Machine Learning techniques applied to Astronomy: The Merging Systems Identification (MeSsI) Algorithm.

Martín de los Rios (ICTP-SAIFR/IFT-UNESP)

October 31, 2019

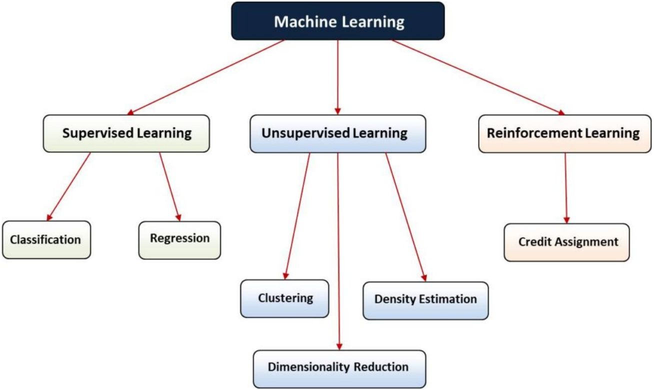
martindelosrios13@gmail.com
https://martindelosrios.netlify.com/

Table of contents

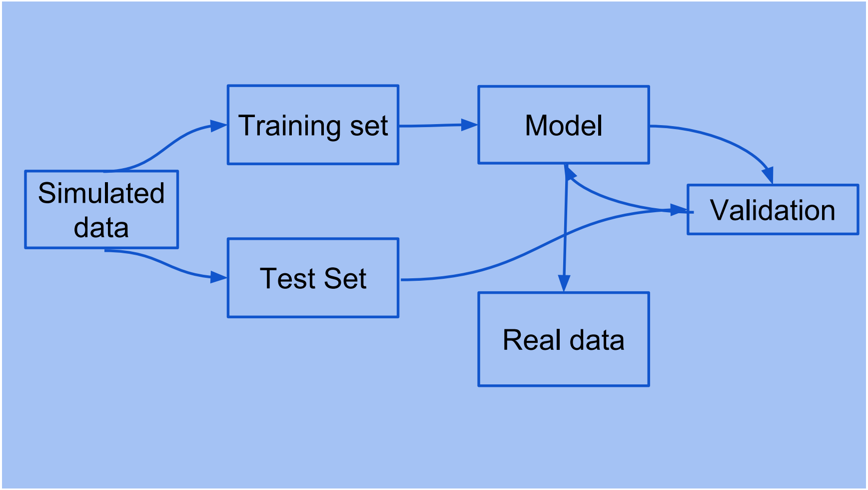
1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.
3. Future Work

> A computer program learns to perform a task T , based on a training E and taking into account a measure P of its performance, if this measure P when making T improves with the training E.

Thomas M. Mitchell. Machine Learning.


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Density Estimation
Dimensionality Reduction
Credit Assignment

Supervised Learning.

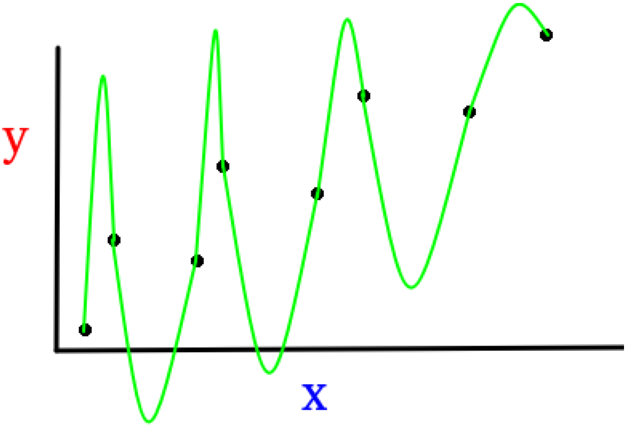


	a	b	...	z
1				
2				
...				
n				

	a	b	...	z
1				
2				
...				
n				

$$y = f(x)$$

Machine Learning


y
x

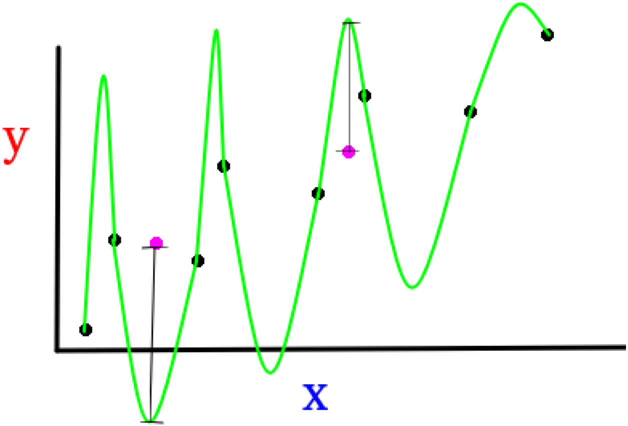
		a	b	...	z
Training-set	1				
Test-set	2				
	...				
Training-set	n				

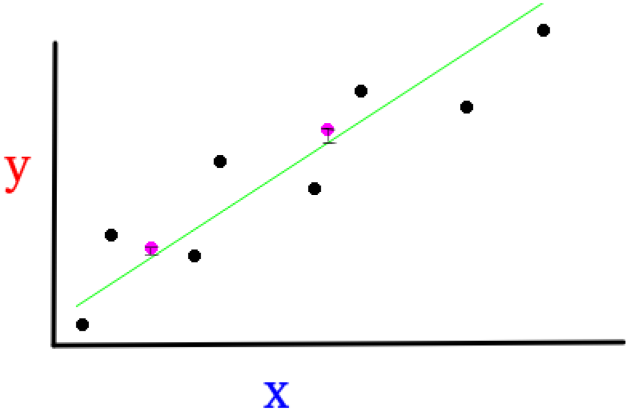
$$\min|y_i - f(x_i)|$$

Training

$$y_i - f(x_i)$$

Testing


y
x


y
x

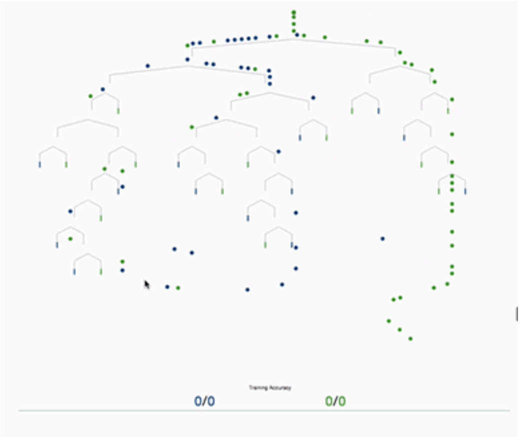
		a	b	...	z
Training-set	1				
Test-set	2				
	...				
Training-set	n				
New data		?	b	...	z

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.

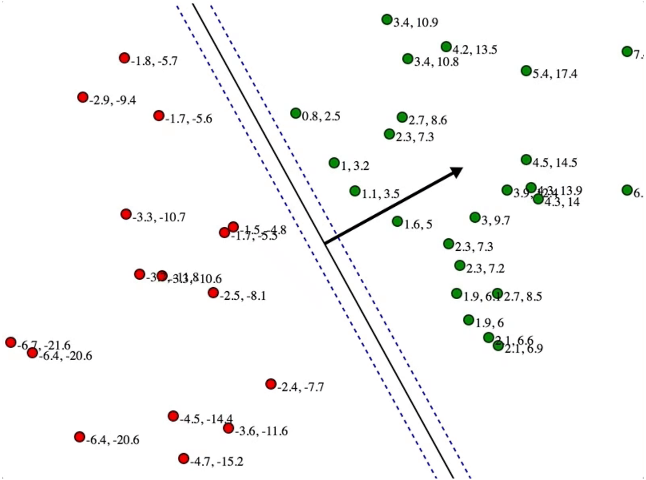
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.

3. Future Work

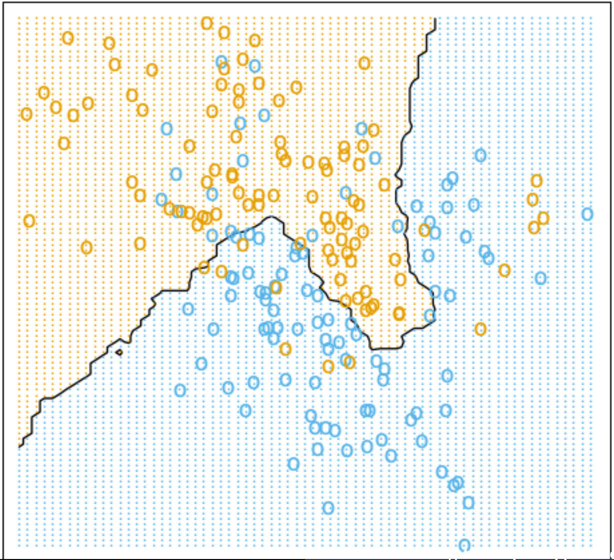
Random Forest



Support Vector Machines



K-Nearest Neighbour



Deep Learning

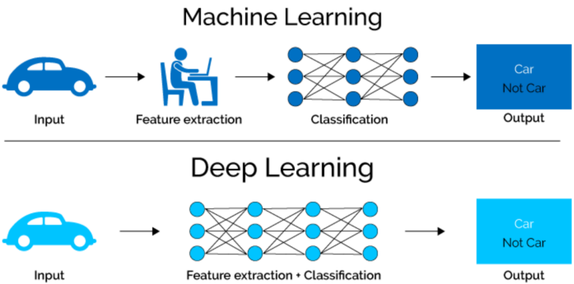


Figure 1: Machine Learning VS Deep Learning

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.

2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.

3. Future Work

Classification Problems

- The S-PLUS: a star/galaxy classification based on a Machine Learning approach. Costa-Duarte et al. (1909.08626)
- A Neural Network Gravitational Arc Finder based on the Mediatrix filamentation Method. Bom et al. (1607.04644)
- Circumventing Lens Modeling to Detect Dark Matter Substructure in Strong Lens Images with Convolutional Neural Networks. Diaz Rivero & Dvorkin. (1910.00015)
- Distinguish standard and MOG with ML and weak lensing. Peel et al. (1810.11030)

Regression Problems

- A Hybrid Deep Learning Approach to Cosmological Constraints From Galaxy Redshift Surveys.Ntampaka et al. (1909.10527)
- An improved cosmological parameter inference scheme motivated by deep learning. Ribli et al. (1806.05995)
- Cosmological parameter estimation from large-scale structure deep learning. Pan et al. (1908.10590)
- Estimating Cosmological Parameters from the Dark Matter Distribution. Ravanbakhsh et al. (1711.02033)

Other applications

- CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks. Mustafa et al. (1706.02390)
- From Dark Matter to Galaxies with Convolutional Networks. Zhang et al. (1902.05965)
- Learning to Predict the Cosmological Structure Formation.He et al. (1811.06533)
- CosmicNet I: Physics-driven implementation of neural networks within Boltzmann-Einstein solvers. Albers et al. (1907.05764)

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.

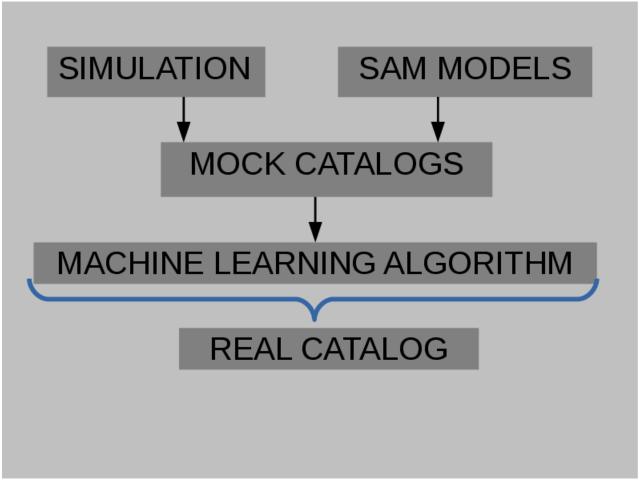
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.

3. Future Work

Mon. Not. R. Astron. Soc. **000**, 1–7 (2015) Printed 17 December 2015 (MN LaTeX style file v2.2)

The MeSsI (Merging Systems Identification) Algorithm & Catalogue.

Martín de los Rios⋆, Mariano J. Domínguez R.⋆, Dante Paz, Manuel Merchán [1,2,3].
[1] *Instituto de Astronomía Teórica y Experimental (CCT Córdoba, CONICET, UNC), Laprida 854, X5000BGR, Córdoba, Argentina.*
[2] *Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina.*
[3] *Consejo Nacional de Investigaciones Científicas y Técnicas, Rivadavia 1917, C1033AAJ Buenos Aires, Argentina.*


SIMULATION
SAM MODELS
MOCK CATALOGS
MACHINE LEARNING ALGORITHM
REAL CATALOG

Clusters identification.

- We construct a mock catalogue based on the results of the application of the SAM Model of Guo et al. 2010 to the Millenium simulation.

Clusters identification.

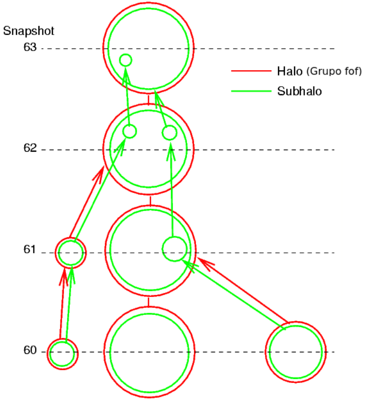
- We construct a mock catalogue based on the results of the application of the SAM Model of Guo et al. 2010 to the Millenium simulation.
- We Perform a friend-of-friend algorithm (*Merchan & Zandivares 2002*) to the mock catalog in order to identify the clusters.

Clusters identification.

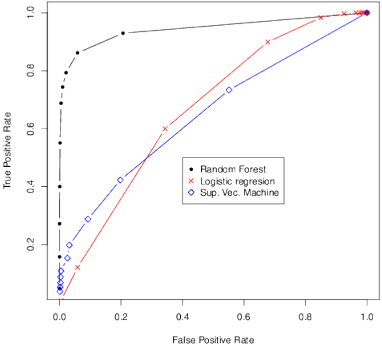
- We construct a mock catalogue based on the results of the application of the SAM Model of Guo et al. 2010 to the Millenium simulation.
- We Perform a friend-of-friend algorithm (*Merchan & Zandivares 2002*) to the mock catalog in order to identify the clusters.
- We assign each identified cluster with a fof-group in the simulation.

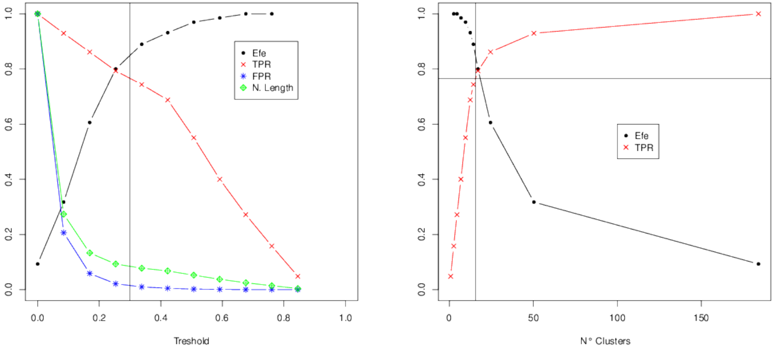
Study of the merger trees.

- Based on the subhalos merger trees, we study the merger tree for every galaxy cluster in the mock catalog.



- Dressler-Shectman test.
- Non gaussianity test.
- Color.
- Number of galaxies.




Efe
TPR
FPR
N. Length
Treshold
Efe
TPR
N° Clusters

Study of the identified substructures.

- We ran a second stage of a random forest algorithm, applied to the galaxies.

Study of the identified substructures.

- We ran a second stage of a random forest algorithm, applied to the galaxies.
- In order to define substructures we identify clumps of galaxies based on their proximity, using a mixture of gaussians weighted by the probability of each galaxy of being part of a substructure, calculated with the RF.

Study of the identified substructures.

- We ran a second stage of a random forest algorithm, applied to the galaxies.
- In order to define substructures we identify clumps of galaxies based on their proximity, using a mixture of gaussians weighted by the probability of each galaxy of being part of a substructure, calculated with the RF.
- After that, we estimate the velocity dispersion and the virial radius.

Application of the MeSsI Algorithm to spectroscopy catalogues

- We find 12 Clusters with high probability of been in a merger in the SDSS DR7.

Application of the MeSsI Algorithm to spectroscopy catalogues

- We find 12 Clusters with high probability of been in a merger in the SDSS DR7.
- We find 7 Clusters with high probability of been in a merger in the WINGS Clusters.

Application of the MeSsI Algorithm to spectroscopy catalogues

- We find 12 Clusters with high probability of been in a merger in the SDSS DR7.
- We find 7 Clusters with high probability of been in a merger in the WINGS Clusters.
- We find 15 Clusters with high probability of been in a merger in the HECs Clusters.

https://github.com/Martindelosrios/MeSsI

```
library('devtools')
install_github('MartindelosRios/MeSsI')
```

Example

```
# Loading the MeSsI library.
library('MeSsI')

# Loading the data
data('GalaxiesDataset')

# Let's see the structure of this dataset
str(GalaxiesDataset)

# As you can see this dataset already have all the properties of the galaxies precomputed.
# We will remove this properties and start with a dataset with only the angular positions (ra, dec),
#  the redshift (z), the identification of the cluster to which the galaxy belongs (id), the color (color)
#  and the r apparent magnitude (mag).

cat <- GalaxiesDataset[, (1:6)]
colnames(cat)[1] <- 'id'
str(cat)

# Then we just can apply the messi functions to this catalog, optionally given a name to the folder where al
#  outputs file will be saved.

messi(cat, folder = 'test')
```

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.
3. Future Work

Astronomy & Astrophysics manuscript no. merger1_v3
December 4, 2017

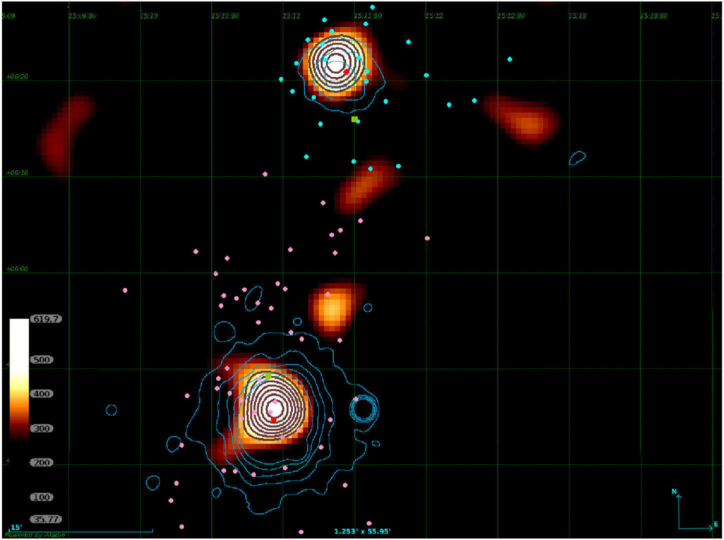
I. Analysis of candidates for interacting galaxy clusters

A1204 and A2029/A2033

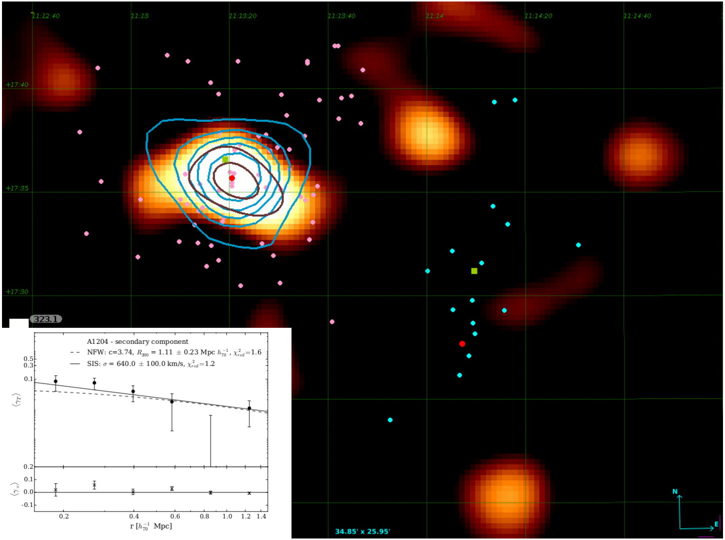
Elizabeth Johana Gonzalez ⋆[1,2], Martín de los Rios[1,2], Gabriel A. Oio[1,2], Daniel Hernández Lang[4], Tania Paez Tagliaferro[1,2], Mariano J. Domínguez R.[1,2], José Luis Nilo Castellón[3,4], Héctor Cuevas L.[4], and Carlos A. Valotto[1,2]

[1] Instituto de Astronomía Teórica y Experimental, (IATE-CONICET), Laprida 854, X5000BGR, Córdoba, Argentina.
[2] Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina.
[3] Instituto de Investigación Multidisciplinario en Ciencia y Tecnología, Universidad de La Serena. Benavente 980, La Serena, Chile.
[4] Departamento de Física y Astronomía, Facultad de Ciencias, Universidad de La Serena. Av. Juan Cisternas 1200, La Serena, Chile.

A2029/2033


619.7
500
400
300
200
100
35.77
15'
Powered by Aladin
1.253° x 55.95'
N
E

A1204


11:12:40
11:13
11:13:20
11:13:40
11:14
11:14:20
11:14:40
+17:40
+17:35
+17:30
323.1
A1204 - secondary component
NFW: c=3.74, R_{200} = 1.11 ± 0.23 Mpc h_{70}^{-1}, χ^2_{red}=1.6
SIS: σ = 640.0 ± 100.0 km/s, χ^2_{red}=1.2
$\langle \gamma_T \rangle$
$\langle \gamma_\times \rangle$
r [h_{70}^{-1} Mpc]
N
E
34.85' x 25.95'

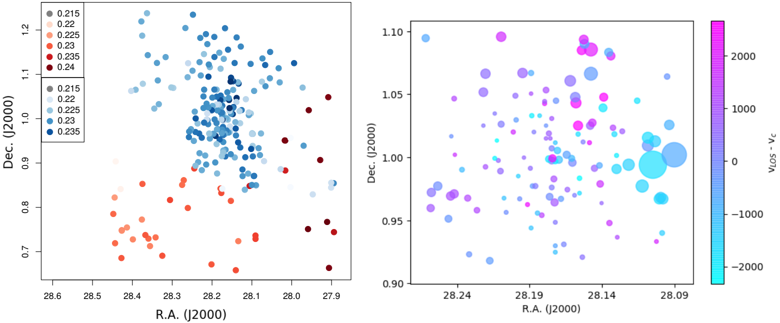
1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.
3. Future Work

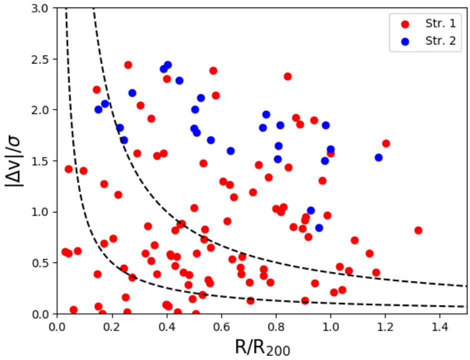
II. Analysis of candidates for interacting galaxy clusters:

A267, a merging fossil group.

Elizabeth Johana Gonzalez ⋆[1,2], María Jose Kanagusuku[1,2], Martín de los Rios[1,2], Gabriel A. Oio[1,2], Daniel Hernández Lang[4], Tania Aguirre Tagliaferro[1,2], Mariano J. Domínguez R.[1,2], José Luis Nilo Castellón[3,4], Héctor Cuevas L.[4], and Carlos A. Valotto[1,2]

[1] Instituto de Astronomía Teórica y Experimental, (IATE-CONICET), Laprida 854, X5000BGR, Córdoba, Argentina.
[2] Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina.
[3] Instituto de Investigación Multidisciplinario en Ciencia y Tecnología, Universidad de La Serena. Benavente 980, La Serena, Chile.
[4] Departamento de Física y Astronomía, Facultad de Ciencias, Universidad de La Serena. Av. Juan Cisternas 1200, La Serena, Chile.


0.215
0.22
0.225
0.23
0.235
0.24
0.215
0.22
0.225
0.23
0.235
Dec. (J2000)
R.A. (J2000)
Dec. (J2000)
R.A. (J2000)
v_{LOS} - v_c



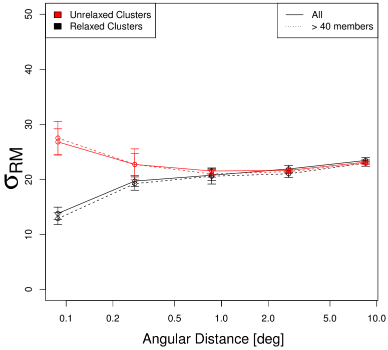
- $\approx 40\%$ Fossil groups have a major merger $z < 0.8$.
- $\approx 15 - 25\%$ Fossil groups have a major merger $z < 0.3$.

Noble et al. 2013

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.
2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.
3. Future Work

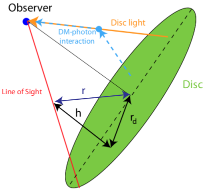
Faraday Rotation Measure dependence with galaxy clusters dynamics

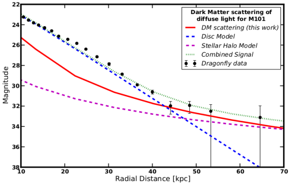
F.A. Stasyszyn[1,2⋆] & M. de los Rios[1,2,3]
[1] *Instituto de Astrofísica Teórica y Experimental (IATE), Laprida 854, Córdoba, Argentina*
[2] *Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argen*
[3] *Consejo Nacional de Investigaciones Científicas y Técnicas, Rivadavia 1917, C1033AAJ Buenos Aires, Argentina.*



Unrelaxed Clusters
Relaxed Clusters
All
> 40 members
σ_{RM}
Angular Distance [deg]
Projected Distance [Mpc]

Davis & Silk (1410.5423)



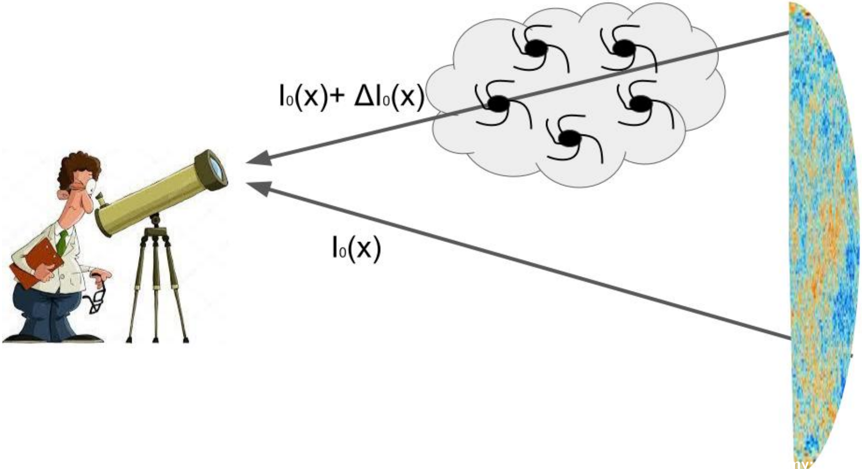


- Boœm et al. (1404.7012)
- Schewtschenko et al. (1512.0677)
- Stadler & Boœm (1802.06589)

$$\sigma_{DM\gamma} \leq 2.25x10^{-6}\sigma_{Th}(mDM/GeV)$$

DM-γ interactions (work in progress)

In collaboration with Dra. Celine Bœhm




CMB
SZ
?
?

1. Introduction to Machine Learning techniques.
   - Some important methods.
   - Machine Learning in Astronomy.

2. The MeSsI Algorithm
   - Machine learning algorithm applied for identification of substructures.
   - A2029/2033 & A1204.
   - A1204.
   - A267.
   - Statistical analysis of the magnetic fields in merging clusters.
   - DM-γ interactions.

3. Future Work

Future Work

- Perform astrophysical test over our sample of colliding cluster candidates of the catalog looking for impose some constraints on dark matter particle properties.

Future Work

- Perform astrophysical test over our sample of colliding cluster candidates of the catalog looking for impose some constraints on dark matter particle properties.
- Study the physical properties of the galaxies that belong to the identified substructures.

Future Work

- Perform astrophysical test over our sample of colliding cluster candidates of the catalog looking for impose some constraints on dark matter particle properties.
- Study the physical properties of the galaxies that belong to the identified substructures.
- Construct a catalog of relaxed clusters in order to study the dark matter equation of state (Serra & Dominguez 2011).

Future Work

- Perform astrophysical test over our sample of colliding cluster candidates of the catalog looking for impose some constraints on dark matter particle properties.
- Study the physical properties of the galaxies that belong to the identified substructures.
- Construct a catalog of relaxed clusters in order to study the dark matter equation of state (Serra & Dominguez 2011).
- Study the candidate to merging cluster A376 (HSC SUBARU Images).

Future Work

- Estimate the CMB Masked Bispectra using Machine Learning techniques.

Future Work

- Estimate the CMB Masked Bispectra using Machine Learning techniques.
- Estimate the Dark matter Distribution of spiral galaxies using Deep Learning.

Future Work

- Estimate the CMB Masked Bispectra using Machine Learning techniques.
- Estimate the Dark matter Distribution of spiral galaxies using Deep Learning.
- Identifying backsplash galaxies in the phase-space using Machine Learning.


THANK YOU