The Law of Starformation in Disk Galaxies

J.Köppen

Observatoire de Strasbourg

Does the available Hα-data from disk galaxies allow to distinguish between a SFR depending only on local current gas density and a more complicated SFR which also depends on galactocentric distance. What do the gas fractions tell us? These questions are answered with a practical Bayesian method we developed and which might be useful to address similar questions in other types of galaxies.

1. Introduction

Whether the observed properties of specific galaxies are to be explained or models of the theoretical evolution of galaxies are computed, the most crucial ingredient is the star formation history. The starformation rate (SFR) certainly depends on the physical parameters of the interstellar gas, but which parameters are involved and what its functional form is, is still rather unclear.

As young objects are found in regions of enhanced gas density, it seems most likely that the SFR increases with the current local gas density. Schmidt (1959, 1963) finds a quadratic dependence on the mass density of H I gas: $\Psi \propto \rho^2$. Madore et al. (1974) argue for a dependence on surface density, and Donas et al. (1987) derive a linear dependence of the total SFR of a galaxy and the mass of its atomic gas. But there is also evidence that the molecular gas is more responsible for the SFR (e.g. Guibert et al. (1978), Rana & Wilkinson (1986) find exponents larger than 1 for the molecular mass density).

There may well be other quantities the SFR depends on: Talbot (1980) proposes that the SFR in the disk of a galaxy is proportional to the frequency with which the gas passes through spiral arms. Considerations of the stability of gas disks (cf. Kennicutt 1989) also give rise to an explicit dependence on radial distance. Wyse & Silk (1989) generalize this approach by also allowing a dependence on the gas surface density. Dopita (1985), Dopita & Ryder (1994), and Ryder & Dopita (1994) suggest a dependence on the *total* mass surface density.

Furthermore, there are both observational and theoretical indications for a minimum density or pressure for the SFR (cf. Kennicutt 1989, van der Hulst et al. 1993, Elmegreen & Parravano 1994, Kennicutt et al. 1994, Wang & Silk 1994, Chamcham & Hendry 1996). Above that threshold density, Kennicutt (1989) finds a SFR depending nearly linearly on total gas density (exponent of 1.3 ± 0.3).

Theory has not been able to give firm predictions of the SFR much beyond providing possible identifications of plausible physical processes in order to explain the observed trends (cf. Franco 1990). A rather attractive concept is that the SFR is in fact the consequence of a self-regulation in a network of processes in the interstellar medium in the galactic environment. Several scenarios for such an equilibrium have been discussed (e.g. Cox 1983, Franco & Cox 1983, Franco & Shore 1984, Dopita 1985, Firmani & Tutukov 1992, Köppen et al. 1995, 1998), but a unique and precise prediction independent of observations is hard to get.

Which is the true law of starformation? Evidently, it is the one which gives the best fit to the data. But suppose one has 10 data points. While a simple prescription with a single fit parameter may already give an acceptable fit, it is surely improved by introducing another free parameter, and it could be made even better with a third one, ... and so on. Finally we come up with a perfect fit using 10 free parameters – which is bound to encounter a general disbelief! How many parameters can be extracted from data? Statistical tests (like χ^2-test) can only tell us when to reject too poor a fit, but there is nothing to indicate when a fit is too good. Usually, this decision is made more or less subjectively, based on one's feeling and experience: The metaphor of Occam's razor – of leaving out any complexity in a model unnecessary to explain the facts – is well known as a philosophical concept, but can one devise a practical, objective, and mathematically correct method to aid in this judgement?

Yes! It is the Bayesian approach to statistics that permits to formulate Occam's razor without introducing any artificial assumptions or subjective constructs, but in a natural and mathematically fully consistent way (Köppen & Fröhlich, 1997). This method is applied to our principal question: Can we with the existing observational data decide between a SFR law which depends only on local gas surface density $\Psi \propto g^x$ with a fixed or free parameter x and a law which also has an explicit radial dependence $\Psi \propto g^x/r^y$ ($y = 1$ corresponds to a flat rotation curve in the concept of Wyse & Silk)?

2. A Bayesian Method

The basic difference between 'classical' and Bayesian probability theory lies with the meaning of probability: in the classical picture it is defined as the *relative frequency* of occurrence of an event, and thus is the result of a large number of identical experiments. In the Bayesian view probability is the *degree of belief* of a hypothesis, which has the advantage that it can be applied also to single events. On either of these definitions, one has built mathematical theories of probability and statistics – which turn out to be equivalent. These two views are merely two aspects of the same thing, like the two sides of a coin. More on the theory and practice can be found in Jeffreys (1983), Erickson & Smith (1988), Fougère (1990), O'Hagan (1994).

The probabilities are computed with Bayes' Theorem (which is *strictly* valid in either view!): Let us consider a set of N hypotheses H_i, only one of which can be true. Then the probability for the k-th hypothesis, taking into account the data D, is

$$p(H_k|D) = \frac{p(H_k) \cdot \Lambda(D|H_k)}{S} \tag{1}$$

where $S = \sum_{i=1}^{N} p(H_i)\Lambda(D|H_i)$ is a normalizing factor. The prior probabilities $p(H_k)$ represent the investigator's degree of belief or his knowledge from previous measurements. The likelihood $\Lambda(D|H_k)$ is a measure of how well the predictions of H_k match the data, which incorporates the distribution functions of the random errors. Thus, Eq. (1) describes how a measurement improves our knowledge: It states that the *posterior* probability for a hypothesis is proportional to the product of its probability assigned before the observation and the likelihood $\Lambda(D|H_k)$ of the data D.

In practice, one compares a few specific models, and so the set of hypotheses is not exhaustive. Therefore, only ratios of probabilities can be calculated, but as long as one sticks to the same set of models, explicit calculation of the normalizing sum S is unnecessary. It suffices to compute the numerator in Eq.1, called the *Bayes factor*.

The correct formulation of the priors must include all previous knowledge. Since this is not easily derived from mathematical axioms, the classical approach prefers not to include the prior information at all, and is content by maximising the likelihood, and therefore is unable to compute the true posterior.

When a hypothesis contains free parameters λ, Bayes' Theorem gives the posterior density $p(H_k, \lambda|D)$. This is integrated over the space of the parameters to get the probability for the hypothesis irrespective of the parameter's actual value:

$$p(H_k|D) \propto \int p(H_k, \lambda|D) d\lambda = p(H_k) \cdot \int p(\lambda)\, \Lambda(D|H_k, \lambda)\, d\lambda \tag{2}$$

where $p(\lambda)$ is the prior probability density for the parameter. Since this prior is normalised over the parameter space ($\int p(\lambda)\, d\lambda = 1$) the posterior decreases, if the parameter space is increased much beyond the volume where the likelihood density $\Lambda(D|H_k, \lambda)$ contributes significantly. In this way, the increased freedom to get a good fit may well more than compensate any increase of the likelihood itself due to the better fit. It is this feature, naturally occurring in the Bayesian approach by taking seriously the prior information, which allows a mathematically consistent formulation of Occam's razor.

Equation 2 also shows that problems with many free parameters quickly lead to a heavy load of computer resources, because of the evaluation of multiple integrals. Wherever possible, analytical evaluation of some of the integrals thus is highly recommended.

Thinking about the formulation of the priors I found to be a most wholesome and enlightening exercise to make oneself aware of what one knows or not, or pretends to do so. For our question, we have used the following recipe. The **parameter priors** are taken from Jeffreys (1983) formulations for the absence of any prior information: Parameters which are simple numbers such exponents, a uniform prior density $p(\lambda) = const.$ is taken. Dimensioned parameters (scalelengths, timescales, ..) follow a prior uniform in $\ln(\lambda)$, i.e. $p(\lambda) \propto 1/\lambda$.

The **theory priors** would be the place to formulate a possible (personal) preference of theories with a smaller number of free parameters. We take all hypotheses – irrespective of the number of their parameters – as equally preferable: $p(H_k) = const.$

For the treatment of **nuisance parameters** which are parameters of the problem or the model but whose values are of no interest to us – here: the factor between Hα-brightness and SFR, and the size of the scatter in the data due to noise or intrinsic fluctuations of the SFR – and practical aspects of the integrations see Köppen & Fröhlich (1997, 1998).

3. Results on Hα-density

To answer our main question, we take for a dozen galaxies observational data on the surface brightnesses of Hα (Kennicutt 1989), H I, and CO from the literature. The gas density is obtained as the sum of atomic and molecular hydrogen, using the same CO-H_2 conversion recipe. The distribution function of the random scatter (noise or genuine fluctuations) is assumed to be a Gaussian in the relative deviations. Figure 1 shows the contour lines, in the space of the two exponents x and y, within which 90 percent of the integrated probability is found. Though some galaxies – with few data or with strong fluctuations – show large confidence regions but low weight, most objects are peaking near $x \approx 1...2$ and $y \approx 0$).

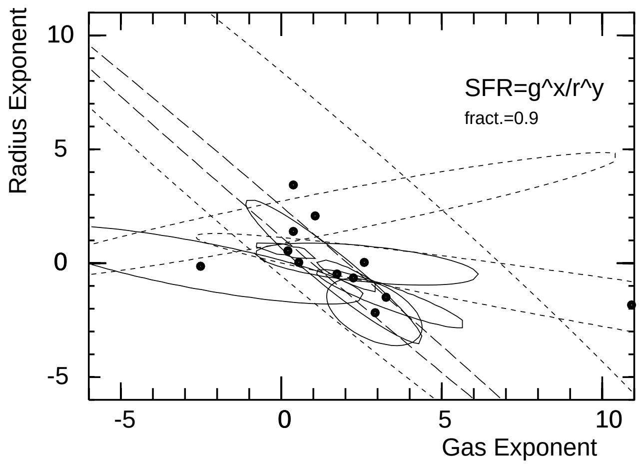

Figure 1: The confidence regions for individual galaxies in the parameter space within which 90 percent of the total posterior probability is found.

Since our method yields proper probabilities (and densities), they can be used in a straightforward manner to calculate joint probabilities. Figure 2 shows the confidence region for the

SFR-law	individual	common value
g	1	
g^2	$1.9 \cdot 10^{-6}$	
g/r	$1.3 \cdot 10^{-15}$	
g^x	0.59	0.061
g^x/r	$2.0 \cdot 10^{-13}$	$4.1 \cdot 10^{-12}$
g/r^y	$1.5 \cdot 10^{-5}$	0.033
g^2/r^y	$4.5 \cdot 10^{-7}$	0.002
g^x/r^y	$9.8\,10^{-6}$	0.002
x_{best}	—	0.73
y_{best}	—	-0.06

probability that the same pair (x, y) of parameters fits the data of all 12 galaxies. This region is quite small, with a peak corresponding to no radial dependence but a nearly linear dependence with gas density. The figure also shows that the choice of the CO-H_2 conversion does not greatly change this finding.

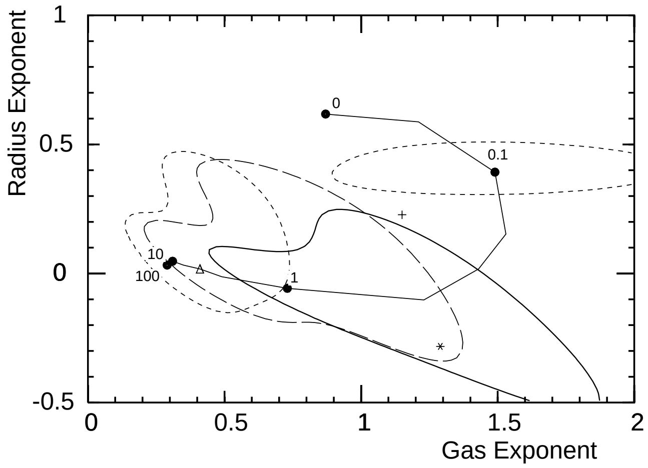

Figure 2: The 90 percent confidence region of the joint likelihood for all galaxies and how it depends on the CO-H_2 conversion factor X. The solid line refers to the standard value $X = 3.68$, short dashes indicate $X/3.68 = 0.1$ and 100, and long dashes are done with the metallicity-dependent prescription of Arimoto et al. (1996).

Likewise, probabilities for various sub-hypotheses can be obtained. In the Table we collect the joint Bayes factors for various laws with no, one, or two free parameters. This is done for the parameter either allowed to vary between individual galaxies or to have a value common for all objects. The ranges of the parameters are as shown in Fig. 1. All Bayes factors are normalized relative to the simple linear law $\Psi \propto g$:

One notes that the most likely laws are the simple linear law and the dependence on gas density where each galaxy has its own optimal exponent. Laws like g^2 or g/r are in quite strong disagreement with the observational data, and so are the other laws.

If one reduced the assumed parameter range for x of $(-6...11)$ by say a factor 2, the Bayes factors of laws with common parameter values for all galaxies would be approximately doubled – the parameter prior is applied only once. On the other hand, for the laws whose parameters are allowed to vary individually among the 12 galaxies the prior is applied for each object, so the factors would increase by about $2^{12} \approx 4000$. In this manner, these laws could be made more probable. But this merely reflects their much larger freedom for making the fit as compared to laws where the same parameter value is to be used for all objects.

4. The Gas Fraction

The study has been extended to include the information contained in the stellar population that has been born up to now (Köppen & Fröhlich, in prep.). Radial profiles of the optical surface brightness – preferentially in the red – are taken from the literature. For an initial study, we do not make use of the radial variation of the colours. The information enclosed in the gas fraction has been found to be far more important.

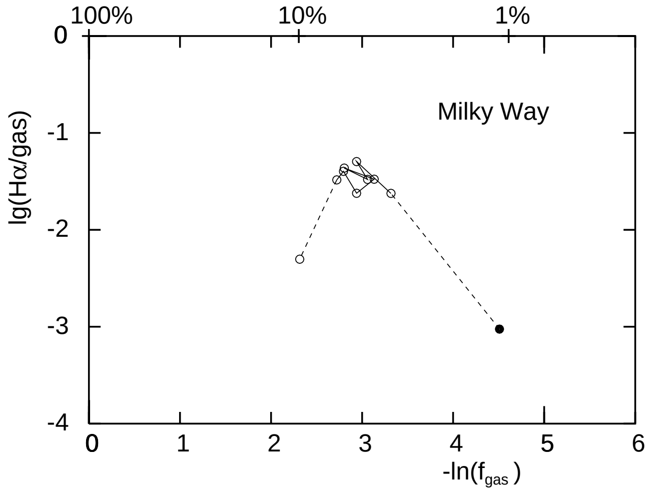

Figure 3: The ratios of Hα surface brightness and surface density g of the total gas as a function of the local gas fraction $f_{\rm gas}$ in the Milky Way. The black dot indicates the inner 3 kpc.

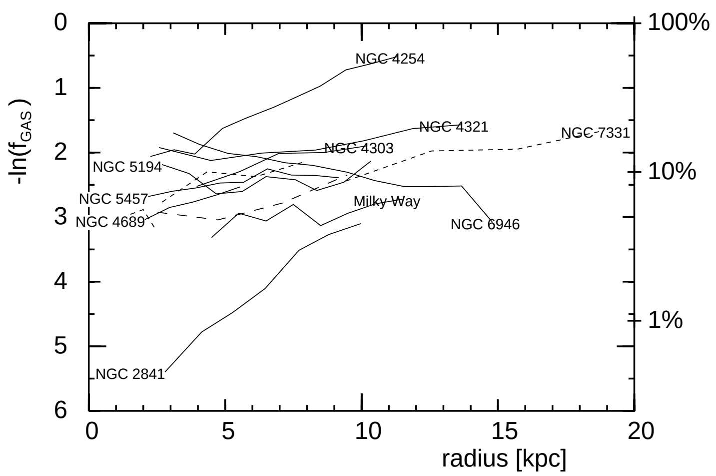

Figure 4: The gas fraction as a function of galactocentric distance.

First, a direct analysis of the data from each galaxy is done, as shown in Fig. 3 for the Milky Way as a typical example: Most of the galactic disk (from 3 to 12 kpc) has not only a nearly constant gas fraction of about 7 percent, but also a nearly constant ratio of Hα-brightness and (total) gas density – i.e. the current SFR is simply proportional to the gas density. Outside this region of the disk, the SFR is much lower, and the gas fraction quite different: e.g. inside 3 kpc (filled circle) there is little gas. With a few exceptions, the other galaxies show quite similar behaviour. This suggests that in the ring of the disk where most of 'normal' starformation occurs, the SFR is rather close to a simple linear dependence of the gas density.

In nearly all galaxies the gas fraction increases only very gently towards the exterior, as shown in Fig. 4.

That the linear SFR has been responsible for all previous star formation in the disk, is the simplest explanation for this shallow variation of the gas fraction: Assuming a closed-box model with initial gas density $g_0(r)$ for each radial ring, one gets with a general SFR $\Psi = C(r) \cdot g^x$ for an age t:

$$(g/g_0)^{1-x} - 1 = (x-1) \cdot t \cdot C(r) \cdot g_0^{x-1} \tag{3}$$

A nearly constant $f_{\rm gas} = g/g_0$ requires

$$t \cdot C(r) \cdot g_0^{x-1}(r) \approx const. \tag{4}$$

If all sections of the disk have the same age t, the linear dependence of the SFR on gas density ($x = 1$ and $C(r) = const.$) provides an easy explanation.

The Bayesian method is applied to the comparison of the observed radial profiles of stellar and gas density with models where the disk is formed by gas infall and where gas is allowed to flow outward in radial direction, and starformation follows the two-parameter law as before. Figure 5 shows the contour lines in the space of parameters (x, y) of the probability density integrated over all remaining (four) parameters. These first results show that the 90 percent confidence regions are much smaller that those obtained from the Hα-data, and they keep very close to the simple linear law ($x \approx 1...2$ and $y \approx 0$). This finding appears to be not very sensitive to other model parameters, such as infall and gas flows, though the probability obtained does vary, of course.

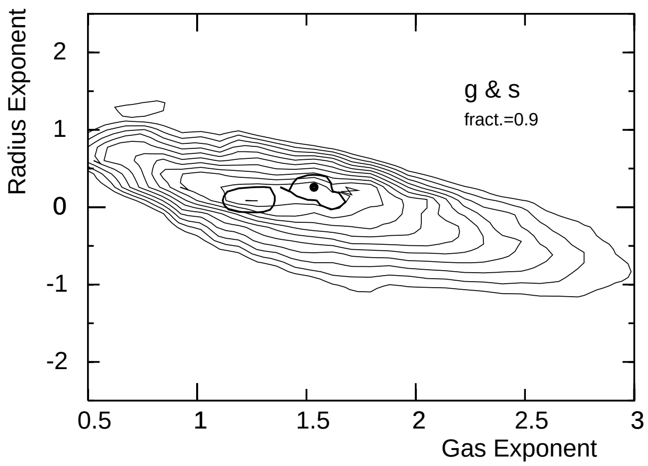

Figure 5: Contours of the joint probability density for all galaxies of models which accrete gas decreasing exponentially in time and allow radial inflow of gas. The density varies by a factor 100 between contour lines. That the confidence region is divided in two islands is due to the coarseness of the integration grids.

5. Conclusions

The result that the starformation law in the disks of galaxies is the trivial linear dependence, may be somewhat sobering or even disappointing, but the Bayesian approach tells us that this is all that can be deduced with confidence from the currently available data. This pertains to the Hα-surface brightnesses as an indicator for the present SFR, which might be affected by systematic effects such as extinction and the degree of ionization boundedness of the H II regions. But even more so it applies to the gas fractions which measure the average starformation in the past. To detect finer dependences of the SFR on parameters of the ISM appears to require either a much larger and more accurate data set or (more likely) specific and local studies based on other types of information.

With respect to irregular and dwarf galaxies, one may envisage to unravel the history of the starburst activity or to find out which spatial and temporal relations between star forming regions may be deducible. The method outlined here can be applied for any comparison of models with data, providing from the observations themselves a good indication of whether a certain model is justified or might come close to an over-interpretation.

References

Arimoto N., Sofue Y., Tsujimoto T., 1996, PASJ, 48, 275
Chamcham K., Hendry M.A., 1996, MNRAS, 279, 1083
Cox D.P., 1983, ApJ 265, L61
Dopita M.A., 1985, ApJ 295, L5
Dopita M.A., Ryder S.D., 1994, ApJ, 430, 163
Donas J., Deharveng J.M., Laget M., Milliard B., Huguenin D., 1987, A&A, 180, 12
Elmegreen B.G., Parravano A., 1994, ApJ, 435, L121
Erickson G.J., Smith C.R., (eds.), 1988, Maximum-Entropy and Bayesian Methods in Science and Engineering, Kluwer, Dordrecht
Firmani C., Tutukov A., 1992, A&A, 264, 37
Franco J., 1990, in *Chemical and Dynamical Evolution of Galaxies*, ed. F.Ferrini, J.Franco, F.Matteucci, ETS editrice, Pisa
Franco J., Cox D.P., 1983, ApJ, 273, 243
Franco J., Shore S.N., 1984, ApJ, 285, 213
Fougère P.F., (ed.), 1990, Maximum Entropy and Bayesian Methods, Kluwer, Dordrecht
Jeffreys H., 1983, Theory of Probability, 3rd ed., Clarendon Press, Oxford
Kennicutt R.C., 1989, ApJ, 344, 685
Kennicutt R.C., Tamblyn P., Congdon C.W., 1994, ApJ, 435, 22
Köppen J., Fröhlich H.-E., 1997, A&A, 325, 961
Köppen J., Theis Ch., Hensler G., 1995, A&A, 296, 99
Köppen J., Theis Ch., Hensler G., 1998, A&A, in press
Madore B.F., van den Bergh S., Rogstad D., 1974, ApJ, 191, 317
O'Hagan A., 1994, Kendall's Advanced Theory of Statistics, Vol. 2B, Bayesian Inference, Edward Arnold, London
Rana N.C., Wilkinson D.A., 1986, MNRAS, 218, 497
Ryder S.D., Dopita M.A., 1994, ApJ, 430, 142
Schmidt M., 1959, ApJ, 129, 243
Schmidt M., 1963, ApJ, 137, 758
Talbot R.J., 1980, ApJ, 235, 821
van der Hulst J.M., Skillman E.D., Smith T.R., et al., 1993, AJ, 106, 548
Wang B., Silk J., 1994, ApJ, 427, 759
Wyse R.F.G., Silk J., 1989, ApJ, 339, 700