Journal of Scientific Exploration, Vol. 29, No. 3, pp. 425–448, 2015 0892-3310/15

RESEARCH ARTICLE

A Review on the Relation between Population Density and UFO Sightings

Julio Plaza del Olmo

plazaoj@gmail.com

Submitted June 15, 2014; Accepted March 23, 2015; Published September 15, 2015

Abstract—In the search for patterns of the UFO phenomenon, the relationship of sightings to population density has historically shown contradictory results. After more than 40 years of studies, there is not yet a clear conclusion on whether the relationship between the number of reports and population density is direct or inverse. We have reviewed some of these works and found out how to reconcile all of them. We found that there is a direct relationship between the number of sightings and population density; however, the increase in number of reports is not proportional but sublineal with respect to the increase in population.

Introduction

The study of Unidentified Flying Objects (UFOs) or UFO phenomena has been an activity focused on finding patterns that may help us understand whether the phenomenon is real and what its nature is. One of these patterns has been the relationship between the number of UFO reports and population density. Since UFO reports are dependent on the presence of witnesses, common sense dictates that more populated areas should produce more UFO reports than unpopulated areas. However, the first person to study this relationship was Jacques Vallée (1966), who reached the conclusion that there was an inverse correlation, that is unpopulated areas produced more reports than populated areas. This conclusion led him to think that there was some kind of intelligence behind the phenomenon that made it avoid populated places and be more active in isolated places.

After this initial study, the relationship has been addressed on many occasions by different authors (see references throughout). However, the results in these works seem to lead to contradictory conclusions. Some show the inverse relationship found by Vallée, but other studies show a direct relationship, in which populated areas provide more UFO reports than less populated areas.

These studies, however, have important differences, especially regarding the variables taken into account: Number of reports (N), Population (P), Number of reports per capita (N/P), Population density (P/S, or δ), Reports per square kilometer (N/S), and some other subtleties such as the kind of cases taken into account, which makes it harder to find a direct comparison and to understand the origin of the different conclusions. A summary of these works, spanning a time period of 48 years between 1966 and 2014 can be seen in Table 1.

TABLE 1
Summary of Studies Regarding Population Density and UFO Sightings

Authors / reference	Variables	Correlation	Comments
Vallée 1966	N vs δ	Inverse	Landings in France
Vallée 1968	N/P vs δ	Inverse	All kinds of sightings
Condon 1968	N vs δ	Direct	Non-urban areas
Bonabot 1971	N vs δ	Inverse	Landings
Saunders 1975	N vs P	Direct	All kinds of sightings
Poher & Vallée 1975	N vs δ	Inverse	Landings
Poher 1976	N vs δ	Direct	All kinds of sightings
Ballester Olmos 1976	N/P vs δ	Inverse	Landings
López et al. 1978	N/S vs δ	Direct	All kinds of sightings
Fernández & Manuel 1980	N/S vs δ	Direct	Landings
Fernández & Manuel 1980	N vs δ	Direct	Sightings in the Comunidad Valenciana
Weiller 1980	N vs P	Direct	All kinds of sightings
Verga 1981	N/δ vs δ	Inverse	Landings in Italy
Ballester & Fernández 1987	N vs δ	None	Landings. UFO and IFO cases
Breysse 1993	N vs P	Direct	All kinds of sightings
Ballester Olmos 2014	N vs δ	Direct	Photo and video images
Rospars 2014	N/P vs δ	Inverse	UFO and IFO cases

N = Number of reports. P = Population. S = Area. N/P = Reports per inhabitant. N/S = Reports per unit area. δ = Population density.

In this work we are going to review those studies, and we will show that it is possible to reconcile their results. The methods used in the above-mentioned works as well as in this one are explained in basic terms in Appendices A and B.

As we will see, the use of different variables in each study is the origin of the different conclusions reached by different authors. In particular, it is important to keep in mind that an inverse relationship in the number of reports per capita with population density (N/P vs δ) does not necessarily imply an inverse relationship of the total number of reports with population density (N vs δ).

Historical Review

Vallée and the First Negative Law

The interest in the geographical distribution of UFO reports goes back to 1966, when Vallée (1966) argued against the hypothesis of a psychological origin of UFO sightings, put forward by Georges Heuyer, to explain the 1954 wave in France. The development of a *psychosis* should follow strict rules, it is not a random phenomenon, and populated areas like Paris, Lille, Marseille, or Bordeaux would have had to have the right conditions for the propagation of rumors.

Vallée analyzed 200 landing reports, most of them occurring in France during 1954. After plotting these landings on a map (Figure 1 Left), Vallée argued that these did not concentrate in populated areas, and finally stated what he called the *First Negative Law*:

> The geographic repartition of the landing sites in 1954 is inversely correlated with population density.

Later, in 1968, Vallée published another paper (Vallée 1968), analyzing more than 8,000 sightings in the U.S. He reported that the number of reports per capita was higher in low-population areas, confirming the rural character previously noted in 1966. A plot of the number of reports per capita versus population density, grouped by states of the U.S., showed an inverse relationship (Figure 1 Right). Note that the variables are not the number of reports (N), but the number of reports per capita (N/P). Here we find one of the first misunderstandings with respect to the variables, since the statement of the first negative law referred to the total number of reports.

Also in 1968, the Condon Report (Condon 1968) had a paragraph pointing out that:

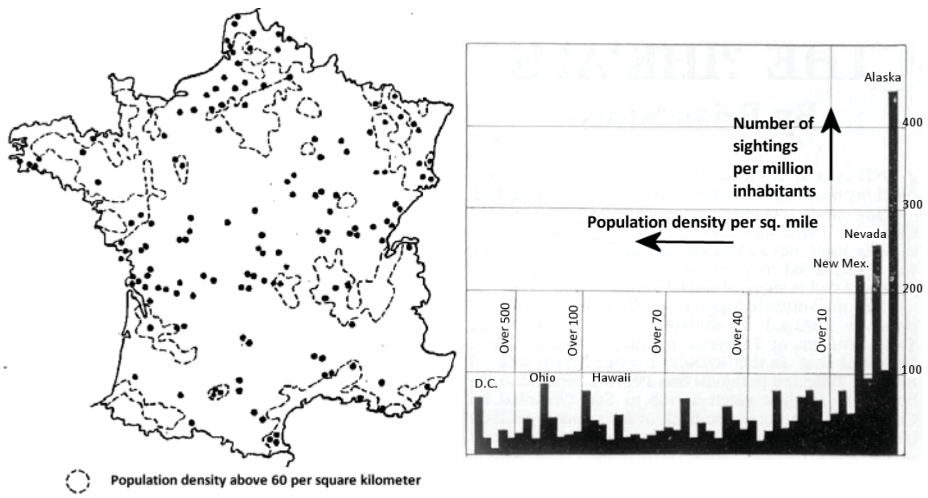

Figure 1. Left: UFO Landing sites in France, 1954. Reproduced from Vallée (1966). Right: Distribution of UFO sightings in U.S. states (Vallee 1968).

> The geographical distribution of reports correlates roughly with population density of non urban population. Very few reports come from the densely populated urban areas. Whether this is due to urban sophistication or to the scattering of city light is not known, but it is more probably the latter.

This partially contradicts Vallée, but agrees in that urban areas do not significantly contribute to the total of UFO reports. However, an interesting reason is sketched to account for that: City lights would eclipse lights from UFOs.

Studies During the 1970s

Using a similar graphical analysis as Vallée, Jacques Bonabot (Bonabot 1971) made a simple analysis of the distribution of UFO events in Belgium. Landings and close encounters seemed to follow Vallée's first negative law. But he also noted that flying or distant objects seemed to be more frequent near populated places like Brussels and Liege.

David Saunders (Saunders 1975) did excellent work on a huge database of U.S. cases, UFOCAT, seeking multiple correlations with up to 14 different variables, population and geographical area among them. He took into account about 18,000 reports scattered in more than 3,000 counties. In January 1975 he presented his results at the 13th Aerospace Sciences Meeting, showing that the main correlation was between number of reports and population, and it was a direct one.

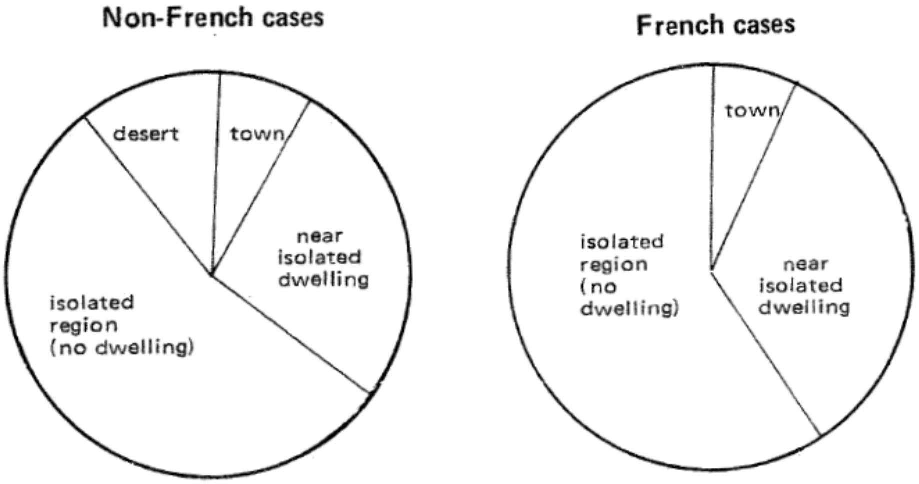

Figure 2. Distribution of Type I cases as a function of population density. Reproduced from Poher and Vallée (1975).

Also at this meeting, Poher and Vallée (1975) presented a work on basic patterns in UFO observations. The distribution of reports as a function of the location of the sightings is reproduced in Figure 2. More than 70–80% of the reports occurred in or near isolated places.

These results are only part of a larger statistical study done by Poher between 1971 and 1976 (Poher 1976). In that work, he also analyzed the distribution of 220 sightings and 40 landings, summarized in Table 2. The landing reports are too few to draw conclusions. As for the sightings, it looks like a direct correlation between number of reports and population density could be possible. However, the division of population density into only three bins of different sizes for the population density seems poor and inadequate for drawing conclusions.

TABLE 2
Number of Reports vs. Population Density (taken from Poher 1976)

Density [Pers/km^2]	Total (220)	Landings (40)	Landing with occupants (16)
$\delta > 80$	110	48%	31%
$50 < \delta < 80$	64	20%	25%
$\delta < 50$	46	32%	44%

This possible direct relationship was also mentioned by Poher (1975), albeit the main argument was to show a dependence on visibility and atmospheric conditions. Poher compared two groups of French departments: a first group of 1,200 hours of sunlight on average, and a second group with about 2,500 hours of sunlight on average. This difference is caused by atmospheric conditions. Both series showed a direct correlation with population density, but the second one also showed a higher number of reports.

A review of French cases published in the magazine *Lumières dans la Nuit* was carried out by Bettini et al. (1977). They noted that departments such as Nord and Pas de Calais, which have high population density, contained the majority of reports.

Ballester Olmos (1976) used a database of 200 landings in Spain in 1976. He grouped 48 provinces into 8 groups of 6 areas. The relationship between reports per inhabitant and population density had a negative correlation, although with a low statistically significant level. It is worth remarking that Ballester Olmos took into account an effect that Saunders had already warned against, the *investigator effect*: Local UFO investigative groups can generate more reports than would otherwise be predicted for that area.

Finally, in 1978, a long work was presented by López et al. (1978) at the First National Congress of Ufology in Spain. The main focus was the development of a model to predict when and where a UFO would be seen. The model supposed that UFO phenomena occurred randomly, and the number of cases was determined by parameters like area, population, climate, and orography. The theoretical number of reports was compared to real values that were also corrected to take into account the *investigator effect*. López et al. finally found a direct relationship between population density and UFO sightings for all types of reports, and also specifically for landing reports. Theoretical and real values correlated quite well, meaning that UFOs basically had a random pattern.

Later Studies

Following the methodology of López et al., Fernández Peris and Manuel Garijo (Fernández & Manuel 1980) produced another excellent work on UFO patterns. They focused on landing events in the Spanish wave of 1974. Even though their results showed a direct relationship (i.e. number of reports per unit area increasing with population density), they reasoned that when comparing the theoretical values (given by the López et al. model) against the real ones, the number of reports was higher than predicted by the model for low-population areas, and lower than predicted in high-population areas.

Thus, their final conclusion was that a negative law was correct, since low-population areas produced more reports than expected.

However, this kind of "inverse relationship" is different from the one originally stated by Vallée. Vallée is referring directly to the behavior of the variables (N vs δ, or N/P vs δ). If Fernández and Manuel analyzed the same variables in the same way, the relationship is direct for both the model and the real data, and thus contrary to Vallée's negative law. However, the comparison made by Fernández and Manuel is between the *proportionality* of the model and their data, and thus they are comparing something completely different from Vallée. Any conclusion on such a comparison is not valid with respect to Vallée's negative law.

Fernández and Manuel also made a second interesting study (Fernández & Manuel 1980b). Instead of analyzing large areas like countries, they studied the distribution of sightings in the regions of Valencia, in Spain. A total of 208 cases were distributed in smaller regions, and although the number of reports was low, analysis showed high values of the correlation coefficient for population and population density. On the other hand, the correlation between reports and surface area was low, and the final conclusion was that it had no effect on the sightings. They went into further detail with a *finer structure*, and the distribution was studied as a function of the population of the village or town where the sighting occurred. Once again, there was a remarkably high correlation.

In 1980, Weiller (1980) used the correlation between reports and population to hypothesize about the existence of an unknown meteorological effect associated with UFO sightings. He used a database of nearly 2,000 reports in France scattered across departments. The correlation using the whole database is weak, but positive.

Weiller then introduced a luminosity index for each department, based on meteorological data. It is based on the fraction of sky covered by clouds in each department, and thus is very similar to the definition of sunlight hours used by Poher (1975). With this new index, he selected the 22 most luminous ("least cloudy") and the 22 least luminous ("most cloudy") departments. He found that the most luminous group contained more sightings than the least luminous. The correlation coefficients improved remarkably. But also, when taking into account a multiple correlation with population and luminosity index as variables, he deduced that it is also a factor contributing to UFO reports.

This dependence on luminosity agrees with Poher (1975) using sunlight hours as a variable. The definitions of luminosity and sunlight hours are based on meteorological conditions that affect visibility, or the distance at which objects may be seen. Both Weiller (1980) and Poher (1975) show

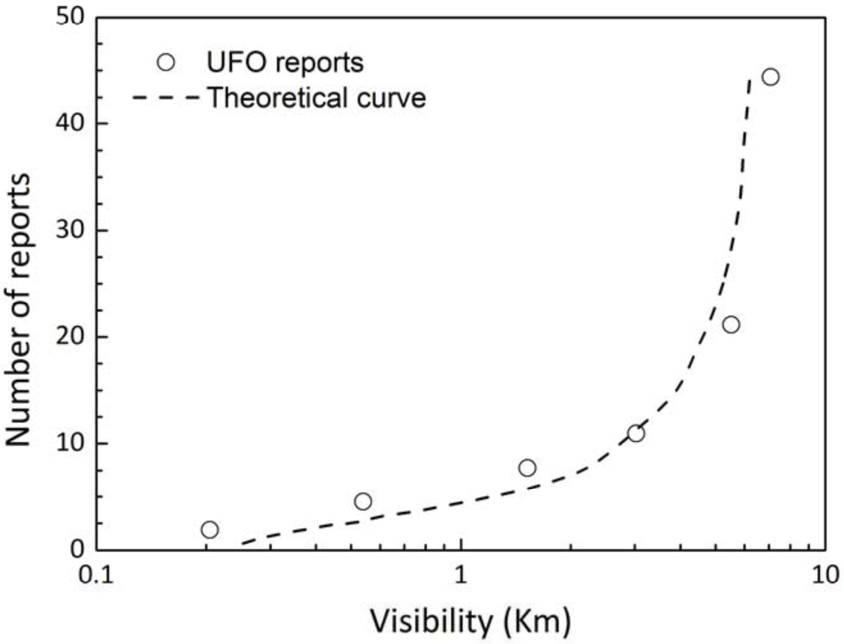

Figure 3. UFO reports versus visibility and atmospheric transparency. Reproduced from Poher and Vallee 1975.

coherent results with those obtained by Poher and Vallée (1975), where they analyzed the relationship between the number of reports and visibility (Figure 3). The results are totally consistent with sightings produced by misperceptions of stars, planets, airplanes, etc. Cloud cover or low visibility conditions prevent these stimuli from being seen; even though they do not account for all sightings, they represent an important fraction of UFO reports.

One year later, Maurizio Verga made a review on the subject, and also conducted a study on a database of 326 landings in Italy (Verga 1981). He noted the contradictory results, and assumed it was due to the different methodologies used, including the definitions of *populated* and *unpopulated* areas. He also noted that many different factors could affect the data count. Therefore, in his final remarks, Verga stated that the majority of studies were useless to debate about. However, he also noted that about 90% of the reports came from places with *a high probability of witness presence*.

In 1987, Ballester Olmos and Fernández Peris published a book on landing reports (Ballester & Fernández 1987). They updated Ballester's landings database used in previous studies, and selected only cases with a high strangeness, as defined in Ballester and Guasp (1981). With the help of collaborators throughout Spain and Portugal, a rigorous study was done in each case and finally two catalogs were produced:

- **LANIB** (new catalogue of LANding reports in the IBerian peninsula): Landing cases with a high strangeness factor as defined in Ballester and Guasp (1981) (230 cases).
- **NELIB** (catalogue of NEgative Landings in the IBerian peninsula): Landing cases for which an explanation was found (310 cases).

NELIB was thought of as a *control database*. If the UFO phenomenon is real, some kind of difference has to be found when compared to databases containing real UFO cases. Ballester and Fernández found a correlation coefficient of 0.62 for NELIB, meaning a direct relationship between number of reports and population density, with a moderate correlation. For LANIB they found a correlation coefficient of 0.21, which was interpreted as a lack of correlation.

Breysse (1993) tested two different hypotheses: whether the number of reports was a function of surface area, or dependent on population; and also whether this dependence was proportional ($N \sim S$, $N \sim P$) or quadratic ($N \sim S^2$, $N \sim P^2$). After analyzing two databases from the U.S. (643 reports in 51 states or territories) and France (346 reports in 95 departments), he came to the conclusion that surface area was not correlated to the number of reports. On the other hand, the best correlations were between number of reports and population, and they were direct correlations (Table 3). However, the data did not allow differentiating between the two models, proportional or quadratic.

Just recently, Ballester Olmos published some basic statistics on FOTOCAT (Ballester 2014), showing the distribution of 11,060 UFO images (photos and footage) by continent. The result was a positive correlation between the number of reports and population density.

Finally, the last study to date was presented at the CAIPAN International

TABLE 3
Correlations for Two Databases (reproduced from Breysse 1993)

	N vs S	N vs P	N vs δ
US (N = 643)	0.063	0.887	0.386
France (N = 346)	0.145	0.391	0.126

Workshop by J. P. Rospars (Rospars 2014). His analysis on reports per unit area and reports per capita versus population density in France showed that probability of reporting an observation increases with population in large areas, whereas the probability for an inhabitant to report an observation decreases when the density increases. As we will see, this last study offers the best description for the real dependence of the number of reports on population density.

Review and Recalculation of Results

So far, we have reviewed several studies and analysis on the geographic distribution of reports, showing different and apparently contradictory results. Intuition tells us that the more people, the more probability somebody can witness a UFO event. However, this seems to be true when all kind of reports are considered, but the relationship seems the opposite when only landings are studied.

As Verga pointed out, there are multiple factors that can affect the number of reports. Population changes over the years, and cities grow in size absorbing towns that become suburbs. We have already mentioned the *investigator effect* as another factor that can affect the results, and some of the studies effectively show an increase in number of reports in regions where these groups existed.

We can even extend the argument to a *hot spot effect*: In FOTOCAT, the state of New Mexico is the sixth least populated per square kilometer of the 50 states in the U.S. However, it is the sixth contributor in number of reports, most likely due to the fame of the Roswell incident. A similar effect can be presumed for Nevada (9th least populated, 16th contributor) and the famous Area 51.

To be able to compare different areas, the time interval of the databases should be the same for every region accounted for, and the efficiency in collecting reports should be the same. Using FOTOCAT again as an extreme example, countries like Bosnia-Herzegovina have existed for a short time compared with countries like France and Spain. Also, it does not seem reasonable to have only one entry for a country like Bangladesh with 160 million inhabitants.

Poher and Weiller showed that even visibility and different meteorological conditions throughout the year have an effect on reports from different geographical areas.

All these effects add noise to the results and make a direct quantitative comparison between different databases difficult. Thus, only trends can be compared. However, the main issue in past studies is that they used different variables to describe the relationship between number of reports and population. Saunders already noted that taking Vallée's first negative law as stated in 1966 literally was different with respect to the relationship N/P vs. δ that he plotted in 1968. Other investigators have used more exotic variables like the Number of reports per unit area (N/S), or the Number of reports per inhabitant and unit area. Therefore, it is not surprising that different results have come up during these years, since the variables explored have been different each time.

For that reason, we have calculated the same set of correlation coefficients for most of the studies we have reviewed, and also added new data using databases not taken into account before. Three correlation coefficients were calculated for each database:

- Number of reports vs Population (N vs P)
- Number of reports vs Population Density (N vs δ)
- Number of reports per inhabitant vs Population Density (N/P vs δ).

The results are shown in Table 4, which also shows values and results for some of the studies for which we could not calculate the correlation coefficient. The reported values are shown instead.

We can see a very clear and unmistakable trend:

- The Number of reports correlates **directly** with Population.
- The Number of reports correlates **directly** with Population Density.
- The correlation with Population Density is weaker than with Population.
- The Number of reports per capita correlates **inversely** with Population Density. Even if not all p-values are statistically significant, and most correlations are weak, they all show a negative sign, except for López et al. (1978).

Looking at Table 4, we can now see that all reported data fit coherently when compared with the proper variables of the other studies. It is also worth remarking that there is no difference in correlation trends when analyzing all kind of reports, landings, or images. Past studies did not show contradictory results. The issue was that **different variables were being compared**. There is only one exception: The original 1966 claim of an inverse correlation between N and δ is the only contradiction with the data presented. We will look further into it in the next section. Also, we will look into the fact that N/P vs δ is an inverse correlation and what it means with respect to the N vs δ direct correlation.

TABLE 4
Summary of Results of the Works Analyzed in this Paper

Source	Reports (N)	r (N vs P)	p-value	r (N vs δ)	p-value	r (N/P vs δ)	p-value	Comments
Vallée 1966	151	0.307	0.007	0.245	0.026	−0.307	$7\cdot10^{-4}$	Landings in departments of France (population from Weiller 1980)
		0.286	0.11	−0.007	0.49	−0.429	0.033	Landings in regions of France (population taken from Weiller 1980)
Vallée 1968	8,260	-	-	-	-	(Inverse)	-	Deduced from graph
Saunders 1975	18,122	(0.723)	-	-	-	-	-	Data from simple correlation calculations
Poher 1975	-	-	-	(Direct)	-	-	-	Deduced from graph
Ballester 1976	200	-	-	-	-	−0.40	0.0033	Landings in provinces of Spain
Bettini et al. 1977	299	0.35	0.0025	0.21	0.048	−0.37	0.001	Departments of France
López et al. 1978	1,721	0.79	$2.1\cdot10^{-11}$	0.71	$9.1\cdot10^{-9}$	0.009	0.52	All kind of sightings. Provinces of Spain
	237	0.49	$4.0\cdot10^{-4}$	0.40	0.0040	−0.23	0.069	Landings in provinces of Spain
Fernández & Manuel 1980a	49	-	-	-	-	−0.53	0.004	Landings in provinces of Spain
Fernández & Manuel 1980b	208	0.74	$1.1\cdot10^{-5}$	0.67	$1.1\cdot10^{-4}$	−0.53	0.0027	Regions of the Comunidad Valenciana
		(0.91)	-	-	-	-	-	Villages in the Comunidad Valenciana
Weiller 1980	1,919	0.54	$1.4\cdot10^{-8}$	0.15	$5.3\cdot10^{-6}$	−0.31	0.0014	All sightings in France except for Paris
	483	0.81	$1.9\cdot10^{-6}$	0.77	$1.0\cdot10^{-5}$	−0.09	0.34	Most luminous departments in France (22 departments)
	398	0.79	$5.4\cdot10^{-6}$	0.77	$1.5\cdot10^{-5}$	−0.12	0.30	Least luminous departments in France (22 departments)
Verga 1981	326	0.67	0.0012	0.45	0.023	−0.31	0.087	Landings. Population by regions of Italy
Ballester & Fernández 1987	205	0.37	0.009	0.14	0.17	−0.57	$1.5\cdot10^{-6}$	LANIB, high-strangeness landings in Spain
	310	0.60	$1.9\cdot10^{-6}$	0.47	$4.0\cdot10^{-4}$	−0.29	0.0026	NELIB, negative landing cases (IFO) in Spain
Breysse 1993	643	(0.887)	-	(0.386)	-	-	-	Cases in states of the U.S.
	310	(0.392)	-	(0.126)	-	-	-	Cases in departments of France
Ballester 2014	11,060	0.70	0.040	0.73	0.059	−0.91	$5.4\cdot10^{-4}$	Photo & video. Sorted by continents
CUCO	8,298	0.79	$3.7\cdot10^{-12}$	0.71	$4.3\cdot10^{-9}$	−0.06	0.34	All kind of sightings. Provinces of Spain
ALLCAT	953	0.60	$1.7\cdot10^{-6}$	0.46	$4.2\cdot10^{-4}$	−0.45	$1.6\cdot10^{-4}$	Landings in Spain
FOTOCAT	2,785	0.81	$2.9\cdot10^{-13}$	0.32	0.039	−0.45	0.024	Photo & video in states of the U.S.

Correlation coefficients (*r*) were recalculated when possible, taking logarithms of the variables. Values in brackets are those reported or deduced from the references, for which *r* could not be calculated.

First Negative Law

> The geographic repartition of the landing sites in 1954 is inversely correlated with population density.

The only support for this claim is a map of France showing the landing sites (reproduced in Figure 1 Left) that seems to *qualitatively* support the statement. Let us have a closer look at Figure 1. Enclosed by dashed lines are areas with a population density greater than 60 inhabitants per km^2. We can count 159 landings, 46 of them inside the *high-density* areas, and 113 outside them. That is, 28.9% of landings occurred in high-density areas. On the other hand, in a rough calculation counting the number of pixels in each zone, we can estimate that high-density areas represent only 26% of France.[1]

Let us suppose that UFO landings occur uniformly throughout the country. Let us also assume for the moment that population density has no effect at all. That is, every landing has the same probability of being witnessed regardless of the population of the area. Under those assumptions that we can label as the *null hypothesis*, the number of reported landings must be proportional to the area taken into account. If we take an area representing 26% of the total, then an average of 26% of the landings (41 landings) should occur in that area. If for any reason UFOs tried to avoid those areas, then this value would decrease accordingly, leading to a negative law. With a lower probability of a UFO visiting the area, the number of UFO landings should be lower than the null hypothesis prediction. However, what we find is that the actual value (46 landings, 28.9%) is greater than expected by our null hypothesis, meaning that some factor is favoring their sighting. Since we are considering only two regions that differ in population density, this factor could be the one increasing the probability of a UFO being witnessed in the >60% inhabitants per km^2 areas, and thus we should have a direct relationship.

We have to remark that this difference between the null hypothesis and the actual number of landings is *not statistically significant* (p-value = 0.4). This means that this data does not support either a direct or inverse relationship. On the other hand, the same data does not support Heuyer's *psychosis hypothesis* either, which was the main point being made by Vallée at that time.

This analysis was based only on Vallée's graph. It suggests that there is not any inverse relationship between sightings and population density, but it is not especially accurate, and yields a not statistically significant result. Therefore, we have made a second analysis looking for a more accurate quantification of the relationship.

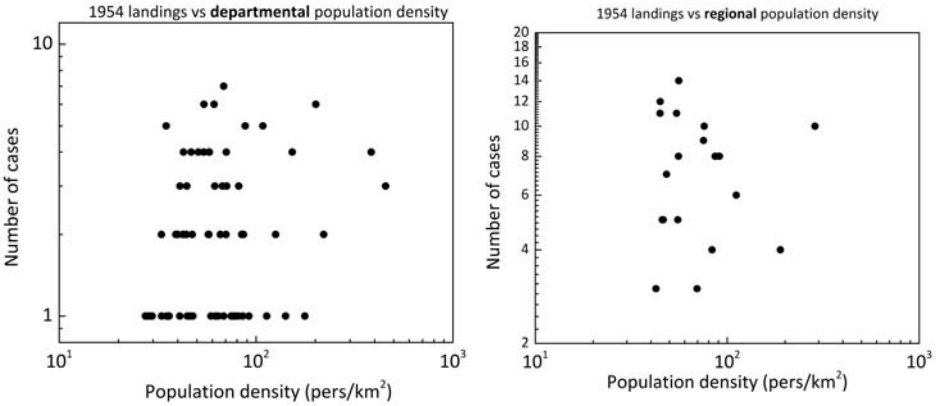

Figure 4. Left: Number of landings versus population density in French departments. Right: Number of landings versus population density in French regions. Landings data from Vallée (1966), population data from Weiller (1980).

We were able to unambiguously locate 151 of the landings. Since the 1966 paper did not contain data about population and regions, we used that of Weiller, circa 1973 (Weiller 1980). 151 landings are very few data points, and divided into 90 departments means many of the areas have 1, 2, or no datapoints at all. Looking for a different distribution, we also tried to group the data by regions, where each region includes several departments. Plots of number of landings versus population density can be seen in Figure 4.

The correlation coefficients were calculated for N vs P, N vs. δ, and N/P vs. δ, and are shown in Table 4, along with the p-values assuming as an alternative hypothesis the correlations being greater than (i.e. direct) or lower than (i.e. inverse) zero. Results are summarized in Table 5.

TABLE 5
Correlations for Two Different Geographical Distributions of 151 Landings in France, Taken from Vallée (1966)

	r	p-value
Correlations by regions		
N vs P	0.286	0.11
N vs δ	–0.007	0.49
N/P vs δ	**0.429**	**0.033**
Correlations by departments		
N vs P	**0.307**	**0.007**
N vs δ	**0.245**	**0.026**
N/P vs δ	**–0.307**	**0.0007**

Bold p-values are statistically significant at 0.05 level.

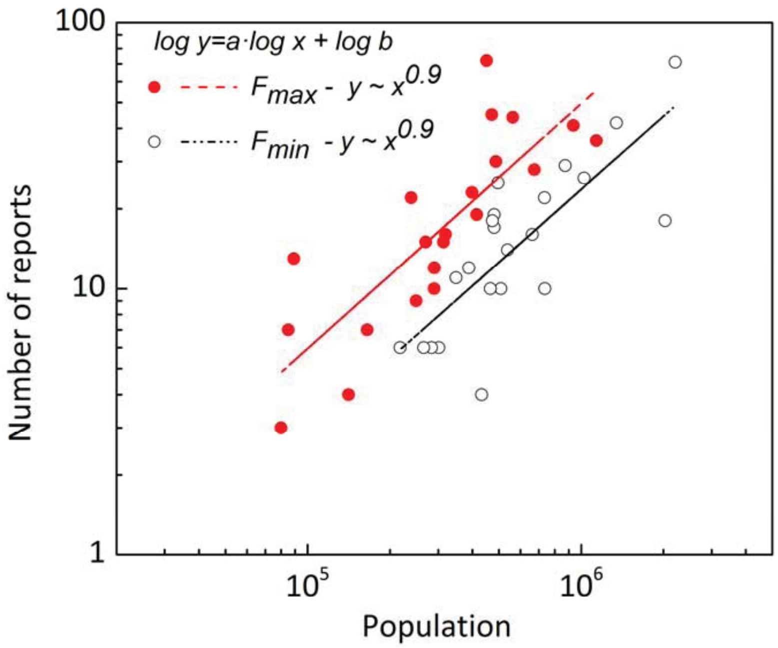

Figure 5. Number of reports versus population in French departments, re-elaborated from Weiller (1980). Open circles represent the 22 least luminous (F_{min}) departments. Closed circles represent the 22 most luminous (F_{max}) departments.

Table 5 shows that the results in a regional distribution are not statistically significant, except for the N/P vs δ inverse correlation. The departmental distribution shows that all correlations are statistically significant at the $p = 0.05$ level, even if they are by a small margin.

We must conclude that the negative law, as it was stated, had no basis since an inverse correlation between Number of Cases and Population Density could not be found in any case.

Sub-Lineal Relationship between Number of Reports and Population

Let us look now at the study by Weiller (1980). Figure 5 reproduces his results after the separation of 2,000 reports into two series. F_{min} corresponds to the 22 departments with fewer hours of light per year. F_{max} corresponds to the 22 most luminous departments. In both series, not only the light available because of geographical latitude, but also meteorological data relative to cloud cover were taken into account. Both series show a clear direct relationship between number of reports and population.

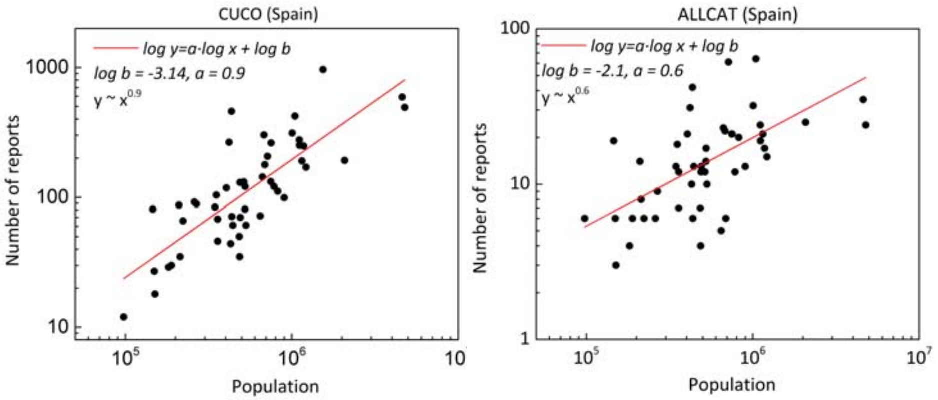

Figure 6. Left: Number of reports versus population in Spanish provinces for CUCO. Right: Number of reports versus population in Spanish provinces in ALLCAT (2013).

Using the logarithmic scale, Weiller also calculated the slope of the linear regression, whose value is lower than 1. As discussed in Appendix B, this value corresponds to a direct function with *sub-lineal growth*, i.e. the proportion in which the number of reports grows is lower than the proportion in which population increases.

We have looked for this sub-lineal growth in other databases, to check for the reproducibility of the results. First, we used the data from (CUCO), a database including all types of reports, divided by Spanish provinces (more than 8,000 reports in 50 provinces). The cases are dated over several decades (60s, 70s, 80s, and 90s). We used population data from 1986, from the Spanish National Institute of Statistics (INE). After plotting the data in logarithmic scale, the linear regression again shows a sub-lineal growth (Figure 6 Left). ALLCAT, a database of 953 landings in Spain, was also examined using the same population data as in CUCO. Figure 6 Right shows the sub-lineal growth.

We did the same with data from U.S. states in FOTOCAT (2,785 cases in 50 states and Washington, D.C.). Population data have been obtained from Wikipedia. Again, a *sub-lineal growth* is observed (Figure 7 Left). Next, we checked the data from Fernández and Manuel on the regional database centered on Comunidad Valenciana (Fernández & Manuel 1980b), and the sub-lineal relationship showed up again.

Finally, NELIB was also explored as the *control group* (as originally thought by Ballester and Fernández). The sub-lineal growth, once again, is clearly seen (Figure 7 Right). Table 6 summarizes the values of correlation coefficient, p-values, and lineality coefficient *a*, given by

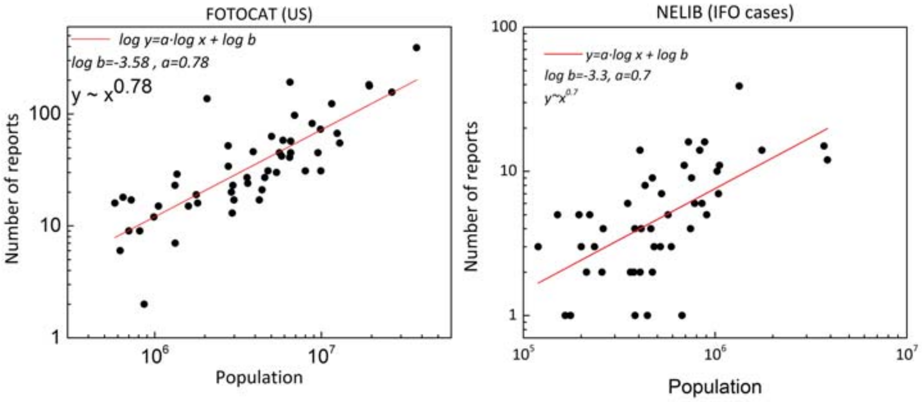

Figure 7. Left: Number of reports versus population in U.S. states for FOTOCAT (catalog of images and video footage). Right: Number of reports versus population in NELIB (catalog of IFO cases).

$$Y = b \cdot X^a$$

for all these databases.

It is worth remarking on the different sorts of databases used: Weiller's and CUCO are databases of all kinds of reports. Weiller's data was divided into two regions of high and low luminosity (hours of day per year). ALLCAT is a landings database, and FOTOCAT is a photo and video/film footage database. We have regional data covering much smaller areas compared to the other databases (Fernández and Manuel 1980b). And we also explored NELIB, a negative (IFO) landings database.[2]

Despite all their differences, all of them behave in the same way. This result allows us to understand now why the number of reports per inhabitant is inversely correlated with population density. Since we have established a relationship of the type $N \sim P^a$, with $a < 1$, if we calculate the quantity N/P we get:

$$N/P \sim P^{a-1}$$

Since $a - 1 < 0$, the relationship becomes inverse. Also, as population density and population are proportional ($P = \delta \cdot S$), this inverse relationship is inherited by population density. Thus, all the results shown in Table 4 agree with a sub-lineal growth, whose main feature is that even if the number of reports grows with population (or population density) it grows in a lower proportion. This is exactly what Rospars (2014) is showing in his results: an increase of total reports, as well as a decrease of reports per capita with population density, due to sublineality.

TABLE 6
Summary of Correlation Coefficients and Lineality for Different Data and Catalogs

Source of data	r (log N vs log P)	p-value	Lineality (a)
Weiller 1980 (all data)	0.54	$1.4 \cdot 10^{-8}$	0.6 ± 0.1
Weiller 1980 (22 most luminous departments)	0.81	$1.9 \cdot 10^{-6}$	0.9 ± 0.2
Weiller 1980 (22 least luminous departments)	0.79	$5.4 \cdot 10^{-6}$	0.9 ± 0.2
Fernández & Manuel 1980b (Comunidad Valenciana)	0.74	$1.1 \cdot 10^{-5}$	0.6 ± 0.2
NELIB (Ballester & Fernández 1987)	0.60	$1.9 \cdot 10^{-6}$	0.7 ± 0.2
CUCO	0.79	$3.7 \cdot 10^{-12}$	0.9 ± 0.1
ALLCAT	0.60	$1.7 \cdot 10^{-6}$	0.6 ± 0.1
FOTOCAT—U.S. states	0.81	$2.9 \cdot 10^{-13}$	0.78 ± 0.08

Isolated Places, Population Density, and Homogeneity

Poher and Vallée showed in 1975 that more than 75% of sightings took place in isolated places (Figure 2). How does that reconcile with the results from previous sections?

When we speak about population density we are making a homogenization in the area of a department, province, or state. We assign a homogeneous population density to each of them, but the population is actually concentrated in villages, towns, and cities, whereas the space between them (roads, countryside) is in fact empty, uninhabited, or isolated. We are considering that in a region, every square kilometer is occupied by the same number of inhabitants, but actually most of it is uninhabited.

Even inside a city we can find isolated places. The city of Madrid has a population density of about 5,000 inhabitants per square kilometer. But in a park like Parque del Retiro with an area of 1.18 km^2, at 21:00 h on a winter night, it is difficult to find 5,000 people walking around the place. We should understand population density in the way Verga wrote about it:

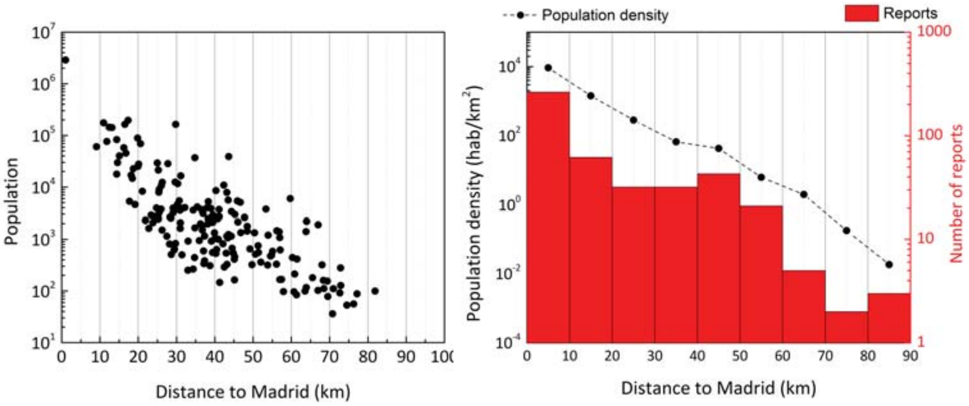

Figure 8. Left: Distribution of population as a function of the distance of villages of Madrid province to the city center, 1996 population (source: INE). Right: Distribution of UFO reports in the province of Madrid as a function of distance to city center.

an area with high population density is an area with *"high probability of witness presence."*

Let us consider the province of Madrid, with the data available in CUCO. Figure 8 Left shows the distribution of population in the villages and towns of the province, according to 1996 demography, as a function of the distance to Madrid city center. There is a clear and simple trend, and the farther a village is, the lower its population. Some towns near the capital (<10 km) were absorbed years ago, and they do not appear anymore as independent towns, but are considered suburbs of Madrid, and their population is included in that of Madrid.

The greater the distance from Madrid, the lower the population of the villages, and the easier it is to find isolated places. According to Poher and Vallée, these places should generate a higher number of UFO reports. We grouped the reports depending on their distance from Madrid in bins of 10 km: from 0 to 10 km, 10 to 20 km, etc., and the trend we find again is to have more reports in places closer to Madrid, which have a *"higher probability of witness presence"* (Figure 8 Right), including places close to the urban areas of Madrid.

Let us remember that such a *fine structure* of the population distribution was also studied by Fernández and Manuel (1980b) for Valencia. They also showed a higher concentration of reports in the vicinity of the three main cities: Alicante (22 cases), Valencia (35 cases), and Castellón (10 cases).

Discussion and Conclusions

We have reviewed several works that studied the patterns in the geographical distribution of UFO reports. These studies had different and apparently contradictory results because they compared different variables. Correlations were recalculated to compare the same variables, and we have shown that all the results are coherent, consistent, and show the same trends.

We have to remark that these trends are also followed by NELIB, a database of negative landings (i.e. IFO cases). Following the idea behind the compilation of this database, we can see that there is no difference between UFO and IFO cases.

There are at least two possible interpretations:

- UFO databases contain a high number of cases that can be solved due to misperceptions, fakes, hoaxes, or any other mundane causes. If there is a real UFO phenomenon, its pattern is hidden within its noise.
- UFO cases are indistinguishable from IFO cases. It is just straightforward to think that all UFO reports can be explained in terms of mundane causes, even if there is not enough information to find that cause.

Regarding the first interpretation, we have to remember that the LANIB database gathered cases with a high strangeness factor, but does not show any difference from other databases.

The so-called First Negative Law stated that the geographical distribution of landings in 1954 was inversely correlated with population density. However, the data available did not allow for making that statement. After checking the data, we found that the correlation of number of reports and population density was direct, although at a low statistically significant value.

A basic and more accurate description of the geographical pattern of UFO reports would be that more reports come from more populated areas, but reports grow in a slower proportion than population does.

Historically, the inverse relationship was interpreted as some kind of intelligence in the UFO phenomenon that avoided populated areas. However, the direct relationship is only a consequence of a higher probability of witness presence in an area, and compatible with a uniformly random distribution of UFO events.

We have checked that this direct relationship is valid for large distributions like continents, counties, or provinces, and also for micro-distributions like villages in a single province. One of the points to take into account is that population density represents a homogenization of an area, and it does not mean that in every square kilometer there are a constant number of people. Therefore, we must understand that in populated places, the probability that any person may witness a UFO event is higher than in low-populated places.

We have been comparing trends, but we have not tried to compare values directly, since as explained in previous sections, there may be many factors contributing to the number of reports in different areas and most likely adding noise. In this regard, Rospars did directly compare values for UFO and IFO distributions belonging to the same catalog, and found no differences when considering departments, but did find highly significant differences when considering the density of French communes.

Is it possible to go into a *finer* structure to find or reject differences? Some analyses have been done with villages, the smallest geographical entities that are usually defined. Density (whatever the field of study) is a variable defined to average over a representative portion, and so reducing the area of averaging would make the value of density meaningless at some point.

Another of the findings of this paper is that even if the total number of sightings increases as population increases, this growth is not lineal, but sub-lineal. That is, doubling the population means that the probability of any person witnessing a UFO should double, too. But the actual number of sightings increases to less than double. What is the origin in this reduction in the proportion of witnessed events as population grows? Is it because fewer people can see UFOs in more populated areas? Is it more difficult to witness those events because of any factor not yet taken into account? Poher, Vallée, and Weiller noted that a lower visibility due to meteorological conditions reduces the probability of a UFO being witnessed, and Condon also suggested the possibility that city lights are reducing the visibility. Are city lights reducing the probability of witnessing events that can be reported as UFOs?

We think that the features of the relationship between UFO sightings and population density are finally well-known, and it should be possible to test different models to reproduce them. López et al. produced a semi-empirical model based on different probabilities: area, population, climate, and orography. Their model assumed a uniformly random UFO phenomenon, and differences in number of reports come from differences in parameters not related to UFOs themselves: It can be regarded as a *null-hypothesis model*.

Breysse studied the relationships using proportional and quadratic (super-lineal) models, but those models are not supported by our results showing a sub-lineal growth. Toulet tried a *contagion* model (Toulet 1974) with interesting results, but it did not take into account population density.

López et al.'s model based on probabilities seems to be the right way to go. An improvement accounting for a lower visibility due to an increase in public lights in more populated areas could be interesting to introduce. Also, adding Toulet's contagion model could be a good basis for Monte Carlo simulations, and the study of UFO waves.

Acknowledgements

The author thanks Vicente-Juan Ballester Olmos for all the information, discussions, comments, and references.

Notes

[1] A rough calculation of the number of pixels in a scanned image is not difficult to produce with image editing software. Higher-resolution images may yield more accurate values. We encourage other researchers to perform their own estimation.

[2] We have to point out that NELIB is included in ALLCAT, and therefore it is not surprising that they yield similar values.

Appendix A: Correlation Coefficient

The mathematical tool most used in the works reviewed here is Pearson's correlation coefficient (Pearson Correlation). This tool allows the study of correlations between the data to be able to establish with some certainty whether a variable is dependent on another.

The correlation coefficient takes a value between −1 and 1. A negative correlation means that the relation is inverse: The dependent variable decreases when the independent variable is increased. A positive value means that the relation is direct: The dependent variable increases when the independent variable is also increased. The correlation (or anti-correlation in the negative case) is stronger when it is closer to 1 (or −1). On the other hand the closer to 0 the weaker the correlation.

From the correlation coefficient, however, the exact mathematical relationship between both variables cannot be deduced. But it can be shown that when the relationship between two variables X and Y is proportional (or lineal)

$$Y = a \cdot X + b \tag{1}$$

the correlation coefficient is closer to 1 (or −1, if $a < 0$). Since there is no way to a priori determine whether a relationship is lineal, a more general relationship can be supposed, such as

$$Y = b \cdot X^{a} \tag{2}$$

Taking logarithms at both sides, and applying their properties, we arrive at:

$$\log Y = a \cdot \log X + \log b \tag{3}$$

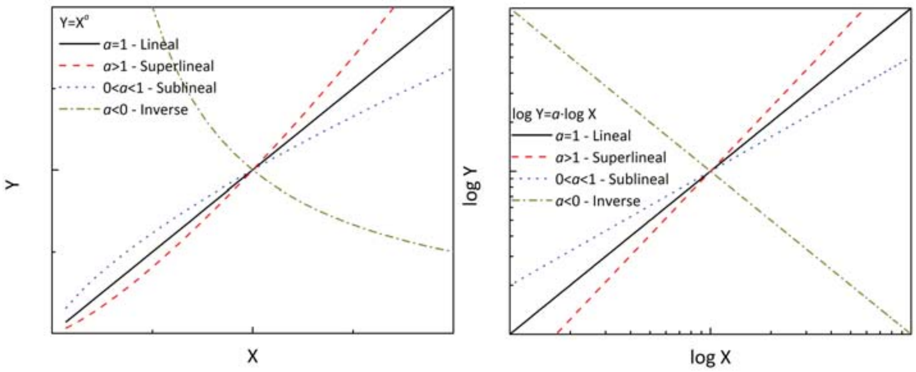

Figure 9. Left: Direct relationships (lineal, sub-lineal, and super-lineal), and inverse relationship between *X* and *Y*. Right: The same relationships between log *X* and log *Y*.

If there is a lineal relationship between log X and log Y, the correlation coefficient would tend to its maximum value (or minimum if $a < 0$), and the correlation is stronger than that without taking logarithms.

Appendix B: Lineality and Sub-Lineality

Equation 1 represents a lineal relationship between X and Y. Equation 2 is a more general equation, which includes a particular case Equation 1 when $a = 1$. When the relationship is lineal, it means that Y increases in the same proportion in which X is increased. That is, if X is doubled, Y doubles.

When $a > 1$, Y increases in a proportion higher than the increase in X. It is a super-lineal relationship. And when $a < 1$, but $a > 0$, the proportion in which Y increases is lower than the proportion in which X is increased, and it is a sub-lineal relationship. In these three cases, the correlation coefficient between log X and log Y is positive, showing a direct relationship.

In the last case, when $a < 0$, we find an inverse relationship, and thus, the correlation coefficient between log X and log Y is negative. Figure 9 shows these four cases graphically.

References Cited

ALLCAT (2013). UFO Landings in Spain and Portugal. Compiled by V. J. Ballester Olmos (unpublished).

Ballester Olmos, V. J. (1976). Are UFO Sightings Related to Population? *Proceedings of the 1976 CUFOS Conference*, edited by Nancy Dornbos, Center for UFO Studies, Northfield, pp. 15–24.

Ballester Olmos, V. J. (2014). UFOs by continent. FOTOCAT. http://fotocat.blogspot.com.es/2014_03_02_archive.html

Ballester Olmos, V. J., & Fernández Peris, J. A. (1987). Enciclopedia de los Encuentros Cercanos con OVNIS. Barcelona: Plaza y Janés, pp. 45–46, 349–353.

Ballester Olmos, V. J., & Guasp, M. (1981). Los OVNIS y la ciencia. Barcelona: Plaza y Janés.

Betttini, Bourrou, Desbois, Gittou, Lenoir, & Venturini (1977). Décomposition d'un doute. *Lumières dans la Nuit*, *167*, 15–22.

Bonabot, J. (1971). Un étude delicate (2 partie). Visiteurs Spatiaux. *GESAG Bulletin, 25*.

Breysse, D. (1993). Project Bécassine. Analyse statistique de la base de donnees. *Lumières dans la Nuit*, *319*, 26–33.

Condon, E. (1968). Summary of the Study. In *Scientific Study of Unidentified Flying Objects* edited by Daniel S. Gilmore, New York: E. P. Dutton, p. 22.

CUCO Catálogo Unificado de Casuística OVNI (Unified Catalog of UFO Cases) of all-kinds of UFO cases in Spain, Portugal, and Andorra. Compiled by J. P. González and supported by Fundación Anomalía, with data from 23 different catalogs.

Fernández Peris, J. A., & Manuel Garijo, L. (1980a). Estudio analítico de las observaciones OVNI tipo I ocurridas en España peninsular durante el año 1974.

Fernández Peris, J. A., & Manuel Garijo, L. (1980b). Distribución del fenómeno OVNI en el País Valenciano: Bases para un estudio comarcal.

FOTOCAT. Worldwide UFO Photograph and footage catalog compiled by V. J. Ballester Olmos. http://fotocat.blogspot.com

INE Instituto Nacional de Estadística. http://www.ine.es

López D. G., Ares de Blas, F., & Salaverría Garnacho, A. (1978). Bases para una modelación teórica del fenómeno OVNI. Actas del primer congreso nacional de ufología. Barcelona: CEI.

Pearson Correlation. Pearson product-moment correlation coefficient. Wikipedia. http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

Poher, C. (1975). Etudes et réflexions à propos du phénomène OVNI. *L'Aéronautique et l'Astronautique*, *52*(3), 69–74.

Poher, C. (1976). Etude statistique des rapports d'observations du phenomene OVNI, 1971–1976, pp. 42–48. http://www.geipan.fr/typo3conf/ext/dam_frontend/pushfile.php?docID=1627

Poher, C., & Vallée, J. (1975). Basic Patterns in UFO observations. AIAA Paper 75-42. 13th Aerospace Sciences Meeting, January.

Rospars, J. P. (2014). Qualitative expert evaluation and quantitative characterization of official reports on alleged unidentified aerial phenomena in France (1970–1979). Paper presented at International Workshop CAIPAN "Collecte et Analyse des Informations sur les phénomènes Aéroespatiaux Non indentifiés" organized by GEIPAN Centre National d'Etudes Spatiales (CNES), Paris, France 8–9 July 2014.

Saunders, D. (1971). People, Places, and UFOs: A Multiple Correlation Analysis. UFOCAT Results—4 (draft).

Saunders, D. (1975). Extrinsic factors in UFO reporting. AIAA Paper 75–43. 13th Aerospace Sciences Meeting.

Toulet, F. (1974). Les observations d'OVNIs obeissent-elles a un processus "contagieux"? *Phénomenès Spatiaux*, 40–42(June–Dec.), 8–10.

Vallée, J. (1966). The Patterns behind the UFO landings. *Flying Saucer Review* (Special Issue 1 The Humanoids, October–November), 8–27.

Vallée, J. (1968). Analysis of 8,260 UFO sightings. *Flying Saucer Review* ,*14*(3), 9–11.

Verga, M. (1981). Il punto sulla distribuzione geographica. http://www.ufo.it/documenti/geografica.pdf

Weiller, A. R. (1980). Existe-t-il un phénomène atmosphérique inconnu associé à une certaine classe d'ovnis? *La Météreologie VIe Série*, *22*, 97–108.