarianciamátrix szinguláris vektorait² tartalmazza – ezek a főkomponensek; $\mathbf{\Sigma}$ egy $D \times N$ méretű mátrix, ami a főátlóban a szinguláris értékeket³ tartalmazza csökkenő sorrendben; \mathbf{V}^T pedig egy $N \times N$ méretű mátrix, és $\mathbf{X}^T\mathbf{X}$ sajátvektoraiból áll.

Kiindulhatunk a kovarianciamátrix faktorizációjából is:

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T, \quad (3.2)$$

ahol kihasználtuk, hogy \mathbf{C} a szimmetrikus, $D \times D$ méretű kovarianciamátrix; \mathbf{U} szintén $D \times D$ méretű, és a kovarianciamátrix szinguláris vektoraiból, azaz a főkomponensekből, mint oszlopvektorokból képzett mátrix; $\mathbf{\Sigma}$ pedig a szinguláris értékeket csökkenő sorrendben tartalmazó diagonális, ugyancsak $D \times D$ méretű mátrix.

Megjegyezzük, hogy a kovarianciamátrix szingulárisérték-dekompozíciója helyett elvégezhető lenne a sajátérték-dekompozíció is, viszont a pszeudoinvert meghatározása a gyakorlatban numerikusan stabilabbnak bizonyul.

Az eredeti adatpontokat tehát a főkomponensek bázisába transzformáló operátor ekkor az \mathbf{U}^T mátrix, és az új adatmátrix

$$\mathbf{Y} = \mathbf{U}^T\mathbf{X} \quad (3.3)$$

alakban írható fel, ahol az \mathbf{Y} mátrix $D \times N$ -es, és az oszlopaiban tartalmazza a főkomponensek bázisán kifejtett adatvektorokat.

Természetesen az egész eljárással az a célunk, hogy az eredetileg D dimenziós adatvektorok helyett $L \ll D$ dimenziós adatvektorokat vezessünk be, melyekből az eredeti vektorok a lehető legkisebb adatvesztéssel rekonstruálhatóak legyenek. Vezessük be

²Egy \mathbf{M} mátrix baloldali (jobboldali) szinguláris vektorai az $\mathbf{M}\mathbf{M}^*$ ($\mathbf{M}^*\mathbf{M}$) mátrix sajátvektora, ahol a csillag az adjungáltat jelöli.

³A nem nulla szinguláris értékek az $\mathbf{M}\mathbf{M}^*$ vagy $\mathbf{M}^*\mathbf{M}$ mátrix nem nulla sajátértékeinek gyökei.

az \mathbf{U}_L mátrixot, mely az eredeti \mathbf{U} mátrix $D \times L$ méretű csonkolt változata, azaz csak az első L oszlopot tartalmazza. Ezek pont az első L legnagyobb főértékhez tartozó sajátvektorok, tehát joggal várhatjuk, hogy az eljárásból adódó adatvesztés minimális lesz. A csonkolt bázison az adatvektorok így kódolandók:

$$\mathbf{Y}_L = (\mathbf{U}_L)^T \mathbf{X}, \quad (3.4)$$

ahol az \mathbf{Y}_L mátrix immár csak $L \times N$ méretű.

3.1.3. A duális PCA

Amennyiben az adatpontok száma jóval kisebb a dimenziók számánál, azaz $N \ll D$, célszerű $\mathbf{X}\mathbf{X}^T$ helyett $\mathbf{X}^T\mathbf{X}$ faktorizációját venni, hiszen ez a mátrix ekkor határozottan kisebb a kovarianciamátrixnál. Ekkor

$$\mathbf{X}^T\mathbf{X} = \mathbf{W}\Sigma\mathbf{W}^T, \quad (3.5)$$

ahol most az összes jobboldalon szereplő mátrix $N \times N$ -es. Redukáljuk ezután a dimenziók számát $L < N < D$ -re, vagyis \mathbf{W}_L legyen a \mathbf{W} első L oszlopából álló mátrix, és Σ_L pedig egy $L \times L$ -es diagonális mátrix, mely a főátlóban az L legnagyobb főértéket tartalmazza. Az állítás – melynek bizonyításától itt most eltekintünk – az, hogy a főkomponensek bázisát ekkor az eredeti adatmátrix segítségével az alábbi módon kapjuk:

$$\mathbf{U}'_L = \mathbf{X}\mathbf{W}_L\Sigma_L^{-1}, \quad (3.6)$$

ahol \mathbf{U}'_L a sima PCA esetében bevezett \mathbf{U}_L mátrixnak felel meg, azzal azonos méretű, így (3.4) továbbra is érvényes marad $\mathbf{U}_L = \mathbf{U}'_L$ helyettesítéssel.

Ritka, hogy az adatpontok száma lényegesen kisebb legyen, mint az eredeti lineáris tér dimenziója. Később látni fogjuk azonban, hogy az iteratív PCA eljárás inicializáláshoz a duális PCA eljárás az optimális.

3.1.4. Az átlag levonása

A főkomponens vektorok meghatározhatók úgy, hogy előzőleg az átlagot levonjuk az adatvektorokból, és úgy is, hogy ezt nem tesszük meg. Ha az átlag levonását előzetesen nem tettük meg, akkor a kapott első főkomponens általában nagyon hasonló lesz a közvetlenül számolt átlaghoz. A két eljárás közötti jelentős elvi különbséget az okozza, hogy a főkomponensek definíció szerint ortogonálisak egymásra. Ha tehát az átlagot nem vonjuk le, hanem főkomponensként kezeljük, úgy az összes többi főkomponens erre merőleges lesz.

Galaxispektrumok esetében a tapasztalatok szerint mindkét eljárás nagyjából azonos eredményre vezet, de ez más jellegű adatokra nem feltétlenül van így.

3.2. A főkomponens-analízis korábbi alkalmazásai

A főkomponens-analízis korai csillagászati alkalmazásaként Francis et al. (1992) kvazárok spektrumait elemezték. Connolly et al. (1995b) néhány galaxispektrum-sablon főkomponens-analízisét végezték el, és megállapították, hogy a spektrumok az első három főkomponensből alkotott bázison kifejtve már nagyon jól rekonstruálhatók. Bevezették a főkomponens-bázisokon való a_i ($i = 0, 1, \dots, D$) kifejtési együtthatókból számolható $\phi_{\text{KL}} = \tan^{-1}(a_2/a_1)$ szöveget, mely segítségével a kék és vörös galaxisok spektruma egyetlen paraméterrel jellemezhető. Ez utóbbi eredmény igen fontos, hiszen a normált galaxispektrumok sokdimenziós tere igen jó közelítéssel kétdimenziósra redukálható! Ugyanők megvizsgálták a spektrumok különböző módon elvégzett normálásának hatását is az eredményül kapott főkomponensekre. Azt találták, hogy bár a normálás magukra a főkomponens vektorokra hatással van, de az ésszerű fizikai szempontok szerinti normálási módokkal kapott különböző galaxisklasszifikációs eredmények nem függenek a normálás konkrét módjától.

A valódi mérési adatok zajosak, továbbá bizonyos hullámhossztartományokban hiányosak is lehetnek. Ennek a problémának a megoldására Connolly & Szalay (1999) a korábbi eljárás javított változatát dolgozta ki. Ők a főkomponens-analízist többször egymás után végezték el, úgy, hogy a spektrumok hiányos vagy zajos részeit mindig az előző futtatásból kapott bázison történő kifejtés alapján rekonstruálták. Az eljárás konvergensenk bizonyult, melyet később SDSS galaxis- és kvazárspektrumokra is alkalmaztak (Yip et al. 2004a;b), és a galaxisokat az első három saját együtthatójuk alapján osztályozták. Az alkalmazott eljárás hátránya volt, hogy a teljes adatmennyiség többszöri beolvasását igényelte, sok memóriát használt, valamint az elvégzendő mátrixfaktorizálások mérete lassú futást eredményezett.

Mások a PCA-t spektroszkópikus és fotometrikus vöröseltolódás becslésére (Glazebrook et al. 1998, Budavári et al. 1999; 2000), vagy az ég emissziójából adódó háttér levonására (Wild & Hewett 2005) is használták.

Wild et al. (2007) megmutatták, hogy miként használható a főkomponens-analízis a sztellárkontinuum 4000 Å-ös törésének vizsgálatára, és a saját együtthatók segítségével miként definiálhatóak a 2.4.2. fejezetben ismertetett H δ és D $_n$ (4000) indexeknek megfelelő, a galaxisok csillagképző aktivitását jellemző mennyiségek. Kiemelték, hogy a PCA alapú eljárás automatikus, nem igényli az indexek előzetes definiálását és kalibrálását, valamint a hagyományos abszorpciós vonalindexekkel szemben robusztusabb.

Györy et al. (2011) a galaxisok erős emissziós vonalainak ekvivalens szélességét tekintették lineáris teret kifesztítő adatvektoroknak, és végezték el az eloszlás főkomponens-analízisét. PCA segítségével reprodukálták a különböző morfológiai típusok korrelációját az emissziós vonalak vonalarányaival, és megmutatták, hogy a PCA térben is elvégezhető

a 2.4.4. fejezetben bemutatott csillagképző, illetve aktív magvú galaxisok szeparációja.

3.3. Egy robusztus, iteratív PCA algoritmus *

Említettük, hogy a robusztus PCA eljárások a kovarianciamátrix többszöri kiszámításával és faktorizációjával járnak. Mivel a kovarianciamátrix mérete $D \times D$, nagy felbontású spektrumok esetében a faktorizációs lépés maga jelentős számítási igényvel bír. Sajnos a duális PCA sem jöhet szóba akkor, ha sok spektrumról van szó. Mivel a PCA lényege, hogy a faktorizáció után a kapott sajátbázist az első néhány komponensre csonkoljuk, felmerül a kérdés, hogy kidolgozható-e olyan algoritmus, ami eleve kisebb bázissal, következésképpen kisebb kovarianciamátrixszal dolgozik.

Budavári et al. (2009) publikáltak egy algoritmust, mely négy fontos területen is előrelépést jelent. Az algoritmus robusztus a kilógó adatpontokra nézve, kezeli a zajos és hiányos adatvektorokat, eleve csonkolt bázissal dolgozik, és iteratíván közelíti meg az eredményt. Ez utóbbi kitétel azért fontos, mert nincsen szükség a teljes adathalmaz memóriába töltésére ahhoz, hogy a kovarianciamátrixot meghatározzuk, hanem elegendő egyszerre egyetlen adatvektort beolvasni. Ez utóbbi előnyös lehet igen magas dimenziójú adatok analízisének, például képeknek. Az algoritmus teljes implementációja a dolgozat szerzőjének munkája.

Magát az algoritmust itt hely hiányában nem fejtjük ki részletesen, az Budavári et al. (2009) cikkében megtalálható. Kitérünk viszont néhány fontos szempontra, ami az eredeti cikk megírásakor még nem volt nyilvánvaló.

Az első fontos kérdés az algoritmus inicializálása, vagyis a kiindulási bázis meghatározása. Mint látni fogjuk ez különösen azokban az esetekben lesz fontos, amikor hiányos adatokkal dolgozunk, hiszen a hiányzó részek kipótlása mindig az aktuális bázis alapján történik. Ha a kezdeti bázis valamilyen pont a hiányos területeken torzult, ez a torzulás a végeredménybe is képes továbbgyűrűzni. Galaxispektrumok esetében jellemzően a legrövidebb és leghosszabb, valamint az ég erős vonalainak megfelelő hullámhossztartományok problémásak. A legrövidebb és leghosszabb hullámhosszakkal kapcsolatban akkor merül fel gond, ha a PCA során maximalizálni szeretnénk a lefedett hullámhossztartományt, ezért a spektrumok változó vörösetelődását kihasználva a spektrográf által mért tartományonál kékebb tartományokra is igyekszünk kiterjeszteni az analízist. Ekkor majdnem minden spektrum két végén kimaszkolt tartományok lesznek.

A 4. fejezetben bemutatandó kompozitspektrum-atlasz megkonstruálásakor a lefedett hullámhossztartomány maximalizálása érdekében a következőképpen jártunk el: Meghatároztuk a legnagyobb és legkisebb vörösetelődást, ami a minta esetében szóba jöhetett, és kiválasztottuk a 100 legkisebb és a 100 legnagyobb vörösetelődású, valamint hozzáadtunk 100 véletlenszerűen választott galaxist, és ezek alapján inicializáltuk a bázist. Az

inicializálást egyszerű duális PCA-val végeztük el. A tapasztalatok szerint ez az eljárás a bázisspektrumok szélein is jó eredményt adott.

A másik fontos kérdés, hogy mi módon egészíthetők ki a spektrumokban levő hiányos szakaszok az aktuális PCA bázis alapján úgy, hogy ez a robusztusságot biztosító súlyfüggvény alkalmazhatóságát ne befolyásolja. A probléma onnan ered, hogy az egyes spektrumokhoz tartozó súlyt úgy határozzuk meg, hogy képezzük az adatvektor és annak legjobb kifejtése közötti különbség négyzetét, melyet r^2 -tel jelölünk. Amennyiben az eredeti adatvektor sok hiányos részt tartalmaz, és a bázis alapján kell kiegészíteni, akkor ezeknek a részeknek a járuléka a súlyfüggvény paraméteréhez zérus lesz. A probléma megoldására a következő eljárást dolgoztuk ki: Amennyiben p főkomponensre vagyunk kíváncsiak, az algoritmust nem p , hanem $p + q$ dimenzióval futtatjuk. Ez az érték még mindig jóval kisebb, mint a vektortér eredeti többbezes dimenziója (p tipikusan 10 körüli, q pedig p kétszerese). Az egyes adatvektorokat a legkisebb négyzetek módszere szerint fejtjük ki az aktuális p sajátvektoron, majd még egyszer kifejtjük őket, de most már a $p + q$ sajátvektoron. A hiányos szakaszok r^2 -hez adódó járulékát ezután a p , illetve a $p + q$ elemű bázisoknál való kifejtések különbségéből származtatjuk.

Hagyományos, hiánytalan adatvektorok esetében az adatvektorokat az aktuális PCA bázison egyszerű skalárszorzat segítségével ki tudjuk fejteni. Ez hiányos adatok esetében természetesen nem működik, és ezért fordulunk a legkisebb négyzetek módszeréhez. Ennek további előnye, hogy az adatvektorok egyes elemeit külön-külön súlyozni is tudjuk. Ennek például a spektrumok binenkénti hibájának figyelembevételekor van jelentősége. Kimaskolt binek esetében a súlyt egyszerűen zérusnak vesszük.

4. fejezet

SDSS galaxisok kompozitspektrum-atlasza *

Ebben a fejezetben galaxispektrumok kompozitjainak atlaszát mutatjuk be. Az atlasz az Sloan Digitális Égtérképezés 7-es adatkiadásán alapul (SDSS DR7). A galaxisokat a színindexük, központi magjuk aktivitása és a csillagkeletkezési aktivitásuk szerint osztályozzuk, hogy nagy jel-zaj arányú ($S/N \approx 132 - 4760$, $\Delta\lambda = 1 \text{ \AA}$ felbontásnál) és nagy felbontású átlagspektrumokat származtassunk. Az átlagoláshoz egy olyan algoritmust használunk, mely robusztus a kilógó adatpontokkal szemben. A kompozit spektrumok mellett kiszámítjuk a különböző osztályokba eső galaxispektrumok eloszlásának első öt főkomponensét is abból a célból, hogy jellemezzük az egyedi spektrumok szórását az átlagok körül. A kompozit spektrumok kontinuumát BC03 csillagpopulációs szintézis-modellekkel illesztjük, hogy a lefedett hullámhossztartományt az SDSS spektrográfjainak hullámhossztartományán túlra is kiterjesszük. Meghatározzuk a kompozitok leggyakrabban használt származtatott paramétereit is: a legelterjedtebb szűrőrendszerekben mért integrális magnitúdókat, az emissziósvonal-erősségeket és a kontinuumok abszorpciós indexeit (Lick-indexek). Spektrumok ezreinek összeadásával több hónapos integrálási idő érhető el, ami különösen alacsony zajú kompozitokat eredményez. A vöröseltolódás variációi nem csak azt teszik lehetővé, hogy a kompozitok által lefedett tartományt az SDSS spektrográfok hullámhossz-lefedettségén túl kék és vörös irányba kiterjesszük, hanem azt is, hogy nagyobb felbontást érjünk el. A spektrumatlasz a következő Internet címen érhető el: <http://www.vo.elte.hu/compositeatlas>.

4.1. Bevezető

A nagy spektroszkópiai felmérések előrehaladtával a spektrumok összeadásának technika az extragalaktikus asztrofizika területén széles körben elterjedté vált. A módszer segítségével nagy jel-zaj arányú kompozit spektrumok készíthetők, melyek lehetővé teszik

olyan spektrális jellemzők tanulmányozását is, melyeket egyedi objektumok megfigyelése esetében a zaj elmos. Korábbi munkák egyes objektumosztályok átlagolására és analízisére koncentráltak, például a vörös óriásgalaxisokra (Eisenstein et al. 2003), nagy vöröseltolódású galaxisokra (Abraham et al. 2004), halmazokban található galaxisokra (Dressler et al. 2004), különböző inklinációjú csillagképző galaxisokra (Yip et al. 2010), vagy kvazárookra (Vanden Berk et al. 2001).

Minthogy az SDSS extragalaktikus spektroszkópai programja befejeződött, érdemes összesíteni a begyűjtött információt. Ebben a munkában különböző típusú galaxisokra koncentrálnak, és bemutatunk egy átfogó atlaszt, mely számos szempont alapján osztályozott galaxisok robusztusan átlagolt kompozit spektrumaiból és főkomponens spektrumaiból áll. Az atlasz elkészítéséhez az SDSS eredeti megfigyeléseit használjuk abban a formában, ahogy azokat a 7-es adatkiadásban redukálták és publikálták (Abazajian et al. 2009).

4.1.1. A galaxisok osztályozásának áttekintése

A galaxisokat automatikusan osztályozzuk egy továbbfejlesztett séma szerint, mely Lee et al. (2008) munkáján alapul. Ők a galaxisokat színindexük, központi magjuk aktivitása, csillagkeletkezési aktivitásuk és morfológiájuk alapján csoportosították.

Mi a galaxisokat három színindex osztályba osztjuk: vörös, zöld és kék galaxisok, ahol a zöld a Strateva et al. (2001) definíciója szerinti ún. „zöld völgy” galaxisait jelenti. A galaxisokat csillagkeletkezési és központi aktivitásuk alapján további hat osztályba csoportosítjuk: passzív, mérhető $H\alpha$ emisszió, csillagképző-, LINER-, valamint Seyfert-galaxisok; a hatodik csoportba azok az aktív magvú galaxisok kerülnek, melyek csillagképződésre utaló jelenségeket is mutatnak, ezekre a AGN + HII jelöléssel hivatkozunk. A csillagképző-, LINER- és Seyfert-galaxisokat a Baldwin–Phillips–Terlevich-diagram (BPT) segítségével osztályozzuk (Baldwin et al. 1981). Az osztályozáshoz a Kewley et al. (2001), Kauffmann et al. (2003c) és Kewley et al. (2006) által meghatározott szegregációs görbéket használjuk. A 4.3. fejezetben részletezett okokból a morfológiai osztályozástól eltekintünk. Az említett fejezetben részletesen ismertetjük a galaxisosztályozási sémát is.

4.1.2. A robusztus átlagolás és a főkomponensek

A spektrumok átlagának meghatározásához a Budavári et al. (2009) által kidolgozott robusztus statisztikai eljárást használjuk. Ez a módszer nem csak egy nagy, folytonos adatfolyam¹ robusztus átlagát és szórását adja meg, hanem az első néhány főkomponenst is (mintegy melléktermékként) meghatározza. Bár minket elsősorban az átlagspektrumok

¹Adatfolyam alatt olyan adathalmazt értünk, mely nem fér be a memóriába, így csak iteratív, az egyes adatvektorokat a háttértárról egyesével betöltve dolgozható fel.

érdekelnek, a saját spektrumok segítségével nyújtanak a spektrumok egy-egy galaxisosztályon belüli variációinak jellemzésében.

Az elmúlt két évtizedben a főkomponens-analízis a csillagászati spektrumok analízisének széles körben használt eszközévé vált. A PCA alap ötlete, hogy a sokdimenziós adathalmazból egy olyan ortogonális bázist származtatunk, ahol a bázisvektorok az eredeti eloszlás legnagyobb szórású irányaiba mutatnak. Korábban már mások megmutatták, hogy a PCA hatékonyan használható nagy dimenziószámú optikai és UV galaxispektrumok osztályozására (Connolly et al. 1995b). Gyakorlati helyzetekben – amikor zajos és a megfigyelési körülmények miatt egyes hullámhossz intervallumokban kimaszkolt, hiányos adatokkal kell dolgozni – kifinomultabb technikára van szükség (Connolly & Szalay 1999). Ez utóbbi robusztus technikát alkalmazták Yip et al. (2004a;b) az SDSS galaxis- és kvazárpektrumai esetében is. Budavári et al. (2009) egy olyan újfajta algoritmust fejlesztettek ki, ami kombinálja a korábbi eljárások robusztusságát és a kimaszkolt adatok kezelésére való képességét azzal, hogy a számításokat iteratívan végzi el, ami képessé teszi szokatlanul nagy adathalmazok feldolgozására is. Ez az új eljárás jelentősen lecsökkenti a számítási- és memóriaigényt, valamint lehetővé teszi, hogy magasabb dimenziójú adatokkal (pl. nagyobb felbontású spektrumokkal) is dolgozhassunk.

4.1.3. A spektrumatlasz felhasználási lehetőségei

Az itt bemutatandó spektrumatlasz két fő felhasználási lehetőségét emeljük ki. A sablon alapú *fotometrikus vöröseltolódás-becslő* eljárások (photo-z) sikerének kulcsa, hogy a megfigyelt galaxisok szélessávú spektrális energiaeloszlását megfelelő spektrumsablonokkal tudjuk illeszteni. Erre a célra gyakran modellekből származtatott spektrumsablonokat használunk, de ezek nem mindig képesek figyelembe venni a valódi spektrumok fontos jellegzetességeit. Ezen kívül a modellekből számolt integrált színindexeknek messzemenően konzisztenseknek kell lenniük a megfigyelésekkel. Alternatívaként használhatóak egyedileg megfigyelt, empirikus spektrumok is, de nem egyszerű olyan konzisztens, nagy jel-zaj arányú sablonkészletet létrehozni, mely az összes galaxistípust megfelelően reprezentálja. A mi kompozitjaink a photo-z eljárásokhoz szükséges nagy jel-zaj arányú, megbízható és reprezentatív sablonkészletet nyújtanak.

Az általunk összeállított spektrumatlasz másik fő felhasználási lehetősége szintetikus katalógusok generálása. Az atlasz összeállításához felhasznált egyedi spektrumokat kifejlesztjük a különböző galaxisosztályok főkomponens-analíziséből származó bázisokon, hogy meghatározzuk a kifejtési együtthatók eloszlásának momentumait. A kifejtési együtthatók eloszlására külön-külön Gauss-görbét illesztünk, és az illesztési paramétereket (m , σ) a saját spektrum-bázisokkal együtt publikáljuk. A kifejtési együtthatók eloszlásának utóbbi paraméterei segítségével, az átlag spektrumok és a főkomponensek lineárkombinációiból egyszerűen generálhatók újabb spektrumok, így az atlasz által lefedett fizikai paraméterek

tere tovább bővíthető.

4.2. Felhasznált adatok

Analízisünkhöz az SDSS 7-es adatkiadásából (DR7) származó galaxisspektrumokat használjuk a `spectro` feldolgozó program 2.6-os verziója szerint redukálva. Minden spektrum (beleértve a többszöri megfigyeléseket is) felhasználásra kerül, melyet a `spectro` program galaxisként azonosított, és több mint 3800 érvényesen megmért spektrális binel rendelkezik.

A fotometriai adatokat az SDSS DR7 legjobb észleléseiből vettük, az SDSS photo képfeldolgozó programja szerint redukálva (Stoughton et al. 2002).

A K-korrekciós értékeket, és az abszolút magnitúdókat a New York University értéknövelt galaxis katalógusából² vettük (Blanton et al. 2005). A K-korrekciók $z = 0,1$ -re vonatkoznak, az abszolút magnitúdók az egyezményes Λ CDM modellen alapulnak, mely paraméterei: $H_0 = 100 \text{ km s}^{-1}$, $\Omega_M = 0,3$ és $\Omega_\Lambda = 0,7$.

Az egyedileg megfigyelt galaxisok vonalerősségeit az MPA/JHU értéknövelt katalógusból vesszük (Kauffmann et al. 2003a, Tremonti et al. 2004, Brinchmann et al. 2004).

Ha másképp külön nem jelöljük, a galaxisok fényesség értékei Petrosian-magnitúdókban vannak megadva. A hullámhossz adatok mindenütt vákuumban értendők.

4.3. Az osztályozási séma

A galaxisok osztályozási sémáját Lee et al. (2008) munkáját követve alakítjuk ki, de több további osztályt is bevezetünk. Az ő sémájukban a galaxisok morfológia (korai és késői típusok), színindex (vörös és kék), valamint nukleáris aktivitás (passzív, csillagképző, LINER és Seyfert) voltak osztályozva. Ezt a sémát kiegészítjük azzal, hogy a galaxisokat a $H\alpha$ emisszió szerint is csoportosítjuk (teljesen passzív galaxisok, illetve azok, melyek mutatnak nebuláris emissziót, de az nem elegendő ahhoz, hogy a BPT diagram segítségével osztályozzuk őket). Bevezetünk továbbá egy átmeneti osztályt, mely a csillagképző, illetve aktív galaxisok közötti tartományt fedi le követve Kewley et al. (2001; 2006), valamint Kauffmann et al. (2003c) munkáit, akik ezt az osztályt „kompozitnak” nevezték. Mi ezekre az „AGN + HII” megjelölést használjuk, hogy elkerüljük a félreértést a *kompozit spektrumokkal* összefüggésben. A színindexek szerinti osztályozás kibővítésre kerül, hogy a „zöld völgy” galaxisait külön kezelhessük. A 4.3.3. fejezetben bemutatottak szerint minden galaxisosztály esetében térfogatlimitált vágásokat alkalmazunk. A legtöbb galaxist tartalmazó osztályok esetében további, színindexvágásokon alapuló osztályokat vezetünk be, erről a 4.3.4. fejezetben lesz szó. Az így bevetett osztályokra a „finomított csillagképző” vagy „SF”,

²NYU Value-Added Galaxy Catalog

illetve a „finomított vörös” vagy „RED” megjelöléseket használjuk. A kompozitok kiszámításához felhasznált galaxisok számát, a lefedett hullámhossztartományt és a becsült jel-zaj arányt a 4.1. táblázatban foglaljuk össze.

Lee et al. (2008) munkájával szemben mi elkerüljük a galaxisok morfológiai osztályozását a következők okok miatt: A morfológiai osztályozás a galaxisok radiális színindex profiljain alapul, melyeket az SDSS photo programja nem tudott tökéletesen megmérni a teljes spektroszkópiai mintára. Emiatt a morfológiai szempontok szerint kialakított osztályok mintegy 60 %-kal kevesebb galaxist tartalmaznának; csak azokat, melyek radiális színindex profilja megfelelő minőségben került meghatározásra. A morfológiai információk felhasználása tehát olyan mértékben csökkentené a felhasználható spektrumok számát, hogy az nem tenné lehetővé a célul kitűzött jel-zaj arányú és felbontású kompozitok származtatását. Ezen kívül a morfológiai osztályozhatóság függ a galaxisok méretétől és magnitúdójától is, ami az adathalmaz nem kívánt statisztikai torzulásait okozná.

A következő alfejezetekben tovább részletezzük az osztályozási kritériumokat.

4.3.1. Színindex szerinti osztályozás

Vöröseltolódásukat és $g - r$ színindexüket figyelembe véve a galaxisokat három színindex osztályba rendezzük:

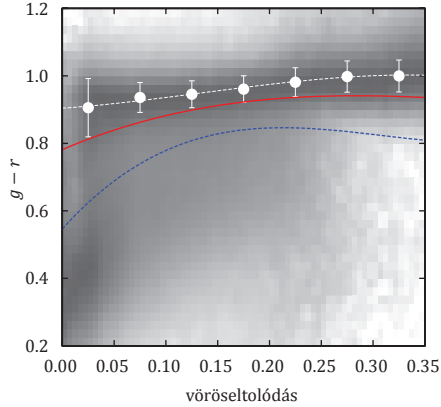
- Vörös galaxisok: Elsősorban idős csillagpopulációval rendelkező, vagy erős belső ex-tinkció által bevörösödött galaxisok.
- Zöld galaxisok: A „zöld völgybe” tartozó galaxisok, melyekről úgy véljük, hogy átmeneti régiót képeznek az aktív csillagképző és a passzív galaxisok között (Strateva et al. 2001, Bell et al. 2004, Baldry et al. 2004).
- Kék galaxisok: Leginkább aktívan csillagokat létrehozó galaxisok.

A 4.1. ábra az általunk definiált színindex osztályokat mutatja. A folytonos vörös görbe feletti galaxisok „vörösek”, a kék görbe alatti galaxisok „kékek”, a „zöld” galaxisok a két görbe között helyezkednek el. A görbék úgy lettek meghatározva, hogy a galaxisok $g - r$ színindex szerinti eloszlására Gauss-eloszlást illesztettünk $\Delta z = 0,05$ nagyságú vöröseltolódás binékben, $0,0 \leq z \leq 0,35$ között. A szaggatott fehér vonal a vörös sorozat közepe, a folytonos vörös görbe ennek $1,2\sigma$ távolságra levő burkolója, míg a kék görbe $3,5\sigma$ távolságra van a vörös sorozattól. Az ábra megkonstruálásához az összes SDSS galaxist felhasználtuk, függetlenül azok látszólagos vagy abszolút magnitúdójától.

A 4.1. ábrán levő színes görbéket a vörös sorozatra illesztett Gauss-eloszlások σ paraméterei határozzák meg, ahogy azt az előzőekben bemutattuk. Mivel a vöröseltolódás növekedésével a fotometriai hiba is növekszik, a vörös sorozat szórása is nagyobb. Ez azt

osztály	λ_{\min} [Å]	λ_{\max} [Å]	szám	S/N
vörös P	3350	8900	10431	1884
vörös H α	3350	8950	61237	3996
vörös SF	3550	8950	481	225
vörös A+HII	3350	8950	726	329
vörös L	3250	8850	901	483
vörös S	3350	8950	577	347
vörös összes	3350	8950	86690	4760
zöld P	3350	8900	2095	791
zöld H α	3350	8950	19246	1858
zöld SF	3550	8950	1533	482
zöld A+HII	3350	8900	1067	431
zöld L	3250	8800	327	233
zöld S	3350	8950	942	442
zöld összes	3350	8950	26613	2148
kék P	3350	8650	312	242
kék H α	3350	8900	26550	1672
kék SF	3550	8950	14304	1088
kék A+HII	3350	8900	2725	541
kék L	3250	8800	215	132
kék S	3350	8900	1157	386
kék összes	3350	8900	40504	1856
minden szín P	3350	8900	12838	1670
minden szín H α	3350	8950	107033	3911
minden szín SF	3550	8950	16318	1035
minden szín A+HII	3350	8950	4518	610
minden szín L	3250	8850	1443	470
minden szín S	3350	8950	2676	569
minden szín összes	3350	8950	153807	4464
RED 1	3350	8950	17913	1777
RED 2	3350	8950	22934	2458
RED 3	3350	8950	46470	3939
RED 4	3350	8950	17546	2303
RED 5	3350	8950	2469	629
SF 1	3550	9000	301	192
SF 2	3550	9000	2862	650
SF 3	3550	9050	7073	1000
SF 4	3550	9000	4436	798
SF 5	3550	9000	1948	513

4.1. táblázat. A kompozit spektrumok alapvető paraméterei aktivitási és színindexosztályok szerint.



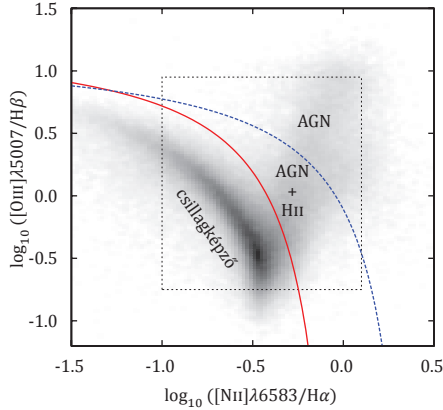
4.1. ábra. A színdexek szerinti vágások definíció. Satírozás: a galaxisminta eloszlásának logaritmus, vöröseltolódás intervallumonként normálva. Vörös galaxisok: a folytonos vörös görbe fölött; kék galaxisok: a szaggatott kék görbe alatt; zöld völgy galaxisok: a két színes vonal között. A feltüntetett adatpontok és a fehér szaggatott görbe a vörös sorozat csúcsát, és a vöröseltolódás függvényében azokra legjobban illeszkedő görbét jelölik.

eredményezi, hogy a színes görbék $z \approx 0,2$ körül visszafordulnak, és jobban megszorítják a „zöld völgy” és kék galaxisokat, több szórást engedve a vörös galaxisoknak. Mivel relatíve kevés zöld völgy és kék galaxis figyelhető meg ilyen magas vöröseltolódásokon, ennek nem lesz számottevő hatása a kompozit spektrumokra, de így több vörös galaxis lesz használható az analízis céljaira. Jelen munkában az összes alminta vöröseltolódás-vágása $z \leq 0,17$ -nél kisebb lesz.

4.3.2. Aktivitás szerinti osztályozás

A csillagkeletkezés és nukleáris aktivitás szerinti osztályozás kialakításánál Kewley et al. (2001) és Kauffmann et al. (2003c) munkáit vesszük alapul. A LINER/Seyfert szeparáció Kewley et al. (2006) dolgozatát követi. Aktivitás szerint az alábbi hat osztályt definiáljuk:

- Passzív galaxisok: nincsen mérhető $H\alpha$ emisszió.
- mért $H\alpha$: A hidrogén emissziója mérhető, de a többi vonal nem elegendő a galaxis további osztályozásához.
- Csillagképző galaxis: világos csillagképződésre utaló jellegzetességek, de nincsen aktív mag.
- AGN + HII galaxisok: AGN és extrém csillagontó jellegzetességek egy időben.



4.2. ábra. Az SDSS galaxisokból álló mintánk BPT-diagramja. A sátrózás azon SDSS galaxisok számsűrűségének négyzetgyöke, melyek emissziós vonalai elegendően jó jel-zaj viszonytal lettek detektálva (ld. szöveg). A folytonos vörös vonal a csillagképző galaxisok és az aktív magvú galaxisok közötti szegregációs görbe. A szaggatott vonalú négyszög azt a területet jelöli, melyet a 4.15. ábra is mutat. Ld. a szöveget a hivatkozásokhoz.

- LINER galaxisok: tiszta LINER jellegzetességek (Heckman 1980).
- Seyfert-galaxisok: tiszta Seyfert jellegzetességek.

A 4.2. ábrán felrajzoljuk a mintánkban szereplő galaxisok BPT-diagramját. A szegregációs görbe (folytonos vörös vonal), mely elválasztja a csillagképző galaxisokat (a görbe alatt) és az AGN-eket (a görbe fölött) a következő:

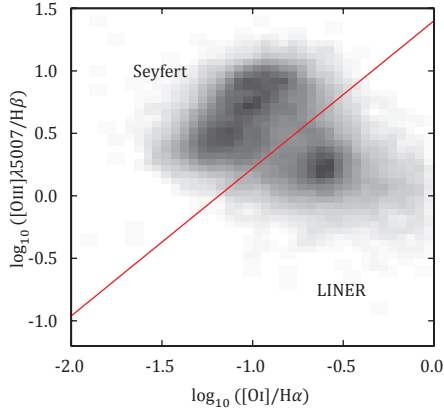
$$\log_{10} \frac{[\text{OIII}]}{\text{H}\beta} = 0,61 \left[\log_{10} \frac{[\text{NII}]}{\text{H}\alpha} - 0,05 \right]^{-1} + 1,3.$$

Az extrém csillagotó galaxisok felső határa (kék szaggatott vonal) pedig:

$$\log_{10} \frac{[\text{OIII}]}{\text{H}\beta} < 0,61 \left[\log_{10} \frac{[\text{NII}]}{\text{H}\alpha} - 0,47 \right]^{-1} + 1,19.$$

Ez utóbbi vonal választja el a tiszta AGN-eket (a görbe fölött) az olyan AGN-ektől, melyek heves csillagkeletkezési jellegzetességeket is mutatnak (a két görbe közötti terület).

A 4.3. ábra azokat a galaxisokat mutatja, melyek aktív magvúként (csillagképző jelleg nélkül) lettek osztályozva. A következő vágást alkalmazzuk, hogy megkülönböztessük a



4.3. ábra. Seyfert-galaxisok és LINER-ek szeparációja. A sátrózás azon SDSS galaxisok eloszlásának négyzetgyökét mutatja, melyeket aktív magvúként (csillagkeletkezési jelleg nélkül) osztályozunk a 4.2. ábra segítségével. A vonal a bimodális eloszlás két felét választja el. Ld. a szöveget a hivatkozásokhoz.

LINER-eket (a vonal alatt) a Seyfert-galaxisoktól (a vonal felett):

$$\log_{10} \frac{[\text{OIII}]}{\text{H}\beta} = 1,18 \log_{10} \frac{[\text{O I}]}{\text{H}\alpha} + 1,4$$

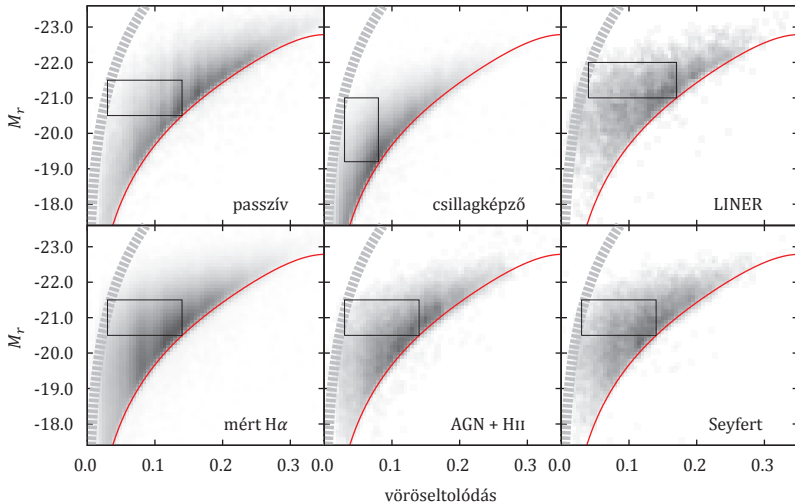
Csak azok a galaxisok lettek mért $\text{H}\alpha$ -ként klasszifikálva, melyeknél a vonal erőssége legalább $S/N \geq 3$ jel-zaj viszony mellett lett megmérve, míg az AGN (akár AGN + HII , LINER vagy Seyfert) besoroláshoz megköveteltük a $\text{H}\alpha$, $\text{H}\beta$, $[\text{O I}]$, $[\text{OIII}]$ és $[\text{NII}]$ vonalak minimum $S/N \geq 3$ jel-zaj viszonyát. Azon galaxisok számára, melyek nem rendelkeznek elegendően nagy jel-zaj viszonyú megfigyeléssel két további csoportot definiálunk: teljesen passzív galaxisok detektálható emissziós vonalak nélkül, valamint passzív galaxisok mérhető $\text{H}\alpha$ emisszióval.

Bár kevés van belőlük, kizártuk azokat a galaxisokat, melyek extrém vonalarányokat mutatnak, azaz a 4.2. ábrán kívül esnek.

4.3.3. Térfogatlimitált minták

A Malmquist-torzítás elkerülése érdekében minden egyes galaxisosztályhoz definiálunk egy vöröseltolódás- és egy abszolút magnitúdó-intervallumot úgy, hogy közben megpróbáljuk a térfogatlimitált mintákba kerülő galaxisok számát maximalizálni.

Bár a vöröseltolódások és az abszolút magnitúdók szűk határok közé kényszerítése lecsökkenti a kompozitok által lefedett hullámhossztartományt, erre mindenképpen szükség van, hogy fizikailag értelmezhető eredményt kapjunk. Az SDSS spektroszkópiai kivá-



4.4. ábra. A vöröseltolódás- és az r szűrőben mért M_r abszolút magnitúdó-vágások definíciói minden aktivítási osztályra. Satírozás: az adott osztályba tartozó galaxisok számsűrűségének négyzetgyöke; folytonos vörös vonal: a halvány látszólagos magnitúdó-határ, mely 17,77 mag-nak felel meg; szürke vonal: a becsült fényes látszólagos magnitúdó-határ, mely 13,5 mag-nak felel meg. A vágások az SDSS *spectro* minta magnitúdóhatárait figyelembe véve, manuálisan lettek meghatározva úgy, hogy az egyes osztályokba minél több galaxis essék.

lasztási algoritmusai (Eisenstein et al. 2001, Strauss et al. 2002, Stoughton et al. 2002) a fő galaxis mintának a halvány magnitúdók felől 17,7 mag-nál (az r színszűrőben véve) éles határt szabnak. Ugyanakkor a fényes magnitúdók oldaláról a minta teljessége sokkal kevésbé kontrollált. A minta hiányosságának egyik fő oka, hogy a hozzánk közel levő galaxisok látszólagos átmérője túl nagy a spektroszkópiai üvegszál átmérőjéhez képest. Amikor a térfogatokat meghatározzuk, úgy becsüljük, hogy a fényes látszólagos magnitúdó határ 13,5 mag körül van az r színszűrőben mérve.

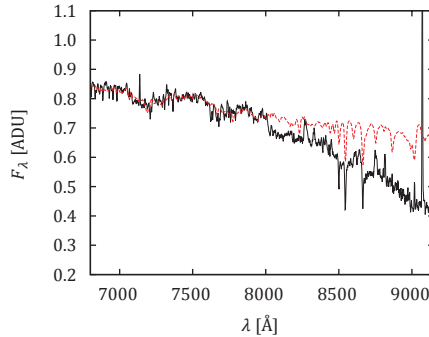
Az általunk definiált térfogatokat a 4.4 ábrán jelöljük, és a 4.2. táblázatban foglaljuk össze.

A nem teljes mintavételezés hatása

Megjegyezzük, hogy az átlagolással számolt kompozit spektrumok kontinuumja fizikai jelentéssel bír olyan értelemben, hogy az valódi csillagpopulációk nem negatív együtthatós lineárkombinációjaként adódik. Ez értelmessé teszi azt, hogy a kompozitokra elméleti populációsintézis-modelleket illesszünk. A galaxisok kompozitszámolásra történő hiányos mintavételezése azonban azt eredményezi, hogy a kompozitok nem lesznek fizikailag

	z_{\min}	z_{\max}	$M_{r,\min}$	$M_{r,\max}$
passzív	0,03	0,14	-20,5	-21,5
mért H α	0,03	0,14	-20,5	-21,5
csillagképző	0,03	0,08	-19,2	-21,0
AGN + HII	0,03	0,14	-20,5	-21,5
LINER	0,04	0,17	-21,0	-22,0
Seyfert	0,03	0,14	-20,5	-21,5

4.2. táblázat. A vöröseltolódás- és abszolút magnitúdó-vágások értékei a különböző galaxisaktivitási osztályok esetében.



4.5. ábra. A hiányos mintavételezés hatása. Bár alacsony vöröseltolódású spektrumok felhasználása lehetővé tenné, hogy a kompozitokat vörös irányba kiterjesszük, de a fényes objektumok hiánya azt eredményezi, hogy a kiátlagolt spektrum hosszú hullámhosszakra lefelé görbül, ami elrontja a kompozitok fizikai értelmezhetőségét. Folytonos fekete görbe: hiányos minta alapján számított kompozit spektrum, szagatott vörös görbe: a legjobban illeszkedő modell. (A kontinuum illesztése a 3500 – 9000 Å tartományon történt.)

értelmezhetők.

Míg az általunk megszabott szűk vöröseltolódás-tartományokon belül a spektrális evolúciót elhanyagoljuk, fel kell hívnunk a figyelmet a spektrumok az objektumok abszolút magnitúdójától való függésére. Például, habár egy alacsonyabb vöröseltolódásbeli vágás lehetővé tenné, hogy a kompozitok által lefedett hullámhossztartományt vörös irányba kiterjesszük, a mintából hiányzó fényes galaxisok miatt ebben a tartományban a halványabb galaxisok dominálnának, melyek (a kék galaxisok esetében legalább is) általában kékebbek. Ez azt eredményezné, hogy a kompozit spektrumok a vörös végükön lefelé kanyarodnának, ahogyan azt a 4.5. ábra is szemlélteti.

4.3.4. Kék és vörös galaxisok finomított osztályozása

A sablon alapú fotometrikus vöröseltolódás-becslő (photo-z) algoritmusok (összefoglalóért ld. Hildebrandt et al. (2010) és a benne foglalt hivatkozásokat) jó minőségű empirikus spektrumokra vagy elméleti modellekre támaszkodva illesztik meg a szélessávú megfigyeléseket. Habár létezik néhány olyan technika is, melyekkel növelhető a modellek megbízhatósága (Budavári et al. 2000), elméleti modellek használata esetén a leggyakoribb probléma, hogy a modellekből számolt színindexek eltérnek a megfigyelt galaxisok színindexeitől. A photo-z technikák rendszerint kis számú sablonnal dolgoznak; egyrészt azért, hogy a számítások bonyolultsága kisebb legyen, másrészt hogy elkerüljék azokat a hamis eredményeket, melyek abból adódnak, hogy nem túl megbízható modelleket túl sok szabad paraméterrel illesztenek.

A legnépesebb galaxisosztályok (passzív vörös elliptikusok és csillagképző kék spirálok) színindexeiben jelentős szórás tapasztalható. Túl kevés illesztendő sablon használata arra vezet, hogy a photo-z által szolgáltatott vöröseltolódás-értékek túl durva becslések lesznek, mivel a sablonok nem elegendőek az objektumok színindexbeli nagy varianciájának jellemzésére. Megbízhatóbb empirikus sablonok használata azonban segíthet a fotometrikus vöröseltolódás-becslés finomításában. Hogy ezt a problémát is érintsük, a legnépesebb galaxisosztályokat tovább finomítjuk a színindexek alapján, és meghatározzuk a finomított osztályok kompozit spektrumait is.

Kék galaxisok

A csillagképző galaxisok (főleg spirálok és irregulárisok) esetében a színindex eredendő szórását több effektus okozza. A galaxisok nyugalmi rendszerben mért színeit elsősorban a csillagkeletkezési ráta szabja meg, de a portartalom és az inklináció is jelentős szerepet játszik (Yip et al. 2010). Mivel a korlátos apertúrával rendelkező üvegszálás spektrográfokkal csak a galaxisok középső, vörösebb régióit figyeljük meg, a galaxisok távolsága és mag-tányér aránya is fontos tényezők a kis apertúrákból meghatározott színindexek szórásának eredetében. A kék galaxisok osztályát a továbbiakban „SF” jelöléssel látjuk el.

Vörös galaxisok

A vörös galaxisok (főleg elliptikusok) esetében a színindexek szórásának az eredete elsősorban egy evolúciós effektus, amit a galaxisok passzívan fejlődő öreg csillagpopulációinak változása okoz (egy átfogó összefoglalóért ld. Renzini (2006) és a benne levő hivatkozásokat). Úgy véljük, hogy az elliptikus galaxisokban levő por minimális mértékben felelős a színindexek szórásáért.

A finomított színindexek szerinti osztályozás esetében a korábbi „passzív” és „mért $H\alpha$ ” kategóriájú galaxisokat együtt kezeljük, és a „RED” jelöléssel hivatkozunk rájuk. Az egyes

szín szerinti osztályokat számokkal jelöljük.

A finomított színindex osztályok

A csillagképző és vörös galaxisok finomított színindex osztályait a galaxisok $g - r$ színindexe alapján definiáljuk annak megfelelően, ahogy azt a 4.6. ábrán jelöljük. A galaxisok eloszlásának hisztogramját Gauss-görbékkel illesztjük, és a színindexbeli határokat az átlagtól $0,5\sigma$ és $1,5\sigma$ távolságokra definiáljuk. Összesen öt-öt finomított színindexosztályt vezetünk be, melyeket a következő táblázat foglal össze:

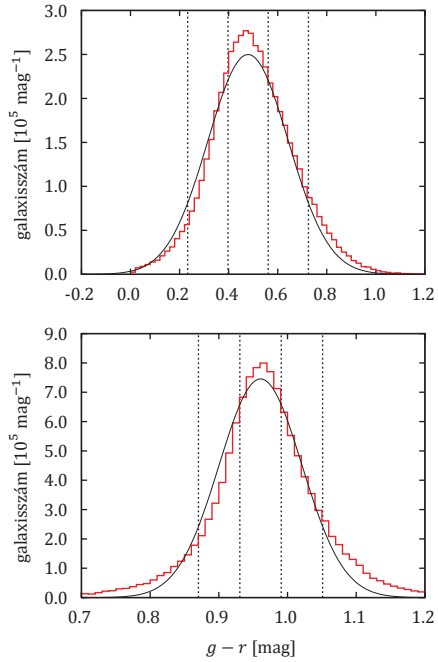
#	határok
1	$g - r < m - 1,5\sigma$
2	$m - 1,5\sigma \leq g - r < m - 0,5\sigma$
3	$m - 0,5\sigma \leq g - r \leq m + 0,5\sigma$
4	$m + 0,5\sigma < g - r \leq m + 1,5\sigma$
5	$m + 1,5\sigma < g - r$

Az m és σ paraméterek a színindexek eloszlására illesztett normális eloszlások átlagát és szórását jelölik. Az illesztett paramétereket az alábbi táblázat foglalja össze:

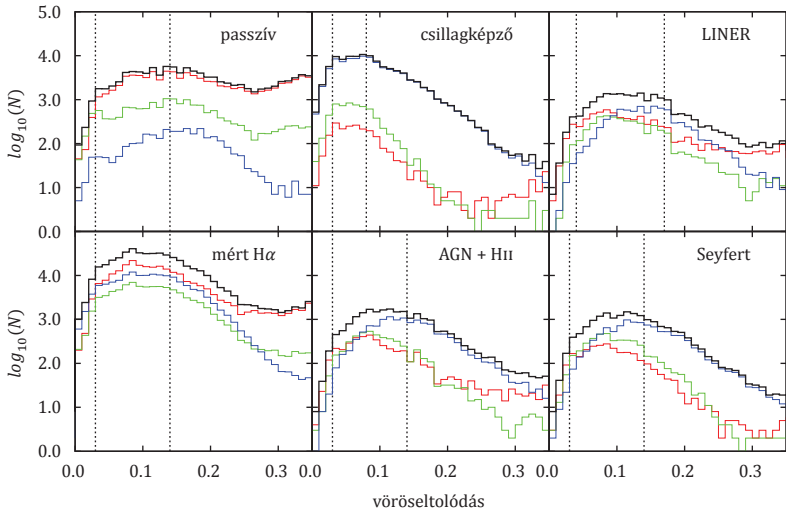
	m	σ
SF	0,48	0,16
RED	0,95	0,06

4.3.5. Vöröseltolódás szerinti eloszlás

A 4.7. ábrán valamennyi aktivitási osztályra felrajzoljuk a galaxisok vöröseltolódás szerinti eloszlását $\Delta z = 0,01$ binékben. A görbék alapján nyilvánvaló, hogy a vörös galaxisokat kivéve a felrajzolt vöröseltolódás-tartományban valamennyi minta az SDSS spektroszkópiai kiválasztási feltételeinek megfelelően fluxuslimitált. A vörös galaxisok mintájának fényes vége $z = 0,38$ -ig térfogatlimitált (Eisenstein et al. 2001, Loh & Strauss 2006). A 4.8. ábra ugyanezeket a vöröseltolódás szerinti eloszlásokat mutatja, de már a térfogatlimitált minták előállításának céljával bevezetett magnitúdóbeli vágások alkalmazása *után*. Megjegyezzük, hogy amennyiben térfogatlimitált minták előállítása a célunk, a minták galaxis-számának maximalizálása nem lehetséges a vöröseltolódás-intervallumok szabad megválasztása mellett; a vöröseltolódás-intervallumok meghatározásánál az abszolút magnitúdó-vágásokat is figyelembe kell venni.



4.6. ábra. A csillagképző (felső panel) és passzív (alsó panel) galaxisok $g-r$ színindexeinek eloszlása. A vörös hisztogramok a mért eloszlást mutatják, a fekete görbék pedig az illetett Gauss-eloszlásokat. A függőleges szaggatott vonalak a színindexvágásokat jelölik az átlagtól $0,5\sigma$ és $1,5\sigma$ távolságokra.



4.7. ábra. Az egyes aktivitási osztályok galaxisainak vöröseltolódás szerinti eloszlása az abszolút magnitúdó-vágások alkalmazása *előtt*. A színek a különböző színindexosztályokat jelölik: vörös – vörös galaxisok; zöld – a zöld völgy galaxisai; kék – kék galaxisok; valamint fekete – az összes galaxis, a színtől függetlenül. A függőleges szaggatott vonalak az aktivitási osztályonként különböző vöröseltolódásbeli vágásokat jelzik.

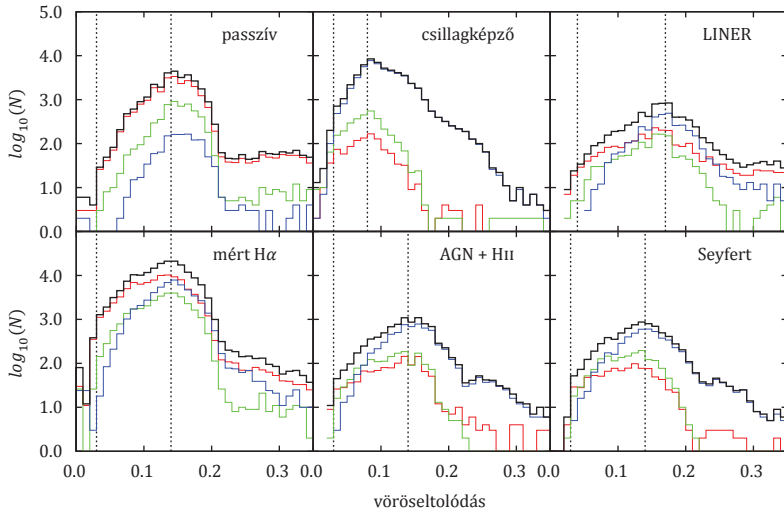
4.4. A kompozit spektrumok

4.4.1. Előfeldolgozás

A kompozitok kiszámítása előtt a spektrumokat a 4.3. fejezetben leírtak szerint választjuk ki, majd elvégezzük az előfeldolgozásukat. Először minden spektrumot korrigálunk a Tejút által okozott extinkcióra, melyhez Schlegel et al. (1998) portérvképét, és O'Donnell (1994) extinkciós görbéjét használjuk. Második lépésben a spektrumokat a 4.4.2. fejezetben leírtak szerint normáljuk. Harmadik lépésben a spektrumokat nyugalmi rendszerbe konvertáljuk, majd az igényelt felbontásnak ($\Delta\lambda = 1, 0, 0,5$, illetve $0,25 \text{ \AA}$) megfelelő rácson újra-mintavételezzük őket.

4.4.2. Normálás

Ahhoz, hogy fizikailag értelmezhető kompozitokat kapjunk, a galaxisspektrumok robusztus normálására van szükség. Mivel a zajos spektrumok fluxusának egy adott hullámhossznál történő normálása nem vezet eredményre, a következő módszert alkalmazzuk: Meghatározzuk a fluxussűrűség mediánját a $4200\text{-}4300 \text{ \AA}$, $4600\text{-}4800 \text{ \AA}$, $5400\text{-}5500 \text{ \AA}$ és



4.8. ábra. Az egyes aktivitási osztályok galaxisainak vöröseltolódás szerinti eloszlása az abszolút magnitúdó-vágások alkalmazása után. A színek a különböző színindexosztályokat jelölik: vörös – vörös galaxisok; zöld – a zöld völgy galaxisai; kék – kék galaxisok; valamint fekete – az összes galaxis, a színtől függetlenül. A függőleges szaggatott vonalak az aktivitási osztályonként különböző vöröseltolódásbeli vágásokat jelzik.

5600-5800 Å hullámhossz-intervallumokban, és az így kapott érték segítségével skálázzuk össze a spektrumokat. A hullámhossz-intervallumok úgy lettek meghatározva, hogy ne tartalmazzanak erős emissziós vonalakat. A spektrumokat úgy skálázzuk össze, hogy a felsorolt intervallumokban véve a fluxussűrűség mediánja mindig egy legyen.

A spektrumok fluxussűrűségét más módon (például az objektumok látszólagos magnitúdóitól, abszolút fényességétől, vagy a megfigyelések jel-zaj arányától függően) nem súlyozzuk vagy skálázzuk. Bár ez az eljárás a kompozitok jel-zaj arányát negatívan befolyásolhatja, az általunk alkalmazott robusztus átlagoló algoritmus ezt a problémát kiküszöböli úgy, hogy az kisebb jel-zaj arányú spektrumokhoz automatikusan kisebb súlyt rendel.

4.4.3. A robusztus algoritmus

A különböző galaxisosztályok kompozit spektrumának meghatározásához a Budavári et al. (2009) által kidolgozott robusztus átlagolási eljárást alkalmazzuk, mivel ez képes arra, hogy az egyedi spektrumok jelentős zaját és változó jel-zaj arányát megfelelően figyelembe vegye. Az algoritmus nem csak arra képes, hogy a spektrumok kimaszkolt részeit is megfelelően kezelve, iteratív módon határozza meg a hullámhossz függvényében vett átlag-

got és szórást, de emellett megadja a spektrum eloszlások első néhány főkomponensét is (ld. 4.4.5. fejezet). A 4.4.6. fejezetben külön kitérünk a robusztus eljárás hatékonyságára az eredményeket a hagyományos átlaggal és mediánnal összehasonlítva.

A robusztus átlagolási eljárással meghatározott nagy felbontású kompozitokat a Függelék 1–4. ábrái szemléltetik.

4.4.4. Nagy felbontás

Mivel a kompozit spektrumok számos, változó vöröseltolódáson rögzített spektrum kombinációjából kerülnek előállításra, az eredeti megfigyelések felbontásánál néhány faktorral nagyobb felbontás is elérhető. Az SDSS nagy jel-zaj arányú ($S/N \geq 30$) galaxispektrumainak esetében a vöröseltolódás mérésének hibája $\sigma_z = 1,6 \times 10^{-4}$, mely lehetővé teszi a kompozitok túlmintavételezését. Másrészt viszont a túlmintavételezés az 1 \AA binelésű, csaknem teljesen zajtalan kompozitokhoz képest zajt okozhat. Ez a zaj a kevés galaxist tartalmazó osztályok esetében (mint például a zöld völgy galaxisai) jelentős lehet.

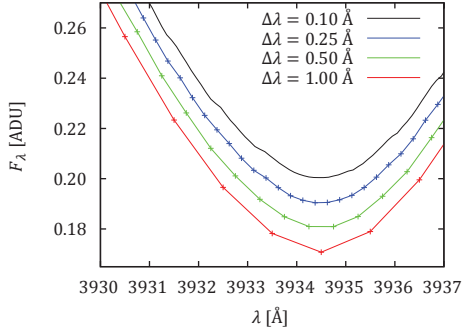
Egy másik fontos tényező, amely a nagyobb felbontású kompozitok előállítása ellen hat az, hogy a galaxisok csillagainak nagy sebességű mozgása kiszélesíti az egyes spektrumok jellegzetességeit. Ahhoz, hogy ezt a problémát elkerüljük, a Doppler-kernel eredeti megfigyelésekből való dekonvolválására lenne szükség. Ezt a témát ebben a dolgozatban részletesebben nem boncolgatjuk.

A 4.9. ábra egy kompozit spektrum néhány spektrális jellegzetességét mutatja különböző túlmintavételezés mellett (1 \AA , $0,5 \text{ \AA}$, $0,25 \text{ \AA}$ és $0,1 \text{ \AA}$ felbontásnál). Megfigyelendő, ahogy a vonalak egyre simábbá válnak a növekvő felbontással. Fontos megemlíteni ugyanakkor, hogy a túlmintavételezett spektrumok nem feltétlenül hordoznak több információt, mivel a széles sebességszperzió eleve elmossa a finom részleteket.

Ugyanakkor egyértelmű, hogy egyszerű lineáris interpolációval nem tudjuk előállítani a nagyobb felbontású kompozitokat a kisebb felbontásúakból. Magasabb rendű görbék ugyan használhatóak lennének arra, hogy az alacsony felbontású kompozitokból sima vonalprofilokat kapjunk, de ez a módszer gyakran műtermékeket produkál. A mi túlmintavételezéses módszerünk kiküszöböli a magasabb rendű interpolációk okozta műtermékek megjelenését, mégis sima vonalprofilokat állít elő.

4.4.5. Főkomponensek

Az egyes galaxisosztályok átlagos spektruma mellett meghatározzuk a hullámhosszfüggő szórást is. Ez utóbbi lehetővé teszi a különösen nagy szórású hullámhossztartományok azonosítását (legtöbbször az emissziós vonalak), valamint a kompozitok zajának hullámhosszbinenkénti meghatározását. Ugyanakkor, amennyiben valakit az egész spektrum változékonysága érdekel (szemben a binenkénti varianciával), a bineket egymástól függet-



4.9. ábra. Egy galaxispektrum-kompozit K vonala különböző felbontások mellett számolva ($\Delta\lambda = 1, 0,5, 0,25$ és $0,1 \text{ \AA}$ - alulról felfelé). A nagyobb felbontású görbék simábbak, mint a $\Delta\lambda = 1 \text{ \AA}$ görbe.

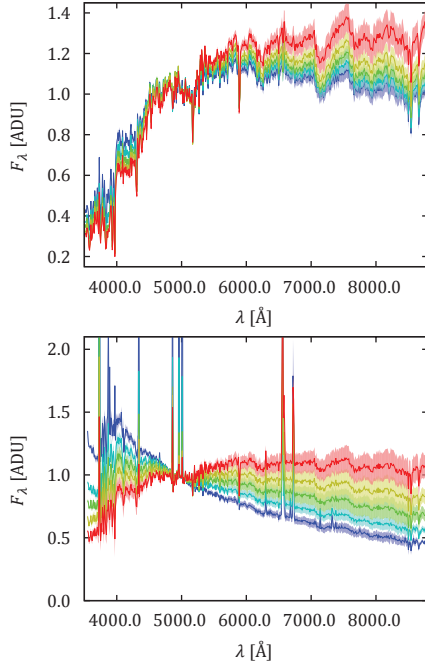
lenül figyelembe vevő szóráspektrum nem elegendő. Ahhoz, hogy az egy galaxisosztályon belüli spektrumok változékonyságát a teljes lefedett hullámhossztartományt alapul véve írjuk le, szükségünk van a binck közötti kovariancia figyelembe vételére is; azaz a spektrumsokaság első néhány főkomponensének meghatározására. Szerencsére az általunk használt robusztus átlagolási algoritmus mintegy az átlagolás „melléktermékeként” szolgáltatja a főkomponenseket.

Meghatározzuk minden egyes galaxisosztály első öt főkomponensspektrumát, majd az egyedi, adott osztályba tartozó galaxisok spektrumait kifejtjük az így nyert bázison. A kifejtési együtthatók eloszlása a legtöbb esetben nagyon közel van egy nulla átlagú normális eloszláshoz. Ebből következően minden galaxisosztály esetében meghatározzuk a kifejtési együtthatók szórását, majd a következő spektrumokat tekintjük:

$$F_{\lambda}^{\pm} = F_{\lambda}^{\text{avg}} \pm k\sigma_{\text{eig}} F_{\lambda}^{\text{eig}},$$

ahol F_{λ}^{avg} az átlagspektrum és F_{λ}^{eig} valamelyik saját-spektrum, σ pedig az ehhez a saját-spektrumhoz tartozó, az adott galaxisosztályba tartozó spektrumok kifejtési együtthatóinak szórása. A k koefficiens egy tényező, amit variálva némileg különböző spektrumok állíthatók elő.

A 4.10. ábrán F_{λ}^{\pm} -t rajzoltuk fel a vörös és kék finomított színindexosztályokra. A vörös galaxisok esetében az első saját-spektrum egyértelműen a galaxis színét határozza meg; a spektrumok meredekségében tapasztalható szórást az első főkomponens csaknem teljes egészében képes leírni. A csillagképző galaxisok esetében a legnagyobb szórást az emissziós vonalak okozzák, tehát a színbeli variációknak megfelelő spektrumok előállításához a második sajátvektort kell felhasználni. Ahogyan a 4.10. ábra szemlélteti, ez a technika sikeresen alkalmazható tetszőleges, az előre definiált osztálybeli átlagok közé



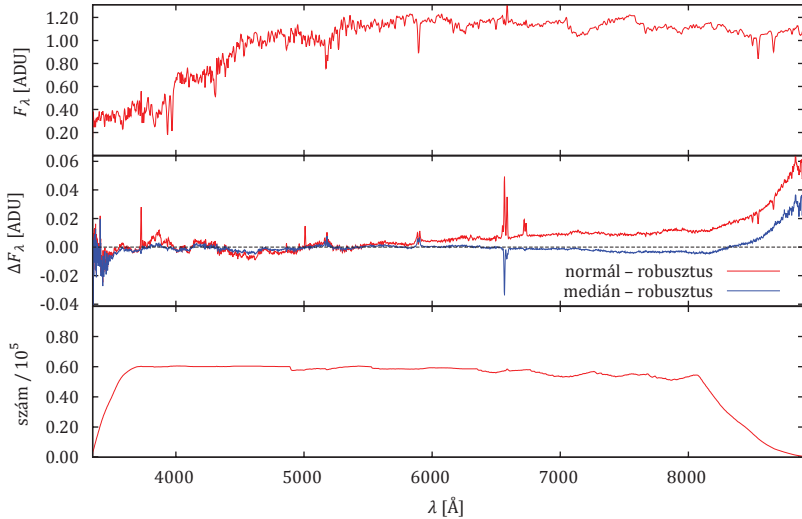
4.10. ábra. A vörös (felső panel) és a kék csillagképző (alsó panel) galaxisok spektrumainak szórása. A folytonos görbék a finomított színindexosztályok (vörösebbtől a kék felé) kompozit spektrumait mutatják. A satírozott területek a 0.5σ (vörös galaxisok), illetve 1σ (kék galaxisok) változékonyságot mutatják az átlag körül.

eső színindexű spektrumok generálására. A módszer egy fontos hátránya, hogy magasabb rendű saját spektrumok az átlaghoz történő hozzáadása némi zajt hoz be.

4.4.6. A robusztus algoritmus minőségének jellemzése

Abból a célból, hogy a robusztus átlagoló eljárás hatékonyságát megállapítsuk, a robusztus kompozitokat hagyományosan átlagolással és medián számolással meghatározott spektrumokkal hasonlítjuk össze. A 4.11. és 4.12. ábrán a robusztus átlag és a nem robusztus kompozitok különbségét rajzoltuk fel (középső panelek). Az ábrákon maga a robusztus kompozit (felső panel) és az egyes hullámhossz binkekhez tartozó átlagfluxus meghatározásához felhasznált spektrumok száma (alsó panel) is fel lett tüntetve.

Érdekes módon a vörös galaxisok esetében a medián sokkal közelebb fut a robusztus átlaghoz, mint a normál átlag, míg a kék galaxisok esetében a különbség nem annyira szig-

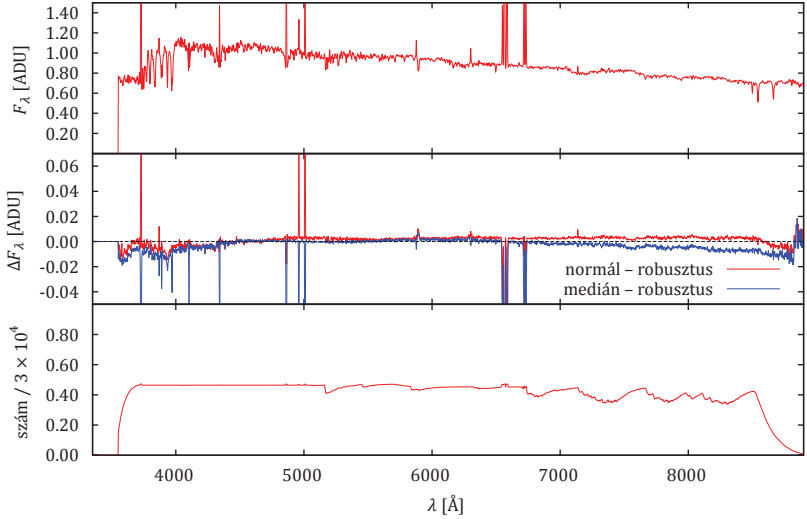


4.11. ábra. A különböző átlagolási eljárások összehasonlítása a mért H α -val rendelkező galaxisok osztályára. A felső panel a robusztus kompozit spektrumot mutatja. A középső panel a robusztus átlag eltéréseit mutatja a normál átlagtól és a mediántól. Az alsó panel görbéje az egyes hullámhosszbinék átlagos fluxusának meghatározásához felhasznált spektrumok számát adja meg.

nifikáns. A robusztus és nem robusztus kompozitok között számottevő eltérés a nagyon rövid és nagyon hosszú hullámhosszaknál, valamint az emissziós vonalaknál található. A kontinuumok közötti eltérés valószínűleg annak köszönhető, hogy a szélső hullámhossztartományokat kevés spektrum alapján határozzuk meg. Úgy véljük, hogy kevés spektrum esetében a nem robusztus eljárások erősen torzított eredményeket adnak. A spektrális vonalak tartományában biztosan állítható, hogy a robusztus algoritmus kisebb súlyt rendel az extrém erős emissziós vonalakat mutató galaxisokhoz.

4.5. Modellillesztés

Az empirikusan származtatott nagy jel-zaj arányú kompozit spektrumaink által lefedett hullámhossztartomány kb. 3350 Å-tól 9050 Å-ig tart. A pontos értékek azoktól a vöröseltolódás határoktól függenek, amiket a galaxisosztályok térfogatlimitált mintáinak definiálásához használtunk; nagyobb vöröseltolódású minták lehetővé teszik a lefedettség kiterjesztését kékebb hullámhosszakra is. Hogy a korlátos lefedettséget kiterjeszthessük, a kompozit spektrumokat Bruzual & Charlot (2003) csillagpopulációs szintézis-modelljéből származó elméleti modellspektrumokkal illesztjük. Három modellhalmazt generálunk há-



4.12. ábra. Ugyanaz, mint a 4.11. ábrán, de a csillagképző kék galaxisok esetében.

rom különböző fémesség értékkel: $Z = 0,008, 0,02$ és $0,05$. Minden halmaz 10 különböző korú, egy időben keletkezett modellspektrumot tartalmaz, melyek kora: $t = 0,005, 0,025, 0,100, 0,290, 0,640, 0,900, 1,4, 2,5, 5,0$, valamint $11,0$ milliárd év. A modellekhez jól ismert kezdeti tömegeloszlás-függvényt (Chabrier 2003) és a „Padova1994” csillagfejlődési trajektóriákat (Alongi et al. 1993, Bressan et al. 1993, Fagotto et al. 1994a;b) használtuk. A kompozit spektrumokhoz legjobban illeszkedő, különböző korú, de azonos fémességű modellek lineárkombinációjának meghatározásához a nem-negatív együttthatós legkisebb négyzetek módszerét használjuk. A por által okozott belső extinkció hatásának megbecsléséhez Charlot & Fall (2000) egyszerű pormodelljét alkalmazzuk, melyet a következő formula ad meg:

$$\tau_{\lambda}(t) = \begin{cases} \tau_V(\lambda/5500 \text{ \AA})^{-0,7}, & \text{ha } t \leq 10^7 \text{ év} \\ \mu\tau_V(\lambda/5500 \text{ \AA})^{-0,7}, & \text{ha } t > 10^7 \text{ év} \end{cases}$$

Az illesztéskor a τ_V optikai mélységet változóként kezeljük, míg μ értékét $\mu = 0,3$ -en rögzítjük. Ebben a bizonyos pormodellben a μ értéke azt szabja meg, hogy az extinkció milyen arányban befolyásolja az 10 millió évnél idősebb, illetve fiatalabb csillagok fényét. A csillagok sebességszóróját megadó σ_v paramétert szintén szabad paraméternek tekintjük. Az illesztést minden egyes modell esetében az összes különböző fémességű modellhalmazzal elvégezzük, és végül a legjobban illeszkedő modellhalmazt használjuk.

A szélessávú színszűrőkben vett szintetikus magnitúdók meghatározásakor, amennyi-

ben a kompozitok hullámhossz lefedettsége nem elegendő, az empirikus spektrumokat a legjobban illeszkedő modellekkel egészítjük ki.

4.6. Fizikai paraméterek meghatározása

A 4.5. fejezetben alkalmazott modell illesztési módszerről ismert, hogy bár a kontinuumot megfelelően becsli, különböző okok miatt nem alkalmas a galaxisokat alkotó csillagpopulációk fizikai paramétereinek pontos becslésére. A galaxisok korának és fémességének meghatározásához kifinomultabb módszerek szükségesek.

Ahhoz, hogy a kompozit spektrumok kontinuumának fizikai paramétereit megbecsüljük, egy sztochasztikus csillagkeletkezési modelleken alapuló Bayes-statisztikai módszert alkalmazunk, ld. 2.3.1. fejezet, valamint Kauffmann et al. (2003a). Az eredményül kapott fizikai paramétereket a Függelék 1. táblázata tartalmazza.

4.7. Kompozitok a „térképen”

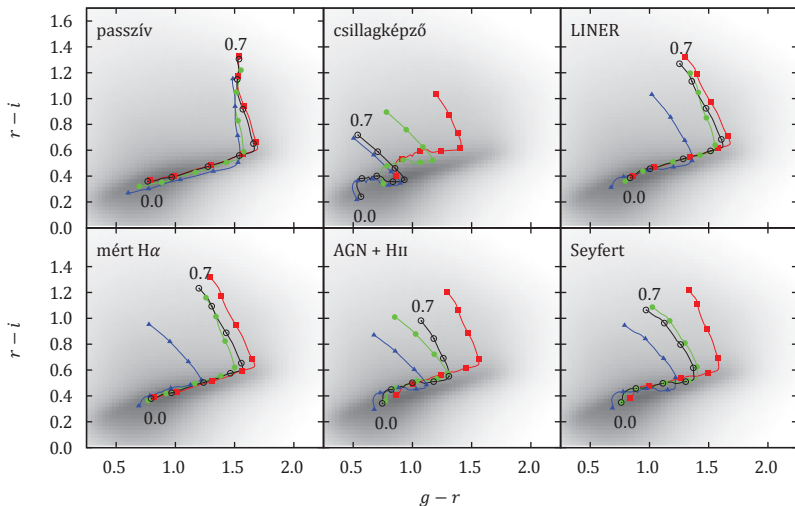
4.7.1. Szín-szín trajektóriák

A 4.13. és 4.14. ábra az összes galaxisosztályra számolt kompozit vöröseltolódás szerint vett szín-szín trajektóriáját mutatja az SDSS $g - r$ és $r - i$ színindexeiben. A trajektóriák a $0,0 \leq z \leq 0,7$ vöröseltolódás-tartományt fedik le, és szépen átfednek a galaxisok megfigyelt eloszlásával. Ez egy jó igazolása annak, hogy a robusztus átlagolási eljárás nem okoz statisztikai torzulásokat a kompozitok színindexeiben.

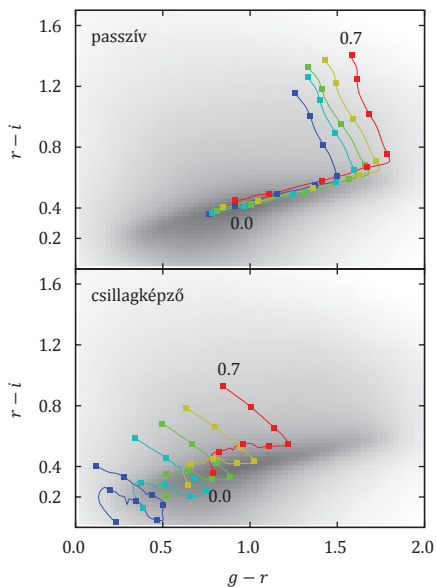
4.7.2. Emissziósvonal-mérések

Az 4.15. ábrán a csillagképző, AGN + HII, LINER és Seyfert-galaxisok BPT diagramját rajzoljuk fel. A kompozitok pozíciói egybe esnek az egyedi mérésekből számolt vonalarányokkal. Ez az eredmény nem triviális: az átlagok aránya nem feltétlenül az arányok átlaga. Szerencsére a mi esetünkben a robusztus átlagolási eljárás nem változtatja meg a vonalarányokat. Ugyanakkor fontos megjegyeznünk, hogy a galaxisok vonalarányainak szórása jóval nagyobb annál, mint amit néhány kompozit spektrum képes jellemezni.

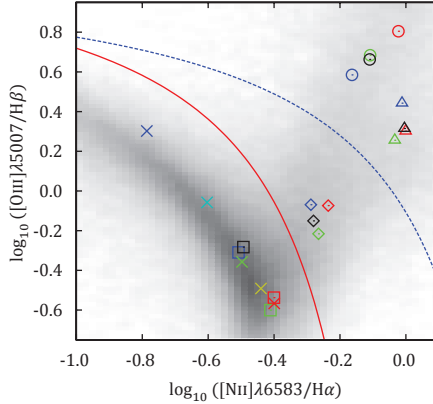
A 4.15. ábra egy érdekessége, hogy a Seyfert-galaxisok egy jól definiált vonal mentén helyezkednek el, mely vonalat a $g - r$ szín paraméterez: a vörösebb Seyferteknek jellemzően nagyobb az $[OIII]/H\beta$ arányuk. Ez konzisztens Kewley et al. (2006) megállapításával, miszerint az AGN-ek csillagpopulációi annál öregebbek, minél távolabb helyezkednek el a BPT diagramon a csillagkeletkezési vonaltól. Az AGN + HII galaxisok esetében hasonló viselkedés nem figyelhető meg. A csillagképző galaxisok esetében a finomított színindexosztályok alapján számolt kompozitokat is ábráztuk. Ebben az esetben szintén világos



4.13. ábra. A kompozitok szín-szín diagramjai $0 \leq z \leq 0.7$ közötti vöröseltolódásokon az SDSS $g-r$ és $r-i$ színindexeire felrajzolva. A színes görbék a 4.3.1. fejezetben definiált különböző színindexosztályok kompozit spektrumait jelzik. A fekete görbék a színindextől független kompozit spektrumok trajektóriái. A görbék menti pontok a $z = 0.0, 0.1$ stb. vöröseltolódásokat jelzik. A háttérsatrozás a spektroszkópiai minimumnál jóval mélyebb minta, a fotometriailag azonosított SDSS galaxisok $3''$ -os apertúrában vett, extinkcióra korrigált színindexeinek eloszlását mutatja.



4.14. ábra. A kompozitok szín-szín diagramjai $0 \leq z \leq 0.7$ közötti vöröseltolódásokon az SDSS $g-r$ és $r-i$ színindexeire felrajzolva. A színes görbék a 4.3.4. fejezetben definiált különböző finomított színindexosztályok kompozit spektrumait jelzik (felső panel: vörös galaxisok, alsó panel: csillagképző galaxisok). A görbék menti pontok a $z = 0.0, 0.1$ stb. vöröseltolódásokat jelzik. A háttérszírozás a spektroszkópiai mintánál jóval mélyebb minta, a fotometriailag azonosított SDSS galaxisok $3''$ -os apertúrában vett, extinkcióra korrigált színindexeinek eloszlását mutatja.



4.15. ábra. A kompozit spektrumok BPT diagramja. A sátrózás az SDSS galaxisok sűrűségeloszlásának logaritmusja. A vörös folytonos görbe a csillagképző és aktív magvú galaxisok közötti szegregációs vonalat jelzi. A kék szaggatott görbe az extrém csillagotó galaxisok határvonala. A kompozitokat pontok jelölik: négyzetek – csillagképző galaxisok; rombuszok – AGN+HII galaxisok; háromszögek – LINER-ek, körök – Seyfert-galaxisok. A különböző színek a különböző színindexosztályokat jelentik, a fekete a színindexektől függetlenül generált kompozitokat mutatja. A színes keresztet a csillagképző galaxisok finomított színindexosztályai alapján meghatározott kompozitok helyét jelölik.

trend figyelhető meg: a kékebb galaxisoknak magasabb az $[OIII]/H\beta$ arányuk, melyhez alacsonyabb $[NII]/H\alpha$ arány tartozik.

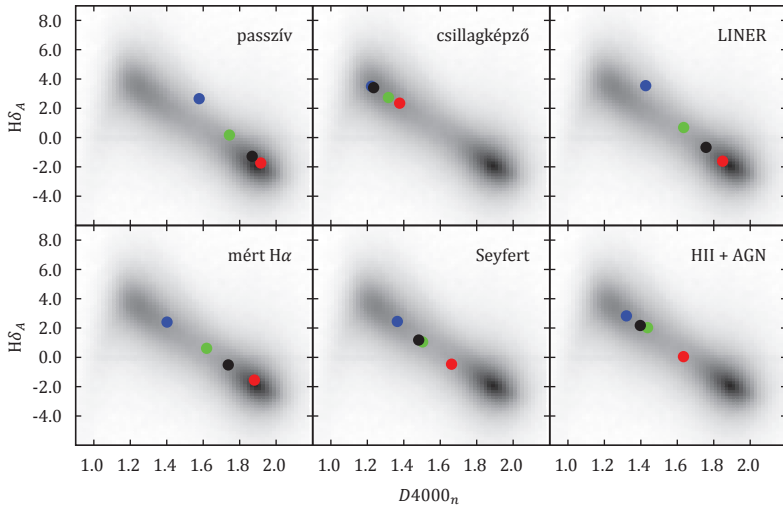
Az emissziós vonalak ekvivalens szélességét a Függelék 2–4. táblázatai foglalják össze.

4.7.3. Sztellárkontinuum-indexek

A kompozitok számos abszorpciósindexét határoztuk meg: Lick-indexek (Worthey et al. 1994, Worthey & Ottaviani 1997), BH-indexek (Brodie & Hanes, ld. Huchra et al. (1996)) és DTT-indexek (Diaz et al. 1989). A 4000 Å körüli törés nagyságát megadó indexek Bruzual A. (1983) és Balogh et al. (1999) munkáiból származnak.

Példaképpen felrajzoljuk a 4.16. ábrát, mely $H\delta_A$ -t mutatja a $D4000_n$ függvényében. Az ábrázolt mennyiségek a nem rég történt csillagkeletkezés jó indikátorai (Bruzual A. 1983, Balogh et al. 1999, Kauffmann et al. 2003a;b). A kompozitok világosan mutatják a színindex – $H\alpha$ – $H\delta_A$ korrelációt. A két érdekes galaxisosztály a passzív kék galaxisoké és a LINER-eké, melyek az egyenes mentén elhelyezkedő pontoktól kissé arrébb vannak, és jó eséllyel csillagkeletkezésen éppen hogy átesett galaxisok.

Az összes galaxisosztályra kiszámolt kompozit spektrum abszorpciós indexeit a Függelék 5–7. táblázatai foglalják össze.



4.16. ábra. A kompozitok nem rég végbement csillagkeletkezésre utaló jelei. A satírozás az összes SDSS galaxis sűrűségeloszlásának négyzetgyökét mutatja. A kompozitok elhelyezkedését körök jelölik. A különböző színűek a színindexosztályokra utalnak, a fekete pontok a különböző aktivitási osztályok azon kompozitjait jelölik, melyek a színindexektől függetlenül lettek meghatározva.

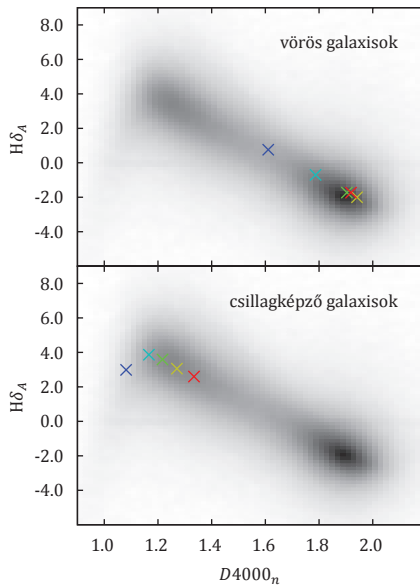
4.8. Empirikus alapú szimulált katalógusok generálása

Az általunk összeállított atlasz egyik lehetséges felhasználási lehetősége, hogy a kompozitok alapján szimulált katalógusokat³ állíthatunk össze. A különböző galaxisosztályok saját spektrumai, az egyedi spektrumok sajátbázisokon vett kifejtési együtthatóit jellemző eloszlások alapvető paramétereivel egyetemben on-line elérhetők. Ez utóbbi paraméterek felhasználásával, az átlag kompozit spektrumok és a saját spektrumok lineárkombinációjaként egyszerűen generálhatók újabb spektrumok, így tovább bővíthető a fizikai paramétertérnek az atlasz által lefedett tartománya.

4.9. Összefoglalás

Ebben a részben bemutattunk egy átfogó kompozitspektrum-atlaszt, melyet az SDSS által megfigyelt galaxisok spektroszkópiái adatai alapján állítottunk össze, valamint kiszámítottunk a kompozitok legfontosabb spektrális jellemzőinek értékeit. Megmutattuk, hogy a kompozitok paramétereik majdnem olyan széles értéktartományokat ölelnek át, mint az

³mock catalogs



4.17. ábra. A kompozitok nem rég végbement csillagkeletkezésére utaló jelei a finomított színindexosztályok esetében. A sátrózás az összes SDSS galaxis sűrűségeloszlásának négyzetgyökét mutatja. A keresztek a finomított színindex-osztályozáson alapuló kompozitok elhelyezkedését jelölik.

egyedi galaxispektrumok esetében, ami az atlaszt nagy számú galaxistípus reprezentatív gyűjteményévé teszi. A kompozitspektrum-atlasz elérhetősége az interneten:
<http://www.vo.elte.hu/compositeatlas>

Fizikai mennyiségek jelölése

jelölés	jellemző egység	mennyiség
A_λ	mag	extinkció
c	km s^{-1}	fénysebesség
DM	mag	távolságmodulus
E	mag	evolúciókorrekció
$E(B - V)$	mag	színindextöbblet
F	$\text{s}^{-1} \text{cm}^{-2}$	fotonfluxus
F_λ	$\text{erg s}^{-1} \text{cm}^{-2} \text{\AA}^{-1}$	spektrális fluxussűrűség
I_λ	$\text{erg s}^{-1} \text{cm}^{-2} \text{\AA}^{-1} \text{sr}^{-1}$	intenzitás
K	mag	K-korrekció
λ	\AA	hullámhossz
m_{AB}	mag	AB magnitúdó
$m_{u,g,r,i,z}$	mag	SDSS magnitúdó az u, g, r, i, z színszűrőkben
\mathcal{M}	\mathcal{M}_\odot	csillagtömeg
$\mu_{u,g,r,i,z}$	mag arcsec $^{-2}$	felületi fényesség az u, g, r, i, z színszűrőkben
ψ	$\mathcal{M}_\odot \text{év}^{-1}$	csillagkeletkezési ráta
$r(\lambda)$	-	színszűrő átviteli függvénye
R	-	spektrális felbontás
S_λ	$\text{erg s}^{-1} \text{cm}^{-2} \text{\AA}^{-1} \mathcal{M}_\odot^{-1}$	fajlagos spektrális fluxussűrűség
SN	-	jel-zaj arány
σ_v	km s^{-1}	sebességszórás
t	10^9 év	kozmikus idő
τ_λ	-	hullámhosszfüggő optikai mélység
W_λ	\AA	ekvivalens szélesség
ξ	\mathcal{M}_\odot^{-1}	kezdeti tömegeloszlás-függvény
z	-	vöröseltolódás
Z	Z_\odot	fémesség (metallicitás)
ζ	Z_\odot	kémiai evolúció

Irodalomjegyzék

- Abazajian, K. N., et al. 2009, *ApJS*, 182, 543
- Abraham, R. G., et al. 2004, *AJ*, 127, 2455
- Alongi, M., Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., Greggio, L., & Nasi, E. 1993, *A&AS*, 97, 851
- Antonucci, R. 1993, *ARA&A*, 31, 473
- Antonucci, R. R. J., & Miller, J. S. 1985, *ApJ*, 297, 621
- Armandroff, T. E., & Zinn, R. 1988, *AJ*, 96, 92
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, *ApJ*, 600, 681
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, *ApJ*, 527, 54
- Bell, E. F., et al. 2004, *ApJ*, 608, 752
- Benn, C. R., & Ellison, S. L. 1998, *NewAR*, 42, 503
- Bica, E., & Alloin, D. 1987, *A&A*, 186, 49
- Blanton, M. R., et al. 2003, *AJ*, 125, 2348
- . 2005, *AJ*, 129, 2562
- Bressan, A., Fagotto, F., Bertelli, G., & Chiosi, C. 1993, *A&AS*, 100, 647
- Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, *MNRAS*, 351, 1151
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Bruzual A., G. 1983, *ApJ*, 273, 105
- Bruzual A., G., & Charlot, S. 1993, *ApJ*, 405, 538
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, *AJ*, 120, 1588
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., Dickinson, M. E., & The Hdf/Nicmos Team. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 191, *Photometric Redshifts and the Detection of High Redshift Galaxies*, ed. R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, 19–+
- Budavári, T., Wild, V., Szalay, A. S., Dobos, L., & Yip, C.-W. 2009, *MNRAS*, 394, 1496
- Calzetti, D. 2001, *PASP*, 113, 1449
- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, 429, 582
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Carrera, R., Gallart, C., Pancino, E., & Zinn, R. 2007, *AJ*, 134, 1298
- Chabrier, G. 2003, *PASP*, 115, 763
- Charbonnel, C., Däppen, W., Schaerer, D., Bernasconi, P. A., Maeder, A., Meynet, G., & Mowlavi, N. 1999, *A&AS*, 135, 405
- Charbonnel, C., Meynet, G., Maeder, A., & Schaerer, D. 1996, *A&AS*, 115, 339
- Charlot, S., & Fall, S. M. 2000, *ApJ*, 539, 718
- Charlot, S., & Longhetti, M. 2001, *MNRAS*, 323, 887
- Colless, M., et al. 2001, *MNRAS*, 328, 1039
- Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., & Munn, J. A. 1995a, *AJ*, 110, 2655
- Connolly, A. J., & Szalay, A. S. 1999, *AJ*, 117, 2052
- Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., & Calzetti, D. 1995b, *AJ*, 110, 1071
- Csabai, I., Connolly, A. J., Szalay, A. S., & Budavári, T. 2000, *AJ*, 119, 69
- Csabai, I., et al. 2003, *AJ*, 125, 580
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, 388, 1595
- Diaz, A. I., Terlevich, E., & Terlevich, R. 1989, *MNRAS*, 239, 325
- Dressler, A., Oemler, Jr., A., Poggianti, B. M., Smail, I., Trager, S., Shectman, S. A., Couch, W. J., & Ellis, R. S. 2004, *ApJ*, 617, 867
- Drineas, P., Mahoney, M. W., & Muthukrishnan, S. 2008, *SIAM J. Matrix Anal. Appl.*, 30, 844
- Eisenstein, D. J., et al. 2001, *AJ*, 122, 2267
- . 2003, *ApJ*, 585, 694
- Faber, S. M., et al. 2007, *ApJ*, 665, 265
- Fagotto, F., Bressan, A., Bertelli, G., & Chiosi, C. 1994a, *A&AS*, 104, 365
- . 1994b, *A&AS*, 105, 29
- Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
- Foster, C., Forbes, D. A., Proctor, R. N., Strader, J., Brodie, J. P., & Spitler, L. R. 2010, *AJ*, 139, 1566
- Francis, P. J., Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1992, *ApJ*, 398, 476
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Girardi, L., Bressan, A., Chiosi, C., Bertelli, G., & Nasi, E. 1996, *A&AS*, 117, 113
- Glazebrook, K., Offer, A. R., & Deeley, K. 1998, *ApJ*, 492, 98
- Gordon, K. D., & Clayton, G. C. 1998, *ApJ*, 500, 816
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- . 2006, *AJ*, 131, 2332
- Györy, Z., & Bell, E. F. 2010, *ApJ*, 724, 694
- Györy, Z., Szalay, A. S., Budavári, T., Csabai, I., & Charlot, S. 2011, *AJ*, 141, 133
- Heavens, A. F., Jimenez, R., & Lahav, O. 2000, *MNRAS*, 317, 965
- Heckman, T. M. 1980, *A&A*, 87, 152
- Hildebrandt, H., et al. 2010, *A&A*, 523, A31+
- Ho, L. C. 2008, *ARA&A*, 46, 475
- Hogg, D. W. 1999, *ArXiv Astrophysics e-prints*
- Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, *ArXiv Astrophysics e-prints*
- Huchra, J. P., Brodie, J. P., Caldwell, N., Christian, C., &

- Schommer, R. 1996, *ApJS*, 102, 29
- Jones, D. H., et al. 2004, *MNRAS*, 355, 747
- Kauffmann, G., et al. 2003a, *MNRAS*, 341, 33
- . 2003b, *MNRAS*, 341, 54
- . 2003c, *MNRAS*, 346, 1055
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, *ApJ*, 556, 121
- Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, *MNRAS*, 372, 961
- Kurucz, R. L. 1992, in *IAU Symposium*, Vol. 149, *The Stellar Populations of Galaxies*, ed. B. Barbuy & A. Renzini, 225+
- LAMOST website. 2011, <http://www.lamost.org>
- Lawson, C. L., & Hanson, R. J. 1974, *Solving least squares problems*
- Le Borgne, J.-F., et al. 2003, *A&A*, 402, 433
- Lee, J. H., Lee, M. G., Park, C., & Choi, Y.-Y. 2008, *MNRAS*, 389, 1791
- Lejeune, T., Cuisinier, F., & Buser, R. 1997, *A&AS*, 125, 229
- Loh, Y.-S., & Strauss, M. A. 2006, *MNRAS*, 366, 373
- Mahoney, M. W., & Drineas, P. 2009, *PNAS*, 106, 697
- Maraston, C. 2005, *MNRAS*, 362, 799
- Maraston, C., Strömbäck, G., Thomas, D., Wake, D. A., & Nichol, R. C. 2009, *MNRAS*, 394, L107
- Martin, D. C., et al. 2007, *ApJS*, 173, 342
- Misselt, K. A., Clayton, G. C., & Gordon, K. D. 1999, *ApJ*, 515, 128
- O'Donnell, J. E. 1994, *ApJ*, 422, 158
- Osterbrock, D. E., & De Robertis, M. M. 1985, *PASP*, 97, 1129
- Panther, B., Heavens, A. F., & Jimenez, R. 2003, *MNRAS*, 343, 1145
- Pearson, K. 1901, *Philosophical Magazine*, 2, 559
- Pickles, A. J. 1985, *ApJS*, 59, 33
- Renzini, A. 2006, *ARA&A*, 44, 141
- Richards, G. T., et al. 2002, *AJ*, 123, 2945
- Salim, S., et al. 2005, *ApJ*, 619, L39
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, *A&AS*, 96, 269
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- SDSS website. 2011, <http://www.sdss.org>
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, *ApJ*, 509, 103
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Strateva, I., et al. 2001, *AJ*, 122, 1861
- Strauss, M. A., et al. 2002, *AJ*, 124, 1810
- Szalay, A. S., Kunszt, P. Z., Thakar, A. R., Gray, J., & Slutz, D. 2000, in *Astronomical Society of the Pacific Conference Series*, Vol. 216, *Astronomical Data Analysis Software and Systems IX*, ed. N. Manset, C. Veillet, & D. Crabtree, 405+
- Tegmark, M., et al. 2004, *ApJ*, 606, 702
- Tremonti, C. A., et al. 2004, *ApJ*, 613, 898
- Urry, C. M., & Padovani, P. 1995, *PASP*, 107, 803
- Vanden Berk, D. E., et al. 2001, *AJ*, 122, 549
- . 2004, *ApJ*, 601, 692
- Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., Cenarro, A. J., Beasley, M. A., Cardiel, N., Gorgas, J., & Peletier, R. F. 2010, *MNRAS*, 404, 1639
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, 63, 295
- Wild, V., Charlot, S., Brinchmann, J., Heckman, T., Vince, O., Pacifici, C., & Chevallard, J. 2011, *ArXiv e-prints*
- Wild, V., & Hewett, P. C. 2005, *MNRAS*, 358, 1083
- Wild, V., Kauffmann, G., Heckman, T., Charlot, S., Lemson, G., Brinchmann, J., Reichard, T., & Pasquali, A. 2007, *MNRAS*, 381, 543
- Wilhite, B. C., Vanden Berk, D. E., Brunner, R. J., & Brinkmann, J. V. 2006, *ApJ*, 641, 78
- Wilhite, B. C., Vanden Berk, D. E., Kron, R. G., Schneider, D. P., Pereyra, N., Brunner, R. J., Richards, G. T., & Brinkmann, J. V. 2005, *ApJ*, 633, 638
- Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, *ApJS*, 94, 687
- Worthey, G., & Ottaviani, D. L. 1997, *ApJS*, 111, 377
- Wyder, T. K., et al. 2007, *ApJS*, 173, 293
- Yanny, B., et al. 2009, *AJ*, 137, 4377
- Yip, C. W., & Csabai, I. 2011, (in preparation)
- Yip, C.-W., Szalay, A. S., Wyse, R. F. G., Dobos, L., Budavári, T., & Csabai, I. 2010, *ApJ*, 709, 780
- Yip, C. W., et al. 2004a, *AJ*, 128, 585
- . 2004b, *AJ*, 128, 2603
- . 2009, *AJ*, 137, 5120

II. rész

Vörös óriásgalaxisok statisztikája *

Bevezető a II. részhez

Ebben a részben a sorrendi statisztika módszerének segítségével vizsgáljuk az SDSS DR7 katalógusban szereplő, spektroszkópiailag is megfigyelt, nagyon fényes ($M_r < -22$) korai típusú galaxisok luminozitásfüggvényét. Megmutatjuk, hogy megadható egy olyan általános, normált luminozitásfüggvény, mely a $0,1 \leq z \leq 0,5$ közötti széles vöröseltolódás-tartományban sorrendi statisztikai megfontolások alapján megfelelően jellemzi az elliptikus galaxisok eloszlását. Numerikusan kiszámítjuk a luminozitásfüggvény tipikus sorrendi statisztikai paramétereit, azaz meghatározzuk egy véletlenszerűen választott N elemszámú minta k . legfényesebb galaxisának várható fényességét és a fényesség szórását. Az így kiszámított statisztikai mennyiségek alapján megmagyarázzuk, hogy miért, és milyen hátfeltételekkel használhatóak a k . legfényesebb galaxisok standard gyertyaként a kozmológiai célú munkákban.

A sorrendi statisztika módszereinek alkalmazására példaként megmutatjuk, hogy egy adott vöröseltolódás-tartományban jó becsléssel megadható a galaxisok száma, ha ismerjük a tartományba eső néhány legfényesebb galaxis abszolút magnitúdóját, valamint a luminozitásfüggvény alakját (amiről feltesszük, hogy nem változik lényegesen z függvényében). Először demonstráljuk, hogy egy adott vöröseltolódás-intervallumban levő galaxisok számanak ismeretében megbecsülhető a k . legfényesebb korai típusú galaxis abszolút magnitúdója. Az eljárás megfordításával a galaxiszám is nagyon egyszerűen meghatározható, például olyan vöröseltolódás-tartományokban, ahol csak a legfényesebb galaxisokról áll rendelkezésünkre spektroszkópiai mérés.

Mivel az általunk használt minta minden fényes galaxist tartalmaz, beleértve a legfényesebb klasztergalaxisokat⁴ (BCG-k) is, a sorrendi statisztikára hivatkozva megindokoljuk, hogy a BCG-k miért tekinthetők egy jól meghatározott Schechter-luminozitáseloszlás statisztikus extrémumainak akkor, ha a galaxisokat vöröseltolódás szerint bineljük, és nem klaszterről klaszterre vizsgáljuk. Úgy véljük, hogy emögött az áll, hogy a galaxishalmazokban levő vörös óriásgalaxisok nem tekinthetők egy általános luminozitáseloszlásból *véletlenszerűen* választott mintának, hanem az eloszlásukat torzítja az a tény, hogy ők egy olyan galaxishalmaz tagjai, amely tartalmaz egy legfényesebb klasztergalaxist. Megmutatjuk, hogy egy egyszerű, de jól megkonstruált statisztikai játékmódel is alkalmas arra, hogy reprodukálja a jól ismert magnitúdórést, mely a galaxishalmazok legfényesebb és második legfényesebb galaxisai között tátong.

⁴BCG = brightest cluster galaxy

5. fejezet

A vörös óriásgalaxisok luminozitásfüggvényének extrémérték- és sorrendi statisztikája

A vörös óriásgalaxisok¹ (LRG) fizikai tulajdonságai lehetővé teszik, hogy belőlük egy széles vöröseltolódás-tartományban konzisztensen evolváló mintát válasszunk. Az LRG-k az Univerzum legnagyobb tömegű csillagpopulációi, és ez a tulajdonságuk az Univerzum anyageloszlásának kiváló nyomjelzőjévé teszi őket, hiszen a mai észlelési technikákkal $z \approx 1,2$ -es vöröseltolódásig spektroszkópiailag is megfigyelhetőek. A spektrumukban a 4000 Å-nél található jelentős törésnek köszönhetően a vöröseltolódásuk viszonylag könnyen meghatározható, még kis jel-zaj viszony mellett rögzített spektrumokból is (Eisenstein et al. 2001). Úgy gondoljuk, hogy az LRG-k $z \geq 2$ vöröseltolódás előtt, vagy annak környékén álltak össze, és a csillagpopulációik jelentős része azóta passzívan fejlődik, számottevő csillagkeletkezés ezekben a galaxisokban az elmúlt sokmilliárd év során nem volt (Gunn & Oke 1975, Ellis et al. 1997, Aragón-Salamanca et al. 1998, van Dokkum et al. 1998, Stanford et al. 1998, Burke et al. 2000, Wake et al. 2006, Maraston et al. 2009). Az, hogy ezek a galaxisok az Univerzum ilyen korai időszakában születtek, korlátot szab a köztük levő korbeltől eltérésre is, ezt a későbbiekben ki fogjuk használni. Ebben a részben bemutatjuk, hogy hogyan állítottunk össze egy konzisztens mintát a Sloan Digitális Égtérképezés 7-es adatkiadásának (SDSS DR7) adataiból, mely mintán a címben szereplő statisztikai eljárások demonstrálhatók.

A sorrendi statisztika (melyet a $k = 1$ speciális esetben szélsőérték-, vagy extrémérték-statisztikának nevezünk) a statisztika azon területe, mely adott valószínűségi eloszlásokból húzott, adott méretű minták k . legnagyobb (legkisebb) elemének várható értékét és momentumait igyekszik meghatározni. A várható értékeket, a magasabb momentumokat

¹LRG = luminous red galaxy

és egyéb statisztikai mennyiségeket az N mintaméret és a k rend függvényében határozzuk meg. Bár a sorrendi statisztika számítási feladatai analitikus megközelítéssel meglehetősen bonyolultak lehetnek, a felírt zárt formulák Monte Carlo-módszerrel nagyon egyszerűen integrálhatóak. Ez tulajdonképpen ekvivalens azzal a megközelítéssel, hogy az adott valószínűségi eloszlás alapján véletlenszerű mintákat generálunk (asztroroztatistikai nyelven megfogalmazva: egy adott luminozításfüggvény alapján szimulált katalógusokat készítettünk), ezek elemeit sorba rendezzük, majd a keresett mennyiségeket a sok véletlenszerűen generált mintára való átlagolással számíttjuk ki. Az extrémérték-statisztika a független, azonos eloszlású, valamint gyengén korrelált valószínűségi változók esetére számos tétellel rendelkezik az extrémumok határeloszlásaira vonatkozóan: a szülőeloszlások a végtelenben való viselkedésük, lecsengési gyorsaságuk alapján véges számú univerzalizációs osztályba sorolhatók, és az azonos osztályba tartozó eloszlások extrémumai azonos eloszlásúak (Fisher & Tippett 1928, Gnedenko 1948). Ezek a tételek természetesen az $N \rightarrow \infty$ esetre vonatkoznak. Mi a statisztikai mennyiségeket numerikusan, véges N -ekre fogjuk meghatározni.

A legfényesebb LRG-k galaxiscsoportok és -halmazok központi régióiban találhatóak. Megfelelő evolúciókorrekció elvégzése után kiderül, hogy az LRG-k valódi fényességének halmazról halmazra vett szórása meglehetősen kicsi, így ezek a galaxisok használhatóak standard gyertyaként (Sandage 1972; 1976, Postman & Lauer 1995, Whiley et al. 2008). Széles körben vitatott téma azonban, hogy vajon ezek a legfényesebb klasztergalaxisok tekinthetőek-e egy általános, a többi vörös elliptikus galaxist is leíró fényességeloszlásfüggvény extrémumaiként, vagy egy külön populációhoz tartoznak (Geller & Peebles 1976, Geller & Postman 1983, Loh & Strauss 2006, De Lucia & Blaizot 2007, von der Linden et al. 2007, Liu et al. 2008, Lin et al. 2010). Az utóbbi eset bizonyos kiválasztási effektusok létét feltételezné, továbbá azt, hogy az őket létrehozó, és az evolúciójukat megszabó fizikai folyamatok eltérnek a többi galaxist alakító folyamatoktól. A főbb statisztikán alapuló érvek egyike amellet, hogy a BCG-k a többi korai típusú galaxis eloszlásától különböző eloszlásból származnak az, hogy a BCG-k és a második legfényesebb klasztergalaxisok magnitúdói között halmazról halmazra vizsgálva egy átlagosan kb. $\Delta M \simeq 0,8$ mag nagyságú magnitúdórés tapasztalható (Loh & Strauss 2006).

Korábbi extrémérték-statisztikán alapuló munkák már vizsgálták a legfényesebb galaxisok fényességeloszlását, és arra a megállapításra jutottak, hogy a legfényesebb, *galaxis-halmazokban* található galaxisok eloszlása különbözik a legfényesebb, *galaxiscsoportokban* található galaxisokétól. Ez utóbbiakról az az állítás fogalmazódott meg, hogy fényességeloszlásuk egy olyan szülőeloszlásból származik, mely a végtelenben bármely hatványfüggvénynél gyorsabban cseng le (Bhavsar & Barrow 1985, Bernstein & Bhavsar 2001). Az említett szerzők azonban nem vették figyelembe, amin azt Györgyi et al. (2008) és Taghizadeh-Popp (2011) megmutatták, hogy N kis értékeire az elméleti határeloszlások

hoz szignifikáns korrekciók járulnak.

Részben az ebben a dolgozatban bemutatott munkára alapozva mutatta meg Paranjape & Sheth (2011), hogy a sorrendi statisztika egyetlen univerzális galaxis-luminozitásfüggvényre alapozva is pontosan leírja a halmazokban található galaxisok fényességeloszlását, beleértve a legfényesebb és második legfényesebb galaxisokat is. Kiemelik, hogy a galaxis-halmazok statisztikájánál az elsődlegesen figyelembe veendő paraméter a halmaz számossága. Az eredmények egy fontos következményeként említik meg, hogy a sötétanyag-halók modelljei is részben átdolgozásra szorulnak.

Az 5.1. fejezetben bemutatjuk az általunk használt adathalmazt, és az eljárást, amivel az abszolút magnitúdókat származtattuk. Az 5.2. fejezetben meghatározzuk a minta fényességeloszlás-függvényét. Röviden bemutatjuk az extrémérték- és sorrendi statisztika alapjait az 5.3. fejezetben, majd kiszámoljuk a mintánk fényességeloszlásának legérdekesebb sorrendi statisztikai mennyiségeit. Az 5.4. fejezetben bemutatjuk, hogy miként használható a sorrendi statisztika távoli vöröseltolódás-tartományokban levő galaxisok számának megbecslésére, ahol már csak a legfényesebb galaxisok figyelhetők meg. Az 5.5. fejezetben összeálltunk egy nagyon egyszerű modellt, ami képes reprodukálni a BCG-k és a halmazokon belüli második legfényesebb galaxisok között megfigyelt magnitúdókülönbséget. Az 5.6. fejezetben elemezzük az extrémérték- és sorrendi statisztika kozmológiai célú lehetőségeit és korlátait, valamint kitérünk a BCG-eket érintő problémákra is.

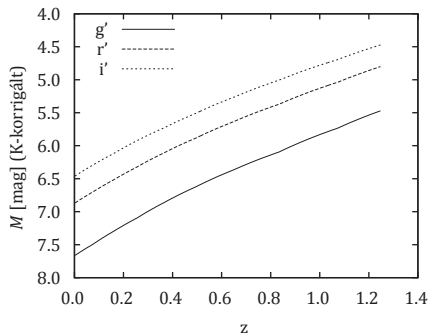
Ebben a részben mindenütt a standard Λ CDM kozmológiát használjuk a következő paraméterekkel: $h = 0,7$, $\Omega_M = 0,3$ és $\Omega_\Lambda = 0,7$. Ha külön nem említjük, az abszolút magnitúdók mindig K - és evolúciókorrigálva értendők, és melyeket a legjobban illeszkedő radiális fényességprofil-modellekből számolt, galaktikus extinkcióra korrigált magnitúdókból származtatunk.

5.1. A felhasznált adatok

Mivel a minta kiválasztását egy csillagpopuláció-evolúciós modellel alapozzuk, és az alapján számoljuk a K - és evolúciókorrekció értékeit, fontos, hogy megtárgyaljuk a modellt, mielőtt rátérünk a mintát meghatározó kiválasztási kritériumokra.

5.1.1. K -korrekció és evolúciókorrekció

Mivel a passzívan fejlődő csillagpopulációk tömeg-fényesség aránya jelentősen változik az idő függvényében, az evolúciós effektus megfelelő korrekciója feltétlenül szükséges akkor, ha a különböző vöröseltolódású galaxisok számsűrűségét a megfigyelt látszólagos magnitúdók alapján akarjuk megbecsülni (Eisenstein et al. 2001, Loh 2004, Loh & Strauss 2006, Wake et al. 2006). Abból a célból, hogy a csillagok evolúciójára korrigáljunk és az össze-



5.1. ábra. Egy passzívan evolváló, $z = 0$ -nál 12 milliárd éves, korai típusú galaxisokra jellemző csillagpopuláció abszolút magnitúdójának változása a vöröseltolódás függvényében. A magnitúdók $1 \mathcal{M}_{\odot}$ -re normáltak. Világosan látszik a korai típusú galaxisok exponenciális fényességváltozása az öregedéskövetkezményeként. Erre az effektusra mindenképpen korrigálni kell, amennyiben a galaxisokat abszolút magnitúdó alapján kívánjuk összehasonlítani.

hasonlítható abszolút magnitúdókat meghatározzuk, Maraston et al. (2009) vörös óriás-galaxisokra legjobban illeszkedő kompozit csillagpopuláció-modelljét alkalmazzuk. Ez a félempirikus modell meglehetősen jól írja le a passzívan fejlődő SDSS-galaxisok $g - r$ és $r - i$ színindexeinek változásait az általunk vizsgált vöröseltolódás-tartományban. Ez a modell azt feltételezi, hogy a vörös sorozat legfényesebb galaxisai egyaránt 12 milliárd évesek lennének, ha $z = 0$ nyugalmi rendszerben figyelnénk meg őket, mivel azóta passzívan fejlődnek, hogy $z \geq 2$ körül kialakultak. (Megjegyezzük, hogy az Univerzum akkoriban olyan fiatal volt, hogy ez csak kis szórást enged meg a csillagpopulációk korában.) Az eredeti modellspektrumok 1 milliárd éves lépésközzel lettek kiszámítva, így mi lineárisan interpoláltunk a köztes időpontokra. Az 5.1. ábrán felrajzoljuk egy K-korrigált (tehát nem vöröseltolódott) passzívan fejlődő csillagpopuláció SDSS g' , r' és i' színszűrőkben vett magnitúdóit.

A Maraston et al. (2009) által megadott receptet követve az összes galaxisra kiszámítottuk a K-korrekción $z = 0$ -ra vonatkoztatott értékét a Fukugita et al. (1996) által közölt u , g , r , i és z SDSS átviteli görbék alapján. Az evolúciókorrekció céljára meghatároztuk a galaxis eredeti vöröseltolódásán vett, $t_1 = 12 \text{ Gyr} - t_{lb}$ korú modell fluxusának és a $z = 0$ -nál vett, $t_2 = 12 \text{ Gyr}$ korú modell fluxusának arányát. Utóbbi képletben t_{lb} a kozmikus visszatekintési időt jelenti.

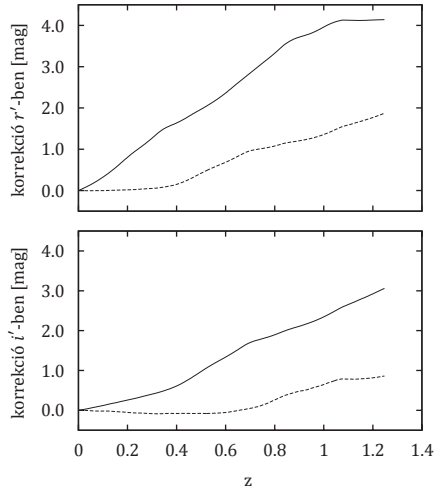
Túl egyszerűnek tűnhet egy olyan folytonosan evolváló modell használata, mely nem számol az azonos vöröseltolódáson megfigyelt galaxisokat alkotó csillagpopulációk korának szórásával, de a gyakorlatban ez sokkal jobban működik, mintha egyenként próbálnánk megfelelő korú modellt illeszteni az egyes galaxisokhoz. Korábbi tanulmányok meg-

állapították, hogy a nagy tömegű elliptikusok korának szórása a 100 millió éves nagyságrendbe esik (Gallazzi et al. 2006). Ha a galaxisok korát egyenként próbáljuk megilleszteni, a K+e-korrigált abszolút magnitúdókból számolt színindexek szórása jelentősen nagyobb-nak adódik, mintha a korábban említett, $z = 0$ -nál 12 milliárd éves modellt használjuk. Úgy véljük, hogy ennek az az oka, hogy a fotometriai hibák jelentősebb eltérést okoznak a kor becslésekor, mint amekkora a rögzített korú modellel való korrekció okozta hiba. Néhány vörös óriásgalaxisban kisméretű galaxisokkal való összeolvadás, vagy lassú akkréció hatására epizódyszerűen beindulhat ugyan némi csillagkeletkezés, de ezek színindexei alig térnek el a teljesen passzívan fejlődő galaxisokétól (Kaviraj et al. 2008). Egyes modellek szerint a galaxisok vörös sorozata folyamatosan épül fel azáltal, hogy a kék sorozaton levő csillagképző galaxisokban a csillagkeletkezés valamilyen oknál fogva befagy, és emiatt azok átvándorolnak a vörös sorozatra (Bell et al. 2004; 2007). Természetesen az általunk alkalmazott egyszerű evolúciókorrekciós modell ezekre a galaxisokra nem alkalmazható, de úgy gondoljuk, hogy az ilyen fiatal vörös galaxisok fényessége nem éri el az általunk összeállított minta esetében megszabott abszolút magnitúdó-határt.

5.1.2. Néhány gondolat az evolúciókorrekcióról

Ismételten ki szeretnénk emelni a korai típusú galaxisok evolúciókorrekciójának jelentőségét abban az esetben, ha fizikailag értelmes eredményeket akarunk kapni. Ahogy a galaxisokat egyre távolibb vöröseltolódásokon figyeljük meg, színindexeik úgy válnak egyre vörösebbé, mivel a spektrumukban megfigyelhető 4000 Å-ös törés egyre hosszabb hullámhosszakra kerül. Ugyanakkor viszont minél távolabbra tekintünk, annál fiatalabb csillagpopulációkat látunk, melyek eredendően fényesebbek és kékebbek. Ez a két effektus az SDSS r és i színszűrője esetén versenyzik, és a $0 \leq z \leq 0,35$, illetve a $0 \leq z \leq 0,6$ vöröseltolódás-tartományokon épp kioltja egymást, ahogyan az az 5.2. ábrán is megfigyelhető. A sima K-korrekció és a K+e-korrekció közötti különbség elérheti a 0,5 magnitúdót is, ami jelentősen torzíthatja egy abszolút magnitúdó-limitált minta együtt mozgó koordinátákban mért sűrűségét.

A közeljövő nagy égfelmérései, mint például a PanSTARRS vagy az LSST, nem fognak spektroszkópiai úton vöröseltolódást mérni. Ebből következően a jövőben nagy szükség lesz az erős evolúciós effektust mutató galaxisok vöröseltolódásának megbízható fotometriai módszereken alapuló meghatározására (Padmanabhan et al. 2005). A csillagpopulációk evolúcióját a legtöbb jelenlegi, sablon alapú, széles körben használt fotometriai vöröseltolódás-becslő algoritmus közvetlenül nem veszi figyelembe (Bolzonella et al. 2000, Blanton et al. 2003, Blanton & Roweis 2007, Hildebrandt et al. 2010). Néhány algoritmus a problémát úgy oldja meg, hogy a vörös galaxisok szélessávú spektrális energiaeloszlását számos, különböző korú LRG sablonnal próbálja megilleszteni, de nem veszi figyelembe a kor és a vöröseltolódás kovarianciáját. Ez az LRG-k esetében nem jó gyakor-



5.2. ábra. Az SDSS r' (felső panel) és i' (alsó panel) színszűrőjére számolt K-korrekció (folytonos görbe), illetve K+evolúció-korrekció (szaggatott görbe) egy olyan passzív evoláló korai típusú galaxis esetén, mely csillagpopulációjának kora $z = 0$ -nál 12 milliárd év. Felhívjuk a figyelmet, hogy a K+e-korrekció értéke majdnem pontosan null a $z \leq 0,35$, illetve a $z \leq 0,65$ tartományokon.

lat, hiszen kizárólag optikai mérések esetében a mért magnitúdók fotometriai hibákból adódó szórása, illetve a galaxist alkotó csillagok változó korából valamint fémességéből adódó szórás ilyen egyszerű eljárással nem választható szét. Megemlítendő továbbá, hogy túl sok sablon használata túl sok illesztendő szabad paramétert von maga után, ami az illesztés értelmezhetőségét kérdéssé teszi.

Léteznek azonban olyan empirikus és félempirikus sablon alapú algoritmusok, melyek igen jó fotometrikus vöröseltolódás eredményeket szolgáltatnak oly módon, hogy a spektrumsablonokat „megjavítják”; ez abban áll, hogy a vöröseltolódás becslését megelőzően a sablonokat valamilyen módon az adatokhoz illesztik (Csabai et al. 2000, Budavári et al. 2000, Blanton et al. 2003). Az így nyert, kissé torzított, „javított” sablonok valamiféle *effektív evolúciós* jellemzőkkel rendelkeznek, mely a sablonok meredekségébe, görbületébe van belekódolva.

5.1.3. Az SDSS DR7 minta kiválasztása

A tanulmányhoz az SDSS DR7 spektroszkópiai adatait (`spectro`) használtuk (Abazajian et al. 2009). Hogy a korai típusú galaxisokat elkülönítsük, az összes olyan `spectro` galaxist gyűjtöttük, amelyek spektrumai az SDSS spektroszkópiai feldolgozó programja

által használt LRG sablonnal mutatták a legjobb keresztkorrelációt, és a keresztkorrelációból megállapított vöröseltolódás konfidenciaszintje legalább $z_{conf} \geq 0,9$ (Eisenstein et al. 2001). Mivel az SDSS spektrográfjai által alkalmazott üvegszalak szűk átmérője a kiterjedt objektumoknak csupán a középső 3"-át észleli – mely spirál galaxisok esetében gyakran csak a galaxis magját jelenti – alacsony vöröseltolódáson észlelt spirálgalaxisok spektruma gyakran hasonlít elliptikus galaxisok spektrumára. Bár a spirálok magja az vörös óriásgalaxisoknál jóval halványabb, alacsonyabb abszolút magnitúdó-vágás mellett összeállított mintákat mégis képesek eltorzítani. Ez a probléma részben orvosolható oly módon, hogy a fotometrikus magnitúdókat és a radiális felületifényesség-profilokat is figyelembe vesszünk, melyek alapján el lehet választani az elliptikus galaxisokat a spirálok magjaitól. Esetünkben elsőként azt követeltük meg, hogy a kiválasztott galaxisok színindexei megfelelően közel essenek Maraston et al. (2009) általunk megbízhatónak tartott evolválódó modelljének trajektóriájához a $g-r$; $r-i$; $i-z$ színindexek terében. A következő feltételt alkalmaztuk:

$$\frac{1}{9}((g-r) - (g_m - r_m))^2 + ((r-i) - (r_m - i_m))^2 + ((i-z) - (i_m - z_m))^2 < 0,015, \quad (5.1)$$

ahol g , r , i és z a galaxisok megfigyelt, galaktikus extinkcióra korrigált, legjobban illeszkedő radiális profilból számolt modell magnitúdói, míg g_m , r_m , i_m és z_m a csillagpopulációmodell (ld. 5.1.1. fejezet) SDSS színszűrőkkel vett konvolúciójából nyert szintetikus magnitúdók a vöröseltolódással paraméterezett trajektória mentén.

Hogy a kiválasztott galaxisokat valóban csak az elliptikusok körére szűkítsük, további feltételeket kellett szabni. Megköveteltük, hogy a mintába kerülő galaxisok r színszűrőben vett radiális felületifényesség-profilja jól illeszthető legyen a de Vaucouleurs-profillal (egész pontosan a teljes illesztett radiális profilban a de Vaucouleurs-tag hozzájárulása legalább 80 % legyen; az SDSS által észlelt galaxisok profilillesztésével kapcsolatban ld. Stoughton et al. (2002)). Ezen kívül feltételül szabtuk, hogy a galaxisoknak legyen legalább 90 %-os konfidenciaszinten megállapított spektroszkópiai vöröseltolódása.

Az ismertetett kiválasztási feltételek alapján automatikusan kapott minta 1000 legfényesebb abszolút magnitúdójú galaxisát szemmel is ellenőriztük, és a hibás méréseket eltávolítottuk. A hibás mérések aránya kb. 1,5 %. A hiba a legtöbb esetben abból eredt, hogy az objektum közel volt egy másik nagyon fényes objektumhoz, vagy ún. „szellemhez”, ami az SDSS képfeldolgozó programját megzavarva hibás magnitúdókhoz vezetett. A kiválasztási eljárás végül 142.762 galaxist eredményezett, melyek vöröseltolódása egészen $z = 0,63$ -ig terjed, és a teljes minta 8.032 négyzetfokot fed le. A minta legfényesebb része ($M_r < -22$ mag) 73.459 galaxisból áll. A spectro adathalmaz teljességi paraméterét a sűrűségek kiszámításánál 0,9-nek feltételezzük.

Az SDSS jól ismert tulajdonságai miatt a mintánk egyenletessége $z < 0,07$ alatt nem tökéletes. Az SDSS fotometriai feldolgozó programja hajlamos a nagy látszólagos átmé-
rőjű galaxisokban több felületi fényesség csúcsot azonosítani, és ez alapján a galaxist több
részre bontani. Mivel minket épp a legnagyobb és legfényesebb galaxisok érdekelnek, ez
a jelenség kis vöröseltolódásokon jelentősen befolyásolja a galaxisok számát. A teljes ere-
deti LRG minta $z = 0,38$ -ig térfogat-limitált (Loh & Strauss 2006), de ha az abszolút mag-
nitúdóra kiszabjuk az $M_r \leq -22$ feltételt, akkor $z = 0,45$ -ig, ha az $M_r \leq -22,5$ feltételt,
akkor pedig $z = 0,51$ -ig teljes mintát kapunk.

5.2. A vörös óriásgalaxisok luminozitásfüggvénye

Ebben a fejezetben a korábban ismertetett LRG minta abszolút magnitúdó-eloszlását ha-
tározzuk meg, és a kapott eloszlást egy Schechter-függvénnyel illesztjük a következő alak-
ban:

$$\phi(M) dM = \phi_0 \exp \{C(\alpha + 1)(M - M^*) - \exp [C(M - M^*)]\} dM, \quad (5.2)$$

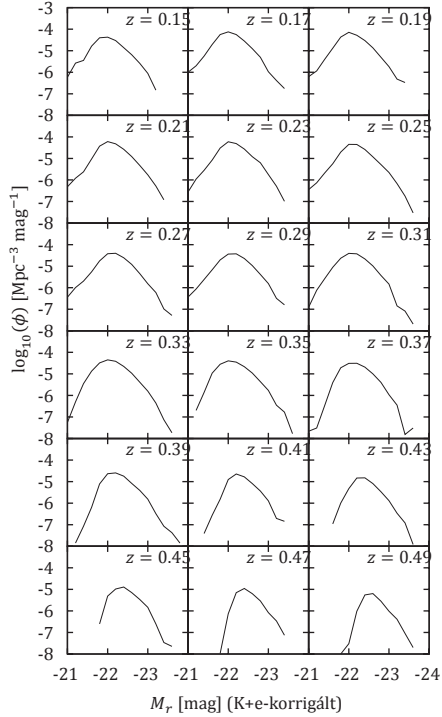
ahol $C = -0,4 \ln 10$.

Az 5.3. és 5.4. ábrákon felrajzoljuk a spektroszkópiailag kiválasztott minta fényességel-
oszlását $\Delta z = 0,02$ -es vöröseltolódás-intervallumokba csoportosítva. A feltüntetett mag-
nitúdók mind vöröseltolódásra, mind evolúcióra korrigálva lettek. Mivel a korai típusú
galaxisok együtt mozgó koordinátákban vett sűrűsége enyhén növekszik a kozmikus idő
függvényében, a magasabban haladó görbék a kisebb vöröseltolódásokhoz tartoznak. Az
 $M_r > -22,5$ utáni (vagyis halvány) levágások a minta magnitúdólimitált volta miatt jelent-
keznek, de mi csak a görbék fényes tartománybeli viselkedése iránt érdeklődünk. Bár az
5.4. ábrán látható görbék bármilyen összeskálázás nélkül is meglehetősen együtt futnak,
abból a célból, hogy a teljes vöröseltolódás-tartományon érvényes luminozitásfüggvényt
származtassunk, a sűrűség enyhe növekedését kiküszöböljük és a görbéket összeskáláz-
zuk.

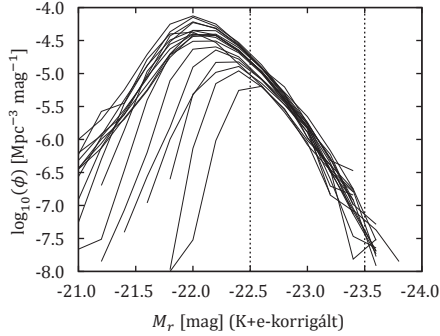
A tanulság kedvéért az 5.5. ábrán ugyanezen galaxisok fényességeloszlását felrajzoljuk
úgy is, hogy az abszolút magnitúdókat csak K-korrekcióval látjuk el (az evolúciókorrekciót
kihagyjuk). Jól látszik, hogy a korábbiakkal szemben most a görbék nem futnak automati-
kusan együtt.

Az 5.6. ábra a legfényesebb korai típusú galaxisok együtt mozgó koordinátákban vett
sűrűségét ábrázolja a vöröseltolódás függvényében. Azért, hogy az ábrázolt minta $z = 0,5$ -
ig térfogat-limitált legyen, csak azokat a galaxisokat vettük számba, melyek r magnitúdója
a $-22,5 \geq M_r \geq -23,5$ tartományba esik. A $z = 0,5$ -nél megfigyelhető levágás a minta
magnitúdókorlátja miatt van, míg az alacsony vöröseltolódáson megfigyelhető levágás ma-
gyarázata az 5.1.3. fejezetben olvasható.

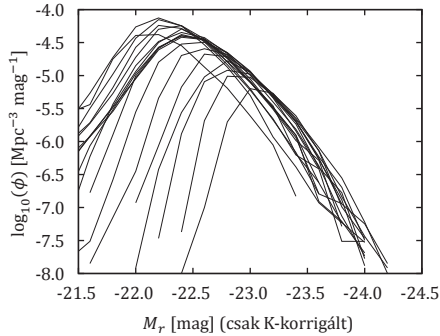
Az együtt mozgó koordinátákban vett számsűrűség megváltozása csak a luminozitás-



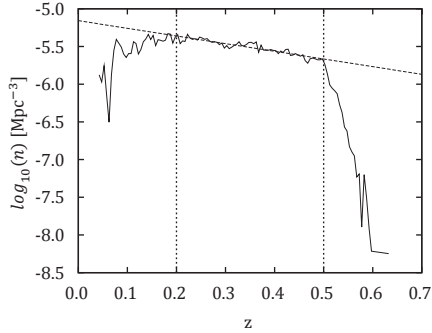
5.3. ábra. A mintában szereplő korai típusú galaxisok fényességeloszlása $\Delta z = 0.02$ vöröseltolódás-intervallumokban a $0,14 \leq z \leq 0,50$ tartományon. A vöröseltolódás-intervallumok középértékét a panelek jobb felső sarkában jelöltük. A minta fényes vége ($M_r \leq -22,5$ mag) a teljes ábrázolt tartományban térfogatlimitált. A kisebb luminozitásoknál megjelenő letörés az SDSS kiválasztási algoritmusa miatt van, mely csak az $m_r \leq 19,5$ magnitúdónál fényesebb vörös galaxisokat jelöli ki spektroszkópiai megfigyelésre. Ld. az 5.4. ábrát is, ahol ugyanezek a görbék egymásra rajzolva kerültek ábrázolásra.



5.4. ábra. A mintában szereplő korai típusú galaxisok fényességeloszlása $\Delta z = 0.02$ vöröseltolódásintervallumokban a $0,15 \leq z \leq 0,49$ tartományban. A minta fényes vége ($M_r \leq -22,5$ mag) a teljes ábrázolt tartományban térfogat-limitált. A kisebb luminozitásoknál megjelenő letörés az SDSS kiválasztási algoritmus miatt van, mely csak az $m_r \leq 19,5$ magnitúdónál fényesebb vörös galaxisokat jelöli ki spektroszkópiai megfigyelésre. A függőleges szaggatott vonalak azt a magnitúdótartományt jelölik, melyen belül a galaxisokat a fényességeloszlás-görbék összeskálázásához összeszámoljuk (ld. a szöveget). A görbék fényes vége meglehetősen együtt halad anélkül, hogy bármilyen összeskálázást végrehajtottunk volna, de az együtt mozgó térfogatra vett számsűrűség enyhe változása szükségessé teszi a görbék összeskálázását ahhoz, hogy egy, a teljes vöröseltolódás-tartományon érvényes luminozitásfüggvényt származtathassunk (vö. 5.7. ábra).



5.5. ábra. A korai típusú galaxisok fényességeloszlása a $0,15 \leq z \leq 0,49$ vöröseltolódás-tartományban, $\Delta z = 0,02$ nagyságú intervallumokban nézve. Az abszolút magnitúdók csak K-korrekcióval lettek ellátva, a spektrális evolúcióra nem korrigáltunk (vö. 5.4. ábra).



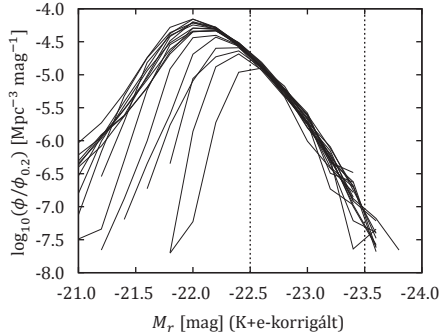
5.6. ábra. A legfényesebb ($M_r \leq -22,5$) korai típusú galaxisok együtt mozgó koordinátákban vett sűrűsége a vöröseltolódás függvényében. A számsűrűség z -vel lassan csökken $z \leq 0,15 \leq 0,5$ között. Az alacsony vöröseltolódásnál megfigyelhető letörés a minta fényes végének hiányossága miatt jelenik meg, ld. 5.1.3. fejezet. Az egyenes illesztése a $0,2 \leq z \leq 0,5$ intervallumon történt, ahogyan azt a függőleges vonalak is jelzik.

függvény ϕ_0 paraméterét befolyásolja. A dolgozat hátralevő részében egy olyan normált luminozitásfüggvényt használunk, amely független a galaxissűrűségtől. A későbbiekben azt is megindokoljuk, hogy a bemutatandó statisztikai módszerek csak az α és M^* paraméterektől függenek. Ahhoz, hogy az 5.4. ábrán szereplő görbéket összekalázzuk, és az 5.7. ábrán látható görbékhez jussunk, a galaxisok számsűrűségének változását a vöröseltolódás gyenge exponenciális függvényének tekintjük. A vöröseltolódás különböző tartományain vett luminozitásfüggvényeket úgy kalázzuk össze, hogy a $z = 0,2$ -nél vett görbével essenek egybe. Az együtt mozgó koordinátákban vett sűrűség megillesztésének eredménye az 5.6. ábrán látható, szaggatott ferde vonallal jelölve. Az illesztés a $0,2 \leq 0,5$ vöröseltolódás-tartományban történt, az ábrán a függőleges vonalak ezt a tartományt jelölik. Az illesztett paraméterek értékei a következők:

$$n(z) = 10^{(-1.02 \pm 0.06) \cdot z + (-5.16 \pm 0.02)} \quad (5.3)$$

Ez a galaxissűrűség körülbelül 0,3 dex-es növekedését jelenti $z = 0,5$ -ről $z = 0,2$ -re, ami mások eredményeivel konzisztens (Bell et al. 2004, Faber et al. 2007).

A normalizálás eredményeként kapott görbéket az 5.7. ábrán rajzoltuk fel. A görbék a fényes tartományban igen jól átfednek egymással, de a legfényesebb galaxisok esetén megfigyelhető némi szórás, amit az ebbe a tartományba eső galaxisok nagyon alacsony száma magyaráz. A luminozitáseloszlások fényes vége azért kalázzható össze ilyen jól, mert α és M^* , bár függenek a vöröseltolódástól, a függés a vizsgált tartományban nagyon gyenge (Bell et al. 2004, Faber et al. 2007). Az összekalázott görbéket kiátlagoljuk, és így származtatjuk azt a luminozitásfüggvényt, ami a teljes vöröseltolódás-tartományon reprezen-



5.7. ábra. A korai típusú galaxisok összeskálázott fényességeloszlásai $\Delta z = 0,02$ nagyságú vöröseltolódás-intervallumokban, az $0,15 \leq z \leq 0,49$ vöröseltolódás-tartományban. Az összeskálázást a $M_r < -22,5$ magnitúdónál fényesebb galaxisokra végeztük el a $z = 0,2$ -nél vett sűrűséget alapul véve. A skálázásra az együtt mozgó koordinátákban vett sűrűség evolúciója miatt volt szükség, amit bővebben az 5.2. fejezetben fejtettünk ki, és az 5.6. ábrán rajzoltunk fel. A görbék most szépen átfednek a $-22,5 > M_r > -23,0$ magnitúdótartományban. A fényes végen megfigyelhető szórás a nagyon fényes galaxisok alacsony száma okozza.

tatív.

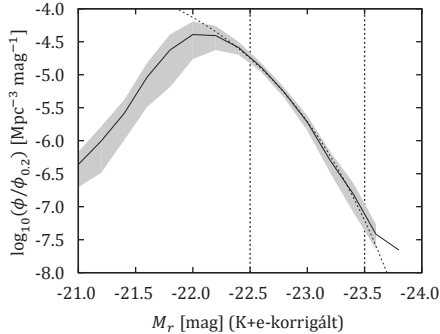
Az 5.8. ábra az átlagolt luminozitáseloszlást (folytonos vonal), és az ehhez illesztett Schechter-függvényt (ld. 5.2. képlet) mutatja. Az illesztés a $-22,5 \leq M_r \leq -23,5$ magnitúdótartományban történt úgy, hogy közben a halvány tartománybeli viselkedést meghatározó α paramétert $\alpha = -1,2$ -ben rögzítettük. Erre azért volt szükség, mert az α elvi okokból sem határozható meg csupán a legfényesebb galaxisok eloszlása alapján. Az illesztés eredményeként kapott paraméterek a következők: $\alpha = -1,20$, $\log_{10}(\phi_0) = -7,06 \pm 0,21$, $M^* = -21,10 \pm 0,04$, ahol ϕ_0 a $z = 0,2$ -nél vett értéket jelenti. A következő fejezetekben ezt a függvényt használjuk arra, hogy bemutassuk a sorrendi statisztika lehetőségeit.

5.3. Extrémérték- és sorrendi statisztika

5.3.1. Az extrémérték- és sorrendi statisztika rövid bemutatása

Tekintsük az $f(x)$ valószínűségi sűrűségfüggvényt. Meg szeretnénk határozni az $f(x)$ -ből véletlenszerűen húzott N elemű minta maximumának (vagy minimumának) várható értékét. Az $f(x)$ -hez tartozó kumulatív eloszlásfüggvényt a szokásos módon írjuk fel:

$$F(x) = \int_{-\infty}^x f(u) du \quad (5.4)$$



5.8. ábra. A korai típusú galaxisok átlagos luminozitáseloszlása. Az átlagolandó eloszlásokat $\Delta z = 0,02$ nagyságú vöröseltolódás-intervallumokra határoztuk meg, majd a $0,15 \leq z \leq 0,49$ vöröseltolódás-tartományban átlagoltunk. Az árnyékolt terület az 5.7. ábrán felrajzolt különböző görbék szórását jelöli. Az átlagolt luminozitáseloszlás jól jellemzi az összes $M_r < -22,5$ -nél fényesebb galaxis magnitúdó eloszlását. Ennél halványabb galaxisokra a minta a nagyobb vöröseltolódásokon már hiányos. A Schechter-függvénnyel történő illesztést a $-22,5 \geq M_r \geq -23,5$ magnitúdótartományban végeztük el, a tartomány határait a függőleges vonalak jelölik.

Ennek felhasználásával annak a valószínűsége, hogy az $f(x)$ -ből húzott x szám értéke kisebb, mint X egyszerűen

$$P(x < X) = F(X). \quad (5.5)$$

Következésképpen, ha van egy N elemű mintánk, amely az egymástól függetlenül húzott $\{x_1, x_2, \dots, x_N\}$ számokból áll, akkor kiszámíthatjuk annak a valószínűségét, hogy az $X_m = \max\{x_i\}$ a mintának egy felső limitje, azaz annak a valószínűségét, hogy a minta összes eleme kisebb X_m -nél:

$$P_m(X_m) = P(x_i < X_m) = P^N(x < X_m) = F^N(X_m). \quad (5.6)$$

Az N elemű minta maximumának valószínűségi sűrűségfüggvényét tehát X_m szerinti differenciálással kapjuk:

$$p_m(X_m, N) = NF^{N-1}(X_m)f(x), \quad (5.7)$$

ahol $f(x)$ az eredeti sűrűségfüggvény (avagy *szülőeloszlás*). Összesen annyit kell tennünk, hogy kiszámítjuk a $p_m(X_m, N)$ -ből húzott X_m változó várható értékét (vagy bármely magasabb momentumát). Ez egy általános $f(x)$ valószínűségi sűrűségfüggvény és véges N esetében analitikusan nem mindig lehetséges. A Monte Carlo-módszerrel végzett numerikus integrálás azonban nagyon egyszerűen célra vezet.

A fent bemutatott eljárást követve felírhatjuk egy minta 2., 3., k . stb. legnagyobb értékének várható értékét és momentumait is. A k . legnagyobb értékhez tartozó valószínűségi

eloszlás a következő:

$$p_{(k)}(X_{(k)}, N) = \frac{N!}{(k-1)!(N-k)!} (1 - F(X_{(k)}))^{k-1} F^{N-k}(X_{(k)}) f(X_{(k)}) \quad (5.8)$$

Az extrémérték-statisztika első főtétele (Fisher–Tippet–Gnedenko-tétel) kimondja, hogy általános esetben egy N elemű minta maximuma (minimuma) három lehetséges eloszlás valamelyikéhez konvergál. Az, hogy éppen melyikhez, az a szülőeloszlás végtelenben való lecsengési ütemétől függ. Ezt a tételt a centrális határeloszlás-tétel extrémérték-statisztikai megfelelőjének szokás tekinteni. Az exponenciálisan lecsengő szülőeloszlásokhoz az $N \rightarrow \infty$ határesetben mindig ugyanaz a szélsőérték-eloszlás tartozik (miután az első két momentumra normáltunk): a $g(x)$ Gumbel-eloszlás (Fisher & Tippett 1928, Gnedenko 1948, Gumbel 1958):

$$g(x) = e^{-x} e^{e^{-x}} \quad (5.9)$$

Fontos megjegyeznünk, hogy a Gumbel-eloszlás csak nagy elemszámú minták határesetében érvényes. Minekutána a konvergencia lassú (általában $\log N$ szerinti, de a gyorsasága az $f(x)$ szülőeloszláson is múlik), kis méretű minták esetében a dupla exponenciális eloszláshoz lényeges korrekciók járulnak. Ezek a korrekciók függenek $f(x)$ -től, és analitikus formában általában nehéz őket megadni; a numerikus számolások viszont általában egyszerűek (Györgyi et al. 2008).

Míg más tudományterületek az extrém értékek statisztikáját általában arra használják, hogy ritka, katasztrófális események valószínűségét becsülik meg vele (erős földrengések, árvizek, tőzsdei fekete péntek), addig a csillagászatban közvetlenül meg tudjuk figyelni ezeket a ritka „eseményeket”: bizonyos típusú objektumok legfényesebbjeit. A távcsövek magnitúdólimitje miatt egy adott vöröseltolódáson túl már valóban csak a legfényesebb galaxisokat, kvazárokat, gamma kitöréseket stb. látjuk. Az extrémérték-statisztika lehetővé teszi számunka, hogy állításokat fogalmazzunk meg a nem detektált objektumok számával, sűrűségével kapcsolatban csupán csak a detektált legfényesebb objektumok alapján, amennyiben valamilyen formában (akár csak modell szinten) rendelkezésünkre áll az adott objektumok fényességeloszlás-függvénye.

5.3.2. Extrémérték- és sorrendi statisztika alkalmazása a vörös óriásgalaxisok esetében

Ebben a fejezetben példát mutatunk arra, hogy az extrémérték- és sorrendi statisztika hogyan használható a korai típusú galaxisok számsűrűségének megbecslésére olyan vöröseltolódásokon, ahol a spektroszkópiai minta magnitúdókorlátja már befolyásolja a közvetlenül mérhető galaxisűrűséget.

Bár a magnitúdóra felírt Schechter-függvény nem normálható, így a teljes magnitúdó-

tartományon nem írható át valószínűségi sűrűséggé, de egy halvány magnitúdólimit bevezetésével ez már megtehető.

A normált, átskálázott és átlagolt r szűrőbeli fényességeloszlásra (ld. 5.2. fejezet és 5.8. ábra) illesztett analitikus görbét véve, és bevezetve egy $M_r < -22$ magnitúdóváagást, Monte Carlo-módszert használva meghatároztuk az N elemű, egymástól függetlenül generált galaxisokból álló mintában a legfényesebb, valamint 2., 5., 10., 20. és 100. legfényesebb galaxis magnitúdójának várható értékét. Az eredmények az 5.9. ábrán láthatóak. Amint a grafikon mutatja, a várható értékek N függvényében kis N -ekre gyorsan növekednek, majd a nagyobb mintaméret tartományában a görbék kilapulnak, és csak lassan növekednek tovább. A Schechter-féle szülőeloszlás esetében ezek a görbék elég nagy N -re $\ln \ln N$ -es viselkedéssel bírnak, hiszen az L_m maximum fényesség környékén igaz a következő:

$$N \delta \left(\frac{L_m}{L^*} \right)^\alpha \exp \left(-\frac{L_m}{L^*} \right) \approx 1, \quad (5.10)$$

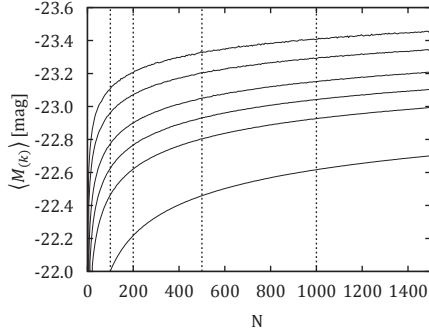
ahol δ egységnyi intervallum L/L^* skáláján. $L_m \gg L^*$ -re ebből

$$\frac{L_m}{L^*} \approx \ln N + O(\ln \ln N) \quad (5.11)$$

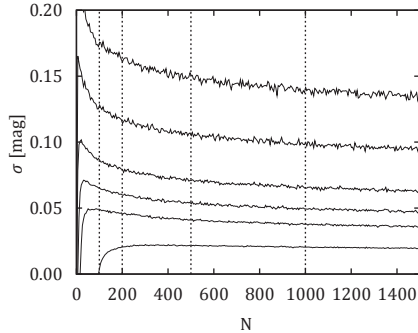
következik. A második logaritmus a magnitúdó definíciójából jön. Az $\ln \ln N$ jellegű görbék nagyon lassú növekedése az oka annak, hogy a legfényesebb klasztergalaxisok sikeresen használhatók standard gyertyaként (Sandage 1972; 1976, Postman & Lauer 1995, Whiley et al. 2008). (Itt most hallgatólagosan feltesszük persze, hogy a BCG-k is ugyanabból az eloszlásból származnak, mint a többi korai típusú galaxis. Ezt a kérdést a későbbiekben még boncolgatjuk.) A legnagyobb fényesség várható értéke k függvényében természetesen csökken, a függése logaritmikus.

$M_{(k)}$ szórása jóval érdekesebben viselkedik, ami lényegesen befolyásolja az extrémérték- és sorrendi statisztika alkalmazhatóságát precíziós asztrostatistikai problémák esetén, továbbá megmagyarázza a csillagászat pár jól ismert „ököl szabályát”. Az 5.10. ábrán felrajzoljuk az N elemű mintában levő legfényesebb, valamint 2., 5., 10., 20. és 100. legfényesebb galaxis várható magnitúdójának szórását. Az egyes görbéket nézve azonnal megállapítható, hogy a szórás N -nel csak nagyon kicsit csökken, viszont k növekedésével a csökkenés igen jelentős. Ez az oka annak, például, hogy a legfényesebb klasztergalaxis magnitúdója többet változik halmazról halmazra, mint a második, ötödik, stb. legfényesebb galaxisoké; vagyis ez utóbbiak miért jobb standard gyertyák, mint a BCG-k (Scott 1957).

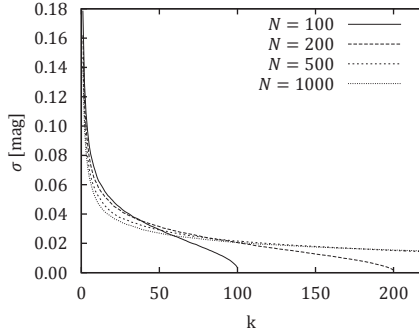
Az 5.11. ábrán $M_{(k)}$ szórását rajzoljuk fel k függvényében az N mintaméret négy különböző értékére. Ahogy az ábrán látszik, k függvényében a szórás nagyon gyorsan csökken. Megfigyelendő az is, hogy a görbék $k \approx N/2$ környékén inflexiós ponttal rendelkeznek.



5.9. ábra. A legfényesebb, valamint a 2., 5., 10., 20. és 100. (felülről lefelé) legfényesebb galaxis $M_{(k)}$ abszolút magnitúdójának várható értéke az N mintaméret függvényében. A minták generálásakor csak $M_r < -22$ -nél fényesebb galaxisokat vettünk. A felrajzolt eredmények közvetlenül az analitikus alakú luminozitásfüggvényből származnak, melyeket Monte Carlo-eljárással számítottunk ki 5000 különböző futás alapján. A függőleges szaggatott vonalak az $N = 100, 200, 500$ és 1.000 nagyságú mintákat jelölik, melyeket az 5.11. és az 5.12. ábrák esetében használunk. Az felrajzolt görbék világosan mutatják a $\ln N$ -es viselkedést.



5.10. ábra. A legfényesebb, valamint a 2., 5., 10., 20. és 100. (felülről lefelé) legfényesebb galaxis $M_{(k)}$ abszolút magnitúdójának szórása az N mintaméret függvényében. A minták generálásakor csak $M_r < -22$ -nél fényesebb galaxisokat vettünk. A felrajzolt eredmények közvetlenül az analitikus alakú luminozitásfüggvényből származnak, melyeket Monte Carlo-eljárással számítottunk ki 5000 különböző futás alapján. A függőleges szaggatott vonalak az $N = 100, 200, 500$ és 1.000 nagyságú mintákat jelölik, melyeket az 5.11. és az 5.12. ábrák esetében használunk. A görbék N függvényében csak alig változnak $N > 100$ esetében.

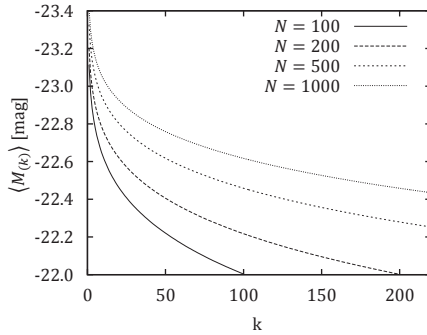


5.11. ábra. Az $M_{(k)}$ abszolút magnitúdó szórása k függvényében, $N = 100, 200, 500$, valamint 1000 elemű minták esetében. A görbéknek $k \approx N/2$ közelében inflexiós pontjuk van. Az inflexiós ponton túli k -kra a szórás nem csökken jelentős mértékben k növekedésével.

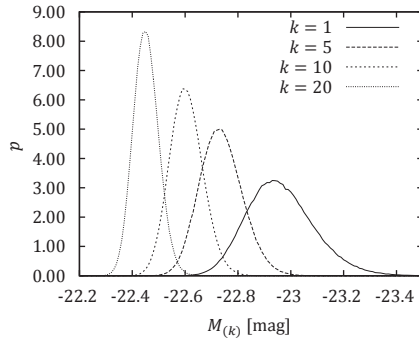
Az inflexiós pont fölötti k értékekre a szórás csak lassan csökken. Összefoglalva ez azt jelenti, hogy ha meg tudjuk mérni az ég egy adott területén levő, adott vöröseltolódás-tartományba eső galaxisok 5-10 %-ának a látszólagos magnitúdóját és a vöröseltolódását, akkor 0,02 magnitúdó pontosságú standard gyertyákhoz jutunk. A mi galaxismintánk esetében magasabb vöröseltolódásokon intervallumonként kb. 1000 galaxis található. Következésképpen, ha az extrémérték- és sorrendi statisztika módszereit ezen a mintán arra kívánjuk alkalmazni, hogy a halványabb galaxisok számát megbecsüljük a legfényesebbek magnitúdói alapján, akkor vöröseltolódás-intervallumonként elegendő lenne az 50–100 legfényesebb objektum spektroszkópiai vöröseltolódását meghatározni. Természetesen az ilyen jellegű spektroszkópiai kiválasztás jó pontosságú fotometrikan becsült vöröseltolódásokat és abszolút magnitúdókat igényel.

Az 5.12. ábrán $M_{(k)}$ várható értékét rajzoljuk fel k függvényében, N négy különböző értékére. Megfigyelendő, hogy a görbék ténylegesen elérik az $M_r = -22$ magnitúóhatárt, amennyiben $k \rightarrow N$ (az ábrán ez csak $N = 100$ és $N = 200$ esetében látszik). Ez csak olyan szülőeloszlások esetében van így, ahol a valószínűségi eloszlásfüggvénynek éles levágása van.

Az 5.13. ábrán a korai típusú galaxisok luminozitásfüggvénye alapján generált minták k . legfényesebb galaxisának magnitúdóját ábrázoljuk $k = 1, 5, 10$ és 20 értékekre. A minta mérete $N = 100$ -ban lett rögzítve, és a szimulációt 500.000-szer futtattuk. A görbék a sorrendi statisztikában gyakran előforduló, enyhén aszimmetrikus eloszlásokat mutatnak. Ha $N \rightarrow \infty$, és $k = 1$, akkor az eloszlásnak a Gumbel-eloszláshoz kell tartania, ha $N \rightarrow \infty$ és $k \rightarrow \infty$, akkor a határeloszlás Gauss-eloszlás. Az általunk ábrázolt, $k = 1$ -hez tartozó eredményt nem lehet jól megilleszteni a Gumbel-eloszlással, mivel N kicsi; ezt már említettük az 5.3.1. fejezetben: a konvergencia a Gumbel-eloszláshoz lassú (Györgyi et al.



5.12. ábra. $M_{(k)}$ várható értéke $M_r < -22$ magnitúdónál fényesebb galaxisok $N = 100, 200, 500$ és 1000 elemű mintájára.



5.13. ábra. Az $N = 100$ elemű véletlen minta legfényesebb, valamint 2., 5., 10. és 20. legfényesebb galaxisának magnitúdóeloszlása. Az eredmények 500.000 lépéses Monte Carlo-futtatásból származnak.

2008, Taghizadeh-Popp 2011). A $k = 1$ -hez tartozó görbe félértékszélessége konzisztens azzal, amit a BCG-kről tudunk, vagyis hogy a magnitúdójuk szórása halmazról halmazra kb. $\Delta M \simeq 0,25$ mag (Sandage 1972; 1988, Aragón-Salamanca et al. 1998).

5.4. Galaxissűrűség meghatározása a legfényesebb óriásgalaxisok alapján

Az 5.14. ábra alapján igazoltnak tekintjük azt a feltevésünket, hogy a sorrendi statisztika alapján, csupán a galaxisok számát, valamint egy rögzített, nem evolválódó luminozitásfüggvényt figyelembe véve megbecsülhető a legfényesebb galaxisok abszolút magnitúdója. Az ábrán a k . legfényesebb galaxisok abszolút magnitúdóját ábrázoltuk két különböző

vöröseltolódás-binelés mellett: $\Delta z_{\text{nagy}} = 0,02$ és $\Delta z_{\text{kicsi}} = 0,005$. A fekete körök a nagy binekben vett legfényesebb magnitúdókat ábrázolják, melyek ezért többnyire fényesebbek, mint az üres négyzetekkel jelölt, kis binekben vettek. A görbék $M_{(k)}$ várható értékét jelölik, melyet a vöröseltolódás-héjakban levő galaxisok számából becsültünk a sorrendi statisztika alapján (vastag görbe – nagy binek, vékony görbe – kis binek). A kis binek esetében ábrázoljuk az $M_{(k)}$ körüli 2σ szórás is (sátrózott terület). A galaxisszámok alapján számított görbék láthatóan jól illeszkednek a mért magnitúdókra azokban a vöröseltolódás-tartományokban, ahol minden $M_r < -22$ -nél fényesebb galaxisról rendelkezésre áll mérési adat.

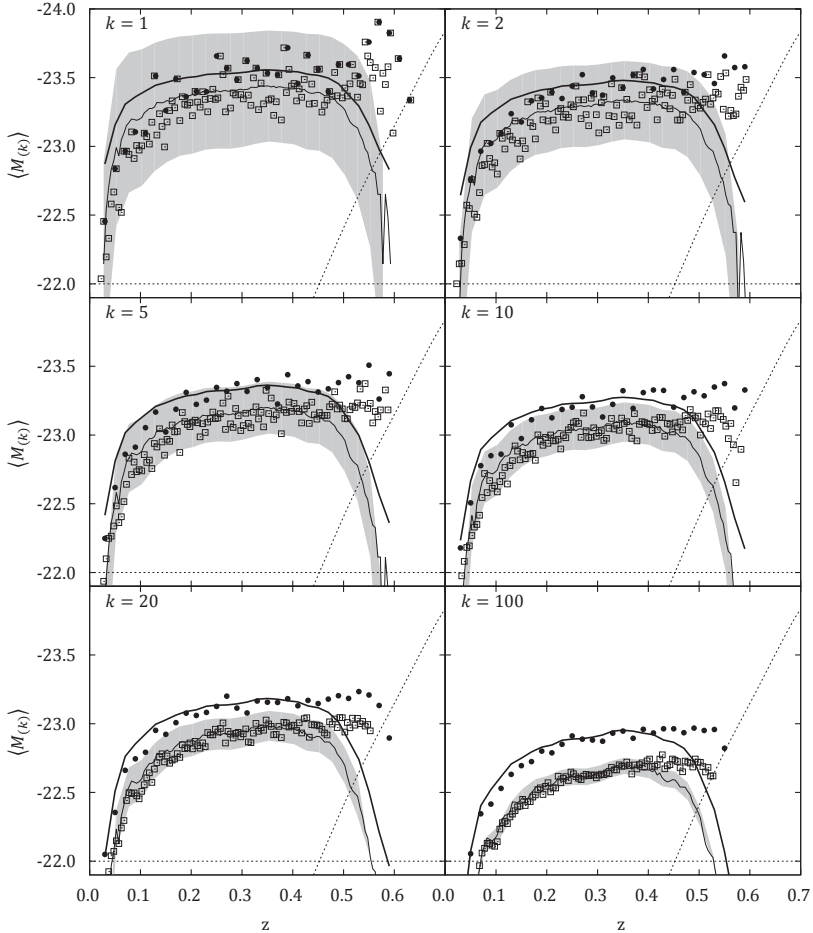
Az 5.9. ábrán felrajzolt görbék invertálásával a vöröseltolódás-héjak k . legfényesebb galaxisának magnitúdója alapján megbecsülhetjük a térfogatban levő galaxisok számát. Az 5.15. ábra az így kapott galaxissűrűségeket mutatja a vöröseltolódás függvényében, k hat különböző értékére. A k kis értékeire a becslés meglehetősen nagy szórással rendelkezik, de a $k = 100$ esetben kapott fluktuációk már összemérhetők a közvetlen galaxis-számlálásból kapott sűrűség fluktuációival. Az ábra megmutatja az eljárás extrapolációs lehetőségét is. A közvetlenül mért galaxisszámok a spektroszkópiai kiválasztási feltételek magnitúdólimitje miatt kb. $z \approx 0,45$ -nél elkezdnek csökkenni, de a minta még $z = 0,55 - 0,6$ környékén is tartalmazza a legfényesebb galaxisokat, így a galaxissűrűséget ebben a vöröseltolódás-tartományban is meg tudjuk határozni.

Kiemeljük a módszer extrapolációs képességeit olyan vöröseltolódásokra, ahol a távcső, vagy a kiválasztási kritériumok magnitúdólimitje nem teszi lehetővé, hogy a galaxissűrűséget közvetlen galaxisszámolással becsüljük meg, ld. 5.16. ábra. A grafikonok a galaxisok sorrendi statisztika alapján számított számsűrűségét mutatják hat különböző abszolút magnitúdó-vágás mellett. A becslést minden esetben a 100. legfényesebb galaxis magnitúdója alapján végeztük, $\Delta z = 0,005$ nagyságú vöröseltolódás-binekben. Halvány magnitúdóvágások esetében a spektroszkópiai kiválasztási feltételek egyértelműen befolyásolják a közvetlenül meghatározott galaxisszámot. Fényesebb vágások esetében a sorrendi statisztikán alapuló becslések viszont már szépen együtt haladnak a mérésekkel. Fontos kiemelni, hogy minden esetben ugyanazt a luminozitásfüggvényt használtuk a sorrendi statisztikai mennyiségek meghatározására, kizárólag az integrálási határon változtattunk.

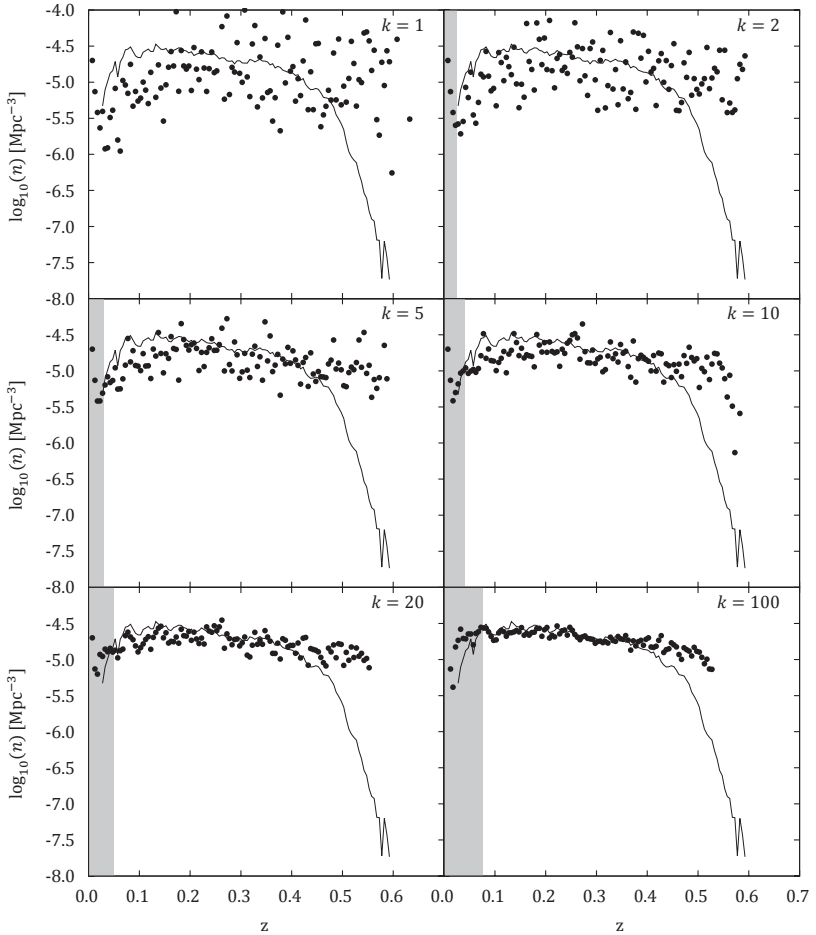
5.5. Klaszterek legfényesebb galaxisainak statisztikája

5.5.1. Korábbi eredmények

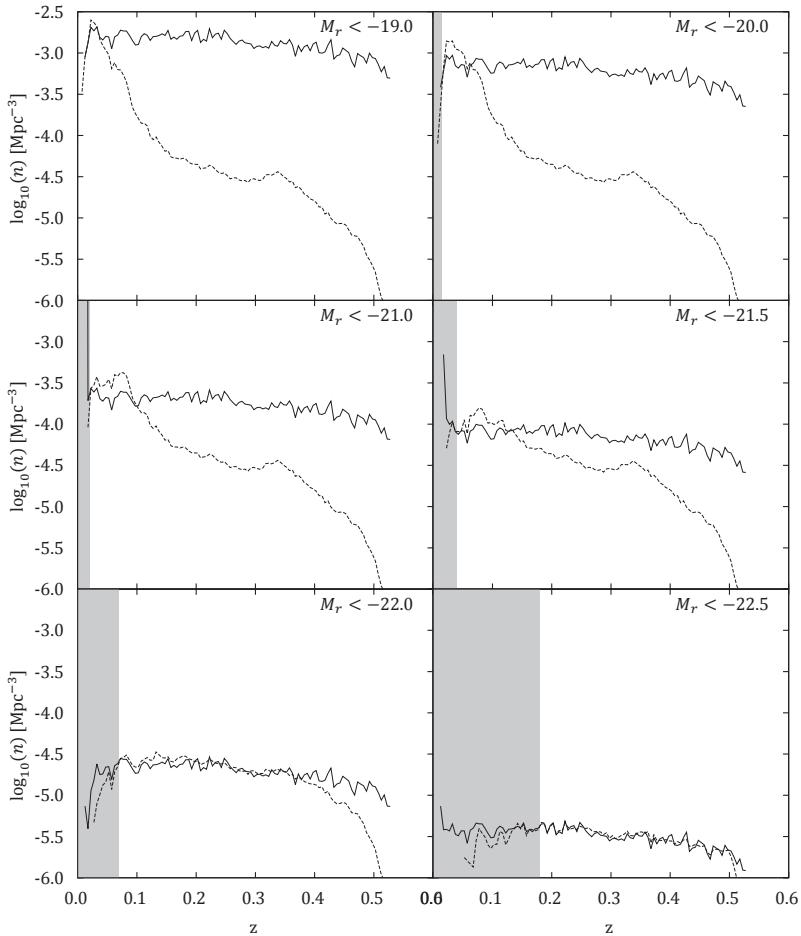
Évtizedek óta tartja magát az a vélekedés (Dressler 1976, Ostriker & Hausman 1977), és újabb keletű tanulmányok is úgy vélik (Loh & Strauss 2006, Lin et al. 2010), hogy a halma-



5.14. ábra. Az első, második, ötödik, tizedik, huszadik és a századik legfényesebb galaxis magnitúdója a vöröseltolódás függvényében, $\Delta z = 0,02$ (fekete pontok), illetve $\Delta z = 0,005$ (üres négyzetek) nagyságú vöröseltolódás héjakban. A folytonos görbék $M_{(k)}$ galaxisok számából sorrendi statisztikával számított várható értékét mutatják. A vastag görbék a széles vöröseltolódás-héjak alapján lettek felrajzolva, a vékonyak pedig a keskenyek alapján. Megfigyelendő, hogy a mért magnitúdók szórása hogyan csökken k növekedésével. A vízszintes szaggatott vonalak a galaxiszámolás abszolút magnitúdó-limitjét jelölik (azaz azt a tartományt, ahol a statisztika még nem extrapoláció), míg a ferde szaggatott vonalak a minta látszólagos magnitúdó-limitjét. A minta azon a vöröseltolódáson válik hiányossá, ahol a szaggatott vonalak metszik egymást. Ez az a pont, ahol a galaxiszámából kalkulált magnitúdók és a mért magnitúdók divergálnak.



5.15. ábra. A $M_r < -22$ mag-nál fényesebb galaxisok számsűrűsége $\Delta z = 0,005$ nagyságú vöröseltolódás-héjak alapján meghatározva. A pontok a vöröseltolódás-héjak legfényesebb, 2., 4., 10., 20. és 100. legfényesebb galaxisainak abszolút magnitúdói alapján lettek meghatározva, sorrendi statisztika segítségével. A folytonos görbék (minden panelen ugyanaz) a közvetlen galaxisszámlással kapott sűrűséget mutatják. Megfigyelendő, hogy a fényesebb galaxisok alapján becsült extrapoláció magasabb vöröseltolódásokon is működik, viszont a becslés zaja jelentősen nagyobb, mint k nagyobb értékei esetén. A sátrózott területek azokat a tartományokat jelölik, ahol a k . legfényesebb galaxis magnitúdója $M_r > -22$ mag-nál halványabb. Ezekben a tartományokban a statisztikai modellek csupán extrapolációk, és nem feltétlenül érvényesek.



5.16. ábra. A sorrendi statisztika extrapolációs lehetőségei. A grafikonok az LRG-k együtt mozgó koordinátákban vett számsűrűségét mutatják a vöröseltolódás függvényében, $\Delta z = 0,005$ nagyságú vöröseltolódás-héjakban számolva. A különböző panelek különböző r abszolút magnitúdó-vágáshoz tartoznak. A szaggatott görbék a közvetlenül számolt galaxissűrűséget mutatják, míg a folytonos görbéket sorrendi statisztika alapján határoztuk meg az egyes vöröseltolódás-héjak 100. legfényesebb galaxisainak abszolút magnitúdójából. A sátrózott területek azt a tartományt jelölik, ahol a 100. legfényesebb galaxis halványabb volt, mint az adott grafikon esetében használt magnitúdóvágás – ezeken a részeken a sorrendi statisztikán alapuló becslések nem érvényesek. Az ábrák alapján egyértelmű, hogy halvány magnitúdóvágás esetén a kiválasztási effektusok dominálják a mintát, így a közvetlenül mért galaxissűrűség jócskán eltér a legfényesebb galaxisok alapján számolt sűrűségtől. Alacsony vöröseltolódásokon a minta szennyezettsége is megfigyelhető. A szaggatott görbék magas vöröseltolódásoknál látható letérése a spectro minta magnitúdólimitált voltából következik.

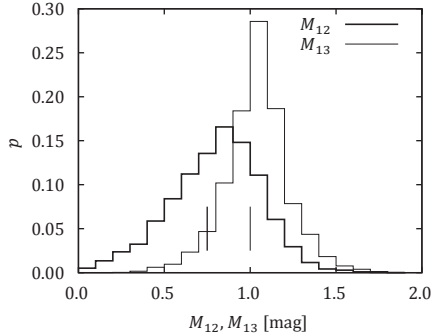
zok legfényesebb galaxisai nem magyarázhatóak ugyanannak a luminozitásfüggvénynek az extrémumaiként, mint ami a halmazokat alkotó többi galaxis eloszlását írja le. A legfőbb érv emellett az az átlagosan $M_{12} \simeq 0,8$ magnitúdós rés, mely a halmazok legfényesebb, illetve második legfényesebb galaxisai között figyelhető meg. Ez a rés nem magyarázható meg úgy, hogy mindkét galaxist ugyanabból a Schechter-függvénnyel leírt eloszlásból *egymástól függetlenül, véletlenszerűen* húzzuk. Ugyanakkor viszont megmutattuk az 5.3.2. fejezetben és az 5.4. ábrán, hogy amennyiben a teljes LRG populációt tekintjük, és nem csak az egyes galaxishalmazokat, akkor egyetlen, Schechter-féle alakban megadott luminozitásfüggvény elegendő a galaxisok statisztikai leírásához, mellyel a k . legfényesebbek magnitúdóját és magnitúdójuk szórását pontosan tudjuk becsülni.

Jelöljük egy adott minta esetében a két legfényesebb galaxis magnitúdójának különbségét $M_{12} = M_{(1)} - M_{(2)}$. Amint azt az 5.3.2. fejezetben megmutattuk, a vörös óriásgalaxisok luminozitásfüggvénye alapján, egymástól függetlenül húzott, *véletlenszerű* minta esetében M_{12} várható értékére bőven igaz, hogy $\langle M_{12} \rangle \leq 0.2$ mag, vö. 5.9. ábra.

5.5.2. A magnitúdórés egyszerű modellje

Abból a célból, hogy ezt az ellentmondást feloldjuk, a következő, nagyon egyszerű modell statisztikai tulajdonságait vizsgáltuk: Egy $V = 1 \text{ Gpc}^3$ térfogatú térrészbe egyenletesen elszórtunk galaxisokat $n = 10^{-4.5} \text{ Mpc}^{-3}$ sűrűséggel. Ezután minden galaxishoz hozzárendeltünk egy $M < -22$ mag-nál fényesebb magnitúdót. A magnitúdókat egymástól függetlenül, véletlenszerűen generáltuk a korábban meghatározott LRG-k luminozitásfüggvénye alapján. Minden olyan galaxist BCG-nek tekintettünk, melyek fényességére igaz volt, hogy $M < -23$ mag. Minden ilyen BCG esetében megkerestük a tőle legfeljebb $r = 10 \text{ Mpc}$ távolságra levő második és harmadik legfényesebb galaxist. Az 5.17. ábra az így talált második és harmadik legfényesebb galaxis, valamint a BCG közötti magnitúdóeltérés eloszlását mutatja. A magnitúdórés várható értéke $M_{12} \simeq 0,75$ mag, illetve $M_{13} \simeq 1,0$ mag, melyek jól egyeznek a megfigyelésekkel.

Bár a modell meglepően primitív ahhoz képest, hogy mit tudunk az Univerzum nagyskálás szerkezetéről, de a véletlenszerűen elszórt pontokat tekinthetjük úgy is, mintha azok az Univerzum egy korai időpontjában – amikor a tömegeloszlás sokkal homogénebb volt a mainál – a későbbi, nagy tömegű galaxisokat felépítő elsődleges sötétanyag-halók középpontjait jelölnék. Az $r = 10 \text{ Mpc}$ -es galaxishalmaz-sugár tekinthető úgy, mint a sötétanyag-halóknak az a régiója, amiből a későbbi galaxishalmaz ki fog alakulni. Természetesen a végső halmaz sugara a halók összeolvadása miatt sokkal kisebb lesz.



5.17. ábra. A szimulált BCG-k körüli $r = 10$ Mpc sugárban fellelhető második (M_{12} – vastag görbe) és harmadik (M_{13} – vékony görbe) legfényesebb galaxisok BCG-től vett magnitúdó különbségeinek eloszlása. A kis függőleges vonalak a várható értékeket jelölik. Az ábra egy 10.000 futásból álló Monte Carlo-szimuláció eredménye.

5.5.3. A magnitúdórés statisztikája

Az 5.5.2. pontban bemutatott modell pontos értelmezéséhez visszatérünk az analitikus megközelítéshez, illetve az 5.3.1. fejezet jelöléseire.

Tekintsük egy véletlenszerűen generált minta k . és l . legnagyobb (legkisebb) eleme közötti különbséget, melyet X_{kl} -lrel jelölünk. Annak valószínűségét, hogy a rés mérete éppen X_{kl} egyszerűen $P_{kl}(X_{kl})$ -lrel jelöljük. Az ehhez valószínűségi sűrűségeloszlást a következő konvolúciós integrál adja meg:

$$p_{kl}(X_{kl}, N) = \int_{-\infty}^{\infty} p_k(t, N) p_l(t - X_{kl}, N) dt, \quad (5.12)$$

ahol N továbbra is a minta méretét, $p_k(t, N)$ pedig az 5.8-ben bevezetett mennyiséget jelöli.

Az előző pontban bemutatott egyszerű modell esetében azt a feltételt kötöttük ki, hogy a BCG-nek egy adott magnitúdónál fényesebbeknek kell lenniük. Ezt természetesen két független esemény feltételes valószínűségének bevezetésével tudjuk leírni:

$$P_{kl}(X_{kl}, N | X_k > X_0) = P_{kl}(X_{kl}, N) P_k(X_k > X_0, N) \quad (5.13)$$

Az ehhez tartozó valószínűségi sűrűségfüggvény pedig:

$$p_{kl}(X_{kl}, X_0, N) = p_{kl}(X_{kl}, N) \int_{X_0}^{\infty} p_k(t, N) dt \quad (5.14)$$

Ez utóbbi mennyiség az eredeti szülőeloszlásból kitaró számolással elvileg meghatározható.

Megválaszolandó kérdés még, hogy mekkorára kell N -et választani. Az extrémérték-statisztikai leírás esetében általában $N \rightarrow \infty$, és kis N -ekre a korábban említettek szerint végesméret-korrektciókat kell bevezetni. A 5.5.2. fejezetben vett modellt vizsgálva azt láthatjuk, hogy legtöbbször kettő, de legfeljebb három nagyon fényes ($M < -22$) galaxis adódik klaszterenként, tehát az $N \rightarrow \infty$ közelítés messze nem érvényes. Emiatt a magnitúdórés meghatározásához egzakt számolásra van szükség, és az extrémek elmélete nem alkalmazható.

5.6. Összefoglalás

Az 5.4. fejezetben, és az 5.14–5.16. ábrákon megmutattuk, hogy az extrémérték- és sorrendi statisztika sikeresen alkalmazható az SDSS adataiból vett LRG minta esetében, és konzisztens magnitúdókat, valamint galaxissűrűséget ad. Szintén megmutattuk, hogy az említett statisztikai módszerek jelentős extrapolációs lehetőséget biztosítanak olyan esetben, ha a kiválasztási effektusok dominálják a mintát. Mivel az SDSS adatbázisa az ég jelentős területét lefedi, természetesen az általunk összeállított minta mindenféle méretű halmazokban levő galaxisokat is tartalmazott, következésképpen ezen klaszterek legfényesebb galaxisait is. Ez két fontos kérdést vet fel, amit az alábbi pontokban kívánunk körüljárni.

5.6.1. A klasztereződés hatása a statisztikára

Abból a célból, hogy a technika extrapolációs lehetőségeit bemutassuk, a mintát vöröseltolódás-héjakra osztottuk – ahelyett, hogy például térbeli binekkel dolgoztunk volna –, és ezen vöröseltolódás-héjak mindegyike tartalmazta a legfényesebb klasztergalaxisokat. Mi történne abban az esetben, ha a mintánkat az ég egy olyan területére korlátoznánk, ahol kizárólag halványabb galaxisok lennének? A halványabb legfényesebb magnitúdókból a sorrendi statisztika megfontolásai alapján az következik, hogy a vizsgált égterületen a galaxisok sűrűsége ennek megfelelően alacsonyabb. Egy (abszolút magnitúdóban) halványabb minta tehát szükségképpen kevesebb galaxisból áll, és feltételezhetően úgy lett kiválasztva, hogy elkerülje a sűrű galaxishalmazokat. Megfordítva, a fényes galaxisok irányába történő megfigyelések szükségképpen sűrű galaxiscsoportokat és -halmazokat fognak érinteni, következésképpen az észlelt galaxisszám is magasabb lesz.

5.6.2. A klaszterek legfényesebb galaxisai

A BCG-eket és a második legfényesebb halmazbeli galaxisokat elválasztó magnitúdórés kérdésében abban látjuk a megoldást, hogy a klaszterek *nem tekinthetők feltétel nélkül véletlenszerű mintáknak*, melyek valamilyen eloszlásból származnak. A minták, melyek tartal-

maznak egy BCG-t, speciálisak. Az 5.5. fejezetben megmutattuk, hogy ha a galaxisok térbeli eloszlását is figyelembe vesszük, akkor a klaszterek statisztikája nagyon könnyen reprodukálható egy univerzális LRG luminozitásfüggvény alapján. Ez arra utal, hogy a BCG-k is ugyanúgy keletkeztek (bár feltehetően hevesebb életet éltek), mint a többi LRG, és a galaxisok kannibalizmusa kizárható forgatókönyv. Mindenesetre a mi modellünk meglehetősen egyszerű, és a BCG-k statisztikájának megértése további vizsgálatokat igényel, melyek N-test szimuláción alapuló összeolvadási fák statisztikai vizsgálatát is érintik.

5.7. Az extrémérték- és sorrendi statisztika lehetséges felhasználásai a jövőben

A dolgozat ezen részében bemutatott statisztikai eljárás az abszolút magnitúdó és a spektroszkópiai vöröseltolódás pontos mérésén nyugszik. A közeljövő nagy égtérképezései – mint a PanSTARRS vagy az LSST – mély fotometriai észleléseket fognak végezni közvetlen spektroszkópiai vöröseltolódás-mérések nélkül. A fotometrikus vöröseltolódás-becslő eljárások relatív hibája kb. $\Delta z \approx 0,01 - 0,02$ (Hildebrandt et al. 2010), ami lényegesen nagyobb az általunk használt vöröseltolódás-héjak vastagságánál. Következésképpen a sorrendi statisztikán alapuló módszerek esetében a fotometrikus becslt vöröseltolódásoknál jóval pontosabb adatokra van szükség. Ettől függetlenül érdekes lehet a fotometrikus égfelméréseket követő spektroszkópiai mérések olyan tervezése, hogy először a fotometriai adatok alapján becslt legnagyobb abszolút magnitúdójú galaxisok kerüljenek megfigyelésre. Ha célunk a galaxissűrűség meghatározása, ez a módszer lényegesen lecsökkenti a vizsgálandó galaxisok számát (vöröseltolódás-héjanként néhány ezer helyett néhány száz), és a galaxisok száma mégis olyan nagy távolságokig meghatározható marad, ahol már csak a legfényesebb galaxisok figyelhetők meg.

A k . legfényesebb galaxisok (k nagy értékei esetén) hasznos standard gyertyák, melyek kiválóan alkalmasak a kozmológiai modellek ellenőrzésére. Fontos azonban, hogy a módszer, legalább is a vörös óriásgalaxisok esetében, érzékeny az evolúció-korrigált abszolút magnitúdó meghatározásakor használt modell jóságára, továbbá a luminozitásfüggvény paramétereinek (α és M^*) megváltozására a kozmikus idő függvényében.

Mennyiségek jelölése

jelölés	jellemző egység	mennyiség
α	-	a Schechter-függvény exponense
ϕ	$\text{Mpc}^{-3} \text{ mag}^{-1}$	galaxissűrűség
k	-	sorrendi statisztika rendje
L	erg s^{-1}	luminozítás
M	mag	abszolút magnitúdó
$M_{(k)}$	mag	k . legfényesebb abszolút magnitúdó
M_{kl}	mag	magnitúdó rész
n	Mpc^{-3}	galaxis-számsűrűség
N	-	statisztikai minta elemszáma
p	-	valószínűség
σ	-	szórás
t	Gyr	kozmikus idő
t_{ib}	Gyr	visszatekintési idő
V	Mpc^3	térfogat
z	-	vöröseltolódás

Irodalomjegyzék

- Abazajian, K. N., et al. 2009, *ApJS*, 182, 543
- Aragón-Salamanca, A., Baugh, C. M., & Kauffmann, G. 1998, *MNRAS*, 297, 427
- Bell, E. F., Zheng, X. Z., Papovich, C., Borch, A., Wolf, C., & Meisenheimer, K. 2007, *ApJ*, 663, 834
- Bell, E. F., et al. 2004, *ApJ*, 608, 752
- Bernstein, J. P., & Bhavsar, S. P. 2001, *MNRAS*, 322, 625
- Bhavsar, S. P., & Barrow, J. D. 1985, *MNRAS*, 213, 857
- Blanton, M. R., & Roweis, S. 2007, *AJ*, 133, 734
- Blanton, M. R., et al. 2003, *AJ*, 125, 2348
- Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, *A&A*, 363, 476
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, *AJ*, 120, 1588
- Burke, D. J., Collins, C. A., & Mann, R. G. 2000, *ApJ*, 532, L105
- Csabai, I., Connolly, A. J., Szalay, A. S., & Budavári, T. 2000, *AJ*, 119, 69
- De Lucia, G., & Blaizot, J. 2007, *MNRAS*, 375, 2
- Dressler, A. M. 1976, PhD thesis, California Univ, Santa Cruz.
- Eisenstein, D. J., et al. 2001, *AJ*, 122, 2267
- Ellis, R. S., Smail, I., Dressler, A., Couch, W. J., Oemler, Jr., A., Butcher, H., & Sharples, R. M. 1997, *ApJ*, 483, 582
- Faber, S. M., et al. 2007, *ApJ*, 665, 265
- Fisher, R. A., & Tippet, L. H. C. 1928, *Proceedings of the Cambridge Philosophical Society*, 24, 180
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Gallazzi, A., Charlot, S., Brinchmann, J., & White, S. D. M. 2006, *MNRAS*, 370, 1106
- Geller, M. J., & Peebles, P. J. E. 1976, *ApJ*, 206, 939
- Geller, M. J., & Postman, M. 1983, *ApJ*, 274, 31
- Gnedenko, B. V. 1948, *Uspekhi Mat. Nauk*, 3, 187
- Gumbel, E. J. 1958, *Statistics of Extremes* (Dover, New York)
- Gunn, J. E., & Oke, J. B. 1975, *ApJ*, 195, 255
- Györgyi, G., Moloney, N. R., Ozogány, K., & Rácz, Z. 2008, *Physical Review Letters*, 100, 210601
- Hilдебрандт, H., et al. 2010, *A&A*, 523, A31+
- Kaviraj, S., et al. 2008, *MNRAS*, 388, 67
- Lin, Y.-T., Ostriker, J. P., & Miller, C. J. 2010, *ApJ*, 715, 1486
- Liu, F. S., Xia, X. Y., Mao, S., Wu, H., & Deng, Z. G. 2008, *MNRAS*, 385, 23
- Loh, Y.-S. 2004, PhD thesis, PRINCETON UNIVERSITY
- Loh, Y.-S., & Strauss, M. A. 2006, *MNRAS*, 366, 373
- Maraston, C., Strömbäck, G., Thomas, D., Wake, D. A., & Nichol, R. C. 2009, *MNRAS*, 394, L107
- Ostriker, J. P., & Hausman, M. A. 1977, *ApJ*, 217, L125
- Padmanabhan, N., et al. 2005, *MNRAS*, 359, 237
- Paranjape, A., & Sheth, R. K. 2011, ArXiv e-prints, in preparation
- Postman, M., & Lauer, T. R. 1995, *ApJ*, 440, 28
- Sandage, A. 1972, *ApJ*, 178, 1
- , 1976, *ApJ*, 205, 6
- , 1988, *ARA&A*, 26, 561
- Scott, E. L. 1957, *AJ*, 62, 248
- Stanford, S. A., Eisenhardt, P. R., & Dickinson, M. 1998, *ApJ*, 492, 461
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Taghizadeh-Popp, M., e. a. 2011, in preparation
- van Dokkum, P. G., Franx, M., Kelson, D. D., & Illingworth, G. D. 1998, *ApJ*, 504, L17+
- von der Linden, A., Best, P. N., Kauffmann, G., & White, S. D. M. 2007, *MNRAS*, 379, 867
- Wake, D. A., et al. 2006, *MNRAS*, 372, 537
- White, I. M., et al. 2008, *MNRAS*, 387, 1253

III. rész

Virtuális Obszervatóriumok

Bevezető a III. részhez

Az elmúlt évtized nagyobb asztrofizikai megfigyelési és szimulációs projektjei (de beszélhetnénk más tudományterületekről is a genetikától az óceánkutatáson és az Internet szerkezetének kutatásán át a geofizikáig) olyan mennyiségben szolgáltatott mérési adatokat, melyek lényegesen átfomálták a tudományos adatok feldolgozásának, tárolásának, a kutatókhoz való eljuttatásának, kiértékelésének és vizualizációjának módszereit (Szalay & Gray 2006). Jim Gray, a számítógépes adatbázisok Turing-díjjal elismert úttörője egyenesen paradigmaváltásról beszélt: a korábbi három fő tudományos paradigma – kísérlet, elmélet és szimuláció – mellett egy új módszer, az adatintenzív kutatás, adatbányászat jelent meg (Szalay & Blakeley 2009). Az egyre nagyobb teljesítményű adatgyűjtő műszerek, illetve soha nem látott méretű szimulációk által szolgáltatott adatmennyiség minden évben megduplázódik (vö. Moore-törvény). A rengeteg adat egyre olcsóbban tárolható, és a nagy adattárházak az Internet segítségével összekapcsolhatóvá váltak. Ez a növekedési ütem természetszerűleg maga után vonta az adatkezelési módszerek újragondolását (Gray & Szalay 2006, Bell et al. 2007). Szalay Sándor és kutatócsoportja a baltimore-i Johns Hopkins Egyetemen nem csak az asztrofizikai adatkezelést forradalmasította azzal, hogy az SDSS égtérkép adataiból adatbázist és az Internet segítségével bárki számára elérhető adattárházat épített, de egyszersmind más tudományterületek számára is utat mutatott. A csillagászat területén az SDSS adatbázisát és eszköztárát továbbgondolva született meg a Virtuális Obszervatórium (VO), mely mára széleskörű nemzetközi együttműködéssé nőtte ki magát (Szalay & Gray 2001; 2006).

A Virtuális Obszervatórium alapötlete, hogy a nagy, sok százmillió dolláros projektek keretében gyűjtött adatokat szolgáltató, nagy kutatóközpontokban felépített adattárházakat úgy kell kialakítani, hogy az adatokhoz az Interneten keresztül, szabványos adatlekérdezési protokollokat és adatformátumokat használva bárki hozzáférhessen. Egy lépéssel továbbmenve: az adattárházakat össze kell kapcsolni, és fel kell készíteni arra, hogy *automatikusan* együtt tudjanak működni, ezzel is elősegítve a több forrásból származó adatokkal dolgozó kutatók munkáját. Az ilyen jellegű fejlesztések nagyságrendekkel csökkenthetik az egyidejűleg több műszer adataival dolgozó csillagászok adatkezelésre fordított munkaidejét, akik így a számítástechnikai problémák megoldása helyett a tudományos kérdésekre fókuszálhatnak.

A nagy adattárházaknak az adatok tárolásán és nyers formátumban való kiszolgálásán túl meg kell oldani azok hatékony feldolgozását is. Mivel a hálózatok adatátviteli sebessége nem nő a felhalmozódó adatmennyiséggel arányos ütemben, ezért mindenképpen szükség van arra, hogy magát a számítást vigyük közel az adatokhoz, és ne az adatokat

a számítást végző processzorokhoz. Ugyanakkor ügyelni kell arra is, hogy az adatintenzív problémák esetében a számítógépek IO rendszere² a szűk keresztmetszet, és nem az adatfeldolgozást végző processzorok teljesítménye. Emiatt újra kell gondolni az optimális hardverről kialakított nézeteinket, és a számítási teljesítmény hajszolásán túl figyelembe kell venni az adatok elérésének sebességét, valamint az egyre fontosabbá váló környezetvédelmi szempontokat is.

A modern csillagászat kezdetén az adatok publikálására elegendőek voltak a folyóiratok. Ezek az információk ma is, akár az eredeti nyomtatott formában, akár a elektronikusan, könnyen elérhetőek. Nagy kérdés, hogy a jövőben mi lesz a sorsa a csak elektronikus formában létező nagy adatbázisoknak. Az amerikai Nemzeti Tudományos Alapítvány már ma előírja, hogy az általa támogatott kutatások során gyűjtött adatokat a kutatás befejeztével teljesen publikussá kell tenni. Az adattárházak üzemeltetése azonban komoly hardver és szoftver eszközöket igényel, a számítástechnikai technológiák pedig rohamlőptekkel fejlődnek, cserélődnek. Nagy kérdés, hogy elegendő-e pusztán az adatok archiválása a jövő kutató nemzedékei számára, vagy – és feltehetően ez a helyes módszer – az adatok feldolgozásához fejlesztett rengeteg szoftvereszköz jövőbeli felhasználhatóságát is biztosítani kell.

Az asztrofizikai Virtuális Observatórium – nem is mint konkrét szoftvertechnológia, sokkal inkább mint nemzetközi együttműködés – ezekre a kérdésekre igyekszik megadni a válaszokat.

A csillagászati adatok feldolgozásán túl a Virtuális Observatórium alapötlete más tudományterületen is alkalmazható. Példaképpen megemlíjtük Matray et al. (2007) által, az ELTÉn folyó kutatás keretében kidolgozott, Internet alapú hálózatanalitikai mérések tárolására létrehozott adattárházat, mely a VO-hoz hasonló felhasználói funkciókkal rendelkezik.

Ebben a részben áttekintjük az adattárházakat, és bemutatunk pár konkrét szoftvertechnológiát, mely a negyedik paradigma számára elengedhetetlen. Ezután kitérünk a sokdimenziós adatbázisok indexelése területén végzett saját munkánkra, majd röviden bemutatjuk a fotometriai katalógusok keresztazonosítására kidolgozott megoldásunkat a *SkyQuery*, és az SDSS galaxisspektrumainak publikálására és előfeldolgozására készített, a Virtuális Observatóriumba szorosan illeszkedő *Spektrum Szolgáltatást*.

²IO = input-output, a kifejezést a permanens háttértár vezérlőrendszerére használjuk.

6. fejezet

Adattárházak a tudományban

Adattárház alatt olyan számítógépes célrendszert értünk, mely elsődleges feladata nagy mennyiségű adat tárolása és a hozzá kapcsolódó alkalmazások pillanatnyi adatigényének rövid határidejű kiszolgálása, illetve ezen alkalmazások futtatása. Mindkét kitétel lényeges: az adattárház nem archívum, vagyis az adatokhoz való hozzáférést rövid határidővel (jellemzően legfeljebb néhány perc) nyújtja, amit az adatok pörgő lemezeken való tárolása biztosít. Másfelől az adattárházakban az adatokat úgy szervezzük, katalogizáljuk, hogy azok az előre pontosan meg nem jósolható alkalmazások igényeinek is megfeleljenek. Magyarán lehetőleg optimális, de mindenképpen univerzális sémákat követünk. Az adattárházak új generációja nem csak az adatok rendszerezését és gyors elérését biztosítja (nagy teljesítményű és kapacitású IO rendszer), de képes kiszolgálni a komoly számítási igényeket is (sok és erős CPU¹, netán GPU²) (Szalay et al. 2010, Thakar et al. 2011).

Bár a tudományos kísérletekből nyert adatok számítógépes feldolgozásának a számítástechnikával egyidős története van, a széles kutatói kör számára elérhető adattárházak megjelenése csak a kétezres évek elejére tehető. A csillagászat terén az első ilyen rendszerek már korábban megjelentek (Simbad, NED, ADS stb.), de az első szigorú értelemben vett adattárház, mely a felhasználóknak a keresési lehetőségen túl számítási kapacitást is biztosított az SDSS SkyServer³ volt (Szalay et al. 2000, Gray et al. 2001, Stoughton et al. 2002).

Az adattárházak megépítése olyan bonyolult szoftvertechnológiákon alapszik, melyek egyedi, a konkrét célra történő kifejlesztésére nincsen mód. Lévéen speciálisan tudományos célra készült adatbázis-kezelő rendszerek ma még nincsenek (bár ilyen irányú törekvések már vannak, ld. Stonebraker et al. (2009)), az eddigi megoldások mind üzleti célra készült szoftverek, szinte kivétel nélkül relációs adatbázis-kezelők alapjaira épülnek.

¹CPU = Central Processing Unit

²GPU = Graphical Processing Unit, a grafikus processzorok általános célú számolási feladatokra is használhatóak.

³<http://skyserver.sdss.org>

6.1. A relációs adatbázis-kezelők áttekintése

A relációs adatbázis-kezelő rendszerek⁴ (RDBMS), és az ezek futtatásához szükséges hardverek részletes bemutatására sajnos nincsen lehetőség egy elsősorban csillagászati témájú disszertáció keretein belül, de néhány fontos jellemzőre kitérünk azért, hogy az általunk kifejlesztett csillagászati célú adatbáziskezelő-kiegészítések értelmét meg tudjuk világítani. Adatbázisok témájában ajánljuk Ullmann & Widom (2008) és Hector-Molina & Ullmann (2008) könyveit.

A relációs adatbázisok (Codd 1970) alapvető adatstruktúrái a táblák, melyek sorokból (rekordok), és oszlopokból (mezőkből) állnak. Az oszlopok névvel, adattípusal és mérettel rendelkeznek, ezek határozzák meg a táblák logikai struktúráját. Egy adatbázis-tábla oszlopaikat előre definiálni kell, és a tábladefiníciók az adatok betöltése után csak korlátozottan módosíthatók. A táblák tetszőleges számú rekordot tartalmazhatnak, de minden rekord meg kell feleljen a tábla mezőinek. Az egyes rekordok általában egy egyedi, egész számból álló azonosítókulccsal (általában pusztán egy sorszámmal) vannak ellátva.

A relációs adatbázisokból az *SQL nyelven*⁵ megírt *lekérdezések* segítségével nyerhetők ki az adatok (Chamberlin & Boyce 1974). Az alapvető lekérdezések általában igen egyszerűek: megadható, hogy mely táblák mely oszlopaiknak értékeit, milyen sorrendben és milyen szűrési feltételekkel kívánjuk visszakapni. Tudjuk képezni táblák Descartes-szorzatát, mely logikailag egy új, a két tábla rekordjainak rendezett páraiból álló tábla lesz. Lehetőség van továbbá a tábla aggregálására, azaz a tábla soraiban szereplő értékek valamilyen statisztikai mennyiségének (átlag, összeg stb.) képzésére. Ezekből az egyszerű műveletekből persze igen komplex problémákat megoldó lekérdezések építhetők.

A táblák oszlopaikat ún. *indexekkel* láthatjuk el, melyek az adott oszlopok alapján gyorsabb keresési lehetőséget biztosítanak. Egy táblára több index is kerülhet, ezek közül egy optimalizáló algoritmus képes kiválasztani a konkrét lekérdezés végrehajtásához legmegfelelőbb indexet. Az indexek közül táblánként egy index kitüntetett szereppel bír: az *elsődleges kulcs*, vagy *klaszterezett index* azt is meghatározza, hogy az adatok milyen adatstruktúrában (Comer 1979, Copeland & Khoshafian 1985), és milyen sorrendben kerülnek tárolásra a háttértárolón. A táblák definícióinak és az indexeknek az összességét az adatbázis *sémájának* nevezzük.

A jelenleg elérhető relációs adatbázis-technológiák mind lemezes háttértárolók alkalmazására épülnek, és ez nagyban befolyásolja belső logikájukat is. Az adatbázisokban tárolt adatok mérete legtöbbször a rendszert futtató szerver központi memóriakapacitásának százszorosa-ezerszerese, miközben az adatok lemezzről való betöltése meglehetősen lassú a memóriaműveletek sebességéhez képest. Általában igaz, hogy a lemezek szek-

⁴RDBMS = Relational Database Management System

⁵SQL = Structured Query Language = Strukturált Lekérdezőnyelv

venciális olvasása nagyságrendekkel gyorsabb, mint a véletlenszerű adatelérés. Emiatt kiemelten fontos, hogy az adatok milyen struktúra szerint kerülnek rögzítésre; erről a későbbiekben még ejtünk szót.

6.1.1. Néhány szó az SQL nyelvről

Az SQL nyelv ún. *deklaratív nyelv*, ami annyit jelent, hogy maga a lekérdezés nem tartalmazza az adatok elérésére szolgáló algoritmust, hanem azt az adatbázis-kezelő maga alkotja meg egy *lekérdezésoptimalizáló* modul segítségével. Az adatbázis-kezelőkben különböző adatstruktúrákra és keresési feladatokra optimalizált algoritmusok kerültek implementálásra, és ezeket az szoftver az aktuális lekérdezésnek leginkább megfelelő módon kombinálja. A fejlett rendszerek optimalizáláskor számos szempontot képesek figyelembe venni: az adatok háttértáron való szervezését, a rendelkezésre álló indexeket, a táblákban tárolt számértékek statisztikáját, valamint az optimális memória, processzor és hálózat használatot.

Egy igen egyszerű lekérdezés lehet a következő:

```
SELECT A.oszlop1, A.oszlop3 FROM A
WHERE A.oszlop1 > 10 AND A.oszlop2 < 10
```

Értelemszerűen ez visszaadja az A tábla sorait úgy, hogy a visszaadott sorok az *oszlop1* és *oszlop3* oszlopokból állnak; azonban csak azok a sorok lesznek az eredménytáblában, melyek a WHERE kulcsszó utáni feltételeknek megfelelnek. Egy lekérdezéssel két táblát is összekapcsolhatunk:

```
SELECT A.oszlop1, B.oszlop2 FROM A, B
WHERE A.ID = B.masikID
```

Itt a FROM kulcsszó után két tábla szerepel, így ezek Descartes-szorzatát kell érteni. A WHERE kulcsszó utáni feltétel szűri ki azokat a rekordpárokat, melyek a végső eredményhalmazba bekerülnek. A második példa esetében ugyan elméletileg a szervernek meg kellene határoznia a teljes Descartes-szorzatot, de a lekérdezésoptimalizáló algoritmus felismeri, hogy a feltételben olyan egyenlőség van megkövetelve, mely a táblák rekordjainak azonosítóira vonatkozik (A.ID, illetve B.masikID), és így a lekérdezést nem egymásba ágyazott ciklusokkal (ami a Descartes-szorzat-képzés algoritmikai megfelelője lenne), hanem ún. *hash-join* művelettel valósítja meg, mely az előbbinél nagyságrendekkel gyorsabb⁶.

⁶Valóban, az egymásba ágyazott ciklusok $O(N_A N_B)$ műveletet igényelnek, míg a hash-join lineárisan skálázódik, bár a hash műveletek az egyszerű ciklusoknál költségesebbek.

6.1.2. Relációs adatbázis-kezelők tudományos célú alkalmazása

A relációs adatbázis-kezelők üzleti adatbázisok megvalósítására készülnek, hiszen a dollármilliárdokat elérő szoftverfejlesztési költségek csak piaci megoldások esetében finanszírozhatók. Ez természetesen alapvetően befolyásolja a rendszerek fő fejlesztési irányait is. Szerencsére az egyik fő fejlesztési irány, hogy az adatbázis-kezelők a tranzakciókezelésen túl (gyors, kevés adatot igénylő műveletek) adattárház-alkalmazások számára is használhatóak legyenek (sok adatot igénylő, nagy számítási igényű, sokáig tartó műveletek). Tudományos alkalmazások esetében minket szinte kizárólag ez utóbbi alkalmazás érdekel.

Az adatbázis-kezelők számos olyan alapvető szolgáltatást nyújtanak, melyek tudományos adatbázisok esetében is elengedhetetlenek. A korábban már említett optimalizált háttértár-, processzor- és memóriahasználton túl kiemeljük az adatok konzisztenciájának automatikus biztosítását és biztonsági másolatok készítésének lehetőségét. Az adatbázis-kezelők gyártói nagy figyelmet fordítanak megoldásaik skálázhatóságára, egyes termékek a hordozható számítógépektől kezdve a nagy teljesítményű szerverekig bármilyen hardveren optimálisan futnak, ugyanakkor könnyen használható, SQL nyelven programozható felhasználói felületet biztosítanak. Fontos tudni viszont, hogy az SQL-alapú relációs adatbázis-kezelők jelenleg csak egy fizikai szerveren futnak optimálisan. Ha az adatbázisba szervezendő adatmennyiség több szerver használatát, és azok hálózaton történő összekapcsolását igényli, akkor szükségképpen egyedi megoldásokat kell alkalmazni az optimális teljesítmény elérése érdekében, ld. 9.4. fejezet.

Hátrányként a szoftverek árán túl (bár a piacon elérhető termékek jelentős részéhez egyetemi kutatási célból ingyen hozzá lehet jutni) azt érdemes megemlíteni, hogy míg hétköznapi használatuk valóban egyszerű, speciális célokra történő alkalmazásuk, saját fejlesztésű modulokkal való kiegészítésük már jóval több számítástechnikai ismeretet igényel, mint ami egy természettudományos egyetemi képzés keretein belül elsajátítható. Ez a nagy adatigényű projektek esetében mindenképpen célirányosan képzett szakemberek alkalmazását teszi szükségessé.

Csillagászati célú felhasználás esetén adódik a kérdés, hogy van-e lehetőség az adatok égi koordináták szerinti indexelésére, hatékony lekérdezésére. Bár az adatbázis-kezelő szoftverek egy része rendelkezik térinformatikai (földi koordináták szerint szervezett) adatok kezelésére szolgáló modulokkal, ezek teljesítménye, pontossága valamint szolgáltatásaik köre nem egyezik meg a csillagászati igényekkel. A 9.5. fejezetben röviden bemutatunk egy programcsomagot, mely ezt a hiányosságot kívánja orvosolni.

6.2. Nem relációs adattárházak

Tudományos adatok esetében gyakran előfordul, hogy a mérési pontok, vagy a szimulációk eredményei egy rácson vannak definiálva. Az ilyen jellegű adatokat általában tömbökben tároljuk. A relációs adatmodell egyik fontos hiányossága pont az, hogy nem támogatja a tömb adatstruktúrát. Néhány termék ugyan rendelkezik egyszerű tömb adattípussal, de egyrészt a tömbök mérete korlátozott, másrészt a tömbök kezelésére optimalizált algoritmusok hiányoznak. A 7.1. fejezetben röviden ismertetünk néhány általunk kidolgozott kiegészítést, mellyel egy relációs adatbázis-kezelő rendszer felkészíthető tömbök hatékony kezelésére, így például kozmológiai szimulációk tárolására.

Szintén speciális esetet jelentenek a gráfadatbázisok. A gráfok a statisztikus fizikától kezdve a szociális hálózatok és biológiai fehérjehálózatok tudományán át az Univerzum nagyskálás szerkezetével foglalkozó asztrofizikai területekig, vagy a számítógépes hálózatokig mindenütt megjelennek. Bár gráfok a relációs adatmodellben egyszerűen reprezentálhatók, hatékony statisztikai analízisük speciális algoritmusokat igényel, melyek ezekben a termékekben nem állnak rendelkezésre.

Az adattárházak harmadik fontos fajtája az ún. *objektumorientált* rendszerek. Lényegük, hogy az adatok tárolásának módja nagyon hasonlít a programozási nyelvek objektumorientált adatmodelljére, így a programozók mentesülhetnek a két adatmodell közötti konverzió fárasztó feladata alól, ami a relációs adatbázisok esetében szinte elkerülhetetlen. Mivel az objektumorientált adatbázisok lekérdezéseinek optimalizálása a relációs rendszerektől eltérően nem kiforrott, ezért ezt a technológiát néhány kivételtől eltekintve (mint például a CERN részecskefizikai adatbázisai) nem használják.

Említettük, hogy a relációs adatbázis-kezelők jelenleg nem, vagy csak alig támogatják több szerver összekapcsolását. Az adatbázisok elosztott rendszerekre történő átalakítását az alapvető algoritmusok újragondolásával kell elvégezni. Mivel szükség van extrém nagy adatmennyiséget kezelő adattárházakra, ezért néhány nagy internetes szolgáltató (Google, Amazon, Yahoo) saját fejlesztéseként megjelentek különböző „noSQL” rendszerek. Ezek közös jellemzője a feladat-végrehajtás gyakorlatilag korlátlan párhuzamosíthatósága, és rengeteg szerveren történő elosztása. Eredményképpen olcsó hardver elemek alkotják, de ezekből nagy kapacitású rendszerek építhetők. A problémát az jelenti, hogy e rendszerek programozására nem áll rendelkezésre az SQL-hez hasonló szabványos, deklaratív lekérdezőnyelv. Következésképpen a „noSQL” rendszereket imperatív módon, az adatok pontos szervezési struktúráját ismerve lehet csak programozni, bár néhány olyan egyszerű keretrendszer rendelkezésre áll ennek egyszerűbbé tételére, mint például a MapReduce (Dean & Ghemawat 2008).

A teljesség igénye miatt megemlítjük a szöveges adattárakat is, melyek természetes nyelven írt szövegeket tesznek kereshetővé, számítógépes algoritmusokkal feldolgozha-

tóvá. Ezek elsősorban a szakirodalom kezelésében nyújtanak segítséget, de léteznek olyan szakterületek is (pl. biológiai fehérjehálózatok), ahol az információk gyakran nem is állnak rendelkezésre strukturált formában, csak a publikációk szövegeiből gyűjthetők ki.

6.3. A jövő adatbázis-technológiái

A technológiák gyors fejlődése miatt a jövőre nézve nehéz jóslatokat adni, de röviden megemlítnék pár lehetséges irányt amerre a tudományos adattárházak fejlődésnek indulhatnak.

A félvezető-alapú, flashmemóriát alkalmazó háttértárak (SSD-k) kapacitásának növekedésével, és áruk csökkenésével valószínűleg el fog érkezni a lemezalapú, mechanikailag sérülékeny tárolórendszerek alkonya (Gray 2007). A flashmemóriák hatalmas előnye a lemezekkel szemben, hogy olvasásuk véletlenszerű adatelérési mintázatok esetében is éppen olyan gyors, mint szekvenciális elérés esetén. Hátrányuk, hogy írásuk lassú (bár a lemezekénél még így is gyorsabb), mivel a flashmemória fizikai jellemzői miatt egyes biteket nem, csak teljes, több kB méretű lapokat lehet írni. A sok írástól a memória „öregszik”, ami a lemezes adattároló rendszerekhez hasonlóan meghibásodásokhoz vezethet. Felmerül a kérdés, hogy az SSD-k szélesebb körű elterjedése mennyiben igényli majd az adattárház megoldások belső algoritmusainak újragondolását. Úgy tűnik, hogy míg az adatírási mintázatokat érdemes lesz az SSD-k jellemzőihez igazítani, az adatok szervezési struktúrája (B-fák és lapok) nem fog lényegesen változni. Ennek fő oka, hogy a processzorok gyorsítótárainak korlátos mérete miatt a jövőben is szükséges lesz az adatokat kis méretű, egy lépésben beolvasható lapok formájában tárolni.

Ma már elérhetőek igen nagy központi memóriával szerelt, sokprocesszoros szerelvek, ám ezek ára a hagyományosnak tekinthető RAID-rendszerekkel⁷ szemben igen borsos. További hátrányuk, hogy a központi memóriát folyamatosan frissíteni kell, így egy áramkimaradás az adatok elvesztését okozza. A már hivatkozott előadásában Gray (2007) kiemelte, hogy a központi memóriában levő adatok mindenféle háttértáron tárolt adatoknál nagyságrendekkel gyorsabban feldolgozhatóak, ám még a memória elérése is lassú ahhoz képest, hogy a modern processzorok milyen számítási teljesítményre képesek. Ha az adatok egy órajelcikluson belüli elérése a cél, azokat mindenképpen a processzor első szintű gyorsítótárában kell tartani. A számítási kapacitás optimális kihasználása érdekében az algoritmusokat ezért úgy kell átalakítani, hogy azok feltétlen figyelembe vegyék az aktuális hardverarchitektúra gyorsítótárainak jellemzőit. Ez a feladat jóval nagyobbban ígérkezik, mint a hagyományos lemezekről az SSD-kre való áttérés.

A relációs adatmodell követő rendszerek esetében már ma is kezd elterjedni, hogy a statisztikai alkalmazásokra tervezett adatbázisok esetében a táblaadatok szervezése a

⁷RAID = redundant array of identical disks, azonos lemezegységekből épített redundáns rendszer

hagyományos B-fákba rendezett lapok helyett oszloponként történik. Ez egyrészt lényegesen gyorsítja a széles táblákon futó, de kevés oszlopot érintő lekérdezések végrehajtását, másrészt lehetővé teszi a modern processzorok gyorsítótárainak jóval hatékonyabb kihasználását (Stonebraker et al. 2005).

A 6.2. fejezetben már említettük a tömb alapú adattárakat, mint a tudományos adatfeldolgozás számára a jövőben elengedhetetlen rendszereket. A teljesen új alapokra helyezett tömb alapú adatbázis-kezelő szoftverek prototípusainak fejlesztése már megkezdődött (Stonebraker et al. 2009), azonban a technológia kiforrása, elterjedése jó néhány évet, akár egy évtizedet is igénybe vehet. Addig is szükség lesz a hagyományos relációs adatbázis-technológiák kibővítésén alapuló tömbkezelő megoldások használatára (Maier & Vance 1993, Baumann 1999, van Ballegooij 2004, Dobos et al. 2011b).

Az adattárházak alapjainak újragondolásakor az adattárolási technológiák fejlesztésével párhuzamosan oda kell figyelni az adatfeldolgozás hatékonyságára is. Ehhez olyan új lehetőségeket is integrálni kell az adatbázis-kezelő rendszerekbe, mint amilyen például a grafikus processzorok általános számítási célokra történő alkalmazása (Szalay et al. 2010).

6.4. Adattárházak optimális hardverei

Amdahl törvénye szerint egy számítási probléma párhuzamosíthatóságát a problémának az a része korlátozza, mely kizárólag szekvenciálisan hajtható végre. Hiába osztjuk szét a párhuzamosítható feladatot száz processzor között, ha a feladat 20 % -a kizárólag szekvenciálisan oldható meg, és az egy óráig tart; ekkor egy óránál rövidebb futási időt nem tudunk elérni (Gray & Shenoy 2000).

Adatfeldolgozásra optimalizált rendszerek esetében a gyenge láncszem az IO alrendszer. Korábban említettük, hogy a lemezek olvasási sebessége csak szekvenciális olvasás esetében optimális, tehát minél nagyobb százalékban tartalmaz a megoldandó számítási probléma adatbeolvasási műveletet, annál inkább az IO rendszertől válik függővé a számítási teljesítmény. Ezen több és erősebb processzor hozzáadásával sem lehet segíteni. A helyzet pontosan ellentétes azzal a felfogással, amit a Beowulf-jellegű⁸ számítási klaszterek képviselnek, ahol a sok processzorral gyakran csak lassú, a hálózaton keresztül elérhető háttértár jut.

A helyzeten úgy lehet változtatni, ha az IO rendszert is elosztottá tesszük, és a klaszter egyes gépeihez közvetlenül csatolt, nagy sebességű IO rendszereket építünk. Egyetlen diszkegység szekvenciális olvasási sebessége a 100 MB s^{-1} nagyságrendbe esik, mely sebesség átlépésének mechanikai korlátja van, nevezetesen a lemezek fordulatszámára. Ma már grafikus processzorok alkalmazásával könnyen építhetők teraflop⁹ számítási telje-

⁸Olcso számítógépekből épített, sokprocesszoros elosztott rendszerek.

⁹1 teraflop = 10^{12} lebegőpontos számítási művelet másodpercenként.

sítményű rendszerek. Egy ilyen rendszer adattal való optimális kiszolgálásához legalább 1000 db 100 MB s^{-1} -os olvasási sebességet biztosító lemezegységre lenne szükség (Bell et al. 2007). Ennél nagyobb számítási teljesítmény kiszolgálására már lemezek tízezrei kellenek. Ilyen nagy rendszerekkel jelenleg csak a legnagyobb adattárház üzemeltetők rendelkeznek (Google, Microsoft, Amazon, Yahoo, eBay).

Szalay Sándor és kollégái a Johns Hopkins Egyetemen megépítették a világ első olyan tudományos céllal létrehozott számítógép-klaszterét, mely elsődlegesen nagy volumenű adatfeldolgozási szempontok figyelembevételével lett megtervezve. A klaszter – Jim Gray-nek dedikálva – a Graywulf nevet kapta. Az 50 egységből álló rendszer teljes másodpercenkénti adatátviteli sebessége 75 GB s^{-1} (Szalay et al. 2009), a teljes hasznosítható tárterület pedig 1,1 PB. A klaszter fogyasztása 56 kW. A dolgozat szerzője által fejlesztett speciális alkalmazások (ld. pl. 9.4. fejezet) ezen a klaszteren futnak.

Baltimore-ban jelenleg két új rendszer építése is zajlik. Az egyik a grafikus processzorokra, kis fogyasztású, hordozható számítógépekbe tervezett CPU-kra és félvezető alapú háttértárakra épülő, energiatakarékos *Amdahl-klaszter* (Szalay et al. 2010). Az Amdahl-klaszter lényege, hogy a Graywulf egységeinél egy nagyságrenddel olcsóbb, önmagukban jóval kisebb teljesítményű elemekből, de egységnyi számítási teljesítményre jutó adatolvasási sebességben közel két és félszeres mutatóval rendelkező számítógépekből épül. A közeljövő nagy adatfeldolgozó klasztere a *Data-Scope* lesz, melyen nagy teljesítményű IO rendszerek és grafikus processzorok integrálásával végezzük majd el Galaxisunk eddigi legnagyobb szimulációjának (Via Lactea projekt) analízisét.

6.5. Számítási felhők

Manapság egyre elterjedtebbé válnak az ún. *számítási felhők*. Megvizsgálandó tehát a kérdés, hogy a tudományos élet mennyiben profitálhat ezek szolgáltatásaiból. A számítási felhőkre érdemes egyfajta szolgáltatásként tekinteni, ahelyett, hogy konkrét technológiát értenénk mögötte. A nagy szerverparkokat üzemeltető vállalatok olyan platformokat dolgoztak ki, melyek segítségével amellet, hogy saját szolgáltatásaikat futtatják a szervereken, a rendelkezésükre álló „felesleges” processzoridőt, tárterületet és hálózati sávszélességet bérbe adhatják (Grossman et al. 2010).

A számítási felhők magasztalása nagy divat, de úgy tűnik, hogy jelenlegi formájukban nem ezek jelentik a tudományos adatkezelési problémák jövőbeli megoldását. A számítási felhők nagyméretű, elosztott, hálózaton keresztül elérhető háttértárakra alapoznak, melyek ellentmondásban állnak a 6.4. fejezetben kifejtett szempontokkal (Thakar et al. 2011). Gazdaságossági tanulmányok is azt mutatják, hogy jelenleg, és feltehetően a közeljövőben is a legköltséghatékonyabb megoldás az lesz, ha a felhő szolgáltatások igénybevétele helyett a nagy tudományos adattárházakhoz saját üzemeltetésű, a célra optimalizált

adatközpontokat építünk.

6.6. Környezetvédelmi szempontok

Nem szabad elmenni a nagy adatközpontok működtetésének környezeti hatásai mellett sem. Bár a számítástechnikai eszközök fogyasztás-teljesítmény aránya az elmúlt években jelentősen javult, az üzemben levő egységek számának rohamos növekedése miatt a vállalati adatközpontok energiafogyasztása jelentős tényezővé vált (Szalay et al. 2010), mely a világ villamosáram-fogyasztásának mintegy 1.1-1.5 % -át teszi ki. Egy-egy nagyvállalati adatközpont fogyasztása kisvárosok teljes energiafelvételével vetekedhet, de egy nagyobb méretű egyetemi rendszer is az 1-2 MW tartományba eshet (Koohey 2008; 2011).

A legnagyobb adatközpontok már ma töreksenek a környezetbarát megoldások megvalósítására, melynek legfontosabb eleme a szerverek által termelt hő hasznosítása; de legalább is a megtermelt hő lehetőleg legkisebb energiafelhasználás melletti elvezetése. Telente, illetve hűvös éghajlaton a termelt hő kiválóan használható épületek fűtésére, míg nyaranta, illetve melegebb éghajlatokon folyók, tavak vizének hűtővízként való felhasználása nyújthat megoldás. Nagyobb szerverközpontok esetében hűtőtornyos megoldások is szóba jöhetnek.

Mivel nagy méretekben a hűtés sokkal hatékonyabban oldható meg, a legnagyobb adattárházak részesedése a globális adatfeldolgozási célú energiafogyasztásból néhány százalékos alatti annak ellenére, hogy a szerverek mintegy 20 %-a ezekben az adatközpontokban található. Sajnos a kisebb szerverközpontok hűtésére a legelterjedtebb módszer aktív légkondicionáló berendezések üzemeltetése. Ez gyakorlatilag megduplázza a számítási igények kiszolgálásához szükséges energiafogyasztást. A környezeti hatások csökkentése érdekében a jövőben mindenképpen ügyelni kell a kisebb rendszerek hűtési infrastruktúrájának környezetbarát megoldására.

7. fejezet

Sokdimenziós adatbázisok építése relációs adatbázis-kezelőkkel *

A tudományos adatok legtöbbször egy többdimenziós paraméterteret töltenek ki. Elelegendő csak a kozmológiai szimuláció tér+idő- koordinátáira gondolni, vagy a fotometriai katalógusok sok színszűrővel mért magnitúdóadataira. Ebben a fejezetben olyan eseteket vizsgálunk, amikor a dimenziószám a tíz nagyságrendjébe esik.

Paraméterbecslő és automatikus klasszifikációt végző algoritmusok implementálása-kor a paramétertereket gyakran kezeljük euklideszi metrikus térként, még akkor is, ha a metrikának szigorú értelemben vett fizikai jelentés nem tulajdonítható. Különböző számítási problémák (nem paraméteres és paraméteres becslési eljárások, korrelációs függvények kiszámítása, sűrűségbecslő algoritmusok, vizualizáció) gyakran igénylik, hogy ezekben a metrikus terekben közeli szomszéd pontokat, vagy adott térfogatban levő pontokat keressünk. Az ilyen jellegű műveletek elvégzéséhez az adatokat speciális adatstruktúrákba kell szervezni, hogy az algoritmusok optimálisan működhessenek. A legtöbb ilyen adatstruktúra jól ismert (Samet 2006), de ezek nagyrészt központi memóriára lettek kidolgozva. Ezért mi az adatstruktúrák relációs adatmodellen alapuló adatbázis-kezelőkre való adaptálását tűztük ki célul. Kutatásaink során több olyan programcsomagot is kifejlesztettünk, mely többdimenziós adatok tárolását, valamint térbeli indexelését, keresését teszi lehetővé.

7.1. Tömbök relációs adatbázisokban

A relációs adatbázis-kezelők bemutatásánál megemlítettük azok egyik hátrányát: nem támogatják a tömb adattípusok (vektorok, mátrixok stb.) tárolását, kezelését, indexelését. Mivel a legtöbb adatbázis-kezelő rendelkezik bináris adatok tárolására fenntartott adattípussal, kézenfekvő egy olyan kiegészítés implementálása, mely a tömbök tárolását bináris típussal oldja meg, ugyanakkor megfelelő függvénykönyvtárat biztosít a tömbök manipu-

lálásához.

Ezzel a céllal dolgoztuk ki az „SqlArray” programcsomagot, mely a Microsoft SQL Server 2008 adatbázis-kezelőhöz nyújt széleskörű tömb-támogatást (Dobos et al. 2011b). A programcsomag támogatja az összes alapvető adattípus tömbként való kezelését, és külön optimalizált függvényeket tartalmaz kisméretű (tipikusan 8 kB-nál kisebb), és nagyméretű (akár 2 GB-os) tömbök kezelésére.

A tömbök tárolási módja úgy lett megválasztva, hogy az kompatibilis legyen a legelterjedtebb matematikai programcsomagokkal (LAPACK, FFTW, Matlab stb.). Néhány alapvető matematikai függvény integrálását el is végeztük, így közvetlenül az adatbázis-kezelőből (mindenféle adatátviteli protokollok közbeiktatása nélkül) hívhatók lineáris algebrai függvények (pl. mátrix faktorizáció), vagy Fourier-transzformáló rutinok.

A programcsomag segítségével lesz kialakítva az INDRA kozmológiai szimuláció adatbázisa (Falck et al. 2011), valamint erre épül majd egy nagy turbulencia szimuláció és a Via Lactea projekt adatbázisa is.

7.2. Keresések térbeli indexekkel ellátott adatokon

Bemutatunk néhány problémát, melyre a térbeli indexek hasznosíthatók. A különböző problémákra különböző dimenziójú és metrikájú, valamint különböző eloszlású adatpontok esetén más és más konkrét megvalósítás optimális. A későbbiekben röviden kitérünk azokra az indexelési módokra, melyekkel csillagászati adatokkal való munkánk során kapcsolatba kerültünk.

N legközelebbi szomszéd megtalálása Ebben az esetben a cél az, hogy egy többdimenziós ponteloszlásból megtaláljuk azokat a pontokat, melyek egy mintaponthoz egy adott metrika szerint a legközelebb esnek (Liu et al. 2003; 2006, Csabai et al. 2007, Dobos et al. 2007). Ez a módszer hasznos lehet galaxishalmazok azonosításakor, valamint olyan paraméterbecslési módszerek esetében, amikor egy mérési pont nem ismert paraméterét a környezetében levő többi pont ismert paraméterei alapján igyekszünk megbecsülni. Ez utóbbira jó példa a fotometrikus vöröseltolódás-becslés esete, amikor egy adott galaxis vöröseltolódását a színtérben hozzá közel eső galaxisok ismert vöröseltolódásából igyekszünk megbecsülni (Abazajian et al. 2009).

Párkorrelációs függvények meghatározása A párkorrelációs függvény jól ismert kozmológiai jelentőséggel bír, de meghatározása nagy adatmennyiség esetén az adatok speciális adatstruktúrába szervezését igényli. Különböző keresési fák segítségével az eljárás nagy mértékben gyorsítható, és többpont-korrelációs esetekre általánosítható (Szalay et al. 2002, Csabai et al. 2007).

Kilógó pontok megtalálása Érdekes kérdés, hogy milyen módon lehet egy nagy adathalmazban valamilyen szempontrendszer szerint különlegesnek számító adatokat gyorsan megtalálni (Chaudhary et al. 2002). Csillagászati adathalmazok esetében ez különleges fizikai tulajdonsággal bíró objektumokat jelent, melyek önmagukban is nagyon érdekesek lehetnek. Természetesen valódi mérési adatok esetén a legtöbb „kilógó pont” mérési hiba, ami a problémát tovább nehezíti.

Keresés analitikus feltételek alapján Csillagászati adatok analízise során gyakran előfordul, hogy az objektumokat a paramétereikre kiszabott feltételek alapján osztályozzuk. Ezek a feltételek általában egy lineáris egyenlőtlenség-rendszert jelentenek valamilyen sokdimenziós paraméterterben. Nagy adatbázisok esetében tehát szükség lehet az analitikus feltételek alapján történő keresések felgyorsítására. A különböző indexelési megoldások erre is segítséget nyújthatnak (Csabai et al. 2007).

Interaktív, adaptív vizualizáció Az interaktív vizualizációs programok megvalósításához az adatokat szintén a célnak megfelelő struktúrába kell szervezni. Az interaktivitás megköveteli a gyors válaszidőt. Ehhez adaptív megjelenítési módszerek kellene, vagyis az adatokat „távolról” nézve elegendő azok sűrűségét érzékeltetni, míg a részletekre „ránagyítva” már az egyes adatpontoknak, a finom struktúráknak is meg kell jelenniük. Az ilyen jellegű vizualizációs megoldások mind hierarchikus adatstruktúrára épülnek (Csabai et al. 2007, Szalay et al. 2008).

7.3. Többdimenziós adatok lokalitása

A továbbiak megértéséhez ismét egy kisebb számítástechnikai kitérőt kell tennünk. A számítógép-architektúrák közös jellemzője, hogy a memória, háttértár stb. címzésére egydimenziós címtereket használnak. Ez első olvasatban nem jelenthetne problémát, hiszen az egydimenziós címteret az alpműveletek segítségével egyszerűen többdimenzióssá lehet transzformálni. Egy ilyen transzformáció azonban az eredeti többdimenziós logikai térben egymáshoz közel levő adatpontokat az egydimenziós címtérben szétszórja, azaz elvesztjük az *adatok lokalitását*. A probléma akkor jelentkezik, amikor olyan sok adatunk van, hogy azokat egyszerre nem tudjuk a memóriába olvasni, viszont minden időpontban egymáshoz közeli adatpontokra van szükségünk. Ez a probléma két fontos területen okoz kellemetlenséget: a lemezes háttértárak és a processzorok gyorsítótárai esetében. Mi itt csak az első esettel fogunk foglalkozni.

A lemezes háttértárak egyik legfontosabb teljesítményt befolyásoló tulajdonsága, hogy az adatok szekvenciális olvasása ugyan viszonylag gyors (60-120 MB/s), de a véletlenszerű adatelérés nagyságrendekkel lassabb, ami annak köszönhető, hogy véletlenszerű olvasáskor az olvasófejeket folyton pozicionálni kell. Egy boltban olcsón megvásárolható

1 TB kapacitású lemez szekvenciális beolvasása kb. 4,5 órát vesz igénybe, míg véletlenszerű olvasás esetén ez akár 15-150 napig is eltarthat (Gray 2007)! Kézenfekvő tehát, hogy az algoritmusainkat úgy tervezzük meg, hogy azok a lemezek rossz tulajdonságaihoz igazodjanak. Sokdimenziós adatok kezelése esetén ez azt jelenti, hogy ügyelnünk kell az adatok lokalitásának megtartására: a logikai paraméterterben egymáshoz közel levő adatpontokat a lemezen is egymáshoz közel kell tárolnunk, hogy azokat egy menetben, szekvenciálisan haladva olvashassuk a memóriába.

7.4. Paraméterterek cellázása és indexelése

A lokalitás megőrzésének legegyszerűbb módja az, hogy ha a többdimenziós tereket valamilyen módon cellákra osztjuk, és a cellákba eső pontokat egymást követően tároljuk. Relációs adatbázis-kezelők esetében a gyakorlatban ez a következőképp valósítható meg: Először a tér minden cellájához egy egyedi azonosítót rendelünk, ami legtöbbször egy egész szám. Ez az azonosító voltaképpen egy hash, amit az adatpontok koordinátájából számolunk. Célszerű az azonosító kiszámítását valami nagyon egyszerű függvénnyel definiálni, hogy ezzel is csökkentjük a számításigényt. Második lépésben az adatpontokat tároló táblához hozzáadunk egy új oszlopot, mely a cella azonosítóját tárolja. Nyilvánvaló, hogy sok adatpontnak lesz ugyanaz a cellaazonosítója. Harmadik lépésben pedig úgy módosítjuk a tábla elsődleges kulcsát (ld. 6.1. pont), hogy az első helyen a cellaazonosítók szerint legyen klaszterezve. Ezzel pontosan a kitűzött célt érjük el: az egy cellába eső adatpontok egymáshoz közel, szekvenciálisan lesznek tárolva a lemezen (Csabai et al. 2007, Dobos et al. 2007).

7.4.1. Reguláris cellázás és a cellák lokalitása

A legegyszerűbb cellázási eljárás, ha a többdimenziós teret szabályos ráccsal osztjuk részekre. Ehhez először meg kell határozni a rács pontos kiterjedését (bár olyan megvalósítás is elképzelhető, ahol a rácsolás az egész térre kiterjed). Ha bizonyos dimenziókban az adatpontok eloszlása jellegzetesen más léptékű, mint a többi dimenzió irányába, akkor a rácsolást érdemes ennek megfelelően megválasztani. Az egyes irányokban más léptéket, esetleg logaritmikus léptéket is használhatunk. A reguláris rácsolás nagy előnye, hogy a cellák csúcseinak koordinátái egy képlet alapján számolhatóak, így nincsen szükség a rácpontok külön tárolására.

A szabályos ráccsal történő megvalósítás lehetőséget biztosít hierarchikus cellázásra is. Ekkor általában egy cellából indulunk ki, amit minden dimenzió mentén ketté vágunk, és ezt a szükséges felbontásig folytatjuk. Könnyű belátni, hogy ez a cellázási módszer legfeljebb három dimenzióban kivitelezhető a cellaszám igen gyors szaporodása miatt. Hierarchikus cellázás esetén a cellák azonosítóit úgy kell megválasztani, hogy azok bitmin-

tázata kövesse a cellák hierarchiáját (Csabai et al. 2007, Szalay et al. 2007, Lemson et al. 2011).

Az eddig leírtak elegendőek ugyan ahhoz, hogy az adatpontok megtartsák lokalitásukat, de a cellák maguk ettől még szétszórva helyezkednek el a háttértáron. Természetesen ez akkor bír jelentőséggel, ha a futtatott algoritmusok egyszerre több, egymás melletti cellát is igénybe vesznek. Ez a helyzet például legközelebbi szomszédok keresése esetében akkor, ha a mintapont valamelyik cella szélén helyezkedik el.

A cellák lokalitása biztosítható úgy, ha egy olyan celláról cellára haladó görbe mentén osztjuk ki a cellák azonosítóit, mely görbe lehetőleg minden cellát csak egyszer érint. A különböző bejárás módszerek különböző algoritmusok esetén optimálisak. A két legelterjedtebb bejárás mód az ún. Z-sorrend (Morton 1966), illetve a Peano–Hilbert-bejárás. Lemson et al. (2011) ilyen algoritmusokat implementáltak kozmológiai szimulációk adatbázisban történő tárolásának céljával.

7.4.2. A gömb cellázása

Égi koordináták szerinti keresések esetén az alapprobléma ugyanaz, mint sík esetében: a kétdimenziós paraméterteret kell egy dimenzióra képezni. A gömbi topológiából adódó szingularitási problémák miatt a gömbfelület indexeléskor célszerű elkerülni a polárkoordinátázást, és ehelyett háromdimenziós egységvektorokkal dolgozni.

Két csillagászati körökben elterjedt indexelési módszert említünk. Az egyik a Hierarchikus Háromszögháló¹ (HTM) módszere, melynél egy gömbbe írt oktaéderből kiindulva, a háromszögek oldalainak felezőpontjait a gömb felszínére vetítve a nagy háromszögeket minden lépésben négy kisebb háromszögre osztjuk (Szalay et al. 2007). A másik módszer a HealPix nevet viseli, és a kozmikus háttérsugárzás méréseinek feldolgozására fejlesztették ki (Górski et al. 2005). Ez az HTM váltakozó területű háromszögei helyett azonos területű rombuszokat használ a gömbfelszín rácsozására.

A csillagászatban gyakori probléma, hogy egy adott égterületen belül megkeressük az összes számunkra érdekes objektumot. Az elsődleges szűrési feltétel, amit kezelünk kell, maga az égterületet leíró geometriai alakzat. Ez bonyolult églefedettségű műszerek esetében egészen komplex lehet. A keresési eljárás a HTM esetében a következőképpen alakul: Először a gömbfelszínt háromszögek hierarchikusan egymásba ágyazott rendszerével közelítjük, minden háromszöget egy egyedi szám azonosítóval ellátva. A számok úgy vannak kiosztva, hogy az egyre kevésbé szignifikáns bitek a hierarchiában alacsonyabb szinten levő, egyre kisebb háromszögeket jelentsék. Kereséskor meghatározzuk, hogy a keresési feltétel alakzata mely kis háromszögekkel fedhető le. A kisebb, az alakzat szélén levő háromszögeket általában megtartjuk, míg az alakzat közepén levő három-

¹HTM = Hierarchical Triangular Mesh

szögeket – ha az azokat hierarchikusan befoglaló nagyobb háromszög is az eredeti alakzat része – nagyobb háromszögekkel helyettesítjük. Eredményül egy olyan háromszög-alapú felbontást kapunk, mely ekvivalens egy, a háromszögek azonosítói szerint meghatározott egészszám-intervallumok halmazával. Egy-egy intervallum egy háromszögre – közvetve pedig az összes, a hierarchiában alatta levő háromszögre – vonatkozik. Mivel az adatbázis-szerverek egyik alapfunkciója az egész számok intervallumai szerinti gyors keresés, a probléma megoldásának többi részét már az adatbázis-kezelőre bízhatjuk.

7.4.3. Adaptív cellázás

Adaptív cellázás alatt azt értjük, amikor a cellázás követi az adatpontok eloszlását, és az egyes cellákban közel azonos pontsűrűséget eredményez. Az adaptív cellázási módszerek gyakori hátránya, hogy a cellák csúcsainak koordinátáit külön tárolni kell, mert a pontok eloszlásától való függésük miatt képlettel nem számolhatók.

Egy egyszerűen definiálható, de bonyolultan megvalósítható cellázás, ha meghatározuk az adatpontok egy részhalmazának Voronoi-celláit. A módszer csak alacsony dimenziókban hatékony, viszont a cellák mérete kiválóan követi az adatpontok sűrűségét. A Voronoi-cellázás legközelebbi szomszédok megtalálására ideális.

Egy másik, igen elterjedt adaptív cellázási mód az ún. kD -fák építésének módszere. Ennél a módszernél egy hierarchikus cellázást kapunk. Úgy járunk el, hogy meghatározzuk a legkisebb adatpontokat befoglaló k dimenziós téglatestet, melynek lapjai párhuzamosak a koordinátákkal. Ezután megkeressük a pontok első koordináta szerinti mediánját, és a téglatestet ketté vágjuk. Következő lépésben a két kisebb téglatesttel tesszük ugyanezt, de ekkor a második koordináta szerint végezzük a vágást. Az eljárás tetszőleges ideig folytatható. Az eredményül kapott hierarchikus cellázás kiválóan használható sűrűségbecslésre, korrelációs függvények kiszámítására és N legközelebbi szomszéd keresésére.

A korábban ismertetett programcsomagban mind a kD -fás cellázást, mind a Voronoi-cellázást megvalósítottuk.

8. fejezet

Az Internet felhasználása tudományos célú adatok és szolgáltatások közzétételére

Bár az internet infrastruktúrája eredetileg katonai célú fejlesztések eredménye, az erre épülő World Wide Webről köztudott, hogy a CERN-ben folyó kutatások „mellékterméke”. Az alábbiakban röviden áttekintjük, hogy az elképzelés mennyiben változott azóta, hogy Tim Barnes-Lee megalkotta az első böngészőt abból a célból, hogy tudományos feljegyzéseit más kutatókkal is megoszthassa.

8.1. Adatbázisok a weben

A hálózatok sebessége mára ugyan elérte azt a szintet, ami lehetővé teszi TB nagyságrendű adatok mozgatását is, de a legnagyobb adatbázisok mérete ennél szűk három nagyságrenddel nagyobb. Bár a legnagyobb adatközpontok összekapcsolhatók nagy sávszélességű hálózattal, a rendszereket használó kutatókig ezek a jövőben sem fognak eljutni. Ebből az következik, hogy az Internet ugyan szolgálhat kisebb adatmennyiségek mozgatásának médiumául, de elsősorban arra kell koncentrálni, hogy rajta keresztül a nagy adattárházak adatfeldolgozó szolgáltatásait tegyék elérhetővé, és ne közvetlenül magukat a nyers adatokat. Így magukat az adatokat nem, csak a bemeneti paramétereket és a futási eredményeket kell a hálózaton keresztül mozgatni.

Az adattárházhoz ezért olyan interfészeket kell kialakítani, melyek egyrészt gyorsan megtanulható felhasználói felületet nyújtanak (egybefoglalva a felhasználói funkciókat és a dokumentációt), másrészt pedig olyan, kényelmesen programozható és platformfüggetlen felületekre is szükség van, melyeket a kliensprogramok hasznosíthatnak. Ez utóbbiak segítségével a kutatók az adattárházak szolgáltatásait úgy integrálhatják a saját kódjaikba, hogy a futtatáshoz ne kelljen az adatokat előzetesen letölteni, és maga a futtatás ne igényel-

jen humán beavatkozást.

8.2. Webszolgáltatások

Mély webnek az Interneten elérhető információk azon részét nevezik, melyek nem érhetőek el közvetlenül hivatkozások követésével, hanem űrlapok kitöltését igénylik. Ezek az információk a hagyományos, szöveg alapú indexelő szolgáltatások (Google, Bing stb.) számára gyakorlatilag nem hasznosíthatók, hiszen az emberek számára tervezett űrlapok kitöltését ezek a rendszerek nem tudják megoldani.

Ahhoz, hogy a világ nagy csillagászati adatbázisait összekapcsoljuk, olyan programok által is megérthető keresési felületeket kell kialakítanunk, melyek a fent vázolt problémát áthidalják. A programok számára készült Internetes felületeket a számítástechnikai irodalom *webszolgáltatásoknak* nevezi, és több általános célú technológia is rendelkezésre áll¹. Egy webszolgáltatásra úgy érdemes gondolni, mint egy olyan programkönyvtárra, aminek a függvényei meghívásuk után a távoli rendszeren futnak le, de a függvényhívás eredményeit lokálisan kapjuk meg.

A dolgozat szerzője több, a Virtuális Observatóriumhoz kapcsolódó webszolgáltatás megalkotásában is közreműködött: Spektrum Szolgáltatás (Dobos et al. 2004), Lányom Szolgáltatás (Budavári et al. 2007), ezekről később ejtünk szót.

8.3. Metainformációk, ontológiák, szemantikus rendszerek és az adatok származási története

A tudományos célú adattárházak esetében nagyon fontos a tárolt, szolgáltatott adatok megbízhatósága. Az adatokat akkurátus metainformációkkal kell ellátni abból a célból, hogy azok lehetőleg az adathalmazhoz kapcsolódó dokumentáció tanulmányozása nélkül is értelmezhetőek legyenek. Erre azért van szükség, hogy az adatokat a közeljövőben ne csak emberek, de automatikus szoftverek is képesek legyenek értelmezni.

A metainformációk szükségességére talán a legegyszerűbb példa, hogy az adatbázisokban tárolt számadatok pontos fizikai egységét programok számára értelmezhető formátumban is elérhetővé kell tenni. Az adatbázisokban tárolt mennyiségek pontos jellemzésére nem elegendő azonban a fizikai mértékegység megadása. Gondoljunk csak a csillagászatban használt különböző módon (vákuumban, levegőben) mért hullámhossz értékekre, vagy a hullámhossz helyett a különböző szakterületeken használt más egységekre (vöröseltolódás, sebesség, frekvencia). Hogy a számok jelentését pontosan leírassuk, a tudományának megfelelő *ontológiák* kialakítására van szükség. Az ontológiák legtöbb esetben hierarchikus formában rendszerezik egy *univerzum* (esetünkben a csillagászat

¹SOAP XML webszolgáltatások, REST szolgáltatások

egész témakörének) a fogalmait, entitásait, attribútumait. Az ontológiában szereplő fogalmakból hierarchikus tartalomazonosítókat generálhatunk, melyekkel az adatbázisaink oszlopait címkézhetjük fel, így definiálva pontosan az egy oszlopban szereplő számok jelentését.

Az ontológiák előfeltételét jelentik minden *szemantikus* alapokon nyugvó adatkezelő rendszernek. Egy szemantikus információk alapján működő szoftver például automatikusan azonosíthatja két csillagászati katalógus tábláinak azon oszlopait, melyek fotometriai, illetve koordináta információkat tartalmaznak, kitalálhatja, hogy miként lehetséges a két katalógus keresztazonosítása, és a katalógusok összefésülésével egy több hullámhossztartományt lefedő katalógust hozhat létre.

A világban elszórtan, különböző intézmények rendszerein található adatokat, a számítóközpontok által végzett adatfeldolgozási szolgáltatások felületét az Internet segítségével könnyedén publikussá lehet tenni, de megfelelő metainformációk nélkül ezek az adatok, szolgáltatások mások számára nem, vagy csak igen nehezen használhatóak. Ugyancsak fontos, hogy a közzétett adatbázisokat, szolgáltatásokat a potenciális felhasználók könnyen megtalálhassák. Erről különböző *regisztrációs szolgáltatásoknak* kell gondoskodniuk, melyek portálként funkcionálnak az adattárházak és a kutatók, valamint a kutatók által használt, automatikus adatfeldolgozó rendszerek között.

Azért, hogy az adathasználók biztosak lehessenek abban, hogy komoly kutatásaikat nem megbízhatatlan adatok alapján végzik, visszakövethetővé kell tenni az adatok eredetét, előfeldolgozásának módját, módszereit (Davidson & Freire 2008). Ez a metainformációk pontos megadásán túl az adatok, tudományos publikációk, kutatócsoportokban közkezen forgó egyéb információk (például levelezési listák) egységes keretbe foglalását igényli.

8.4. Az asztrofizikai Virtuális Obszervatórium

A Virtuális Obszervatórium (VO) a világ különböző pontjain üzemelő adattárházak olyan (jövőbeli) rendszere, mely adattárházak a korábbi pontokban ismertetett szempontoknak megfelelnek. A sok éve folyó munka eredményeként már ma is számos *VO-kompatibilis* adatbázis és webszolgáltatás létezik.

A Virtuális Obszervatórium kidolgozásának koordinálását a Nemzetközi Virtuális Obszervatórium Szövetség² végzi, és melynek az ELTÉ-n működő magyar kutatócsoport is tagja.

²IVOA = International Virtual Observatory Alliance – <http://www.ivoa.net>

9. fejezet

Fotometriai katalógusok a Virtuális Obszervatóriumban

9.1. Fotometriai katalógusok keresztazonosítása

Az égboltról készülő térképek különböző műszerekkel, különböző hullámhossztartományokban készülnek. Az objektumok ezért két égtérképen jelentősen különböző módon jelenhetnek meg. Egyfelől különböző lehet az egyes műszerek *seeingje*, tipikus asztrometriai hibája, de egyes objektumok bizonyos hullámhossztartományokban egyszerűen detektálhatatlanná válhatnak. A keresztazonosítási művelet során két vagy több redukált katalógussal dolgozunk, és célunk az, hogy a különböző katalógusokban szereplő, de egy égi objektumhoz tartozó bejegyzéseket egymáshoz rendeljük. Mindezt lehetőleg úgy, hogy minimalizáljuk a hamis egybeesések és kihagyott párosítások számát. A keresztazonosításhoz első lépésben mindössze az objektumok egyes műszerek által mért égi koordinátáit és a koordináták becsült hibáját használjuk, majd a többi paramétert (például a színindexeket) a további szűréseknél hasznosítjuk.

9.2. Katalógusok keresztazonosításának bayesi modellje

Budavári & Szalay (2008) egy olyan általános statisztikai modellt dolgozott ki objektumok keresztazonosítására, mely az objektumok koordinátáinak és azok hibájának figyelembe vételén túl lehetőséget biztosít további paraméterek figyelembe vételére is. Erre példának egymáshoz közel látszó csillagok esetét hozhatjuk fel, ahol a koordináták alapján esetleg nem lehet egyértelműen eldönteni, hogy melyik detektálások állnak párban, de a spektrális típus figyelembe vétele már egyértelművé teszi az azonosítást.

Az említett modell Bayes-statisztikán nyugszik, és konkrét valószínűségek helyett a Bayes-faktorral számol. Két katalógus esetében a Bayes-faktor a következőképpen írható

fel a koordináták segítségével:

$$B_{\text{pos}} = \frac{2}{\sigma_1^2 + \sigma_2^2} \exp \left\{ -\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)} \right\}, \quad (9.1)$$

ahol σ_i a katalógusok asztrometriai hibája, ψ pedig a két detektálás szeparációs szöge. A formula több katalógus esetére egyszerűen általánosítható. Belátható, hogy a Bayes-faktor iteratíván is meghatározható, azaz több katalógust úgy keresztazonosíthatunk, hogy először kettővel indulunk, meghatározzuk a Bayes-faktort, valamint a két koordináta súlyozott átlagából a legvalószínűbb közös koordinátát. A következő katalógust már az így kapott legvalószínűbb koordinátákkal párosítjuk, és a műveletet az utolsó katalógusig ismételtetjük. Ez számítástechnikai szempontból igen fontos eredmény.

A Bayes-faktorra tekinthetünk úgy, mint egy olyan mennyiségre, ami egy állítás bizonyítékának erősségét méri. Közeli detektálások esetében a fenti módon számolt Bayes-faktor értéke $B \gg 1$, ld. id. cikk, 1. ábra. Belátható, hogy ha alsó limitet szabunk a Bayes-faktornak, az egyenértékű lesz azzal, ha csak egymáshoz adott szögtávolságnál közelebb eső detektálásokat fogadunk el lehetséges párokként. A következő pontban ismertetendő algoritmust erre a tényre alapozzuk. Ugyan ez az eredmény naiv megközelítéssel is megkapható, kiemeljük, hogy csak a pontos bayesi modell teremt lehetőséget a modell további egyszerű bővítésére különböző *a priori* valószínűségek bevezetésével.

A priori valószínűséget rendelhetünk például objektumok színindexeihez, és ezzel finomíthatjuk a statisztikai modellt. Tegyük fel, hogy az A katalógusban egyetlen objektum került detektálásra, míg B katalógusban két objektum látszik nagyon közeli koordinátán. Az A katalógus ráadásul olyan hullámhossztartományban készült, melyben gyakori, hogy bizonyos objektumok egyáltalán nem jelennek meg. Mivel az A katalógus rossz seeingje miatt a koordináták alapján mindkét B-beli objektumot azonosíthatnánk az A-beli objektummal, figyelembe vehetjük a szélessávú fotometriai adatokat is. A fotometriai modellből kapott valószínűségből meghatározható a B_{phot} Bayes-faktor is. Mivel a színindexek függetlenek a objektumok koordinátáitól, a fotometriai adatokat is figyelembe vevő statisztikai modell Bayes-faktora egyszerű szorzat alakjában adódik:

$$B = B_{\text{pos}} B_{\text{phot}} \quad (9.2)$$

Az ismertetett eljárás részletes analízisét Heinis et al. (2009) végezték el. Megemlítjük még, hogy a modell könnyen kiegészíthető mozgó objektumok keresztazonosítására is (Kerekes et al. 2010).

9.3. Keresztazonosítási algoritmus relációs adatmodellekhez

Amikor az előző pontban igen kivonatosan ismertetett statisztikai modell segítségével relációs adatbázisokban tárolt adatokat kívánunk keresztazonosítani, két igen fontos dolgot kell figyelembe venni. Mindkettő a relációs adatbázisok alapvető működési elveiből következik, ezért a technikai részletekre egy picit alaposabban kitérünk.

Induljunk ki abból, hogy a keresztazonosítandó katalógusok adatai külön adattáblákban vannak, és a táblák mindegyik sorához tartozik egy egyedi azonosító, továbbá a sorok tartalmazzák a koordinátákat, a koordináták mérési hibáit és egyéb, esetleg szükséges adatokat. Két katalógus összepárosítására így a következő (erősen leegyszerűsített), de belátható időn belül végrehajthatatlan SQL lekérdezés szolgálna:

```
SELECT A.Ra, A.Dec, B.Ra, B.Dec FROM A, B  
WHERE Distance(A,B) <= D
```

Az adatbázis-kezelő ilyenkor képzí az A és B tábla rekordjainak Descartes-szorzatát, majd kiszámítja az összes koordinátapár távolságát – ami százmillió objektumot tartalmazó katalógusok esetén nagyon sokáig tart – majd csak azokat a pontpárokat adja vissza, melyek a D szög-távolságnál közelebb esnek egymáshoz. A probléma nem is a távolságszámoló rutin lassúságában keresendő, hanem abban, hogy a Descartes-szorzat megalkotásához az A tábla minden egyes sorához végig kellene olvasni a B tábla minden sorát. Ez még trükkös Descartes-szorzat képző algoritmusok esetében is nagyon sokáig tart. Megjegyezzük, hogy ebben az esetben az adatbázis-szerver nem tudja a hash-join műveletet alkalmazni, mert itt a két tábla közötti kapcsolatot nem egyedi azonosítók, hanem égi koordináták adják meg.

A megoldás tehát az, hogy eleve elkerüljük a teljes Descartes-szorzat elkészítését, és a koordináták alapján már eleve durva szűréseket alkalmazunk. Ebből a célból a Gray et al. (2006), valamint Budavári et al. (2008) által kidolgozott módszert használjuk. A módszer szerint a gömböt zónákra osztjuk, mely zónákat a deklináció (szélességi koordináta) értékei szerint jelölünk ki. Ha a zónák méretét megfelelően választjuk meg, azaz a $D \approx H$ esetet vesszük, ahol H a zónák magassága, akkor belátható, hogy egy adott zónába eső adatpontokat csak az ugyanabba a zónába, illetve a két szomszédos zónába eső pontokkal érdemes egyáltalán összehasonlítani, mert az egymástól távoli zónákban levő pontok esetében eleve kizárt, hogy megfeleljenek a szigorú, kis szeparációt megkövetelő keresési feltételnek.

Az eljárás a gyakorlatban úgy néz ki, hogy minden zónához hozzárendelünk egy egyedi azonosítót, mely a deklináció értékéből könnyen kiszámítható egész szám. A detektálások koordinátájával egyetemben ezt a zónaazonosítót is tároljuk az adattáblákban, és egy olyan indexet építünk a táblákra, mely a zónaazonosító alapján gyors keresést biztosít. A korábbi lekérdezést úgy módosítjuk, hogy a távolság szerinti vágást egy zónaazonosító

szerinti vágással kombináljuk:

```
SELECT A.Ra, A.Dec, B.Ra, B.Dec FROM A, B
WHERE A.ZoneID BETWEEN B.ZoneID - 1 AND B.ZoneID + 1
AND Distance(A,B) <= D
```

Ránézésre ez nem látszik, viszont a zóna azonosítókra kiszabott feltételt az adatbázisszerver lekérdezősoptimalizáló modulja már fel tudja használni, hiszen az egész számok intervallumaira történő vágás jóval könnyebben kezelhető, mint a gömbi távolságok szerinti. A lekérdezés végrehajtásakor tehát a zónák hármásával fognak betöltődni; mindig három egymás fölött levő zóna. Ezután a szerver csak a zónák alapján előre filterezett, potenciálisan közel levő pontok esetében számítja ki az egzakt távolságot.

9.4. SkyQuery: Egy robusztus, elosztott keresztazonosító rendszer implementálása relációs adatbázis-kezelővel *

A 9.2. és 9.3. pontokban vázlatosan bemutatott keresztazonosító algoritmust a gyakorlatban is implementáltuk egy elosztott adatbázisszerver-klaszter segítségével. Az operációs rendszeren és az adatbázis-szervereken kívül a megvalósításhoz szükséges összes többi programot a dolgozat szerzője készítette Budavári Tamás és Szalay Sándor témavezetése mellett. A rendszer prototípusként működik a Johns Hopkins Egyetemen, egyelőre nyilvánosan nem elérhető.

9.4.1. A Graywulf-klaszter és szoftver

A klaszter – Jim Gray tiszteletére, és a Beowulf klaszterek nevére utalva – Graywulf névre hallgat (Szalay et al. 2009, Simmhan et al. 2009). A rendelkezésünkre álló rendszer öt darab, egyenként nyolc processzorral, 16 GB memóriával és egyenként valamivel több, mint 10 TB háttértárral rendelkező számítógépből áll. Valamennyi gép Windows Server 2008 operációs rendszert és SQL Server 2008 R2 adatbázis-kezelőt futtat.

A klaszter menedzselésére speciális programot készítettünk, mely részletesen nyilvántartja a rendszer teljes konfigurációját (szerverek, RAID egységek, adatbázisok). A pontos konfiguráció nyilvántartására a számítási feladatok ütemezéséhez van szükség. A teljes klaszter egy webböngészőben futtatható alkalmazás segítségével vezérelhető. Ehhez kapcsolódik egy ütemező modul, mely végrehajtási sorokba rendezett feladatok futtatását látja el. Az egyes feladatokat a Windows Workflow Foundation környezetben kell elkészíteni, a keresztazonosító algoritmust is ennek segítségével implementáltuk. Az ütemező egy központi gépen fut, egyetlen példányban. Ez a központi gép konkrét számolásokat nem végez, csak annyi a feladata, hogy az adatbázis-szerverek munkáját koordinálja.

A SkyQuery rendszerében elérhető csillagászati adatbázisok mindegyikét mind az öt szerverre tükröztük. Ez a konfiguráció a lemezterület pazarlásának tűnhet, de egyfelől a lemezterület ma már nem drága, másfelől az így szervezett adatokon a számítási feladatok nagyon könnyen particionálhatók, azaz párhuzamosan futtathatók valamennyi szerveren. További előnye még a konfigurációnak a nagy rendelkezésre állási valószínűség, hiszen bármelyik gép meghibásodásakor a többi automatikusan képes átvenni annak feladatát anélkül, hogy az adatokat mozgatni kellene. A nagy katalógusok adatain túl a SkyQuerybe regisztrált felhasználók is kapnak egy-egy adatbázist, mely saját adataik, illetve az általuk futtatott lekérdezések eredményeinek tárolására szolgál.

9.4.2. A keresztazonosítási probléma megfogalmazása

Az adatbázisszerverek SQL nyelven programozhatók, ám ez a nyelv nem elegendő valószínűségi alapon nyugvó, égi koordináták szerinti kereszt-azonosítási problémák megfogalmazására. Hogy a feladat mégis az SQL lekérdezések tömörségével, deklaratív módon legyen megfogalmazható, kidolgoztunk egy kiegészítést az SQL nyelvhez, mely speciálisan illeszkedik az igényeinkhez. A kiegészítés pontos szintaxisát itt nem áll módunkban ismertetni, de példaként álljon itt a következő lekérdezés, mely a Gaia és SDSS katalógusok egy adott égtérületre szűkített keresztazonosítási feladatát fogalmazza meg:

```
SELECT s.ObjID, s.Ra, s.Dec, g.Ra, g.Dec, x.Ra, x.Dec
FROM SDSS AS s, Gaia AS g
XMATCH BAYESIAN AS x
    MUST s ON POINT(s.Ra, s.Dec), 0.1
    MUST g ON POINT(g.Ra, g.Dec), 0.5
    HAVING LIMIT 10e5
WHERE s.RA BETWEEN -10 AND 10 AND s.Dec BETWEEN -5 AND 5
```

A minta lekérdezésben az XMATCH utasítás teljesen speciális, az SQL nyelvnek nem része. Az ezt követő, MUST kulcsszóval kezdődő sorok azokat a táblákat sorolják fel, melyekre a keresztazonosítást el kell végezni. A MUST arra utal, hogy a megadott katalógusoknak feltétlenül tartalmaznia kell az objektumról készült megfigyelést ahhoz, hogy az objektum bekerüljön a végső eredményhalmazba. Alternatívaként használható a MAY kulcsszó, mely megengedi, hogy az objektum akkor is szerepeljen a végeredményben, ha bizonyos katalógusok (többnyire olyanok, melyek magnitúdó limitje nem elegendően mély) nem tartalmaznak az objektumra vonatkozó megfigyelést. A harmadik hasonló szerepű kulcsszó a DROP, melyet akkor kell használni, ha kifejezetten csak olyan objektumokra vagyunk kíváncsiak, melyek az adott katalógusban nem szerepelnek. Ez utóbbi kulcsszó olyan objektumok keresésére lett bevezetve, amelyek például jól látszanak az optikai tartományban, de valamilyen okból (a galaxisban levő por miatt) ultrabolygóban nem figyelhetőek

meg. A HAVING LIMIT kulcsszavakat követő szám a Bayes-faktorra vonatkozó alsó korlát. A WHERE kulcsszó a szabványos SQL-ből származik, és az azt követő logikai kifejezés esetünkben a keresztazonosítási műveletet az $20^\circ \times 10^\circ$ -os területére korlátozza.

9.4.3. A keresztazonosítási feladat végrehajtása

Az előző pontban ismertetett, speciális kulcsszavakkal kiegészített SQL parancsot természetesen az adatbázisszerverek nem értik meg, ezért azt előzetesen fel kell dolgozni, majd annak alapján egy komplex végrehajtási tervet kell készíteni, a végrehajtási tervet szabványos SQL nyelvű lekérdezések sorozatára kell konvertálni, majd az ütemező modul segítségével végrehajtani.

A speciális SQL kiegészítések szintaktikáját egy ún. *formális nyelvtannal* írtuk le, melyből egy elemző-generátorral *parancselemzőt*¹ készítettünk. A parancselemző a lekérdezés alapján egy *elemzési fát* épít, melyben ekkor már könnyen elkülöníthetőek a komplex lekérdezés különböző elemei: kulcsszavak, tábla nevek, oszlop nevek, logikai feltételek stb. Ezt követi az ún. *névfeloldás* művelete, amikor is a táblaneveket és oszlopneveket azonosítjuk az adatbázisokbeli megfelelőikkel.

Miután a nevek feloldása megtörtént, készen állunk arra, hogy az érintett táblák adatairól előzetes statisztikát készítsünk. Erre azért van szükség, hogy a végrehajtandó feladatokat a rendelkezésünkre álló szervereken egyenletesen osszuk el. A statisztika készítésekor figyelembe vesszük a keresztazonosítási lekérdezés WHERE kulcsszava után szereplő feltételeket is, és meghatározzuk a feltételeknek megfelelő adatsorok számának rektaszenciói szerinti eloszlását. A hisztogram ismeretében a vizsgálandó égeterületet úgy vágjuk a szerverek számának megfelelő részre, hogy lehetőség szerint minden partícióba ugyanannyi objektum essen.

A partíciók meghatározása után nincsen más hátra, mint a keresztazonosításhoz szükséges, immáron szabványos SQL lekérdezések futtatása. Erre a 9.3. pontban ismertetett zónákon alapuló lekérdezéseket használjuk. A feladat végrehajtása katalógusonként több lépésből áll. Első lépésként megvizsgáljuk, hogy a lekérdezés hivatkozik-e olyan adatbázisokra, melyek nincsenek minden szerveren tükrözve. Amennyiben igen, úgy a hivatkozott táblákban szereplő adatokat a nagy adatbázisok mellett gyorsítótárazni kell. Ezután katalógusonként haladva, lépésenként végrehajtjuk a keresztazonosítási műveletet. Végezetül az eredményeket a felhasználó saját adatbázisába másoljuk, és az átmeneti adattárolásra használt táblákat töröljük.

A művelet sor a valóságban meglehetősen összetett, ezért is választottuk a workflow alapú, sokkal robusztusabb megoldást az egyszerű scriptelés helyet. A workflow-k futtatásához szintén egyedileg fejlesztett ütemező program támogatja a hibaüzenettel leálló

¹parser generator, parser

feladatok partíciónkénti újrafuttatását, az egyes feladatok átmeneti felfüggesztését, vagy teljes leállítását. Itt figyelembe kell venni, hogy egy futó feladat egyidejűleg sok szervert érint, párhuzamosan sok lekérdezést, átmeneti adatok számára készített táblát stb. hoz létre. A feladatok leállításakor természetesen minden futó lekérdezést meg kell szakítani, és az átmeneti táblákat el kell tüntetni. Ezek az egyszerűnek tűnő feltételek a rendszert mégis meglehetősen bonyolulttá teszik, ám többéves munka árán egy stabilan működő prototípus áll rendelkezésünkre, mely hamarosan elérhető lesz a kutatóközönség számára is.

9.5. Csillagászati katalógusok lábnyoma *

Csillagászati katalógusok keresztazonosításakor még egy fontos problémát kell kezelni. Ez elsősorban azt az esetet érinti, amikor arra vagyunk kíváncsiak, hogy egy ismert objektum azért nem látszik egy adott katalógusban, mert az a műszer detektálási limitjénél halványabb, vagy azért, mert az adott égitestet a műszer egyáltalán nem figyelte meg. Ahhoz, hogy ezt eldöntsük, pontosan kell ismerni a katalógusok által lefedett égitesteket.

A gömbi geometriák leírására Budavári et al. (2010) kidolgoztak egy programcsomagot. Ennek segítségével a gömb felszínén levő alakzatokat körívek határolta konvex idomok uniójaként lehet leírni. Az alakzatok leírási módja analitikus, tehát nem gömbi pixelek segítségével történik. A programcsomag támogatja unió, metszet, különbség stb. képzését, valamint az alakzatok pontos területének meghatározását.

A programcsomag segítségével elkészítettük a *Lábnym Szolgáltatást*², mely egy webböngészővel használható felületen, valamint egy webszolgáltatáson keresztül teszi elérhetővé számos égtérkép pontos lábnyomát (Budavári et al. 2007, Budavári et al. 2007). Lehetőség van arra is, hogy megkeressük, hogy egy adott égitestet mely projektek keretében figyeltek meg. Megszerkeszthetjük, és mások számára elérhetővé tehetjük saját megfigyeléseink lábnyomát is.

A web interfész és a vizualizációs eszköz a dolgozat szerzőjének munkája. A Lábnym Szolgáltatás a <http://voservices.net/footprint> címen érhető el.

²Footprint Service

10. fejezet

Csillagászati spektroszkópia a Virtuális Obszervatóriumban *

10.1. A spektroszkópiai adatok sajátosságai

Fotometriai adatok publikálásáról és feldolgozásáról már esett szó a 9. fejezetben. Míg a fotometriai adatok esetében a legnagyobb kihívást a különböző katalógusok automatikus és gyors keresztazonosítása jelenti, a spektroszkópiai adatok egészen más okok miatt érdemelnek külön figyelmet.

Spektroszkópiai adatok esetében két fontos tényező van, mellyel számolni kell. Egyrészt a spektrumok sokdimenziósak, ami eleve bonyolultabb tárolási és adattovábbítási struktúrákat igényel. Másrészt a spektrumok esetében számos olyan előfeldolgozási lépés van, melyet a konkrét kutatási cél igényeihez igazítva kell elvégezni.

A fotometriai adatok néhány színszűrő segítségével készülnek, így egy komoly galaxis katalógus is legfeljebb néhány száz számot tartalmaz galaxisonként (pl. SkyServer), de a magnitúdóadatokat paraméterteret csak néhány dimenziós. A spektrumok tere ezzel szemben sokezer dimenziós, ami, például hasonló spektrumok keresése esetén, fejlett keresési technikák alkalmazását teszi szükségessé, ld. 10.3. pont.

10.2. Spectrum Services: Egy spektroszkópiai adattárház a Virtuális Obszervatóriumhoz

A Spectrum Szolgáltatás egy webes felhasználói felületen, és párhuzamosan egy webszolgáltatáson keresztül program-kliensek által is elérhető adattárház, mely elsősorban az SDSS spektrumaira épül (Dobos et al. 2004; 2006, Dobos & Budavari 2007). Az adatok kihasználásán túl azonban számos adatfeldolgozási lehetőséget is biztosít, melyeket az alábbiakban röviden bemutatunk.

A szoftver folyamatos fejlesztés alatt áll, mely egyrészt a felhasználói felületet, másrészt az adatfeldolgozó könyvtárakat érinti. Az évek során az adatfeldolgozó függvényeket úgy alakítottuk ki, hogy azok a galaxisspektroszkópia igényeinek minél jobban megfeleljenek. A dolgozat első két részében bemutatott spektrumatlász és galaxisstatisztikai tanulmány, valamint még néhány más munka adatkezelési, adatfeldolgozási feladatait is ezzel a programcsomaggal végeztük el (Yip et al. 2009; 2010, Dobos & Csabai 2011, Dobos et al. 2011a).

Ugyan számos, spektrumok feldolgozására alkalmas programcsomag forog közkézen, mi mégis a saját kód kifejlesztése mellett döntöttünk, ugyanis az adattárházi alkalmazhatóság miatt számos olyan szempontot kellett figyelembe vennünk, melyek miatt más megoldások nem jöhettek szóba. Ide tartozik például, hogy a mi megvalósításunk lehetővé teszi igen nagy mennyiségű spektrum párhuzamosított feldolgozását, valamint adatbáziszerverek használatát a spektrumok tárolására, rendszerezésére.

A Spektrum Szolgáltatás implementálja az IVOA által elfogadott spektrumok reprezentálására, keresésére és elektronikus továbbítására készült szabványokat (Tody et al. 2011, McDowell et al. 2011). A spektrum-adatmodell lehetővé teszi igen komplex, többdimenziós, a csillagászat sok területéről (optikai, röntgen, rádió) származó spektrumok reprezentálását, és széleskörű metainformációkkal való ellátását.

10.3. Gyors keresés spektroszkópiai adatbázisokban

A Spektrum Szolgáltatás sokféle keresési lehetőséget biztosít a spektrumok között. Az egyes objektumok azonosítója és koordinátája szerinti keresésen túl kereshetünk a spektrumok számos paramétere alapján is, mint például a lefedett hullámhossztartomány, vöröseltolódás, spektroszkópiai típus, illesztett modell paraméterek stb. alapján.

Az előbbieken kívül összetett koordináta alapú kereséseket is végezhetünk. Ez utóbbi a Spektrum Szolgáltatás és a Lábnym Szolgáltatás (ld. 9.5. fejezet) összekapcsolásán alapul. A Lábnym Szolgáltatás segítségével összeállíthatjuk azt a gömbi alakzatot (például saját megfigyeléseink lábnymát), melyen belül szükségünk van az összes spektroszkópiai megfigyelésre. A keresési eljárás megvalósítása a 7.4.2. fejezetben ismertetett gömbi geometriai könyvtáron és a hozzá kapcsolódó HTM gömbi indexelési algoritmuson nyugszik.

Kiemeljük a hasonló spektrumok keresésére kidolgozott eljárásunkat is, hiszen ez számos, a dolgozat előző fejezeteiben külön-külön már említett eszközt és algoritmust alkalmaz. Első lépésként vesszük egy katalógus összes galaxisspektromát, a spektrumokat valamilyen módszer szerint normáljuk, majd meghatározzuk a főkomponens-spektrumokat (ld. 3.1.2. fejezet). Ezután az egyes spektrumokat kifejtjük az első néhány főkomponensből alkotott bázison, és a kifejtési együtthatókat egy adatbázisban tároljuk. A kifejtési együtthatók adattábláját olyan speciális térbeli indexszel (ld. 7. fejezet) látjuk el, mely lehetővé

teszi egy adott ponthoz legközelebb eső k szomszéd (a hasonló spektrumok) gyors megtalálását. Amikor egy felhasználó feltölt egy saját spektrumot, melyhez hasonló spektrumokat kíván keresni az adattár többi spektruma között, az adott mintaspektrumot is kifejtjük ugyanazon a főkomponens-bázison, mint amelyen a többi spektrumot már kifejtettük. A kapott kifejtési együtthatók alapján futtatunk egy térbeli indexszel gyorsított keresést a többi spektrum kifejtési együtthatói között, hogy meghatározzuk a főkomponens-együtthatók terében a k legközelebbi szomszédot. Bár a főkomponens-együtthatók terének euklideszi metrikája nem bír konkrét fizikai jelentéssel, ezek a spektrumok mégis nagyon hasonlóak lesznek a mintaként beadott spektrumhoz.

10.4. Spektrumok előfeldolgozása és egyéb alapfunkciók

Galaxisspektrumok analízisekor számos előfeldolgozási műveletet kell elvégezni. A Spektrum Szolgáltatás igyekszik minden szóba jöhető műveletet támogatni, melyek közül felsorolunk néhányat: spektrumok újra-mintavételezése tetszőleges rácson, galaktikus extinkció korrekciója, nyugalmi és megfigyelési rendszerek közötti konverzió, többféle normálási algoritmus, hullámhosszak konvertálása levegő- és vákuumbeli értékek között, valamint konvolúció sebességszórás kernellel. Ez utóbbi funkció modellspektrumok előfeldolgozásakor lényeges.

Az alapfunkciók közül megemlíthetjük még a szintetikus magnitúdók kiszámításának lehetőségét is. Az elterjedt fotometriai szűrők átviteli görbéit összegyűjtve létre hoztunk egy adatbázist, mely a <http://voservices.net/filter/> címen érhető el. A Spektrum Szolgáltatás segítségével a spektrumok valamennyi, az adatbázisban szereplő átviteli görbe szerinti szintetikus magnitúdója meghatározható, feltéve, hogy a spektrumok hullámhossztartománya teljes egészében átfed a fotometriai szűrő tartományával.

10.5. Spektrumok analízise

A Spektrum Szolgáltatás legújabb verziója gyakorlatilag az összes spektrumredukációs művelet elvégzését támogatja, melyet a 2. fejezetben bemutattunk. Fontos kiemelni, hogy mindezek az analitikai eszközök közvetlenül az adattárház felhasználó felületéről érhetőek el, így nincsen szükség az adatok előzetes letöltésére, és a programcsomag installálására. A számításokat a távoli szerverek végzik, a felhasználó kutatóhoz csak a végeredmények jutnak el.

A spektrumok kontinuumra Bruzual–Charlot-modellek alapján (vagy a felhasználó által a rendszerbe feltöltött tetszőleges modellek alapján) illeszthető. A kontinuum levonása után az emissziós spektrumvonalak Gauss-görbékkel illeszthetők, illetve a korábban meghatározott elméleti kontinuum alapján meghatározható ekvivalens szélességük. Ugyan-

csak kiszámítható valamennyi elterjedt abszorpciós index.

A programcsomagba a 3.3. fejezetben ismertetett robusztus, iteratív PCA algoritmus is integrálásra került. Lehetőség van továbbá kompozit spektrumok kiszámítására is. Ez utóbbi két funkció valóban megmutatja az adattárházak nagy előnyét: mindkét esetben kiindulhatunk sok tízezer spektrumból, a végeredmény pedig néhány spektrum lesz. A futtatás végén csak ezt a néhány spektrumot kell letölteni, és nincsen szükség a teljes kiindulási adathalmaz hálózaton való átmozgatására.

11. fejezet

Összefoglalás

A Virtuális Observatóriumokról szóló III. részben áttekintettük az adatintenzív tudományos kutatás közelmúltját, és jelenét. Felvázoltuk azt a jelenlegi és jövőbeli eszköztárat (mind hardver, mind szoftver oldalról), melyre szükség van ahhoz, hogy az egyre nagyobb mérések, kísérletek, szimulációk által előállított, a petabájtot is meghaladó adatmennyiség feldolgozható legyen. A ma rendelkezésünkre álló eszközök közül külön kitértünk a relációs adatbázisokra, és arra, hogy ezekhez milyen speciális modulokat kell fejleszteni abból a célból, hogy azok tudományos célra is alkalmazhatóak legyenek. Külön kitértünk a magasabb dimenziós adatok kezelésére, valamint tömbök adatbázisban való tárolására, és bemutattuk kutatócsoportunk ezen területekhez kapcsolódó munkáját. Ezekben a projektekben a dolgozat szerzője jelentős hozzájárulással bír.

A csillagászati adattárházakra áttérve ismertettük az adatok hosszútávú megőrzésének néhány problémáját, majd elemeztük, hogy az adatbázisokban tárolt információk milyen módon tehetők automatikus szoftverek által is felderíthetővé, értelmezhetővé.

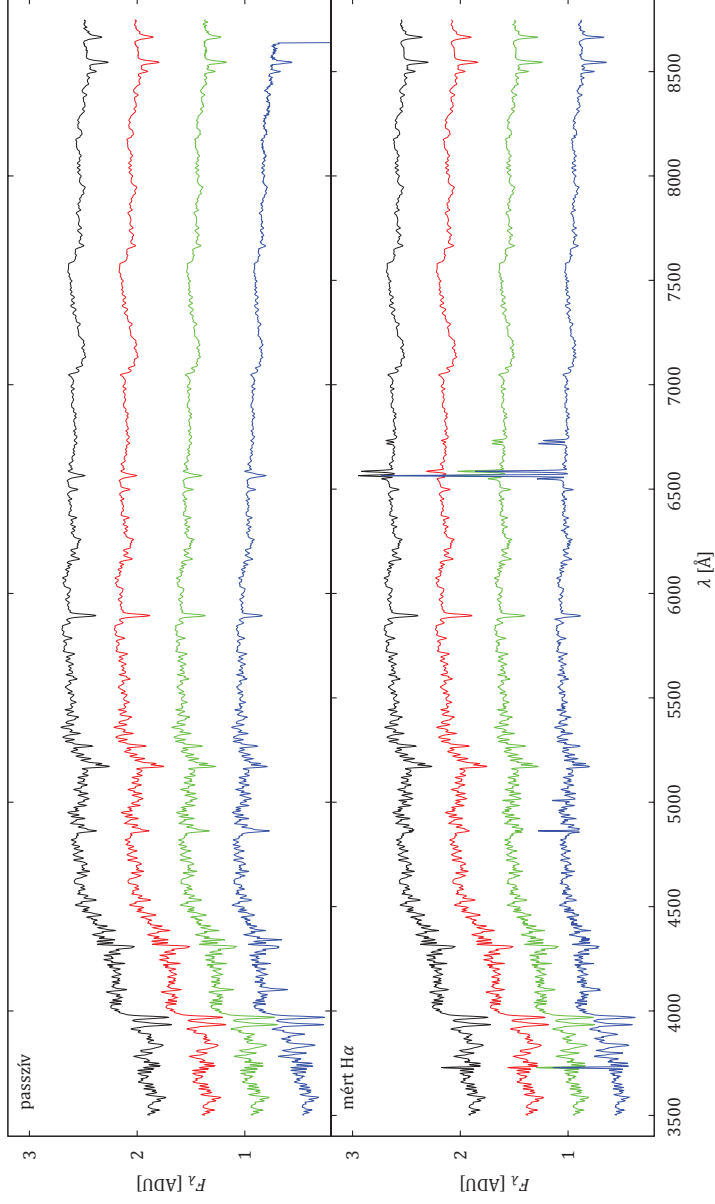
Bemutattunk három olyan Virtuális Observatóriumhoz készült csillagászati szoftvert, mely fejlesztéséhez a dolgozat szerzője jelentős mértékben hozzájárult: A Spektrum Szolgáltatás az SDSS spektrumainak adatbázisba szervezését és egyedi igények szerinti előfeldolgozását végzi el; a Lábnyom Szolgáltatás égtérképek églefedettségi információinak rendszerezésére készült; a még jelenleg is fejlesztés alatt álló legújabb verziójú SkyQuery pedig egy olyan fotometriai katalógusok keresztazonosítására készült eszköz, mely adatbázisszerver-klaszter segítségével néhány perc alatt képes elvégezni olyan számítástechnikai feladatokat, melyekre néhány éve még heteket kellett fordítani.

Irodalomjegyzék

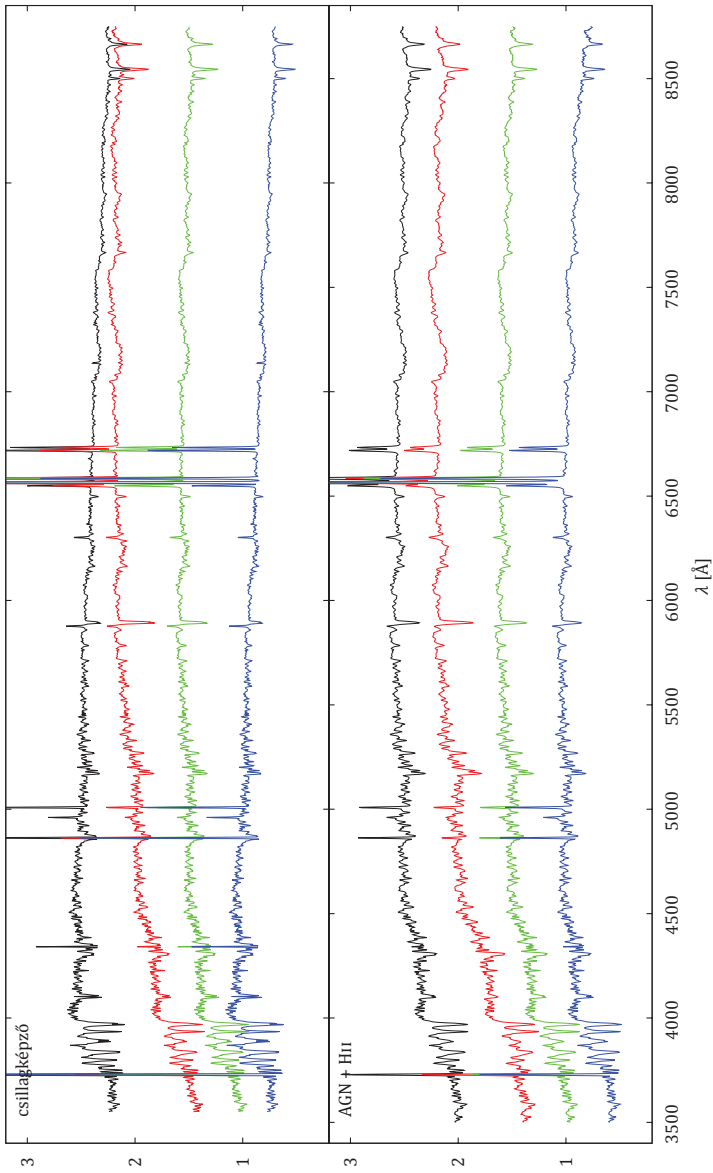
- Abazajian, K. N., et al. 2009, *ApJS*, 182, 543
- Baumann, P. 1999, in *Next Generation Information Technologies and Systems*, 4th International Workshop, NGITS'99, Zikhron-Yaakov, Israel, July 5-7, 1999 Proceedings, 76-93
- Bell, G., Gray, J., & Szalay, A. 2007, eprint arXiv:cs/0701165
- Budavári, T., Dobos, L., Szalay, A. S., Greene, G., Gray, J., & Rots, A. H. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 376, *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, 559-+
- Budavári, T., & Szalay, A. S. 2008, *ApJ*, 679, 301
- Budavári, T., Szalay, A. S., & Fekete, G. 2010, *PASP*, 122, 1375
- Budavári, T., Szalay, A. S., & Nieto-Santisteban, M. 2008, in *Astronomical Society of the Pacific Conference Series*, Vol. 394, *Astronomical Data Analysis Software and Systems XVII*, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis, 165-+
- Budavári, T., Szalay Gyoergy Fekete, A. S., Dobos, L., Greene, G., Gray, J., & Rots, A. H. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 382, *Astronomical Society of the Pacific Conference Series*, ed. M. J. Graham, M. J. Fitzpatrick, & T. A. McGlynn, 75-+
- Chamberlin, D. D., & Boyce, R. F. 1974, in *Proceedings of the 1974 ACM SIGFIDET Workshop on Data Description, Access and Control (ACM)*, 249-264
- Chaudhary, A., Szalay, A. S., & Moore, A. W. 2002, in *DMKD*
- Codd, E. F. 1970, *Communications of the ACM*, 13, 377
- Comer, D. 1979, *ACM Comput. Surv.*, 11, 121
- Copeland, G. P., & Khoshafian, S. N. 1985, in *Proceedings of the 1985 ACM SIGMOD international conference on Management of data, SIGMOD '85 (New York, NY, USA: ACM)*, 268-279
- Csabai, I., Dobos, L., Trencsényi, M., Herczegh, G., Józsa, P., Purger, N., Budavári, T., & Szalay, A. S. 2007, *Astronomische Nachrichten*, 328, 852
- Davidson, S. B., & Freire, J. 2008, in *SIGMOD Conference*, ed. J. T.-L. Wang (ACM), 1345-1350
- Dean, J., & Ghemawat, S. 2008, *Commun. ACM*, 51, 107
- Dobos, L., & Budavari, T. 2007, *Astronomical Society of the Pacific Conference Series*, Vol. 382, Chapter 17: *Web-based Tools - Spectrum and Filter Services for the VO*, ed. M. J. Graham, M. J. Fitzpatrick, & T. A. McGlynn, 147-+
- Dobos, L., Budavári, T., Csabai, I., & Szalay, A. S. 2004, in *Astronomical Society of the Pacific Conference Series*, Vol. 314, *Astronomical Data Analysis Software and Systems (ADASS) XIII*, ed. F. Ochsenbein, M. G. Allen, & D. Egret, 185-+
- Dobos, L., Budavári, T., Csabai, I., & Szalay, A. S. 2006, in *Astronomical Society of the Pacific Conference Series*, Vol. 351, *Astronomical Data Analysis Software and Systems XV*, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 471-+
- Dobos, L., & Csabai, I. 2011, *MNRAS*, 414, 1862
- Dobos, L., Csabai, I., Trencsényi, M., Herczegh, G., Józsa, P., & Purger, N. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 376, *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, 629-+
- Dobos, L., Csabai, I., Yip, C.-W., Budavári, T., Wild, V., & Szalay, A. S. 2011a, *MNRAS* submitted
- Dobos, L., et al. 2011b, in *EDBT Workshops*
- Falck, B., et al. 2011, in *American Astronomical Society Meeting Abstracts #218*, 131.04-+
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, *ApJ*, 622, 759
- Gray, J. 2007, in *Electronic Proceedings of the Conference on Innovative Database Research*.
- Gray, J., Nieto-Santisteban, M. A., & Szalay, A. S. 2006, *Microsoft Research Technical Report MSR-TR-2006-52*
- Gray, J., & Shenoy, P. J. 2000, in *ICDE*, 3-12
- Gray, J., Szalay, A., Thakar, A., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C., & vandenBerg, J. 2001, *Microsoft Research Technical Report MSR-TR-2001-104*, 11
- Gray, J., & Szalay, A. S. 2006, *Nature*, 440, 413
- Grossman, R. L., Gu, Y., Mambretti, J., Sabala, M., Szalay, A. S., & White, K. P. 2010, in *HPDC*, ed. S. Hariri & K. Keahey (ACM), 377-384
- Hector-Molina, G., & Ullmann, J. D. 2008, *Adatbizisrendszerek megvalósítása (Panem)*
- Heinis, S., Budavári, T., & Szalay, A. S. 2009, *ApJ*, 705, 739
- Kerekes, G., Budavári, T., Csabai, I., Connolly, A. J., & Szalay, A. S. 2010, *ApJ*, 719, 59
- Koomey, J. 2011, *Analytics Press*, Oakland, CA
- Koomey, J. G. 2008, *Environmental Research Letters*, 3, 034008
- Lemson, G., Budavari, T., & Szalay, A. S. 2011, in *Lecture Notes in Computer Science*, Vol. 6809, *SSDBM*, ed. J. B. Cushing, J. C. French, & S. Bowers (Springer), 509-526
- Liu, T., Moore, A. W., & Gray, A. G. 2003, in *NIPS*, ed. S. Thrun, L. K. Saul, & B. Schölkopf (MIT Press)
- Liu, T., Moore, A. W., & Gray, A. G. 2006, *Journal of Machine Learning Research*, 7, 1135
- Maier, D., & Vance, B. 1993, in *PODS*, 1-16

- Matray, P., Csabai, I., Haga, P., Steger, J., Dobos, L., & Vattay, G. 2007, in Proceedings of the 3rd Annual ACM Workshop on Mining Network Data, MineNet 2007, San Diego, California, USA, June 12, 2007, 23–28
- McDowell, J., et al. 2011, IVOA web site
- Morton, G. M. 1966, IBM Technical Report
- Samet, H. 2006, Foundations of Multidimensional and Metric Data Structures (Morgan Kaufmann)
- Simmhan, Y., et al. 2009, in HICSS, 1–10
- Stonebraker, M., Becla, J., DeWitt, D. J., Lim, K.-T., Maier, D., Ratzesberger, O., & Zdonik, S. B. 2009, in CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 4-7, 2009, Online Proceedings
- Stonebraker, M., et al. 2005, in Proceedings of the 31st international conference on Very large data bases, VLDB '05 (VLDB Endowment), 553–564
- Stoughton, C., et al. 2002, AJ, 123, 485
- Szalay, A., Budavari, T., Connolly, A., Gray, J., Matsubara, T., Pope, A., & Szapudi, I. 2002, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4847, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J.-L. Starck & F. D. Murtagh, 1–12
- Szalay, A., & Gray, J. 2001, Science, 293, 2037
- . 2006, Nature, 440, 413
- Szalay, A. S., Bell, G. C., Huang, H. H., Terzis, A., & White, A. 2010, Operating Systems Review, 44, 71
- Szalay, A. S., & Blakeley, J. A. 2009, The Fourth Paradigm, ed. T. Hey, S. Tansley, & K. Tolle (Microsoft Research), 5–11
- Szalay, A. S., Gray, J., Fekete, G., Kunszt, P. Z., Kukol, P., & Thakar, A. 2007, CoRR, abs/cs/0701164
- Szalay, A. S., Kunszt, P. Z., Thakar, A. R., Gray, J., & Slutz, D. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree, 405–+
- Szalay, A. S., et al. 2009, in HICSS, 1–10
- Szalay, T., Springel, V., & Lemson, G. 2008, CoRR, abs/0811.2055
- Thakar, A., Szalay, A. S., Church, K., & Terzis, A. 2011, Scientific Programming, 19, 147
- Tody, D., et al. 2011, IVOA web site
- Ullmann, J. D., & Widom, J. 2008, Adatbazisrendszerek - Alapvetes No. ISBN 9789635454815 (Panem)
- van Ballegooij, A. 2004, in EDBT Workshops, 154–165
- Yip, C.-W., Szalay, A. S., Wyse, R. F. G., Dobos, L., Budavari, T., & Csabai, I. 2010, ApJ, 709, 780
- Yip, C. W., et al. 2009, AJ, 137, 5120

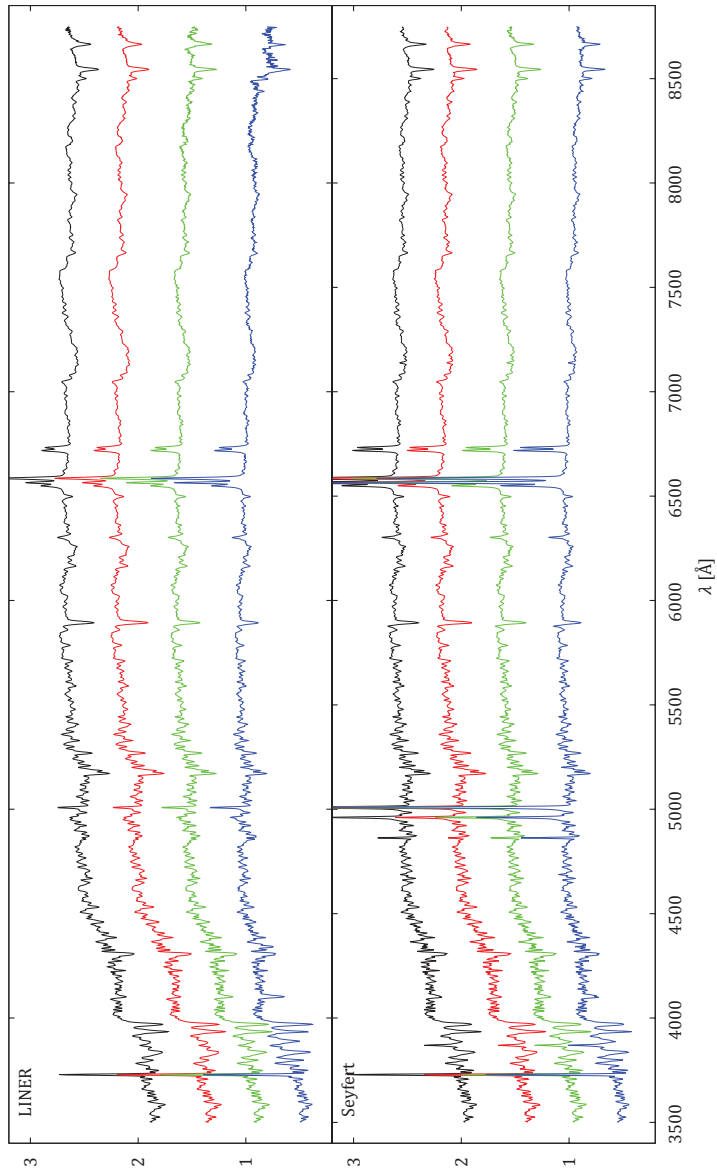
Függelék



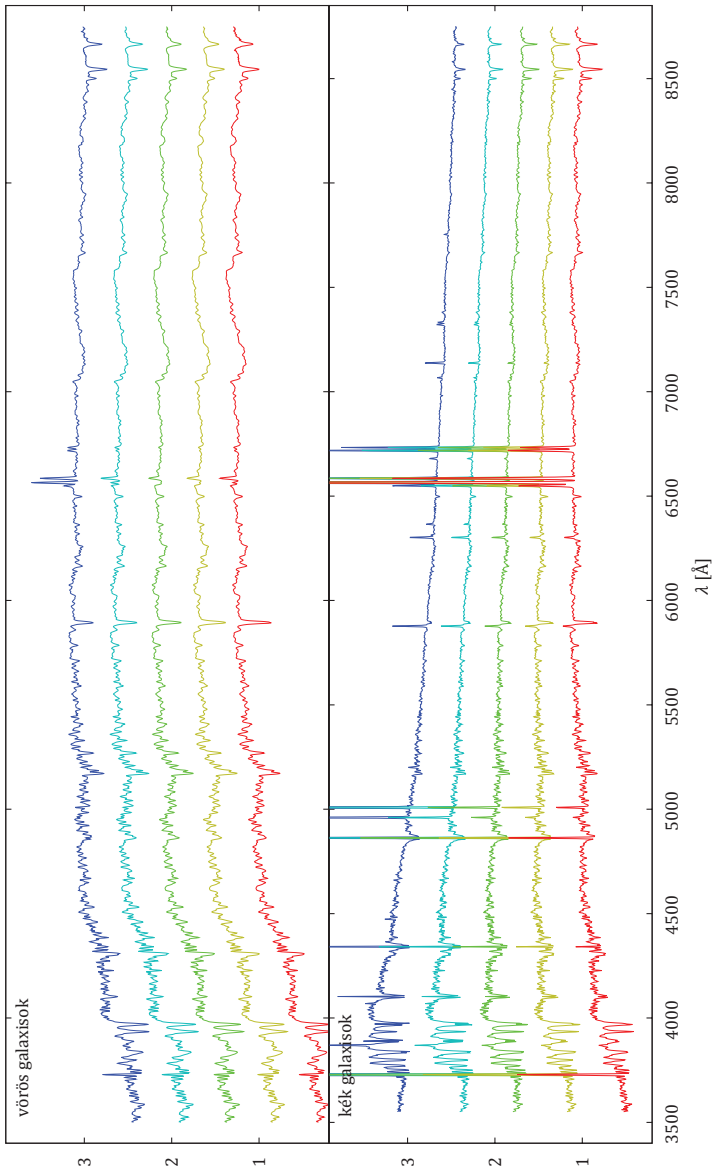
1. ábra. A különböző aktivitási és színindex osztályok kompozit spektrumai. Az aktivitási osztályt a panelek bal felső sarkában tüntettük fel. A tisztánlátás kedvéért a különböző színindex osztályok spektrumai egymáshoz képest 0,5-tel el vannak tolvva az y tengely mentén. A színindex osztályok spektrumait különböző színű görbék mutatják. Felülről lefelé: fekete – összes galaxis, vörös, zöld és kék.



2. ábra. Ld. a 1. ábra felirátát.



3. ábra. Ld. a 1. ábra feliratát.



4. ábra. A finomított színindex osztályok (ld. 4.3.4. fejezet) kompozit spektrumai. A görbékét színékekkel kódoltuk féltőlről (kék) lefelé (vörös) a galaxisok $g - r$ színindexének függvényében.

	t_{form} [Gyr]		γ [Gyr]		t_M [Gyr]		t_r [Gyr]		$Z Z_{\odot}$		τ_V		$\log \text{SSFR}$		$\log \frac{M}{L}$	
	median	mode	median	mode	median	mode	median	mode	median	mode	median	mode	median	mode	median	mode
P	10.33	11.42	0.68	0.89	6.42	6.42	5.31	4.57	1.09	0.89	0.80	0.06	-12.11	-11.40	0.49	0.49
H α	10.33	11.69	0.66	0.97	6.42	6.67	5.55	4.57	1.21	1.09	1.16	0.67	-11.93	-11.40	0.57	0.57
SF	8.69	10.88	0.50	0.81	5.17	4.47	4.08	3.10	1.09	1.71	1.50	1.53	-11.50	-10.70	0.57	0.57
A+H α	9.78	12.78	0.60	0.85	5.92	5.42	5.06	4.57	1.24	1.71	1.93	1.28	-11.58	-11.05	0.57	0.57
L	10.33	12.78	0.66	0.85	6.67	6.67	5.55	4.57	1.28	1.55	1.28	1.28	-11.93	-11.23	0.57	0.57
S	9.78	12.78	0.62	0.77	6.17	5.42	5.06	4.57	1.17	1.71	1.41	1.28	-11.58	-11.05	0.57	0.57
all	10.33	11.97	0.68	0.97	6.42	6.67	5.55	4.57	1.21	1.09	1.04	0.67	-12.11	-11.40	0.57	0.57
P	8.97	10.88	0.62	0.83	5.17	4.67	3.83	3.34	0.89	0.62	0.80	0.06	-11.40	-11.05	0.31	0.31
H α	9.24	11.97	0.58	0.77	5.42	4.67	4.32	3.83	1.01	0.55	1.28	1.16	-11.23	-10.88	0.40	0.40
SF	7.33	4.87	0.44	0.13	4.17	3.92	3.10	2.36	1.01	0.55	1.65	1.28	-10.53	-10.35	0.40	0.40
A+H α	8.15	6.24	0.50	0.30	4.92	4.67	3.83	3.34	0.97	0.55	1.53	1.16	-10.88	-10.53	0.40	0.40
L	9.51	11.97	0.58	0.77	5.67	4.67	4.57	3.83	1.01	0.74	1.28	1.16	-11.40	-11.05	0.49	0.49
S	8.69	6.24	0.54	0.77	5.17	4.67	4.08	3.83	0.97	0.35	1.41	1.16	-11.05	-10.70	0.40	0.40
all	9.24	11.97	0.58	0.77	5.42	4.67	4.32	3.83	1.01	0.55	1.16	1.16	-11.40	-10.88	0.40	0.40
P	7.87	6.78	0.54	0.91	4.17	3.42	2.85	2.12	0.62	0.20	0.92	0.06	-11.05	-10.53	0.14	0.14
H α	7.33	4.87	0.44	0.32	4.17	3.92	3.10	2.36	0.86	0.16	1.28	1.16	-10.53	-10.35	0.31	0.31
SF	4.33	3.24	0.34	0.01	2.42	1.92	1.63	1.38	0.93	0.12	1.28	1.16	-9.82	-9.82	-0.03	-0.03
A+H α	6.51	4.87	0.38	0.03	3.92	2.92	2.85	2.12	0.93	0.16	1.41	1.28	-10.35	-10.35	0.23	0.23
L	7.33	5.42	0.46	0.32	4.17	3.42	3.10	2.85	0.82	0.16	1.16	1.16	-10.70	-10.53	0.23	0.23
S	7.06	4.87	0.42	0.13	4.17	3.42	3.10	2.36	0.89	0.16	1.28	1.28	-10.53	-10.35	0.31	0.31
all	6.78	4.87	0.40	0.32	4.17	3.42	3.10	2.36	0.89	0.16	1.28	1.28	-10.53	-10.35	0.23	0.23
P	10.06	11.15	0.66	0.89	6.17	6.42	5.06	4.57	1.09	0.89	0.80	0.06	-11.93	-11.23	0.49	0.49
H α	9.78	11.15	0.62	0.77	5.92	4.67	4.82	3.83	1.13	1.09	1.16	1.16	-11.58	-11.05	0.49	0.49
SF	4.87	3.24	0.34	0.01	2.67	2.17	1.87	1.63	0.93	0.12	1.28	1.16	-10.00	-9.82	0.06	0.06
A+H α	7.87	5.69	0.48	0.30	4.67	3.92	3.59	2.85	0.97	0.55	1.41	1.16	-10.70	-10.53	0.40	0.40
L	9.78	12.78	0.62	0.85	6.17	6.67	5.06	4.57	1.24	1.52	1.41	1.28	-11.76	-11.05	0.57	0.57
S	8.69	6.24	0.54	0.30	5.17	4.67	4.08	3.83	1.01	0.55	1.41	1.16	-11.05	-10.70	0.40	0.40
all	9.78	11.15	0.62	0.77	5.92	4.67	4.82	3.83	1.13	1.09	1.16	1.16	-11.58	-11.05	0.49	0.49
RED 1	9.24	10.88	0.58	0.67	5.42	4.67	4.32	3.83	0.97	0.35	1.16	1.16	-11.23	-10.88	0.40	0.40
RED 2	10.06	11.15	0.64	0.77	5.92	4.67	4.82	3.83	1.09	0.89	1.04	0.67	-11.76	-11.23	0.49	0.49
RED 3	10.33	11.97	0.68	0.97	6.42	6.67	5.55	4.57	1.17	1.09	1.04	0.67	-12.11	-11.40	0.49	0.49
RED 4	10.60	12.78	0.70	0.83	6.92	6.67	6.04	5.55	1.28	1.17	1.16	1.28	-12.11	-11.76	0.57	0.57
RED 5	10.60	12.78	0.68	0.85	7.17	7.92	6.29	5.55	1.48	1.94	1.53	1.53	-12.11	-11.76	0.74	0.74
SF 1	1.33	1.06	0.44	0.56	0.67	0.42	0.40	0.40	0.86	0.78	1.16	0.92	-8.94	-8.94	-0.63	-0.63
SF 2	2.69	1.87	0.36	0.01	1.42	0.92	0.89	0.89	0.93	0.12	1.16	0.80	-9.47	-9.47	-0.28	-0.28
SF 3	4.06	2.96	0.34	0.01	2.42	1.92	1.63	1.38	0.97	0.12	1.28	1.16	-9.82	-9.82	-0.03	-0.03
SF 4	5.69	3.51	0.36	0.13	3.42	2.67	2.36	1.63	1.01	0.78	1.41	1.16	-10.17	-10.00	0.23	0.23
SF 5	7.60	4.87	0.46	0.30	4.67	3.92	3.34	2.36	1.01	0.55	1.65	1.53	-10.70	-10.35	0.40	0.40

1. táblázat. A a kompozitok fizikai paramétereinek Bayes-statisztikai módszerrel kapott eloszlásfüggvényeinek módusai és mediánjai. t_{form} : a folytonos modell komponens kora; γ : a csillagkeletkezési ráta időskálája; t_M : a csillagpopuláció átlagos kora, a csillagmennyel súlyozva; t_r : a csillagpopuláció átlagos kora, az r magnitúdóval súlyozva; Z : fémség; τ_V : a por komponens optikai mélysége; SSFR: fajlagos csillagkeletkezési ráta; M/L : tömeg-fényesség arány.

	λ [Å]	red P	red H α	red SF	red A+HII	red L	red S	red all	green P	green H α	green SF	green A+HII	green L	green S	green all
OII	3727.09	-	4.163	14.122	7.523	7.337	8.480	5.202	-	2.369	13.465	7.680	9.742	9.010	2.847
OII	3729.88	-	-	4.857	8.665	7.729	9.871	-	-	1.467	17.105	8.734	10.975	10.246	1.723
H θ	3798.98	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H η	3836.47	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HeI	3889.00	-	-	1.390	1.111	-	1.261	-	-	-	1.382	-	-	1.197	-
SII	4072.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H δ	4102.89	-	-	2.068	1.235	-	1.131	-	-	-	2.029	1.252	-	-	-
Hy	4341.68	-	1.433	3.889	2.464	-	2.235	1.868	-	-	3.766	1.888	-	1.730	-
OIII	4364.44	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H β	4862.68	-	3.091	9.453	5.607	1.471	4.784	4.087	-	-	9.060	4.745	1.159	3.401	-
OIII	4932.60	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OIII	4960.30	-	-	1.448	1.121	-	5.680	-	-	-	1.308	-	-	5.063	-
OIII	5008.24	-	1.086	5.125	4.198	2.949	18.697	1.998	-	-	4.692	3.252	2.622	16.649	-
OI	6302.05	-	-	1.272	1.228	1.580	1.611	-	-	-	1.247	1.238	1.490	1.496	-
OI	6365.54	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NI	6529.03	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NI	6549.86	-	-	4.204	5.429	5.731	5.875	3.902	-	-	4.150	6.879	3.603	7.439	1.658
H α	6564.61	-	14.908	41.231	26.957	10.188	22.476	19.225	-	3.751	39.508	21.372	6.435	16.527	4.380
NI	6585.27	-	5.677	12.908	13.600	9.450	15.528	7.629	-	2.137	12.745	11.214	6.453	12.834	2.436
SII	6718.29	-	2.053	7.254	4.735	3.932	4.790	2.853	-	-	6.906	4.380	3.617	4.245	1.100
SII	6732.67	-	1.478	5.205	3.748	3.351	4.242	2.088	-	-	4.973	3.682	3.316	3.932	-

2. táblázat. Az emissziós vonalak ekvivalens szélesség értékei (Ångströmben) minden galaxisosztályra. Folyt. a 3. táblázatban.

	λ [Å]	blue	blue	blue	blue	blue	blue	blue	blue	blue	blue	all	all	all	all
		P	H α	SF	A+HII	L	S	S	H α	SF	A+HII	L	S	S	S
OII	3727.09	-	4.163	14.122	7.523	7.337	8.480	5.202	-	2.369	13.465	7.680	9.742	9.010	2.847
OII	3729.88	-	-	4.857	8.665	7.729	9.871	-	-	1.467	17.105	8.734	10.975	10.246	1.723
H θ	3798.98	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H η	3836.47	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HeI	3889.00	-	-	1.390	1.111	-	1.261	-	-	-	1.382	-	-	1.197	-
SII	4072.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H δ	4102.89	-	-	2.068	1.235	-	1.131	-	-	-	2.029	1.252	-	-	-
H γ	4341.68	-	1.433	3.889	2.464	-	2.235	1.868	-	-	3.766	1.888	-	1.730	-
OIII	4364.44	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H β	4862.68	-	3.091	9.453	5.607	1.471	4.784	4.087	-	-	9.060	4.745	1.159	3.401	-
OIII	4932.60	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OIII	4960.30	-	-	1.448	1.121	-	5.680	-	-	-	1.308	-	-	5.063	-
OIII	5008.24	-	1.086	5.125	4.198	2.949	18.697	1.998	-	-	4.692	3.252	2.622	16.649	-
OI	6302.05	-	-	1.272	1.228	1.580	1.611	-	-	-	1.247	1.238	1.490	1.496	-
OI	6365.54	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NI	6529.03	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NII	6549.86	-	-	4.204	5.429	5.731	5.875	3.902	-	-	4.150	6.879	3.603	7.439	1.658
H α	6564.61	-	14.908	41.231	26.957	10.188	22.476	19.225	-	3.751	39.508	21.372	6.435	16.527	4.380
NII	6585.27	-	5.677	12.908	13.600	9.450	15.528	7.629	-	2.137	12.745	11.214	6.453	12.834	2.436
SII	6718.29	-	2.053	7.254	4.735	3.932	4.790	2.853	-	-	6.906	4.380	3.617	4.245	1.100
SII	6732.67	-	1.478	5.205	3.748	3.351	4.242	2.088	-	-	4.973	3.682	3.316	3.932	-

3. táblázat. Az emissziós vonalak ekvivalens szélesség értékei (Ångströmben) minden galaxisosztályra. Folyt. a 2. táblázatból.

	λ [Å]	RED 1	RED 2	RED 3	RED 4	RED 5	SF 1	SF 2	SF 3	SF 4	SF 5
OII	3727.09	2.924	1.803	1.621	1.992	2.323	28.630	21.238	13.033	9.077	7.940
OII	3729.88	1.577	1.254	1.200	1.445	1.531	-	26.363	16.573	11.115	9.387
H θ	3798.98	-	-	-	-	-	-	-	-	-	-
H γ	3836.47	-	-	-	-	-	-	-	-	-	-
HeI	3889.00	-	-	-	-	-	1.618	1.292	1.378	1.343	1.392
SII	4072.30	-	-	-	-	-	-	-	-	-	-
H δ	4102.89	-	-	-	-	-	2.706	2.103	2.069	1.920	1.842
H γ	4341.68	-	-	-	-	-	6.111	4.408	3.916	3.457	3.166
OIII	4364.44	-	-	-	-	-	-	-	-	-	-
H β	4862.68	1.273	-	-	-	-	17.348	11.736	9.573	8.072	7.133
OIII	4932.60	-	-	-	-	-	-	-	-	-	-
OIII	4960.30	-	-	-	-	-	10.865	3.198	1.175	-	-
OIII	5008.24	-	-	-	-	-	35.024	10.803	4.252	2.568	1.997
OI	6302.05	-	-	-	-	-	2.400	1.600	1.239	1.107	1.072
OI	6365.54	-	-	-	-	-	-	-	-	-	-
NI	6529.03	-	-	-	-	-	-	-	-	-	-
NI	6549.86	2.276	-	-	-	1.038	4.771	4.464	4.372	4.355	5.050
H α	6564.61	6.305	2.206	1.123	1.409	1.958	84.884	53.510	41.972	36.237	33.311
NI	6585.27	3.229	1.598	-	1.281	1.657	13.877	13.506	13.443	13.223	13.223
SII	6718.29	1.276	-	-	-	-	12.821	9.729	7.407	5.949	5.322
SII	6732.67	1.033	-	-	-	-	9.335	6.959	5.312	4.353	4.001

4. táblázat. A színindex szerint finomított galaxisosztályok spektrumaiból számolt kompozitok emissziós vonalainak ekvivalens szélességei (Ångströmben).

	red P	red Hz	red SF	red A+Hu	red L	red S	red L	red S	red P	green Hz	green SF	green A+Hu	green L	green S	green all
CN ₁	-0.033	-0.038	-0.052	-0.044	-0.055	-0.031	-0.041	-0.041	0.050	0.032	-0.060	-0.079	0.045	-0.004	0.033
CN ₂	0.023	0.002	-0.037	-0.009	-0.005	0.005	-0.004	-0.004	0.098	0.077	-0.035	0.006	0.086	0.032	0.077
G4327	A	0.875	0.485	0.615	0.728	0.663	0.710	0.728	1.260	1.172	0.468	0.744	1.160	0.861	1.147
G4300	A	3.810	2.905	1.185	2.205	2.718	2.841	2.613	5.492	4.917	1.296	2.811	4.942	3.732	4.834
Fe4383	A	3.467	3.124	2.698	3.120	3.595	2.928	4.945	4.638	4.638	1.737	3.173	4.814	4.351	4.576
Ca4455	A	1.255	1.122	0.580	0.900	1.019	1.017	1.035	1.599	1.509	0.614	1.035	1.415	1.219	1.477
Fe4531	A	3.156	2.720	1.913	3.384	2.614	2.644	2.591	3.511	3.336	1.965	2.612	3.309	2.983	3.302
Ca4668	A	4.718	3.792	1.839	3.065	3.908	3.153	3.491	6.368	5.928	1.951	3.757	6.116	4.445	5.897
H β	A	3.187	0.212	-4.591	-1.915	1.492	-1.474	-0.550	1.913	1.364	-4.348	-1.222	0.673	-0.739	1.240
Fe5015	A	4.695	3.588	-0.179	0.914	1.673	-7.702	2.748	5.008	4.526	0.155	1.825	2.987	-5.896	4.369
Mg ₁	mag	0.055	0.051	0.025	0.048	0.059	0.067	0.049	0.101	0.093	0.027	0.058	0.104	0.079	0.094
Mg ₂	mag	0.159	0.145	0.086	0.127	0.149	0.150	0.139	0.238	0.223	0.090	0.150	0.234	0.178	0.224
Mg b	A	2.671	2.646	1.812	2.557	2.980	2.923	2.585	3.839	3.686	1.884	2.887	4.113	3.347	3.684
Fe5270	A	2.314	2.221	1.517	1.893	2.014	2.082	2.104	2.680	2.624	1.563	2.071	2.532	2.358	2.588
Fe5335	A	2.159	2.144	1.543	1.907	2.019	2.057	2.050	2.382	2.370	1.581	2.017	2.310	2.260	2.327
Fe5406	A	1.369	1.341	0.945	1.176	1.253	1.272	1.278	1.573	1.552	0.972	1.262	1.518	1.449	1.524
Fe5709	A	0.823	0.784	0.594	0.701	0.760	0.803	0.754	0.906	0.890	0.611	0.758	0.871	0.874	0.877
Fe5782	A	0.641	0.707	0.538	0.687	0.710	0.716	0.693	0.756	0.772	0.564	0.728	0.784	0.769	0.763
Na D	A	2.566	2.209	1.282	2.363	2.447	2.121	2.208	3.366	3.266	1.495	2.864	3.471	2.678	3.302
TiO ₁	mag	0.022	0.022	0.016	0.020	0.024	0.025	0.022	0.032	0.031	0.016	0.023	0.032	0.027	0.031
TiO ₂	mag	0.053	0.053	0.041	0.051	0.056	0.056	0.052	0.073	0.070	0.042	0.056	0.074	0.062	0.070
H δ A	A	2.656	2.405	3.501	2.829	3.544	2.447	2.608	-1.280	-0.514	3.409	2.176	-0.668	1.180	-0.445
H δ F	A	-0.344	-0.822	0.128	-0.668	0.334	-1.528	-0.705	-4.969	-4.202	0.026	-1.476	-4.852	-2.984	-4.172
H γ A	A	2.676	2.265	2.411	2.261	3.096	2.084	2.287	0.625	0.934	2.373	1.973	0.823	1.501	0.944
H γ F	A	1.562	0.643	0.225	0.347	1.653	0.101	0.559	-1.092	-0.865	0.211	0.035	-1.279	-0.531	-0.867
CaI ₁₈₄₉₈	A	0.794	0.854	0.757	0.820	0.806	0.842	0.842	1.150	1.155	0.775	1.001	1.122	0.880	1.147
CaI ₁₈₅₄₂	A	2.723	2.741	2.488	2.673	2.615	2.710	2.720	3.334	3.342	2.514	2.939	3.259	2.757	3.323
CaI ₁₈₆₆₂	A	2.719	2.802	2.683	2.777	2.696	2.779	2.798	3.061	3.082	2.691	2.940	2.982	2.759	3.062
Mg ₁₈₈₀₇	A	0.448	0.473	0.343	0.415	0.382	0.449	0.460	0.372	0.382	0.359	0.343	0.339	0.465	0.371
CNF	mag	0.115	0.080	0.028	0.060	0.087	0.051	0.071	0.202	0.166	0.030	0.073	0.161	0.080	0.163
H + K	mag	0.333	0.248	0.183	0.216	0.258	0.229	0.235	0.373	0.336	0.187	0.259	0.329	0.263	0.331
H + K	mag	0.007	0.009	0.006	0.007	0.006	0.005	0.008	0.015	0.014	0.007	0.009	0.013	0.007	0.014
G	mag	0.173	0.141	0.073	0.110	0.134	0.134	0.128	0.232	0.211	0.077	0.132	0.207	0.168	0.206
H β	mag	0.104	0.046	-0.036	0.007	0.073	0.013	0.032	0.102	0.096	-0.032	0.020	0.051	0.026	0.062
Mg	mag	0.067	0.064	0.040	0.060	0.071	0.067	0.062	0.102	0.096	0.042	0.060	0.107	0.079	0.096
MH	mag	0.034	0.036	0.018	0.024	0.023	-0.028	0.032	0.060	0.055	0.020	0.031	0.051	-0.014	0.055
FC	mag	0.062	0.056	0.037	0.049	0.051	0.052	0.053	0.073	0.070	0.038	0.054	0.069	0.040	0.070
NaD	mag	0.057	0.044	0.014	0.044	0.053	0.039	0.043	0.080	0.075	0.019	0.057	0.080	0.053	0.076
D4000 _n	-	1.785	1.572	1.348	1.481	1.617	1.543	1.533	2.070	1.938	1.364	1.576	1.971	1.677	1.924
D4000 _n	-	1.578	1.401	1.223	1.320	1.427	1.365	1.366	1.868	1.737	1.234	1.397	1.758	1.482	1.724

5. táblázat. A kontinuum abszorpciós indexek értékei valamennyire galaxisosztályra. Folyt. a 6 táblázatban.

	blue P	blue Hz	blue SF	blue A+H α	blue L	blue S	blue all	p	H α	all SF	A+H α	all L	all S	all all
CN λ	-0.033	-0.038	-0.062	-0.044	-0.055	-0.031	-0.041	0.050	0.032	-0.060	-0.029	0.045	-0.004	0.033
CN λ	mag	0.023	0.002	-0.037	-0.009	-0.005	-0.004	0.098	0.077	-0.035	0.006	0.086	0.032	0.077
Ca4227	A	0.875	0.785	0.447	0.615	0.728	0.663	0.720	1.260	1.172	0.468	1.160	0.861	1.147
G4300	A	3.810	2.905	1.185	2.205	2.718	2.613	5.492	4.917	1.296	2.811	4.942	3.732	4.834
Fe4383	A	3.467	3.124	1.640	2.998	3.120	3.595	2.928	4.945	4.638	1.737	3.173	4.814	4.576
Ca4455	A	1.255	1.122	0.580	0.900	1.019	1.017	1.035	1.599	1.614	1.035	1.415	1.219	1.477
Fe4531	A	3.156	2.720	1.913	2.384	2.614	2.644	2.591	3.511	3.336	1.965	3.309	2.983	3.302
Ca4668	A	4.718	3.792	1.839	3.065	3.908	3.153	3.491	6.368	5.928	1.951	3.757	6.116	5.897
H β	A	3.187	0.212	-4.591	-1.915	1.492	-1.474	-0.550	1.913	1.364	-4.348	1.222	-0.673	-0.739
Fe5015	A	4.695	3.583	-0.179	0.914	1.673	-7.702	2.748	5.008	4.525	1.155	1.825	2.987	-5.896
Mg λ	mag	0.055	0.051	0.025	0.048	0.059	0.067	0.109	0.101	0.023	0.058	0.104	0.078	0.094
Mg λ	mag	0.159	0.145	0.086	0.127	0.149	0.150	0.139	0.238	0.223	0.090	0.150	0.234	0.178
Mg b	A	2.671	2.646	1.812	2.557	2.980	2.923	2.585	3.839	3.686	1.884	2.887	4.113	3.347
Fe5270	A	2.314	2.221	1.517	1.893	2.014	2.082	2.104	2.680	2.624	1.563	2.071	2.532	2.358
Fe5335	A	2.159	2.144	1.543	1.907	2.019	2.057	2.050	2.382	2.370	1.581	2.310	2.260	2.327
Fe5406	A	1.369	1.341	0.945	1.176	1.253	1.272	1.278	1.573	1.552	0.972	1.262	1.518	1.449
Fe5709	A	0.823	0.784	0.594	0.701	0.760	0.803	0.754	0.906	0.890	0.611	0.758	0.871	0.874
Fe5782	A	0.641	0.707	0.538	0.687	0.710	0.716	0.693	0.756	0.772	0.564	0.728	0.784	0.769
Na D	A	2.566	2.209	1.282	2.363	2.447	2.121	2.208	3.366	3.266	1.495	2.864	3.471	2.678
TiO λ	mag	0.022	0.022	0.016	0.020	0.024	0.025	0.022	0.032	0.031	0.016	0.023	0.027	0.031
TiO λ	mag	0.053	0.053	0.041	0.051	0.056	0.056	0.052	0.073	0.070	0.042	0.056	0.074	0.062
H δ A	A	2.656	2.405	3.501	2.829	3.544	2.447	2.608	-1.280	-0.514	3.409	2.176	-0.668	1.180
H γ A	A	-0.344	-0.822	0.128	-0.668	0.334	-1.528	-0.705	-4.969	-4.202	0.026	-1.476	-4.852	-2.984
H δ F	A	2.676	2.265	2.411	2.261	3.096	2.084	2.287	0.625	0.934	2.373	1.973	0.823	1.501
H γ F	A	1.562	0.643	0.225	0.347	1.653	0.101	0.559	-1.092	-0.865	0.211	0.035	-1.279	-0.531
CaI λ 8498	A	0.794	0.854	0.757	0.820	0.806	0.842	0.842	1.150	1.155	0.775	1.001	1.122	0.880
CaI λ 8542	A	2.723	2.741	2.488	2.673	2.615	2.710	2.720	3.334	3.342	2.514	2.939	3.259	2.757
CaI λ 8662	A	2.719	2.802	2.683	2.777	2.696	2.779	2.798	3.061	3.082	2.691	2.940	2.982	2.759
MgI λ 8807	A	0.448	0.473	0.343	0.415	0.382	0.449	0.460	0.372	0.382	0.359	0.343	0.339	0.465
CNB	mag	0.115	0.080	0.028	0.060	0.087	0.051	0.071	0.202	0.166	0.030	0.073	0.161	0.080
H + K	mag	0.333	0.248	0.183	0.216	0.258	0.229	0.235	0.373	0.336	0.187	0.239	0.329	0.263
G	mag	0.007	0.009	0.006	0.007	0.006	0.005	0.008	0.015	0.014	0.007	0.009	0.013	0.007
Hb	mag	0.173	0.141	0.073	0.110	0.134	0.134	0.128	0.232	0.211	0.077	0.132	0.207	0.168
Hb	mag	0.104	0.046	-0.036	0.007	0.073	0.013	0.032	0.073	0.064	-0.032	0.020	0.051	0.026
MgG	mag	0.067	0.064	0.040	0.060	0.071	0.067	0.062	0.102	0.096	0.042	0.069	0.107	0.079
MgH	mag	0.034	0.036	0.018	0.024	0.023	-0.028	0.032	0.060	0.055	0.020	0.031	0.051	-0.014
FC	mag	0.062	0.056	0.037	0.049	0.051	0.052	0.053	0.073	0.070	0.038	0.054	0.069	0.060
NaD	mag	0.057	0.044	0.014	0.044	0.053	0.039	0.043	0.080	0.075	0.019	0.057	0.080	0.053
D4000 λ	-	1.785	1.572	1.348	1.481	1.617	1.543	1.533	2.070	1.938	1.364	1.576	1.971	1.677
D4000 λ	-	1.578	1.401	1.223	1.320	1.427	1.365	1.366	1.868	1.737	1.234	1.397	1.758	1.482

6. táblázat. A kontinuum abszorpciós indexek értékei valamennyire galaxisosztályra. Folyt. a 5 táblázatból.

	RED 1	RED 2	RED 3	RED 4	RED 5	SF 1	SF 2	SF 3	SF 4	SF 5
CN ₁	mag	-0.002	0.032	0.063	0.079	0.066	-0.047	-0.069	-0.064	-0.064
CN ₂	mag	0.043	0.078	0.110	0.124	0.110	-0.045	-0.049	-0.039	-0.027
Ca4227	Å	1.038	1.209	1.304	1.301	1.286	0.205	0.341	0.429	0.539
G4300	Å	4.289	5.175	5.563	5.580	5.465	0.092	0.397	1.130	1.659
Fe4383	Å	4.177	4.809	5.087	5.108	5.046	1.123	1.194	1.576	2.064
Ca4455	Å	1.418	1.575	1.604	1.574	1.536	0.096	0.361	0.576	0.735
Fe4531	Å	3.190	3.426	3.499	3.483	3.430	1.183	1.625	1.890	2.122
Ca4668	Å	5.160	6.034	6.615	6.927	6.669	0.683	1.198	1.751	2.337
H β	Å	1.407	1.734	1.597	1.434	1.409	-12.371	-6.644	-4.584	-3.640
Fe5015	Å	4.305	4.728	4.825	4.741	4.567	-0.991	-0.991	0.305	1.743
Mg ₁	mag	0.072	0.089	0.109	0.120	0.113	0.019	0.024	0.032	0.039
Mg ₂	mag	0.186	0.218	0.251	0.268	0.256	0.054	0.068	0.083	0.101
Fe5270	Å	3.210	3.626	4.039	4.232	4.085	1.301	1.520	1.763	2.080
Fe5335	Å	2.544	2.690	2.706	2.690	2.669	0.800	1.229	1.495	1.710
Fe5406	Å	2.370	2.447	2.380	2.316	2.320	0.963	1.295	1.525	1.708
Fe5709	Å	1.525	1.594	1.582	1.543	1.525	0.594	0.800	0.933	1.054
Fe5782	Å	0.889	0.923	0.901	0.878	0.897	0.376	0.505	0.578	0.654
NaD	Å	0.770	0.780	0.774	0.776	0.785	0.303	0.412	0.523	0.631
NaD ₂	Å	2.732	3.077	3.575	4.066	4.481	0.015	0.633	1.162	1.898
TiO ₁	mag	0.026	0.030	0.033	0.035	0.034	0.012	0.014	0.016	0.018
TiO ₂	mag	0.062	0.069	0.076	0.080	0.078	0.032	0.034	0.040	0.046
H δ A	Å	0.762	-0.698	-1.724	-1.993	-1.728	3.012	3.859	3.591	3.057
H δ B	Å	-2.713	-4.343	-5.508	-5.856	-5.574	-1.284	0.332	0.260	-0.311
H γ A	Å	1.568	0.893	0.379	0.222	0.341	1.653	2.489	2.463	2.217
H δ F	Å	-0.081	-0.834	-1.457	-1.655	-1.514	-1.718	-0.004	0.301	0.169
CaII 8498	Å	1.078	1.167	1.144	1.127	1.187	0.470	0.561	0.752	0.807
CaII 8542	Å	3.231	3.370	3.318	3.272	3.445	1.967	2.166	2.476	2.581
CaII 8662	Å	3.074	3.114	3.010	2.961	3.182	2.329	2.476	2.682	2.728
MgI 8807	Å	0.410	0.392	0.365	0.342	0.478	0.101	0.200	0.342	0.408
CNB	mag	0.125	0.175	0.213	0.229	0.214	-0.006	0.021	0.027	0.037
H+K	mag	0.310	0.352	0.372	0.375	0.366	0.113	0.163	0.182	0.198
Ca	mag	0.012	0.013	0.016	0.016	0.016	0.002	0.005	0.006	0.007
G	mag	0.193	0.224	0.232	0.227	0.224	0.030	0.052	0.071	0.090
H β	mag	0.067	0.071	0.067	0.065	0.065	-0.166	-0.070	-0.036	-0.020
MgG	mag	0.080	0.093	0.108	0.116	0.110	0.026	0.033	0.039	0.047
MH	mag	0.044	0.054	0.063	0.068	0.065	-0.064	0.000	0.020	0.028
FC	mag	0.066	0.071	0.074	0.075	0.074	0.017	0.029	0.037	0.043
NaD	mag	0.060	0.071	0.085	0.098	0.105	-0.031	-0.006	0.011	0.029
D4000	-	1.814	1.993	2.105	2.146	2.145	1.161	1.273	1.342	1.415
D4000 _n	-	1.611	1.787	1.903	1.941	1.919	1.081	1.166	1.217	1.271

7. táblázat. A kontinuum abszorpciós indexek értékei a finomított színdexek szerint meghatározott osztályokra.

Összefoglaló

Galaxisok spektrumának elméleti alapokon való modellezése a múlt évtized egyik sikertörténete. Miközben egyre jobb és nagyobb felbontású modellek váltak elérhetővé, a Sloan Digitális Égtérképezés korábban sosem látott mennyiségű megfigyeléssel szolgált, melyekkel az elméleti modelleket össze lehet hasonlítani. A modellek megfigyelésekhez való illesztésével, közvetlenül mérhető spektrális indexek bevezetésével a közeli galaxisok fizikai paraméterei könnyen meghatározhatók. Ebben a dolgozatban röviden áttekintettük a csillagpopulációsintézis-modelleket gyakorlati nézőpontból, megvizsgáltuk a csillagközi por modellezése körüli problémákat, és felvázoltuk a galaxisok emissziósvonal-arányok szerinti osztályozásának alapjait. Röviden áttekintettük a galaxispektrumok dimenzió-redukciójára használt szélesebb körben elterjedt eljárásokat, és robusztus főkomponens-analízis segítségével különböző típusú galaxisosztályok kompozit spektrumait származtattuk abból a célból, hogy egy nagy felbontású, nagy jel-zaj arányú spektrumatlaszt hozunk létre. Meghatároztuk az így kapott kompozitok fizikai paramétereit, és megmutattuk, hogy az atlasz lefedi a fizikai paraméterek egyedi galaxisok által kifizített terét. Az atlasz spektrumait elsősorban fotometrikus vöröseltolódás-becslő algoritmusokhoz szánjuk sablonként, olyan jelenlegi és jövőbeli, kizárólag fotometriai adatokat gyűjtő égfelmérésekhez, mint a PanSTARRS vagy az LSST.

Az értekezés második részében az SDSS által megfigyelt vörös óriásgalaxisok fényesség szerinti eloszlását vizsgáltuk. Az spektrális evolúciós effektusok, és a galaxisűrűség időbeli változásának megfelelő korrigálásával meghatároztuk az LRG-k univerzális, $0 \leq z \leq 0.5$ között érvényes luminozitásfüggvényét. Bemutattuk az extrémérték-statisztika (EVS) és sorrendi statisztika (OS) alapjait, majd kiszámítottuk a luminozitásfüggvény alapvető EVS/OS mennyiségeit. Megmutattuk, hogy az LRG-k együtt mozgó koordináták szerint vett sűrűsége vöröseltolódás-héjanként egyszerűen megbecsülhető az adott héj k . legfényesebb galaxisának magnitúdójából. Ez a technika sikeresen alkalmazható a galaxisűrűség meghatározására olyan nagy távolságokban, ahol már csak a legfényesebb galaxisok figyelhetők meg. Érintettük a galaxishalmazok legfényesebb galaxisainak (BCG-k) kérdéskörét is, és megmutattuk, hogy a legfényesebb galaxisok luminozitása, illetve ezek, és a második legfényesebb galaxisok közötti magnitúdó rés nagysága megmagyarázható egyetlen, közös luminozitásfüggvény segítségével is – szemben az elterjedt vélekedéssel, miszerint a BCG-k és a szatellit galaxisok külön populációt alkotnak, és a fényességeloszlásuk különböző.

A harmadik részben a tudományos adatok feldolgozásának legfontosabb problémáit tekintettük át. Bemutattunk néhány technikát, amelyekkel az ismertetett problémák áthidalhatók. Ennek keretében bemutattuk a többdimenziós adatbázisok területén végzett saját munkánkat is. Röviden felvázoltuk a csillagászati Virtuális Observatórium (VO) koncepcióját, és bemutattuk azokat a szoftvereszközöket, melyeket a VO-hoz fejlesztettünk: a Lányom Szolgáltatást, a Spektrum Szolgáltatást, valamint a SkyQuery legújabb verzióját, ami elosztott adatbázisszerver-klasztereken való futtatásra lett tervezve.

Summary

Theoretical modelling of galaxy spectra was one of the success stories of the last decade. While better and higher resolution spectral models became available, the Sloan Digital Sky Survey supplied an unprecedented amount of observations to compare the theoretical models with. By fitting models to the observations, and introducing readily measurable spectral indices, one can easily estimate the physical parameters of nearby galaxies. In this work, we have briefly reviewed the stellar population synthesis models from the practical prospective, explained the issues arising from modelling interstellar dust, and explained the basics of classification of galaxies by emission line ratios. We have briefly reviewed the dimensionality reduction techniques widely used to analyze galaxy spectra, and applied Principal Component Analysis to various classes of SDSS galaxy spectra to compose an atlas of high resolution, high signal-to-noise ratio composite spectra. We have determined the basic physical parameters of these composites and shown that the atlas covers much of the physical parameter space spanned by individual galaxies. We intend the spectra of the atlas to be used as templates for photometric redshift estimation for the ongoing and upcoming large photometric surveys like PanSTARRS and LSST.

In the second part of the thesis, we have investigated the luminosity distribution of luminous red galaxies of the SDSS. By applying the right corrections to spectral evolution and density evolution, we have determined the universal luminosity function of LRGs that is valid between $0 \leq z \leq 0.5$. We have introduced the basics of extreme-value and order statistics and calculated the basic EVS/OS quantities of the luminosity function. We have shown that the co-moving density of LRGs in redshift bins can be successfully estimated from the magnitudes of the k^{th} brightest galaxies of the bins. This technique can be successfully used to estimate galaxy densities at distances where only the brightest objects are observable. We have touched the topic of brightest cluster galaxies and shown that their luminosity and the magnitude gap between the brightest and second brightest galaxies can be explained solely on the basis of a common luminosity function – in contrast to the common belief that BCGs and satellites belong to separate populations and thus, have different luminosity distributions.

In the third part, we have surveyed the most important problems in modern scientific data processing and presented some techniques that can be used to address these problems, and outlined our work in the field of multidimensional databases. We have briefly introduced the concept of the astrophysical Virtual Observatory and explained the tools we developed for it: the Footprint Service, the Spectrum Services, and the latest version of SkyQuery which was designed to run on distributed database cluster systems.