D. Fraix-Burnet
Univ. Grenoble Alpes / CNRS / IPAG

This file is my small contribution to Chapters 3 and 4:

- *Family Traits of Galaxies: From the Tuning Fork to a Physical Classification in a Multi-Wavelength Context*, Rampazzo et al, pp. 189-242
- *The Anatomy of Galaxies*, D'Onofrio et al, pp.243-380

of the book:

From the Realm of the Nebulae to the Society of Galaxies
Dialogues on a Century of Research
D'Onofrio, M.; Rampazzo, R. & Zaggia, S. (Eds.)
Astrophysics and Space Science Library, Vol 45
Springer International Publishing, 2016.

e-mail: didier.fraix-burnet@univ-grenoble-alpes.fr

in Chapter 3

- Dear Didier (*Fraix-Burnet*),

you attempted a phylogenetic approach, called astro-cladistic, to galaxy classification. In the opinion of Sandage "A good classification can drive the physics, but the physics must not be used to drive the classification. Otherwise the process becomes circular." (A. Sandage, ARA&A, 2005, 43, 581). Would you express your opinion about the role of morphological studies in extragalactic astronomy?

It is interesting to recall that the debate about the usefulness of the morphological classification is a rather old one. There was a conference in 1990 ([1]) that confronted the morphological vs a physical classification of galaxies. The debate is still alive and Sandage in 2005 affirms his preference for the morphological classification and rightly noticed that the Hubble classification and the Hubble tuning fork have not yet been replaced by anything else some fifteen years after this conference. So does the proponents for a physical classification have lost the battle? Is the morphological classification the unique and most powerful classification of galaxies?

To my point of view, all parties are right... and wrong. On one side, Sandage is perfectly correct when he says that classification should be driven by the data. Naturalists and biologists have known that for more than 250 years now. They have invented the science of classification (systematics) and have given birth to a branch of statistics that is so successful in several disciplines. On the other side, the proponents of a physical classification are right for fundamental reasons (see for instance [2]): would it be possible that one property alone could depict the diversity of billions of complex systems like galaxies that have evolved during billions of years? Could it be possible that the classification of galaxies, ensembles of billions of stars, could be simpler than the stellar classification? Would it be possible that one century of observations, technological developments, and so many discoveries about our Universe, have not yielded a modified view of the galaxy diversity? Why should we stick to the visible morphological parameter only while we have at our disposal morphologies from X-rays to radio wavelengths, spectra, chemical compositions, stellar populations, central black hole masses, kinematics of stars and gas...? And if really the morphology is well correlated to other properties, why not use one of them that is quantitative, objective, simpler to observe even automatically, to build an equivalent classification?

So our modern physical understanding of galaxies clearly asks for a renewal of the classification of galaxies, a classification that is not based on morphology only and must be driven by the data.

I would add that a classification inevitably drives the physics, and not simply *can* as Sandage writes it. Indeed, there is not necessarily one unique immutable classification. There can be several classifications, each one having a particular purpose and changing according to the discoveries. In this sense a classification can be seen as "driven" (justified) by some physical question. For instance, you may want to classify the entire diversity of observed galaxies, or the different morphologies of galaxies in the visible or in the X-ray, or the different kinds of AGNs, like you want to classify supernovae in the particular goal of probing the shape of our Universe. All of these classifications are perfectly legitimate provided that they are performed correctly and not used for other purposes. The morphological classification of galaxies will remain a classification of morphologies, nothing else. To classify morphologies, you have to use morphological parameters to build the classification. This will necessarily influence the physics you can do using this classification since everything will be related to the morphological parameters.

In conclusion, the physical goal drives the choice of the data, this choice drives the classification which in turn drives the physics you will infer from it. The most dangerous circularity is when you select your parameters from a priori arguments, like because you think they are the most important for instance. Then you end up with a classification that reflects you inputs, and you make no new physical inference.

So why despite after all these years and several attempts, the debate is still alive and no alternative to the Hubble classification and the Hubble sequence have yet emerged? To me, there are two reasons, apart from the simplicity and even the beauty of this traditional scheme.

The first reason is that I am convinced that the debate is based on a misunderstanding of the full classification process. I have noticed also that throughout the literature there is a frequent confusion between the Hubble classification and the Hubble tuning fork.

The second reason is the fantastic difficulty of the task. Each aspect of the classification process must be mastered and requires a great expertise. In addition, apart from a somewhat technical part present at each stage, a new global classification must prove its usefulness, pertinence, generality and completeness from an astrophysical point of view. The difficulty here is that physicists are not used to compare models to multivariate statistical results. So even if the technical difficulties were lifted, the way the astrophysical interpretation is made must evolve.

Let me clarify the full classification process, since this is an essential point in the debate. There are several distinct aspects to build a proper classification, each one being a different topic of statistics:

1. clustering (unsupervised classification): this is the statistical gathering of objects sharing some similarity (with the unique parameter of global morphology, this could be: elliptical, spiral, irregular);
2. taxonomy: this consists in characterising the classes with a full description, and giving them some name (the labels, which are obvious with one parameter like in the Hubble classification);

3. relationships: to understand the origin of the classes, one has to discover the relationships between the classes (this is the role of the Hubble tuning fork diagram);
4. classification (supervised classification): in statistics, this consists in putting new objects into previously established classes. But in more general language, this is the four aspects altogether, which adds some confusion.

The first aspect is the most important one, and depends much on the sample and the parameters. If you want to make a morphological classification, then select objectively the parameters that describe the morphology. When I say objectively, I mean three things: i) all parameters, not only the easiest to obtain, ii) all of them and not the one you think are more important, iii) those that avoid redundancies and too much disturbance (use statistical tools to assess this). Then you must use some of the many clustering techniques developed in statistics. They all have their particularities and limitations, so that they should be seen as exploring tools to navigate inside a multidimensional parameter space.

The second aspect, taxonomy, is very important because there are some strict rules. In particular, when you are multivariate, the labels (names) of the classes should not be related to any particular parameters. For example, a class of galaxies which are blue and elliptical should not be called "blue" or "elliptical" or "blue elliptical". This rule has made the overwhelming success of the Linné's nomenclature in biology for about three centuries. Otherwise, you over-interpret the classification by putting forward a particular property and miss most of the physics.

Let me take an example.The origin of lenticular galaxies (SOs) is not clear [3]: are they formed as such or are they spiral galaxies stripped of their interstellar medium? This formulation of the problem implicitly assumes that the SOs form an homologous class. This obviously cannot be true since at least two origins are physically plausible, making at least two theoretical classes. From a physical point of view, the good question is not about the origin of lenticular *galaxies*, but about the origin of the lenticular *morphology*. Other parameters are required to distinguish the two classes. The (multivariate) classification should give them two different names unrelated to any property, and unambiguously distinguish them in the data if they both exist in reality.

The third aspect, relationships, is useful to understand the origin of the classes. When Hubble established his classification, he tried to understand the physical drivers that generated the observed diversity. He proposed evolution (of the morphology) as the main factor and devised a scheme known as the Hubble tuning fork. This is not an observation, this is an interpretation at the origin of the terms "early-type" and "late-type" galaxies. This is an explanation of his morphological classification. In this approach (that I call clustering approach, see below), the relationships are outside the classification process, they are established afterwards, if required.

Phylogenetic approaches, the subject of astrocladistics, build the relationships directly from the data, and define classes based on the structure of these relationships, taxonomy coming next. Cladistics, also called Maximum Parsimony, is the most general of these approaches. For instance, with the two dichotomous parameters "elliptical" (yes or not) and "bar" (present or absent), a Maximum Parsimony analysis yields a tree that is exactly the Hubble tuning fork. This is obtained without any a priori classification or physical assumption, contrarily to what Hubble did. Not only the relationships are entirely integrated into the classification process, but it is the necessary first step. The only assumption behind Maximum Parsimony is that your objects can be related by some discrete or continuous link, that you will have to interpret at the end. I have demonstrated that this link is not necessarily real or physical, it simply shows the simplest path to follow if you could transform an object into another, even in a thought experiment.

I think that the proponents of a physical classification want to replace the Hubble tuning fork by a new diagram drawn in a space of physical parameters [4, 5]. But these are not new classifications since the classification and the corresponding taxonomy are still those of Hubble. They are looking for the relationships between the Hubble classes (aspect 3), and are thus entirely driven by the morphological classification. This is perfectly correct if one is interested by the morphology of galaxies and its correlation or explanation with other properties. But in no case does it provide a general picture of the galaxy diversity and its origin.

At the end, the classification itself (aspect 4) is necessary to order new observations and to detect new kinds of objects when they cannot be put into the known classes. This is particularly crucial in the case of huge flows of data when the detection of deviant objects must be made automatically, if possible in real time (e.g. [6]). You know that morphology is difficult to determine automatically. This is why some studies have tackled the problem from the start with multivariate data sets. However, since they want to recover the Hubble classification with clustering methods using multivariate objective and quantitative parameters (aspect 1), this results in a entangled process. Let me show why.

On Fig. 1, I depict the different approaches to classification: the Hubble method, the current multivariate studies and the complete process that I have been working on for the last thirteen year.

Hubble proceeded correctly through a clustering method based on similarities, then devised the taxonomy to establish his classification (upper left panel of Fig. 1). It was based on the morphology only, and nowadays many studies propose a monovariate classification based on a particular property of galaxies, even though they are generally dichotomous: radio vs non-radio, infrared, ultraviolet or X-ray galaxies, field vs cluster objects, star-forming... Even it none of these monovariate classifications has the reputation of the Hubble one, it clearly shows that morphology is not enough to understand the physics of galaxies. At the end, there are many monovariate classifications, not necessarily compatible with each other (a galaxy belongs to several classes, some classifications do not include some galaxies). This is startlingly similar to the situation of biology in the Middle Age, that fostered the invention of a new nomenclature by Linné and of the multivariate analysis by Adanson in the eighteenth century.

How to make this classification method evolve in order to become multivariate and thus more representative of the real complexity of galaxies? The panel to the upper right of Fig. 1 shows my interpretation of the attempts done in this direction, mostly during the last ten years or so, by most of the multivariate studies, and the proponents of a physical classification as well. Due to the enormous reputation of the Hubble classification, it occupies a central role in these studies, This leads to the paradox that the taxonomy and classification are still those from Hubble, even though the parameters and the clustering methods are different. Just recall the example of the lenticular morphology confused with a class. In this diagram, the circular process

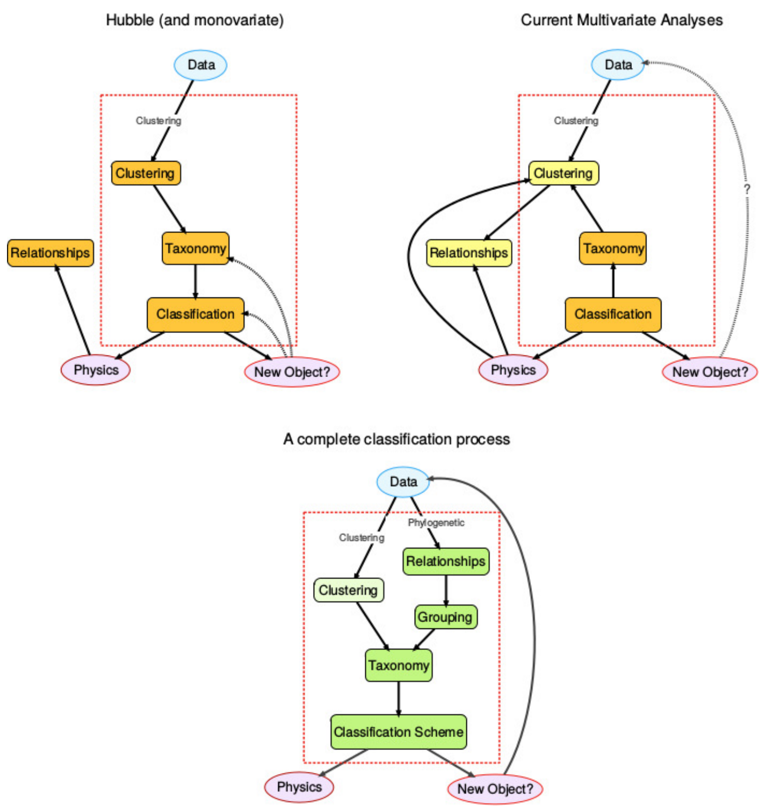

Fig. 1 Classification processes. Upper left: the Hubble classification. Upper right: most attempts to perform multivariate clustering analyses. Bottom: the correct and complete classification process that necessarily results in a new taxonomy and classification scheme.

mentioned by Sandage is obvious! I personally do not see how a new classification scheme could emerge in these conditions, especially a scheme that could give new physical insights to the galaxy diversity and its evolution since the (visual) morphology remains so preponderant.

The bottom panel shows the correct way of doing multivariate classification. The branch called "clustering" is the same as the one followed by Hubble, but since we want to be multivariate, the taxonomy must be entirely reconsidered. The second branch, called "phylogenetic", is the one I have been following, because there are many kinds of links and relationships between different sorts of galaxies. Evolution is the most obvious one, like Hubble rightly thought. So why not begin by establishing these relationships from the data, right at the start? Anyhow, in both approaches, physics is a product of the classification process, so that there is no risk of circularity. But a new taxonomy must be invented.

Coming back to the role of morphological studies, I would say they are only one side of the diversification of galaxies. What is morphology after all? The distribution of the stellar orbits. This is not less, not more important than the metallicities of the different populations of stars, or the presence of the black hole at the center of the galaxies, or the presence of molecular gas. So why concentrate exclusively on this feature? Probably because of the psychological effect of images that are always more artistic than, say, spectra! But we must do physics and understand the Universe as it is.

Milt-parametric approaches of a galaxy classification have been attempted in the past, may you describe what are their limits ?

The first multivariate analyses performed with the purpose to renew the classification of galaxies are from [7] and [8]. There were only few similar studies until 2005, and then several papers are published yearly.

Indeed, most of the multivariate studies are intended to automatically classify observed objects into the Hubble morphological classification [9]. The goal is not to build a new classification, but to use objective and quantitative observables obtained in large surveys (photometry and spectra) since one of the biggest problem with the Hubble classification is that the classes are determined with the eyes. Even the quantitative characteristics for morphology (e.g. [10]) are not easy to determine quickly and automatically and are subject to some caveats [6].

So most of these multivariate studies have a specific goal which is not aimed at building a new classification system. I believe they are all based on the more or less advocated assumption that the morphology contains all the physics, so that using some physical observables should suffice to retrieve the morphological classification. Like Sandage warned in 2005, they end up into a circular process which is clearly visible on the upper right panel of Fig. 1. This is their limit. This is also true for other multivariate studies that pretend to build new classifications [4, 5].

This limit has a more fundamental cause: this is the absence of a nomenclature to design the new classes found in multivariate clustering studies. I am often disappointed when a very serious and often difficult analysis ends up with ... the Hubble morphological classification to describe the results. For instance, a group of galaxies having a majority of elliptical shapes will be named the group of ellipticals for simplicity, throwing away all the methodological and the physical complexity! I know of only two studies that dare to give new names to new classes of galaxies: one by Sánchez-Almeida et al [11] and my own work [12, 13]. These are only first steps toward a new taxonomy for galaxies, but this is really a key point.

It is clear that from the year 2005, there is a strong interest for multivariate analyses which proves the need for a new system to classify galaxies. There is a cultural revolution going on, but contrarily to the debate in the 1990s, a lot of groups have acquired advanced statistical expertise. This is related to the advent of astrostatistics, opening a new era for astrophysics. Regarding classification of galaxies, the bottleneck is the lack of an adapted taxonomy. This requires to get rid of a traditional usage.

May you explain your new astro-cladistic approach and its application to galaxy classification and evolution?

When you think over the problem of galaxies, three words come to mind: classification, diversity and evolution. These naturally points toward the evolutionary biology. So, in 2001, I discovered the phylogenetic tools, and especially the cladistics (or Maximum Parsimony) which is the most general, probably the simplest to implement, even though not the simplest to understand. Astrocladistics[1] was then born.

Astrocladistics is the introduction of phylogenetic methods in astrophysics. For instance, a big question is to find the progenitors of present day galaxies. This is typically a phylogenetic problem. Hence, using phylogenetic approaches to study the formation and evolution of galaxies, that I prefer to call the diversification of galaxies, looks a rather natural idea.

How does it work? Two objects are close not only because they are similar (like in clustering methods) but also because they share some relationship, like a common ancestor. The phylogenetic approach builds these relationships from the data, and then classes are defined from the structure of these relationships (see right panel of Fig. 1). In the Maximum Parsimony technique, all possible arrangements of the galaxies on trees are built, and the one minimising the total number of parameter changes is chosen. This represents the simplest evolutionary scenario.

The use of continuous parameters in Maximum Parsimony is widely debated in biology, yet there is a priori no argument preventing this practice. Indeed, we have demonstrated that it is mathematically justified. To my point of view, Maximum Parsimony can be used for any continuous distribution of points, the notion of "evolution" simply being the continuous variation of the

[1] http://astrocladistics.org/

parameters. In this case, one can see the relationships scheme as representing the relative costs to transform an object into another.

Maximum Parsimony is the most general and powerful technique to relate objects on a tree. It may be compared to the Minimum Spanning Tree method (also known by astronomers as the friends-of-friends algorithm) in the sense that it considers all possible arrangements between the objects and then select the most parsimonious one according to the weights attributed to the edges (branches). But Maximum Parsimony allows for unlabelled internal nodes (in other words unobserved "objects"). In this way, it considerably extends the range of possible tree structures at the expense of the computer time. It is a parameter-based approach, so it is relatively easy to implement accepting error bars and undocumented parameter values. Its main drawback is that it is NP-hard, preventing the analysis of thousands of objects at once. However, we are looking for relationships between classes (phylogeny), not between individuals (genealogy).

There are other phylogenetic techniques, but generally they are specific to the data or processes of biological evolution, so they cannot be applied directly to astrophysics. For instance, there is a class of probabilistic approaches that assume some laws for the gene mutations (usually under a brownian hypothesis). I am convinced that such probabilistic methods can be used for galaxies since we know rather well the transformation processes of galaxies.

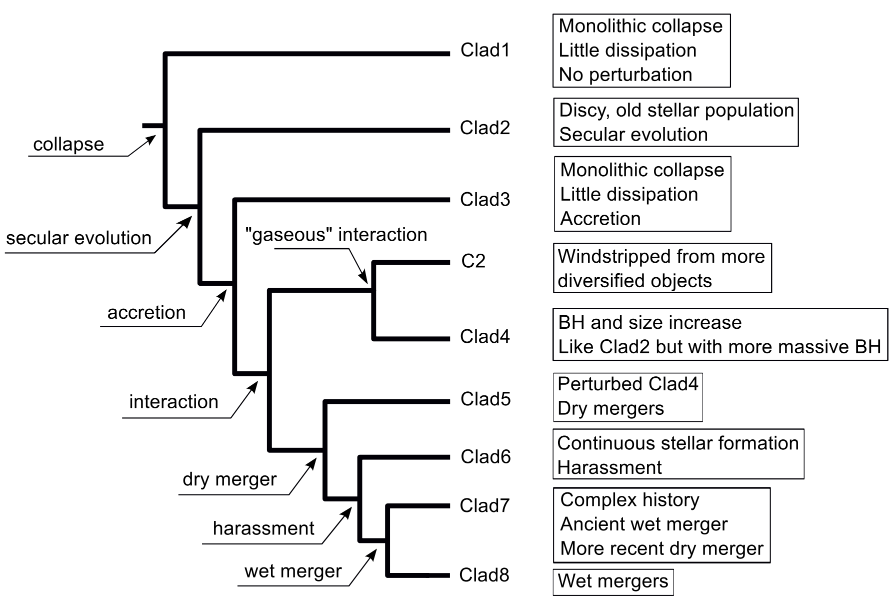

Fig. 2 A first step toward a global classification of galaxies based on their multivariate properties. From the properties of the classes, the history of the formation of the galaxies can be inferred. The internal nodes correspond to transformation events that are at the origin of the classes of galaxies emerging from these nodes [13].

This is a direction I have not yet explored any further and would require the development of some statistical theory of galaxy diversification.

Galaxies tend to present continuous range of values for most properties. The advantage of phylogenetic approaches is that they treat continuous distribution of points more smoothly than clustering techniques. The latter are devised for distinct classes and generally yield rather sharp cuts in the multidimensional space of parameters, cuts which do not appear very realistic. For the same reason, phylogenetic approaches are suited to find evolutionary tracks, kinds of very elongated structures that clustering techniques cannot recover at all.

I am aware that reading phylogenetic trees needs some practice, but be happy since a more realistic representation of galaxy diversification might be a network (a split network also called reticulogram for cladists) because of "hybridisation" processes (mergers essentially) which, I think, may be as common among galaxies as among bacteria!

After nearly fifteen years of development, where does astrocladistics stand? I must say that we spent a lot of efforts to learn, experiment, understand Maximum Parsimony in the context of galaxies, and also for some other astrophysical objects like globular clusters, stars and gamma-ray bursts. We have compared this phylogenetic method with other clustering techniques. We have published papers in order to demonstrate the validity and usefulness of the approach. We have thus obtained a first instance of a true multivariate classification of galaxies (Fig. 2). It does not pretend to be global since only about a thousand galaxies have been considered, and these galaxies are not representative of the whole diversity we know. But this result shows that we are on the right track.

We have used correct taxonomic rules for this classification. The names of the classes are Clad1 to Clad8 and a class C2 coming from a previous study on a independent sample has been included. These classes are fully described using several parameters, and from these characterisation we analysed the distribution of parameter values and correlations within each class, compared the classes, and used models and simulations from the literature to derive a possible history for the galaxies in each class. It is fundamental to understand that since our classes are characterised by many parameters, they are necessarily homogeneous, and very certainly homologous. As a consequence, they have a priori different formation histories which are easier to find thanks to the variety of the parameters.

From the diversification histories, we attempted to identify the events, represented by the internal nodes, that generated the new "species" of galaxies, which are the classes situated below the node in the ladderized representation of the tree given in Fig. 2. Here, the common ancestorship shared by some classes is a common transforming event that induced some new and durable property within a galaxy. This attempt is only tentative and illustrates the power of the tree-like representation that depicts the relationships between the classes.

To be clear, the tree representation results from an objective analysis of the data, it is not an interpretation like the Hubble tuning fork. The interpretation lies only in the inference of the formation histories and the transformation events indicated in Fig. 2.

This classification of a limited sample of the diversity of galaxies has already produced some major results. One of them is the possibility to compare in detail the properties of each class with numerical simulations of galaxy evolution. Why is this possible? Because our classes are homogeneous and built from many parameters, they represent more realistically the complexity of the transforming events that the simulations try to reproduce. This is impossible with a classification made from one parameter only since it is insufficient to constrain the many free parameters of the simulations.

Another important result is that many correlations, or scaling relations, appear to be due to (co-)evolution, not to a causal physical process. It is striking that many correlations vary from one class to the other, and differ from the whole sample. This proves that there are different populations of galaxies, with quite different properties. Is this surprising? Not to me, but the difficulty was to correctly identify these populations.

This second result has strong implications for the fundamental plane of galaxies. This is discussed elsewhere in this book.

What's next for astrocladistics? Everything is now in place to extend the tree of Fig. 2 to cover most of the known diversity of galaxies and invent a new taxonomy. This goal is clearly within reach.

in Chapter 4

- Dear Didier (*Fraix-Burnet*),

you have suggested an alternative interpretation of the Fundamental Plane, one of the most famous scaling relations of elliptical galaxies. May you explain your point of view?

Globally, considering galaxies statistically as an ensemble, the mass of galaxies increases with time, with the evolution of our Universe. This is because gravity is attractive, mass attracts mass. Luminosity, radius and surface brightness also increase with more stars being formed and accreted. Velocity dispersion globally should increase as well due to the increasing occurrence of interactions and internal orbital perturbations. One can discuss the details of these evolutions, but I don't think this raw image is wrong.

If all these quantities vary with a same parameter according to some monotonic functions, then correlations appear. If the functions are linear, then the correlation is linear. In three dimensions, this makes a plane. No need for physics here. For instance, the distance of the Voyager probes to the Earth and the average temperature on Earth since their launch both increase with time. A nice correlation thus exists, but would you conclude that the launch is responsible for the global warming? Certainly not. Time is here the confounding factor. Correlation is not causality, this is a principle to keep in mind.

The fundamental plane is known for a long time, and still resists a clear understanding. My point of view is that we should enlarge the way we consider this scaling relation to solve this mystery. I am saying that the statistical evolution of all the parameters creates a non-causal correlation. This is hard to avoid. We have thus to take it into account before detecting any causal correlation. Let me explain the consequences.

Usually, the fundamental plane is compared to the virial plane which is associated with two important assumptions. First galaxies are virialised. Second, since we do not have access to the mass (involved in the virial relation) but only to the luminosity, an hypothesis must be made on the M/L ratio. The so-called virial plane corresponds to M/L being constant. Unfortunately, the observed fundamental plane is tilted with respect to the virial plane. The most convincing "explanation" is that M/L is not constant. Empirically, it is found that M/L depends on a power of M. Why? To my knowledge, nobody knows. At least there is no emerging consensus from the physics.

In my interpretation [14], M, L, surface brightness, velocity dispersion and radius depend on a hidden factor. Observations suggest that this dependence most probably is a power law but this is not so important. Consequently, M/L has no reason to be constant and can be expressed as a function of the hidden factor, or equivalently as a function of M or L. Exactly as observed. But there is more than that: by constraining the functions with the observations, I can *derive* whether the galaxies are virialised or not.

By simply assuming that mass, luminosity, velocity dispersion, radius and surface brightness are monotonic functions of a same hidden factor, I can *deduce* from the observations: i) its nature, ii) the expressions of the dependence functions, and iii) whether and which galaxies are virialised. There is no other assumption in this interpretation of the fundamental plane.

Now, the dependence functions must be explained. As we have seen, it is hard to avoid evolution as a confounding factor. Why the functions seem to be linear? This should be interpreted in the statistical sense, and I have not yet the precise answer. Somehow, these functions can be seen as statistical laws induced by the physical processes of galaxy diversification. This is similar to the statistical physics which derives macroscopic laws from the statistics at the microscopic level. My suggestion is that we should develop a kind of statistical extragalactic astrophysics, this would be invaluable to better represent and understand the galaxy diversification process, and to fully explain the so-called fundamental plane as well as many other correlations.

Finally we must keep in mind that the fundamental plane is not planar everywhere. This is true only for galaxies that are the most diversified (i.e. that have gone through many transformation events) as we have shown [12]. Again, the correlations are different depending on the class/population of galaxies.

The accurate determination of fundamental galaxy scaling relations as a function of redshift and environment is the target of most important future surveys like those involving LSST. What is your suggestion?

I like to look at the historical evolution of ideas in science. I find quite stimulating to see that the biologists were first confronted to the complexity and diversity of the living organisms. They thus devised classification techniques to deal with that, and then more and more sophisticated statistical tools. But thanks to the progress in technology they came to the heart of the cells so that now they explore fundamental and detailed physical and chemical processes.

Astronomy has made the reverse path. In some sense, we are living within a cell – our Galaxy – and have logically first developed a detailed understanding of the chemical and physical processes within galaxies. We can touch the complexity of galaxies when we try to simulate them. But Hubble discovered many other similar cells, and the progress of technology now provides us with a huge diversity of objects, so that we have in our turn to deal with problems like populations and classification. Statistics in other words.

This is to say that we should probably have to renounce to master both the details and the complexity of galaxies in the same envelope. I understand that the proponents of a physical classification would like to establish the physics before doing some classification, in other words physics before statistics, but this is unrealistic. Physics and statistics are complementary.

I personally do not put too much attention to scaling relations in the sense that they are convenient because we think we can understand them. My strong recommendation is that we should *now* work hard through astrostatistics and astroinformatics to analyse the various and large data sets already available. I think in particular to the SDSS data base, but there are many others even if smaller. Surely enough, the future instruments and surveys are a significant leap toward a data and statistically driven astrophysics. But we can and must learn now to adapt our ways of analysing data, interpreting them and building models to compare with. Future surveys are designed with our present practice, but obviously they will require something different. The LSST has triggered astroinformatics to deal with the incredible data flow. But the astronomers around the world should invest a lot of efforts for astrostatistics. How could you still imagine to classify millions of galaxies from the LSST or equivalent projects with the Hubble morphological classification? How can you envisage to analyse millions of galaxies described by tens or hundreds of parameters with scatter plots between some a priori selected variables?

We need now to learn the data mining tools and the statistics approach will tell us, among many other diagnostics, what are the correlations, that you can call scaling relations, and whether they are causal correlations or spurious ones created by some confounding factor that we have to determine. My astrocladistics studies have revealed many new correlations with two or three parameters because they are projections of relationships found in a higher dimension parameter space. Hence focusing observations on a particular scaling relation might not be the most informative approach since we severely limit the quantity of information that we get with a strong a priori choice. As I have shown, these correlations depend strongly on the population of galaxies. And we have not yet a good multivariate picture of these populations.

I am totally convinced that we have to consider the galaxies as an ensemble of populations, in evolution of course, with interactions between each other and with their environment. This obviously defines a society, and astronomers should not be reluctant to use the adequate tools similar to the ones used in Human and Biological Sciences. In particular, in any scaling relation, there are very probably several hidden factors creating a confounding correlation. We must learn how to deal with that.

References

1. *Morphological and Physical Classification of Galaxies - Proceedings of the Fifth International Workshop of the Osservatorio Astronomico di Capodimonte*, Longo, G., Capaccioli, M., Busarello, G. (Eds.),*Astrophysics and Space Science Library*, vol. 178 (Springer, 1990). URL http://www.springer.com/astronomy/astronomy%2C+observations+and+techniques/book/978-0-7923-1712-8
2. J. Silk, G.A. Mamon, Research in Astronomy and Astrophysics **12**(8), 917 (2012). URL http://stacks.iop.org/1674-4527/12/i=8/a=004
3. G. Cecil, J.A. Rose, Reports on Progress in Physics **70**(7), 1177 (2007). URL http://stacks.iop.org/0034-4885/70/i=7/a=R04
4. C.J. Conselice, Monthly Notices of the Royal Astronomical Society **373**, 1389 (2006). DOI 10.1111/j.1365-2966.2006.11114.x. URL http://adsabs.harvard.edu/abs/2006MNRAS.373.1389C
5. M. Cappellari, E. Emsellem, D. Krajnović, R.M. McDermid, P. Serra, K. Alatalo, L. Blitz, M. Bois, F. Bournaud, M. Bureau, R.L. Davies, T.A. Davis, P.T. de Zeeuw, S. Khochfar, H. Kuntschner, P.Y. Lablanche, R. Morganti, T. Naab, T. Oosterloo, M. Sarzi, N. Scott, A.M. Weijmans, L.M. Young, Monthly Notices of the Royal Astronomical Society **416**, 1680 (2011). DOI 10.1111/j.1365-2966.2011.18600.x
6. R. Andrae, K. Jahnke, P. Melchior, Monthly Notices of the Royal Astronomical Society **411**, 385 (2011). DOI 10.1111/j.1365-2966.2010.17690.x. URL http://adsabs.harvard.edu/abs/2010arXiv1009.2508A
7. B.C. Whitmore, Astrophysical Journal **278**, 61 (1984)
8. M. Watanabe, K. Kodaira, S. Okamura, Astrophysical Journal **292**, 72 (1985)
9. N.M. Ball, R.J. Brunner, International Journal of Modern Physics D **19**, 1049 (2010). DOI 10.1142/S0218271810017160. URL http://adsabs.harvard.edu/abs/2010IJMPD..19.1049B
10. C.J. Conselice, Astrophysical Journal Supplement **147**, 1 (2003). DOI 10.1086/375001. URL http://cdsads.u-strasbg.fr/abs/2003ApJS..147....1C
11. J. Sánchez Almeida, J.A.L. Aguerri, C. Muñoz-Tuñón, A. de Vicente, ApJ **714**, 487 (2010). DOI 10.1088/0004-637X/714/1/487. URL http://arxiv.org/abs/1003.3186
12. D. Fraix-Burnet, M. Dugué, T. Chattopadhyay, A.K. Chattopadhyay, E. Davoust, Monthly Notices of the Royal Astronomical Society **407**, 2207 (2010). DOI 10.1111/j.1365-2966.2010.17097.x. URL http://hal.archives-ouvertes.fr/hal-00487853/en/
13. D. Fraix-Burnet, T. Chattopadhyay, A.K. Chattopadhyay, E. Davoust, M. Thuillard, Astronomy and Astrophysics **545**, A80 (2012). DOI 10.1051/0004-6361/201218769. URL http://dx.doi.org/10.1051/0004-6361/201218769
14. D. Fraix-Burnet, Monthly Notices of the Royal Astronomical Society: Letters **416**(1), L36 (2011). DOI 10.1111/j.1745-3933.2011.01091.x. URL http://dx.doi.org/10.1111/j.1745-3933.2011.01091.x