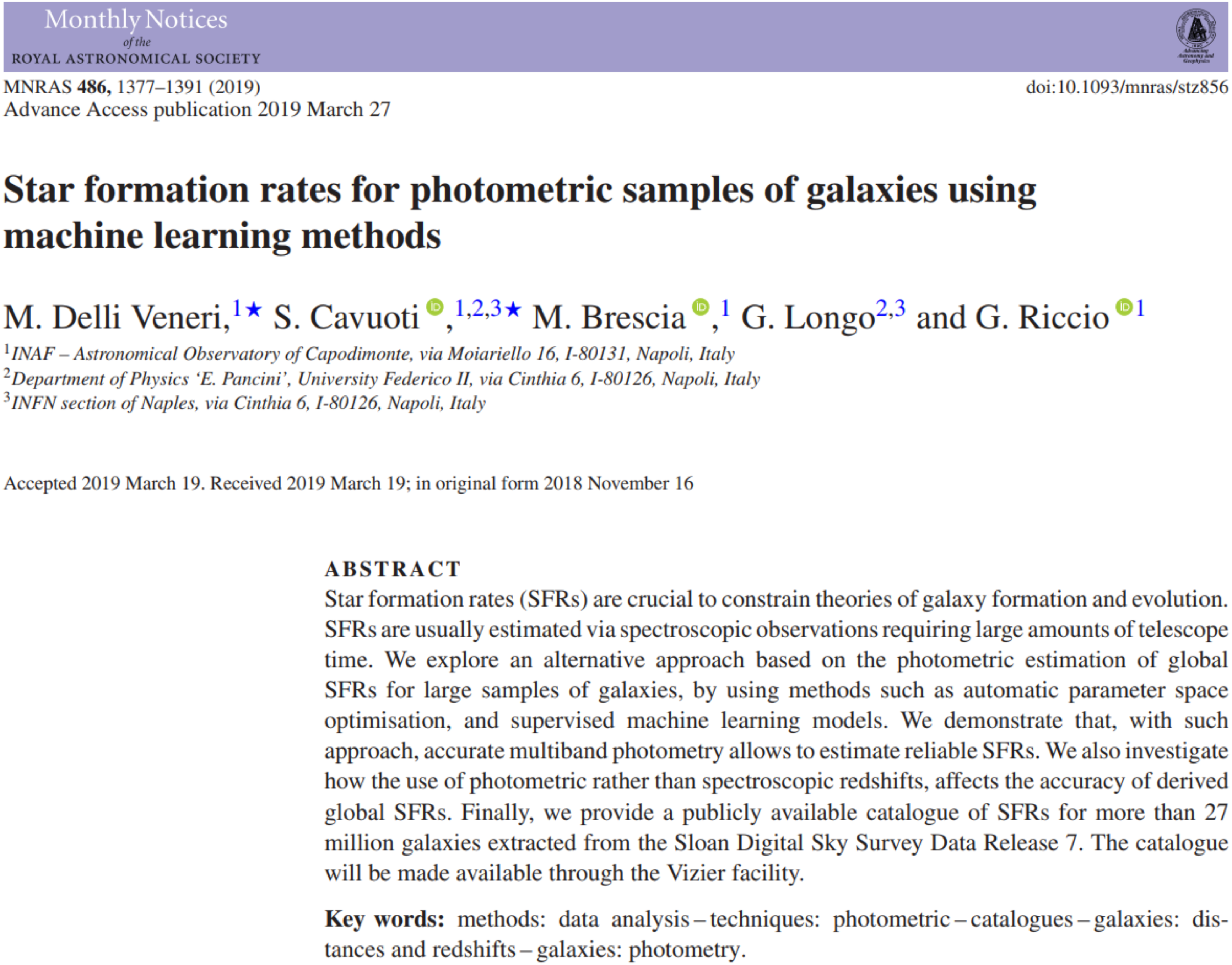
Star Formation Rates for large samples of galaxies using Machine learning methods

Michele Delli Veneri(2),
Stefano Cavuoti(1,2), Massimo Brescia(2),Giuseppe Riccio(2) &
Giuseppe Longo(1)

1 - University of Naples Federico II
2 - Astronomical Observatory of Capodimonte



Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **486,** 1377–1391 (2019)
Advance Access publication 2019 March 27

doi:10.1093/mnras/stz856

Star formation rates for photometric samples of galaxies using machine learning methods

M. Delli Veneri,[1]★ S. Cavuoti ,[1,2,3]★ M. Brescia ,[1] G. Longo[2,3] and G. Riccio [1]

[1] *INAF – Astronomical Observatory of Capodimonte, via Moiariello 16, I-80131, Napoli, Italy*
[2] *Department of Physics ‘E. Pancini’, University Federico II, via Cinthia 6, I-80126, Napoli, Italy*
[3] *INFN section of Naples, via Cinthia 6, I-80126, Napoli, Italy*

Accepted 2019 March 19. Received 2019 March 19; in original form 2018 November 16

ABSTRACT

Star formation rates (SFRs) are crucial to constrain theories of galaxy formation and evolution. SFRs are usually estimated via spectroscopic observations requiring large amounts of telescope time. We explore an alternative approach based on the photometric estimation of global SFRs for large samples of galaxies, by using methods such as automatic parameter space optimisation, and supervised machine learning models. We demonstrate that, with such approach, accurate multiband photometry allows to estimate reliable SFRs. We also investigate how the use of photometric rather than spectroscopic redshifts, affects the accuracy of derived global SFRs. Finally, we provide a publicly available catalogue of SFRs for more than 27 million galaxies extracted from the Sloan Digital Sky Survey Data Release 7. The catalogue will be made available through the Vizier facility.

Key words: methods: data analysis – techniques: photometric – catalogues – galaxies: distances and redshifts – galaxies: photometry.

Star Formation Rates

SFR measures the amount of stars generated each year in a galaxy. It is proportional to the amount of light emitted by the galaxy.

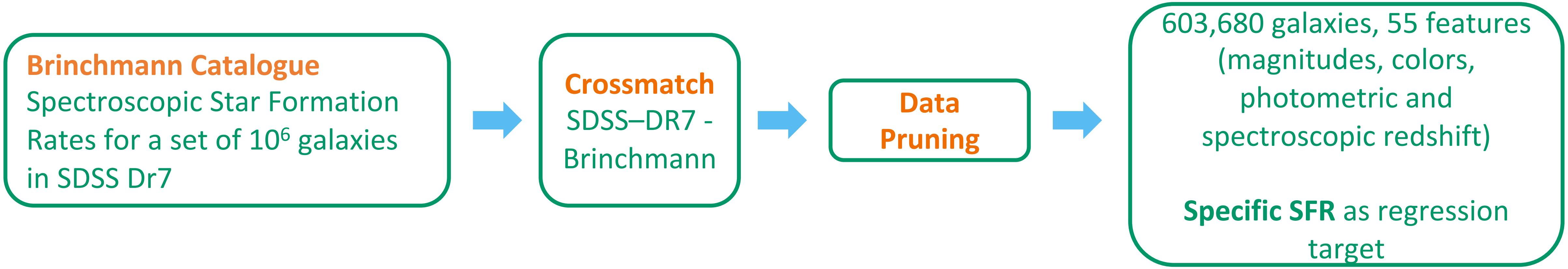
The light, depending on the frequency, probes a specific class of stars present in the galaxy.

Usually the measurement requires:

- Dust absorption calibration;
- IMF and metallicity assumption;
- Redshift correction;
- **Spectroscopic follow up.**

We have developed a ML approach to photometrically derive SFRs for a large subset of the SDSS - DR7.

Data



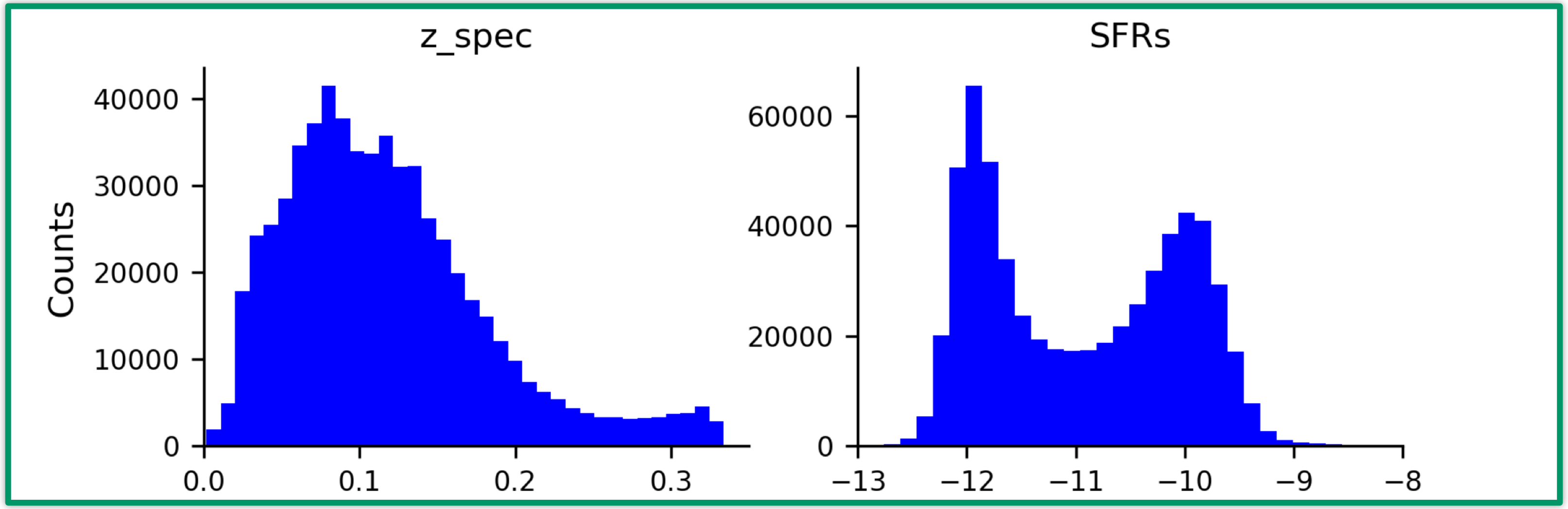
Brinchmann Catalogue
Spectroscopic Star Formation Rates for a set of 10^6 galaxies in SDSS Dr7

Crossmatch
SDSS–DR7 - Brinchmann

Data Pruning

603,680 galaxies, 55 features (magnitudes, colors, photometric and spectroscopic redshift)

Specific SFR as regression target



Methods

Random Forest

MLPQNA

PhiLAB

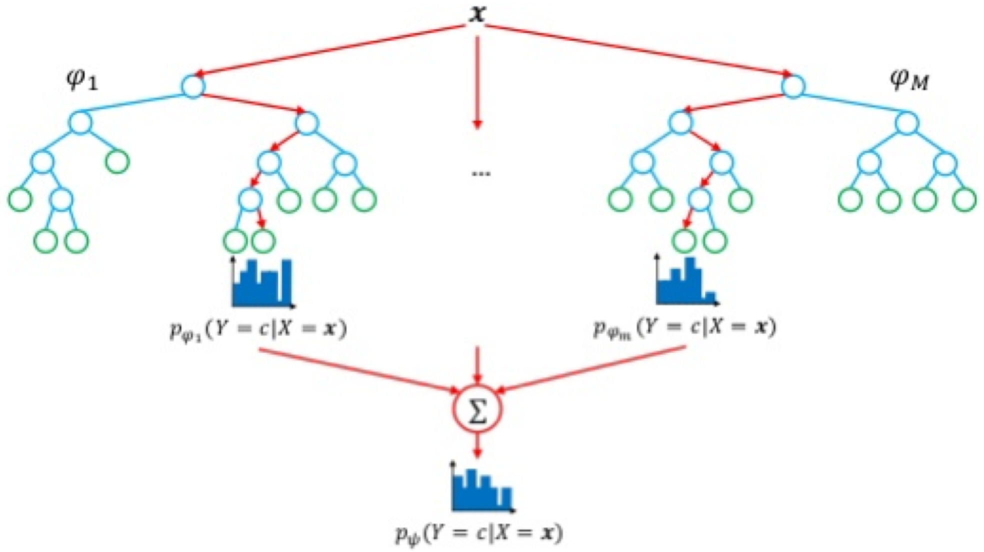
Methods

Random Forest

MLPQNA

PhiLAB

- **Random forest** (or **random forests**) is (are) an ensemble classifier that consists of many decision trees and outputs the class that is the mode of individual trees output.
- The method combines Breiman's "bagging" idea with the random selection of features
- It naturally provides a **Feature Importance Ranking**


$\boldsymbol{x}$
φ_1
φ_M
...
$p_{\varphi_1}(Y = c|X = \boldsymbol{x})$
$p_{\varphi_m}(Y = c|X = \boldsymbol{x})$
Σ
$p_\psi(Y = c|X = \boldsymbol{x})$

Randomization

- Bootstrap samples
- Random selection of $K \leqslant p$ split variables

} Random Forests

- Random selection of the threshold

} Extra-Trees

Methods

Random Forest

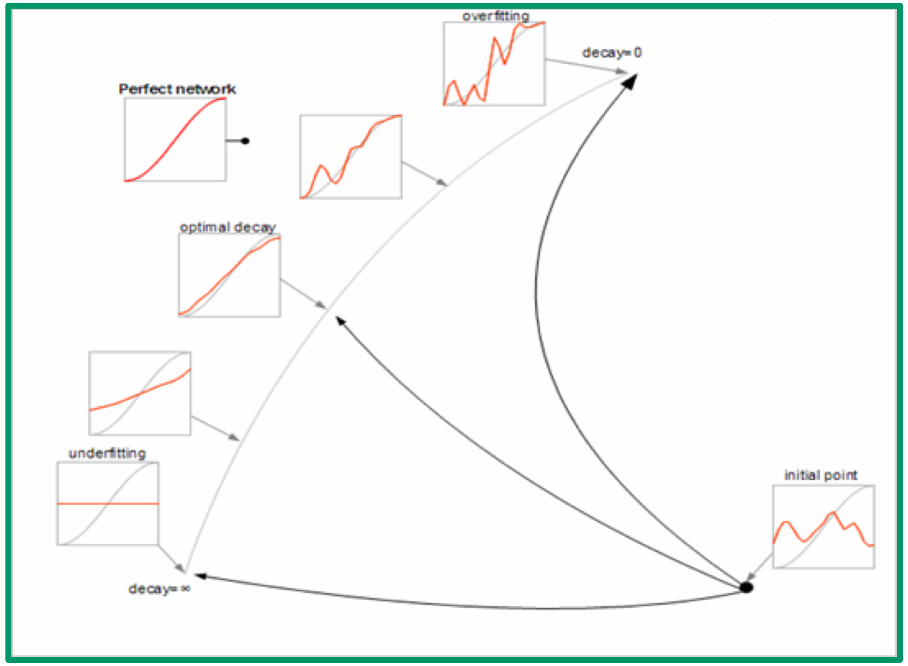
MLPQNA

PhiLAB

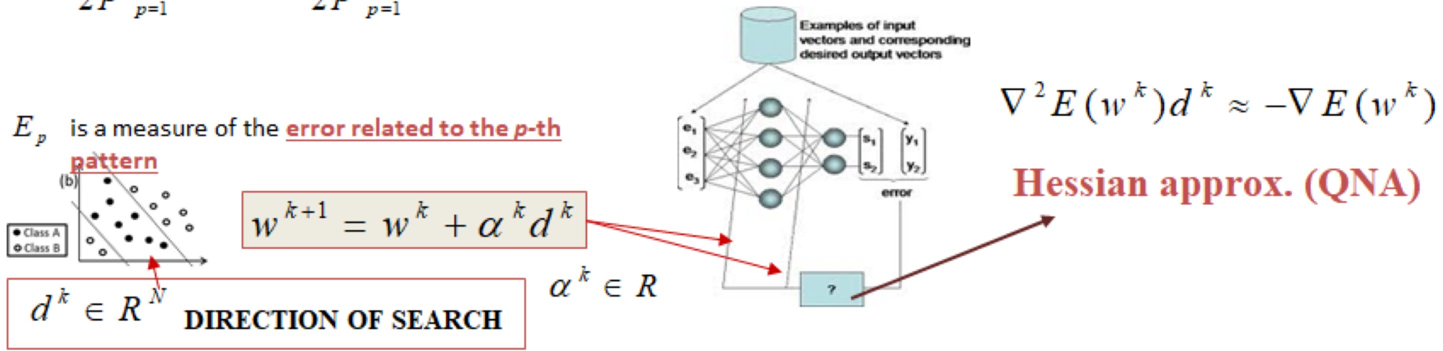
Artificial Neural Network:

- consists of simple, adaptive processing units, called neurons
- the neurons are interconnected, forming a large network
- parallel computation, often in layers
- nonlinearities are used in computations

MLPQNA is a traditional MLP that implements as training algorithm the Quasi Newton Approximation (QNA), Brescia et al. 2013



$$\min_{w} E(w) = \frac{1}{2P}\sum_{p=1}^{P} E_p(w) = \frac{1}{2P}\sum_{p=1}^{P}\left(y(x^p; w) - d^p\right)^2$$



(b)
Class A
Class B
Examples of input vectors and corresponding desired output vectors
e_1
e_2
e_3
s_1
s_2
y_1
y_2
error
?

E_p is a measure of the error related to the p-th pattern

$$w^{k+1} = w^k + \alpha^k d^k$$

$d^k \in R^N$ DIRECTION OF SEARCH

$\alpha^k \in R$

$$\nabla^2 E(w^k) d^k \approx -\nabla E(w^k)$$

Hessian approx. (QNA)

Methods

PhiLAB

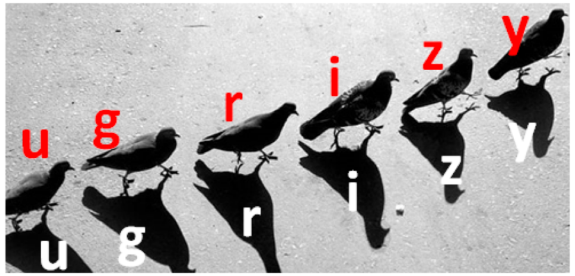
ϕLAB

Able to solve the All-relevant feature selection!

PHiLAB *(Parameter Handling investigation LABoratory)*

Based on two concepts: «**shadow features**» and **Naïve-LASSO regularization** and exploiting Random Forest model as importance computing engine.

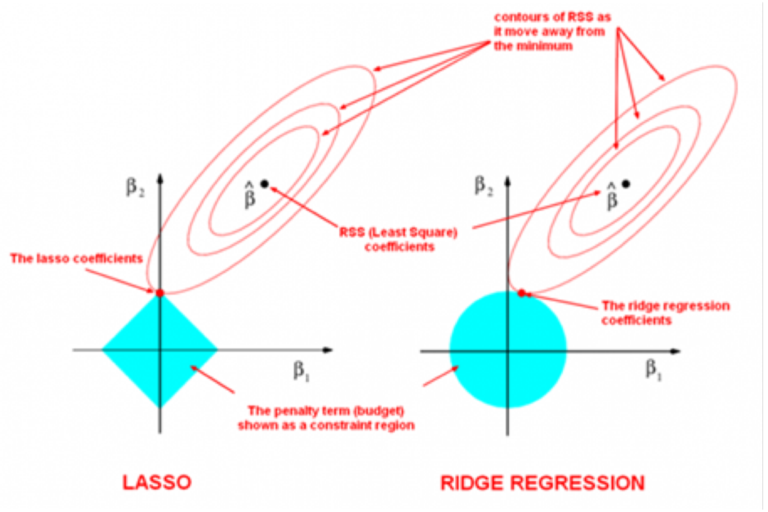
SHADOW FEATURES represent the noisy versions of the real ones and their calculated importance can be used to estimate the relevance of the real features.



A shadow feature for each real one is introduced by randomly shuffling its values among the N samples of the given dataset.

Kursa & Rudnicki 2010, Journal of Statistical Software, 36, 11

LASSO penalizes regression coefficients with an L_1-norm penalty, shrinking many of them to zero. Features with non-zero regression coefficients are "selected".


contours of RSS as it move away from the minimum
β_2
$\hat{\beta}$
β_1
RSS (Least Square) coefficients
The lasso coefficients
The ridge regression coefficients
The penalty term (budget) shown as a constraint region
LASSO
RIDGE REGRESSION

Hara & Maehara 2016, Proceedings of NIPS 2016, Barcelona, Spain

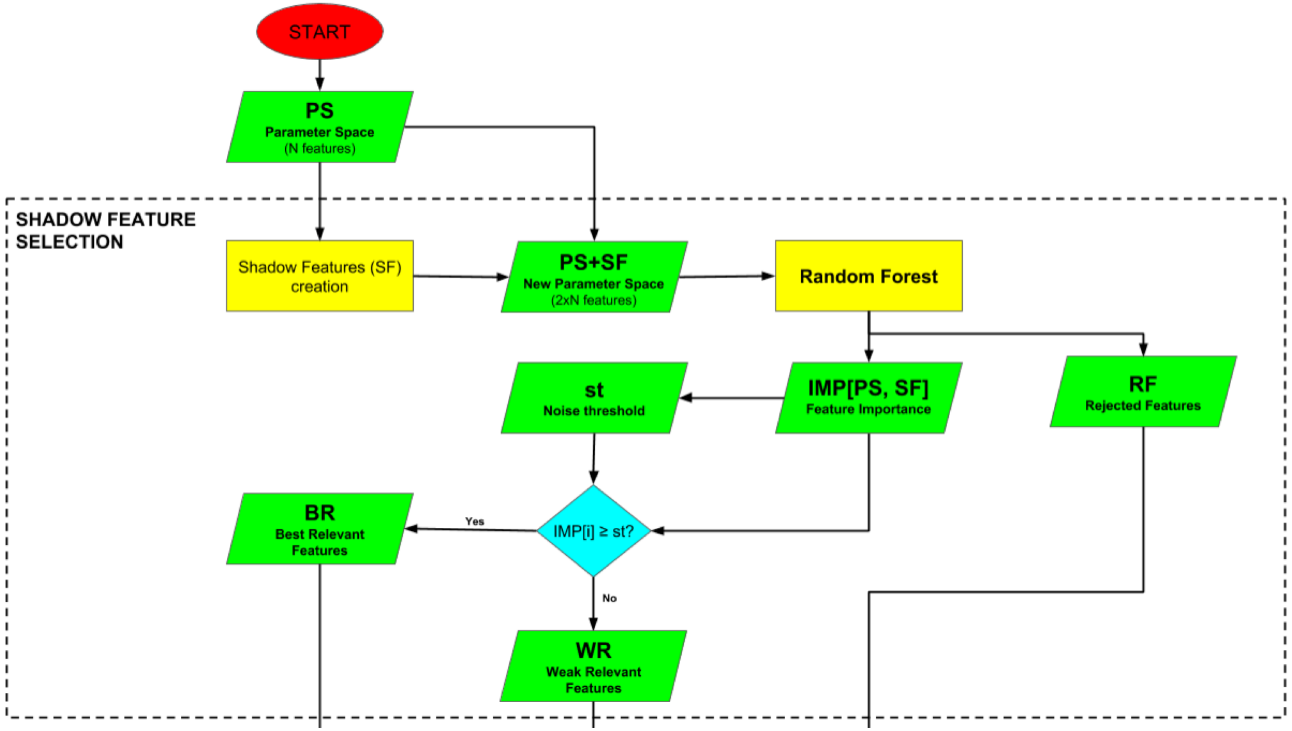
Why all-relevant feature selection is challenging?

Random accuracy fluctuation: the impact of random fluctuation in the prediction/classification accuracy of a learning system. Such effect, common in all real problems, may condition and mask the true importance contribution of a weakly relevant feature.
Does not affect the selection of strong relevant features;

Obscuration of weakly relevance: the detection of weakly relevant features can be completely obscured by the strongly relevant ones.

High-correlation compromise: in the frequent case of important features highly correlated, it is difficult to find the exact relevance contribution of single features. Shall we equally partition their importance and assign the same relevance?

Shadow features method is specialized to solve first issue, Naïve-LASSO the third issue, while both solve the second.



START

PS
Parameter Space
(N features)

SHADOW FEATURE SELECTION

Shadow Features (SF) creation

PS+SF
New Parameter Space
(2xN features)

Random Forest

st
Noise threshold

IMP[PS, SF]
Feature Importance

RF
Rejected Features

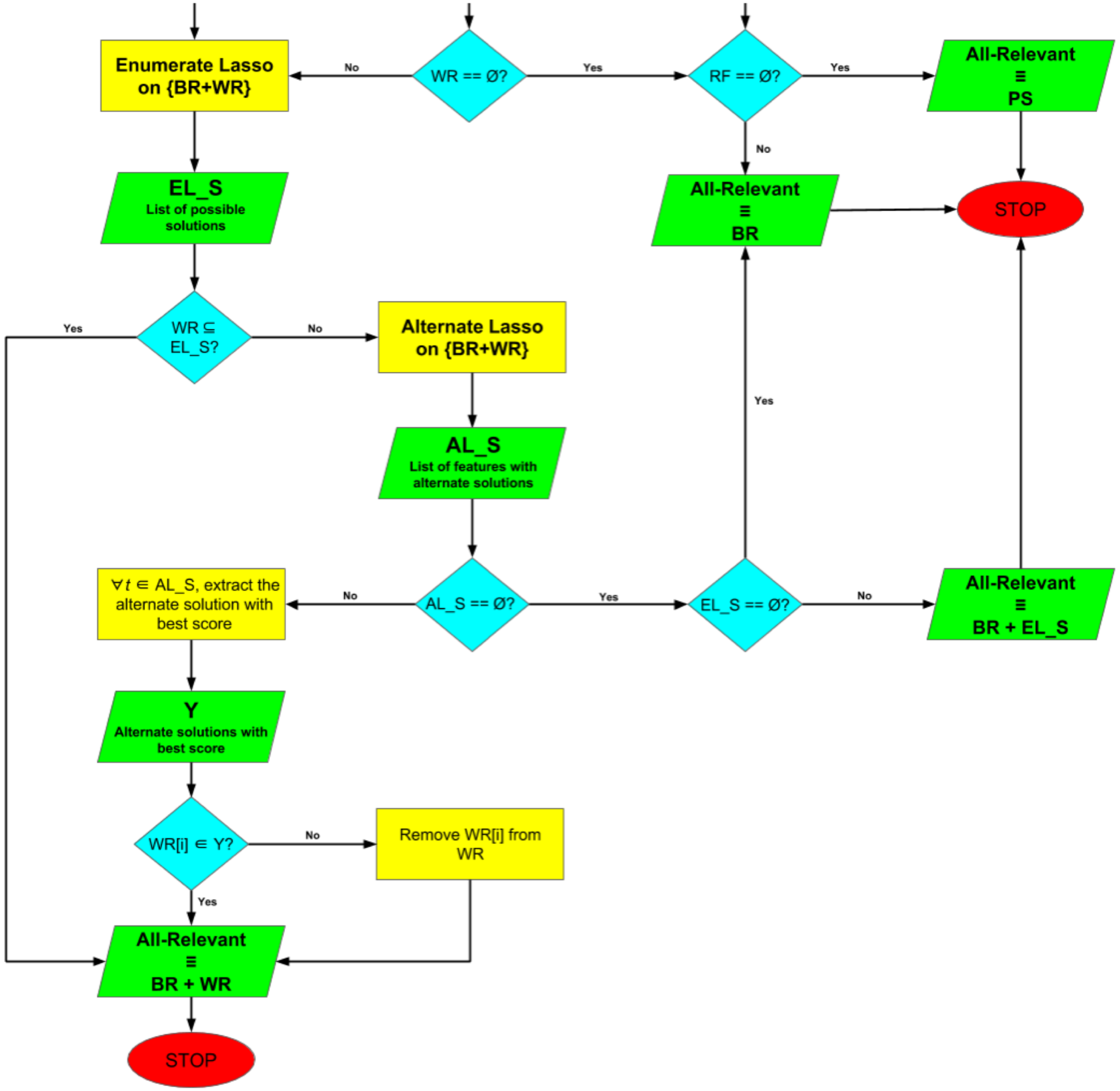
BR
Best Relevant Features

Yes

IMP[i] ≥ st?

No

WR
Weak Relevant Features



Enumerate Lasso on {BR+WR}

No

WR == Ø?

Yes

RF == Ø?

Yes

All-Relevant
≡
PS

No

EL_S
List of possible solutions

All-Relevant
≡
BR

STOP

Yes

WR ⊆ EL_S?

No

Alternate Lasso on {BR+WR}

Yes

AL_S
List of features with alternate solutions

∀*t* ∈ AL_S, extract the alternate solution with best score

No

AL_S == Ø?

Yes

EL_S == Ø?

No

All-Relevant
≡
BR + EL_S

Y
Alternate solutions with best score

WR[i] ∈ Y?

No

Remove WR[i] from WR

Yes

All-Relevant
≡
BR + WR

STOP

ΦLAB voting algorithm

Experiments

K-fold cv

Completeness

Feature selection

RF vs Mlpqna

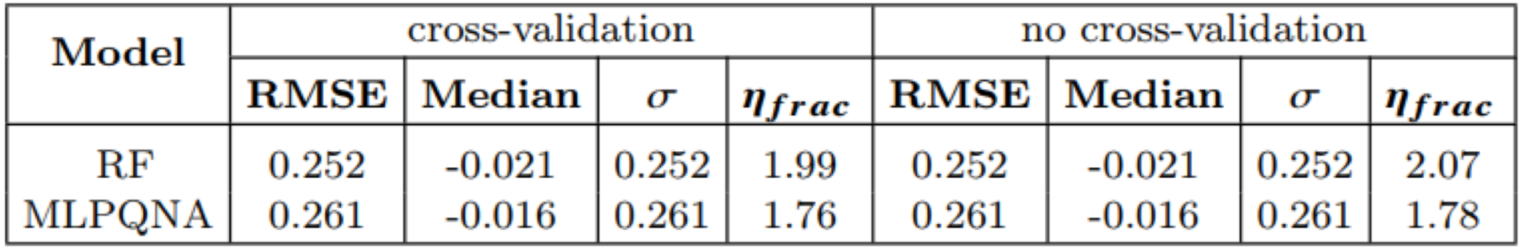
Experiments

K-fold cv

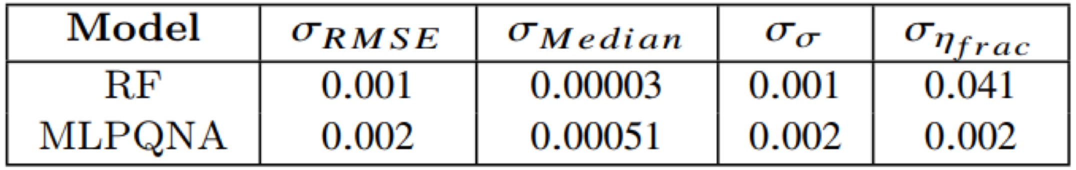
Completeness

Feature selection

RF vs Mlpqna



Model	cross-validation				no cross-validation			
	RMSE	**Median**	σ	η_{frac}	**RMSE**	**Median**	σ	η_{frac}
RF	0.252	-0.021	0.252	1.99	0.252	-0.021	0.252	2.07
MLPQNA	0.261	-0.016	0.261	1.76	0.261	-0.016	0.261	1.78



Model	σ_{RMSE}	σ_{Median}	σ_{σ}	$\sigma_{\eta_{frac}}$
RF	0.001	0.00003	0.001	0.041
MLPQNA	0.002	0.00051	0.002	0.002

No need
Longer training

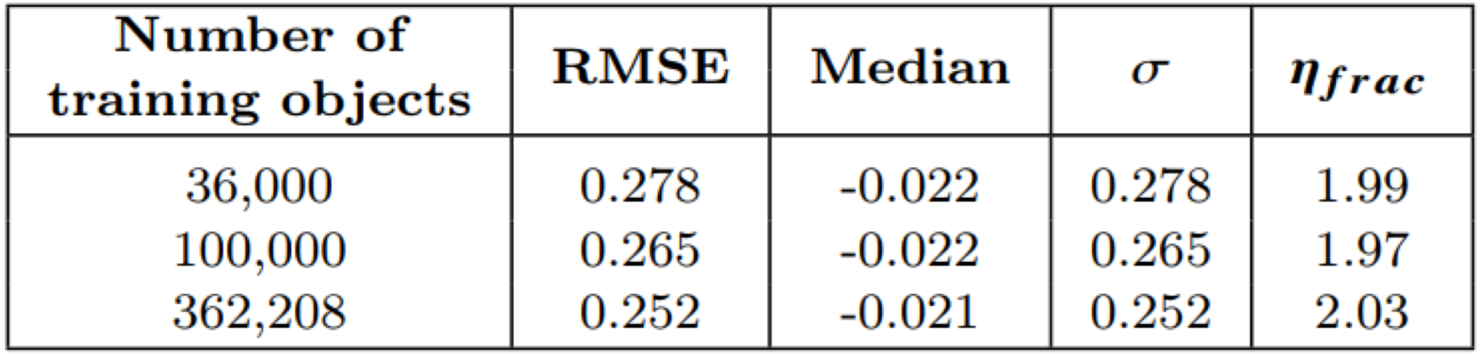
Experiments

K-fold cv

Completeness

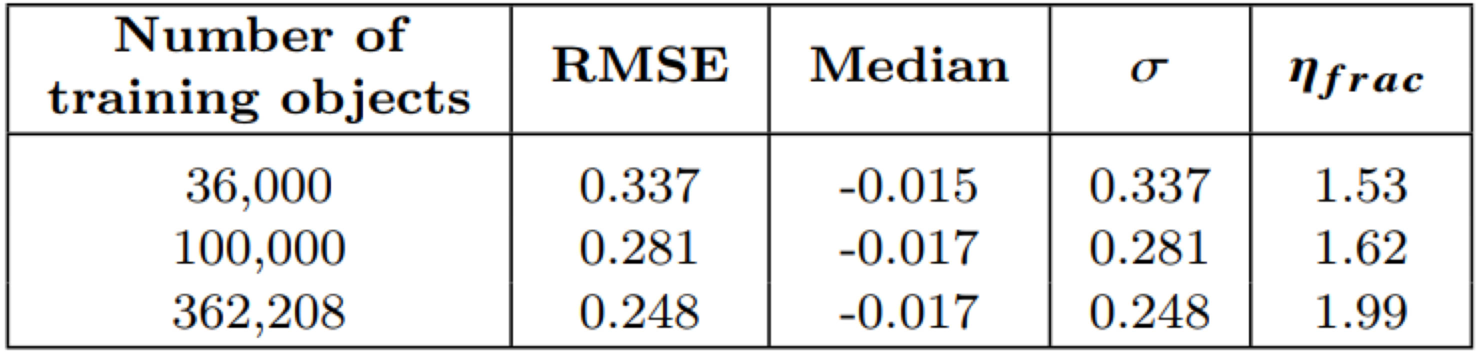
Feature selection

RF vs Mlpqna



Number of training objects	RMSE	Median	σ	η_{frac}
36,000	0.278	-0.022	0.278	1.99
100,000	0.265	-0.022	0.265	1.97
362,208	0.252	-0.021	0.252	2.03

RF



Number of training objects	RMSE	Median	σ	η_{frac}
36,000	0.337	-0.015	0.337	1.53
100,000	0.281	-0.017	0.281	1.62
362,208	0.248	-0.017	0.248	1.99

MLPQNA

Problem not saturated
Need for more samples

Experiments

- K-fold cv
- Completeness
- **Feature selection**
- RF vs Mlpqna

ϕLAB

PHiLAB *(Parameter Handling investigation LABoratory)*

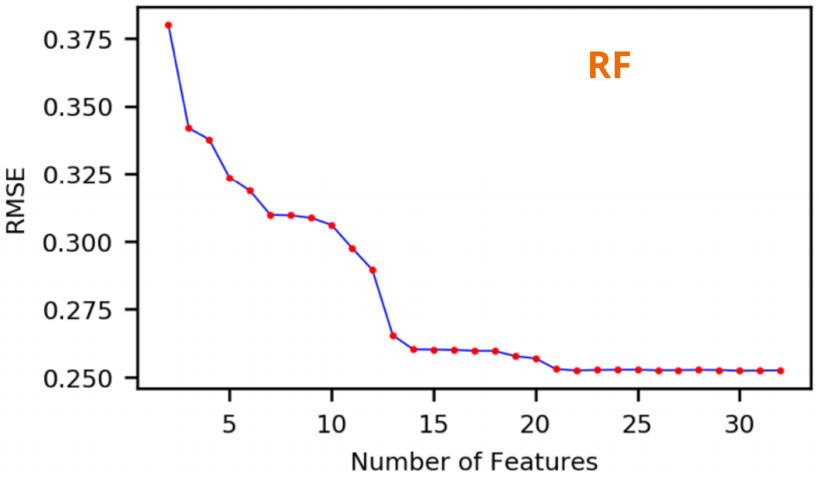
Able to solve the All-relevant feature selection!

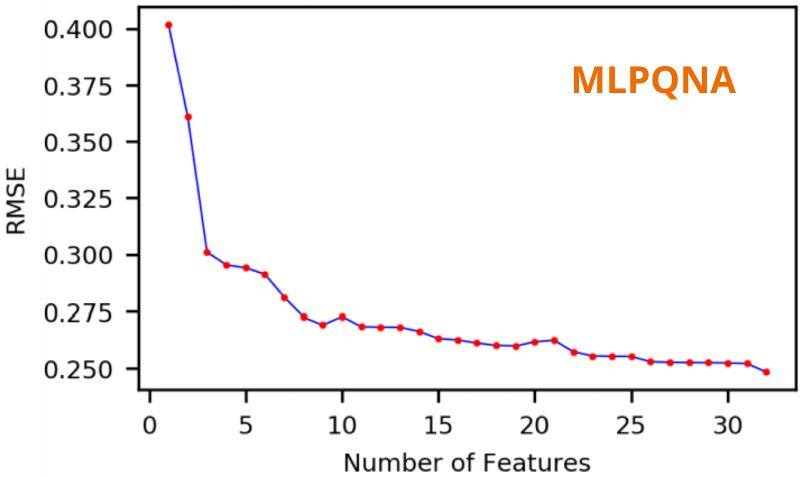
Experiments

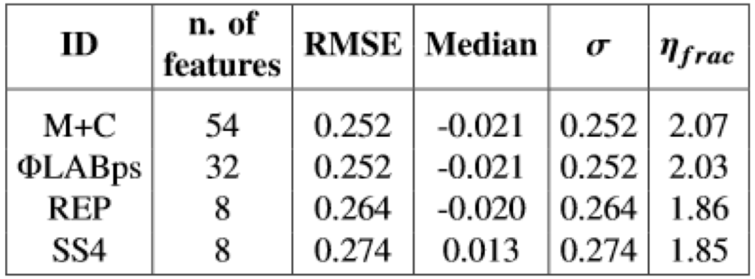
K-fold cv

Completeness

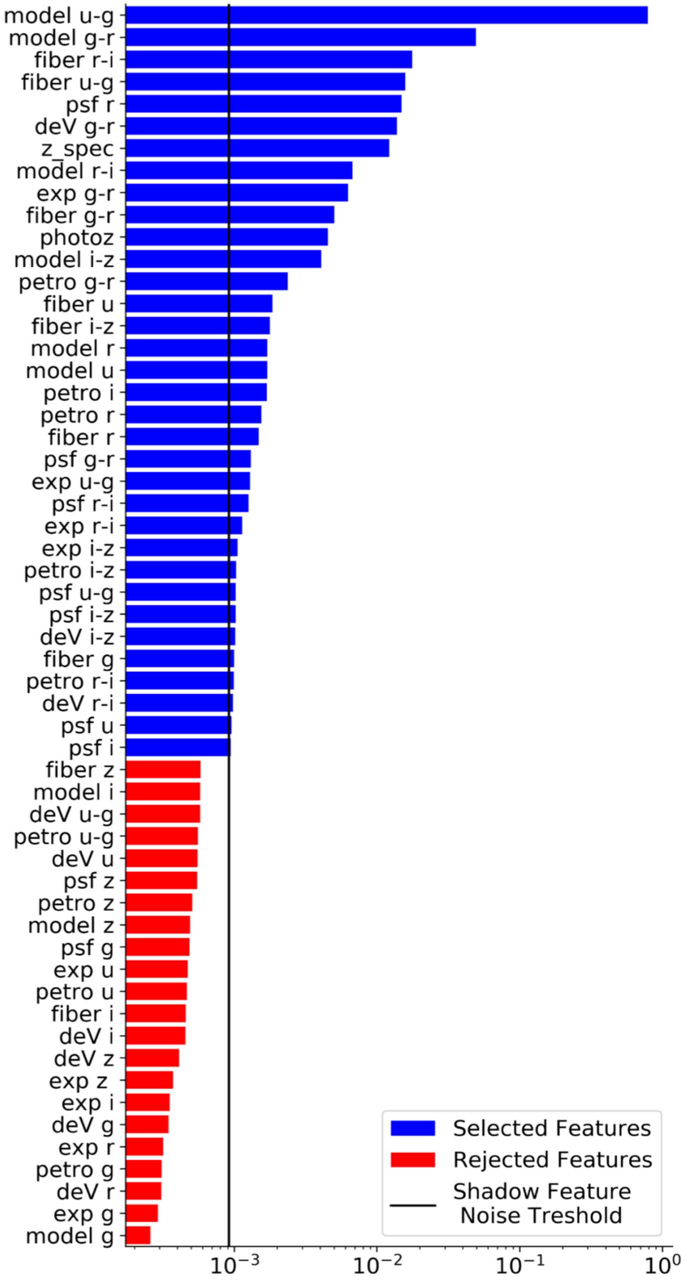
Feature selection



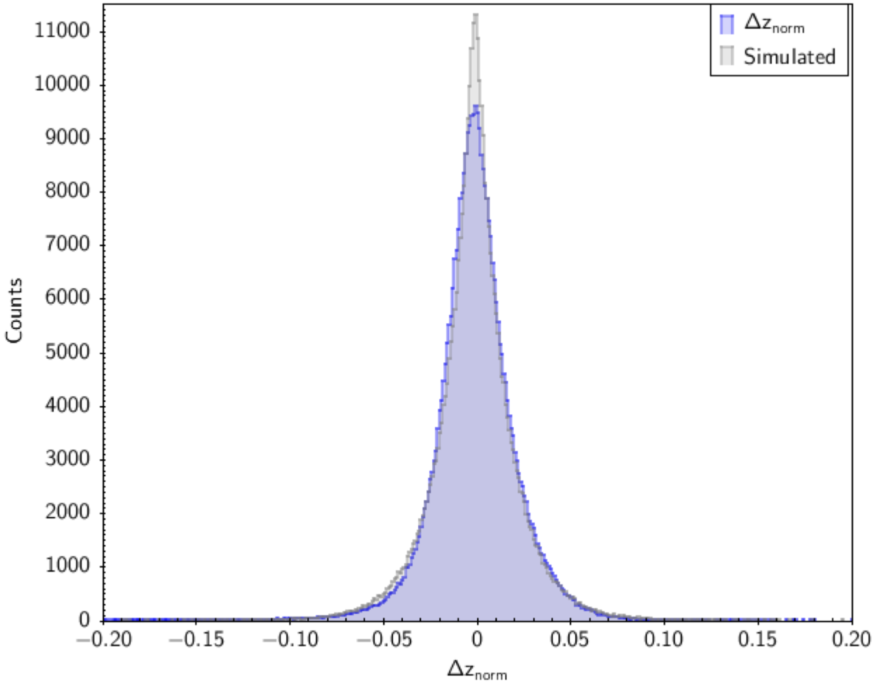




ID	n. of features	RMSE	Median	σ	η_{frac}
M+C	54	0.252	-0.021	0.252	2.07
ΦLABps	32	0.252	-0.021	0.252	2.03
REP	8	0.264	-0.020	0.264	1.86
SS4	8	0.274	0.013	0.274	1.85

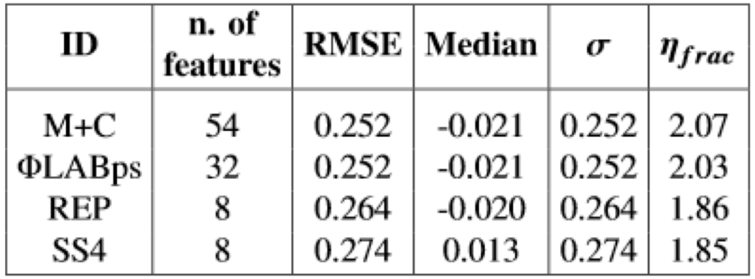


Experiments

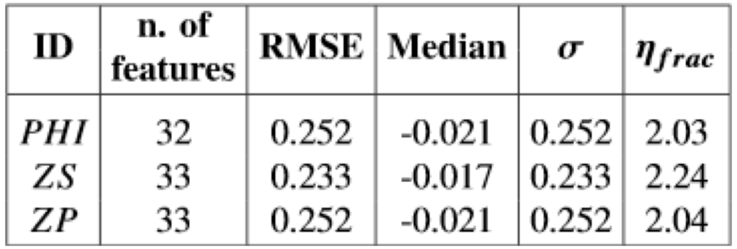


1. Distribution that best fit the Δz_norm distribution (**Kolmogorov-Smirnov Test**);
2. Simulated several distribution with increasing level of photoz measurement accuracy.

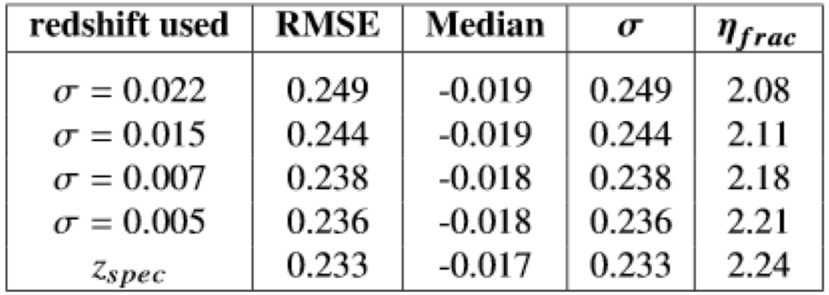
Feature selection



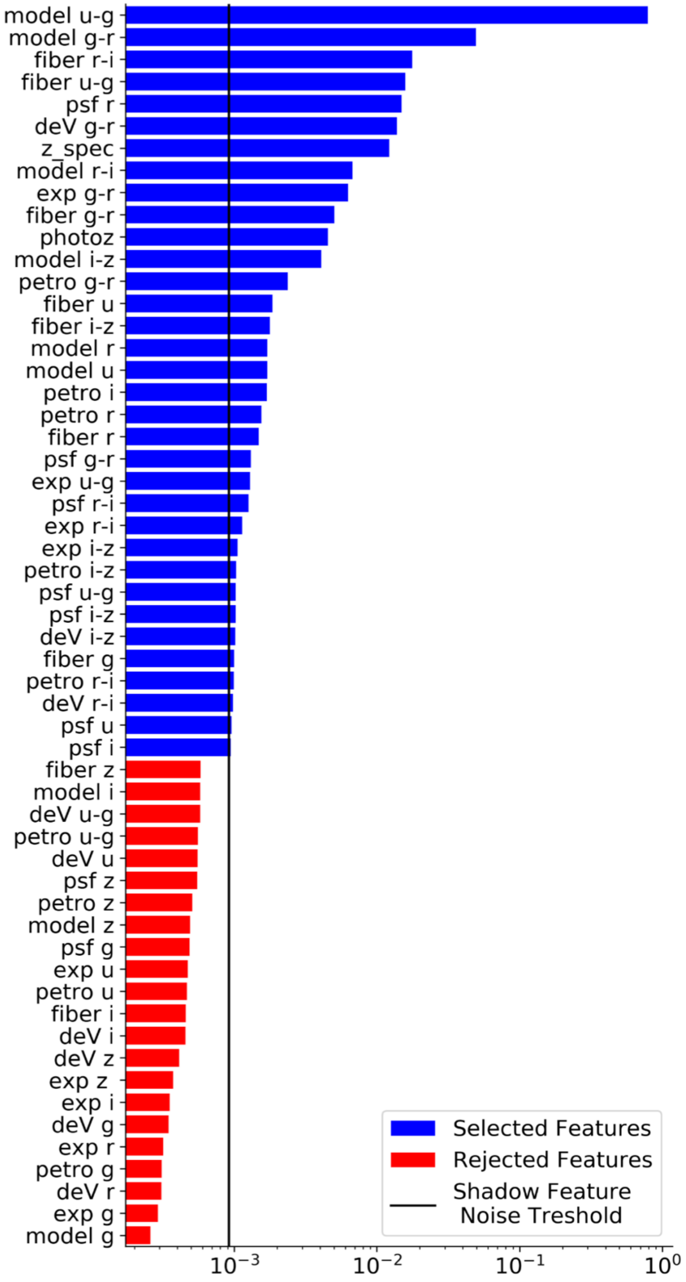
ID	n. of features	RMSE	Median	σ	η_{frac}
M+C	54	0.252	-0.021	0.252	2.07
ΦLABps	32	0.252	-0.021	0.252	2.03
REP	8	0.264	-0.020	0.264	1.86
SS4	8	0.274	0.013	0.274	1.85



ID	n. of features	RMSE	Median	σ	η_{frac}
PHI	32	0.252	-0.021	0.252	2.03
ZS	33	0.233	-0.017	0.233	2.24
ZP	33	0.252	-0.021	0.252	2.04



redshift used	RMSE	Median	σ	η_{frac}
$\sigma = 0.022$	0.249	-0.019	0.249	2.08
$\sigma = 0.015$	0.244	-0.019	0.244	2.11
$\sigma = 0.007$	0.238	-0.018	0.238	2.18
$\sigma = 0.005$	0.236	-0.018	0.236	2.21
z_{spec}	0.233	-0.017	0.233	2.24



Experiments

K-fold cv

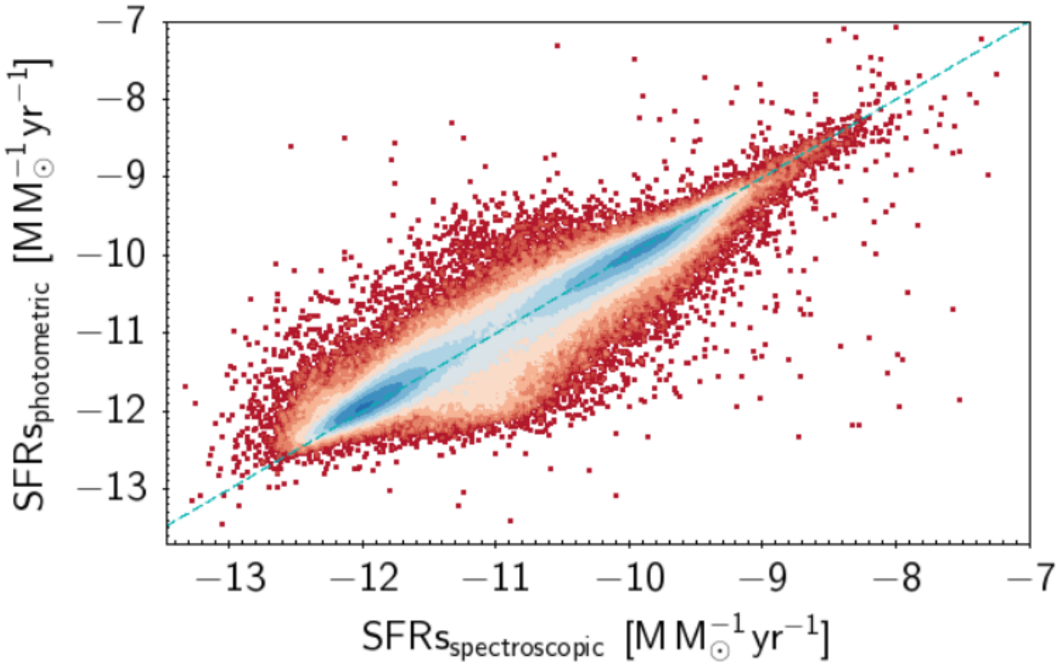
Completeness

Feature selection

RF vs Mlpqna

Model	**RMSE**	**Median**	**σ**	**η_{frac}**
RF	0.252	-0.021	0.252	2.03%
MLPQNA	0.248	-0.017	0.248	1.99%
Stensbo-Smidt et al. 2016	0.274	0.013	0.274	1.85%

Improvement of the literature results

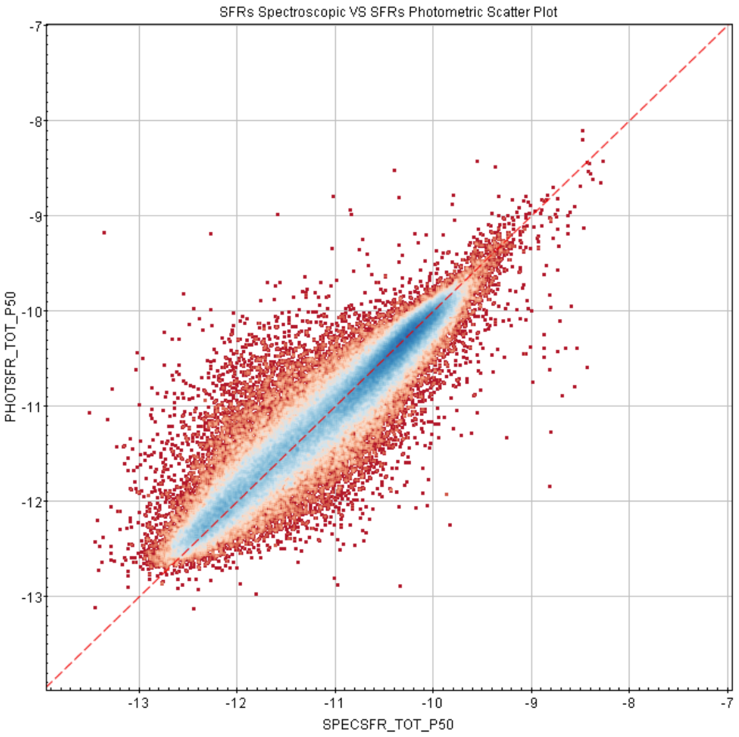


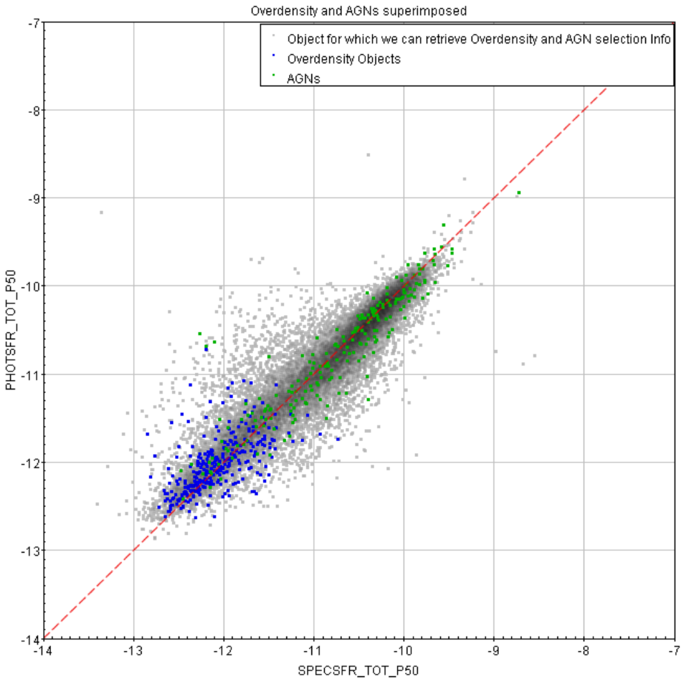
Work in progress

- Filtered the KB for objects on which our BPT classification matched Brinchmann's;
- Added WISE colors and magnitudes;
- Accounted for reddening;
- DR7 -> DR9 with updated SFRs.

KB shrinking from 603,680 galaxies to 196,652

Run	RMSE	Median	η
Old	0.248	-0.017	1.99%
New	0.238	0.003	1.95%





Thank You for the Attention

From this work we built a catalogue of photometric SFRs for 27 million of galaxies available on Vizier through the following link: ftp://cdsarc.u-strasbg.fr/pub/cats/J/MNRAS/486/1377/

Feature Selection with ΦLAB

What's behind the ***ΦLAB*** (*Parameter Handling investigation Laboratory)* project?....the property of **feature importance and relevance** in the context of a parameter space used to approach any prediction/classification task with machine learning methodology.

The **importance** of a feature is the relevance of its informative contribution to the solution of a learning problem.
The **relevance** of a feature can be formally defined as follows:

- Feature x is **strongly relevant** when removal of x from the parameter space always results in degradation of learning accuracy
- Feature x is **weakly relevant** if is not strongly relevant and there exists at least one subset S of features such that learning accuracy on S is worse than S U {x}
- Feature x is **irrelevant** if it is neither strongly nor weakly relevant.

feature selection problem taxonomy:
Minimal-optimal feature selection: selection of the smallest parameter space giving best accuracy. There are plenty of methods proposed in literature, either for prediction and classification problems (PCA, leave-one-out, forward selection, backward elimination, RF, PPS, Naive-Bayes, etc.).
All-relevant feature selection: the identification of the exact parameter space (all features) which are in some circumstances relevant for the problem solution. Basically, finding all relevant features, instead of only the non-redundant or unuseful ones, may help to understand the hidden mechanisms behind the problem. In more philosophical terms, it makes a predictive/classification model as a gray box, instead of merely as a black box!
There are very few methods proposed in literature to solve this type of feature selection.

Methods

ϕLAB

***PHiLAB** (Parameter Handling investigation LABoratory)*

PhiLAB

Able to solve the All-relevant feature selection!

We include two naive LASSO technique in PhiLAB:

1. A-LASSO: creates a list of features alternate to those selected by the standard LASSO, associating to each feature a score reflecting the performance degradation from the optimal solution; PhiLAB selects only the features that achieve the lowest score from the optimal solution.

Trade-off between feature selection performance and flexibility in the analysis of the parameter space. Degrading the score solution to have more flexibility.

2. E-LASSO: enumerates a series of of different feature subsets. The optimal solution of a mathematical model is not always the best solution to the physical problem.

Chance to obtain a better solution to the physical problem

Kursa & Rudnicki 2010, Journal of Statistical Software, 36, 11

Hara & Maehara 2016, Proceedings of NIPS 2016, Barcelona, Spain

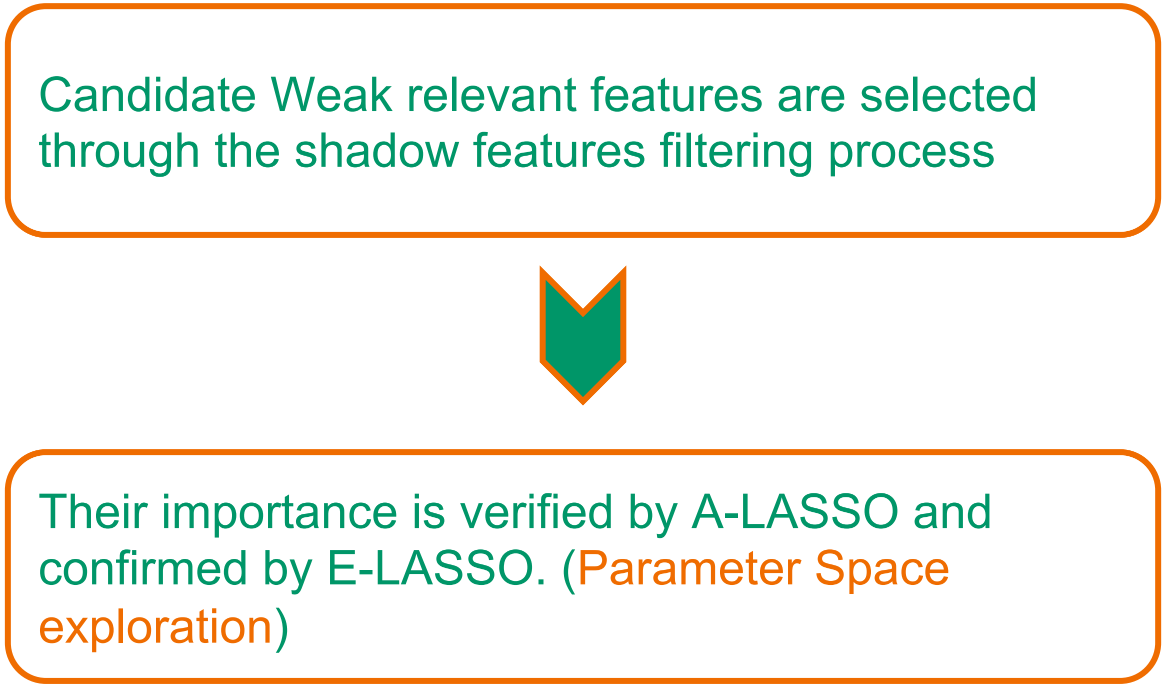
Methods

ϕLAB

***PHiLAB** (Parameter Handling investigation LABoratory)*

PhiLAB

Able to solve the All-relevant feature selection!



Candidate Weak relevant features are selected through the shadow features filtering process

Their importance is verified by A-LASSO and confirmed by E-LASSO. (Parameter Space exploration)

Kursa & Rudnicki 2010, Journal of Statistical Software, 36, 11

Hara & Maehara 2016, Proceedings of NIPS 2016, Barcelona, Spain

ΦLAB voting algorithm

0. Let it be PS={x_1…x_N} the initial complete Parameter Space composed by N real features;

1. Apply the Shadow Feature Selection (**SFS method**) and produce the following items:
   - ➢ **SF**={x_s_1…x_s_N}, the list of shadow features, obtained by randomly shuffling the values of real features;
   - ➢ **IMP[PS, SF]** for each x ϵ PS & for each x_s ϵ SF, the importance list of all 2N features, original and shadows;
   - ➢ **st**: noise threshold, defined as the **max{IMP[SF],** for each **x_s ϵ SF}**;
   - ➢ **BR**={x ϵ PS t.c. IMP[x] ≥ st}, the set of best relevant real features;
   - ➢ **RF**={x ϵ PS, rejected by the Shadow Feature Selection}, the set of excluded real features, i.e. not relevant;
   - ➢ **WR**={x ϵ PS t.c. IMP[x] < st}, the set of weak relevant real features;

2. From the previous step, it resulted that PS ≡ {BR+WR+RF}. Now we consider the **PS_{red}= {BR+WR}**, by excluding the rejected features. In principle it may correspond to the original PS, in case of no rejections from the SFS;
   a) If RF==ø && WR==ø, the SFS method confirmed all real features as high relevant, therefore return **ALL-RELEVANT(PS)**, i.e. the full PS, as the optimized parameter space and **EXIT**.
   b) If RF≠ø && WR==ø, the SFS method rejected some features and confirmed others as high relevant, therefore return **ALL-RELEVANT(BR)** as the optimized parameter space and **EXIT**.
   c) If WR≠ø, regardless some rejections, SFS confirmed the presence of some weak relevant features that must be evaluated by LASSO methods, therefore goto 3;

ΦLAB voting algorithm

3. Given PS_{red}= {BR+WR}, the set of candidate features, apply **E-LASSO method**. It produces:
   - ➢ **EL_S**, a list of M subsets of features, considered as possible solutions, ordered by decreasing score;
   - a) If WR ⊆ EL_S, then all weak relevant features are possible solutions, therefore return **ALL-RELEVANT(BR+WR)** as the optimized parameter space and **EXIT.**
   - b) Else goto 4;
4. Given PS_{red}= {BR+WR}, the set of candidate features, apply **A-LASSO method**. It produces:
   - ➢ **AL_S**, a set of T features, each one with a list of features List(t) considered as alternate solutions with a certain score;
   - a) if AL_S ==ø then no alternate solutions exist, therefore:
     - i. If EL_S==ø then return **ALL-RELEVANT(BR)** as the optimized parameter space and **EXIT.**
     - ii. Else if EL_S≠ø then return **ALL-RELEVANT(BR+EL_S)** as the optimized parameter space and **EXIT.**
   - b) Else extract for each t ϵ T the alternate solution xas, t.c. *Score(xas) = min{Score(y), ∀ y ϵ List(t)};*
   - c) goto 5.
4. For each x ϵ WR:
   - a) If x is alternate solution of at least one feature t ϵ T, t.c. [t ϵ BR || t ϵ EL_S], then retain x within WR set;
   - b) Else reject x (by removing x from WR);
5. Return **ALL-RELEVANT(BR+WR)** as the final optimized parameter space and **EXIT.**