# To Press or not to Press: Relationships between the Earliest Chant Manuscripts Reconsidered*

Jacqueline ELEMANS

The Netherlands

**Abstract:** The occurrence of the isolated pressus in the Gregorian Mass Proprium chants is examined in 11 early manuscripts. The early written tradition for the pressus in isolation is not at all uniform. The frequency of a pressus notation varies widely from one source to another. The author sketched the potential power of standard non parametric statistical theory in order to overcome the problem of contamination that until now hampered the study of relationships between manuscripts. The notation of the pressus in isolation is used as an illustration.

**Keywords:** Gregorian Mass, Gregorian notation, pressus

The aims of this paper are, to point out reasons why in chant variants analysis one should apply more refined statistical methods than have been used in the past, and to demonstrate the heuristics of one such method by means of an example. A full discussion, though necessary, must wait for another occasion.

I came upon this approach when, for a very different reason, I looked into the occurrence of the isolated pressus (hereafter designated as pressus, the qualification isolated being implied) in the Gregorian Mass Proprium chants. I looked at 11 early manuscripts, facsimiles of which are readily available.

Manuscripts consulted:

| My siglum | Solesmes siglum | Identification |
|---|---|---|
| E | Mur3 | Einsiedeln 121 |
| B | Bab1 | Bamberg lit. 6 |
| G | Gal2 | Sankt Gallen, Stiftsbibl. 339 |
| L | Lan | Laon 239 |
| Ch | Cha1 | Chartres 47 |
| MR | Eli | private collection |

* An earlier version of this paper was presented at the IMS conference, Leuven 2002. I thank Peter Jeffery for his interest in the idea, and David Hughes for his critical comments on what he kindly calls my elliptic style of writing but what really is muddy thinking.

| My siglum | Solesmes siglum | Identification |
|---|---|---|
| Mp | Dij1 | Montpellier, Fac. Méd. H159 |
| Bv33 | Ben1 | Benevento, B. Cap. VI 33 |
| Bv34 | Ben5 | Benevento, B. Cap. VI 34 |
| VL | Ben9 | Rome, Vat. Lat. 10673 |
| An | Lav | Rome, Angelica 123 |

To my surprise I found that the early written tradition for the pressus in isolation is not at all uniform. Even the early Sankt-Gallen manuscripts do not agree with one another for more than about 50% of their pressus. I found that in all but 7 of the 99 places where the melodic progression, according to the Graduale Triplex, would allow one to occur, one or other manuscript does in fact write one. There are sufficient lacunae that in any of these 7 places one or other manuscript might indeed have used one. In principle therefore, all 99 places are potential pressus candidates.

Moreover, I found that the frequency of a pressus notation varies widely from one source to another, as can be seen in *Table 1*. These observations raise intriguing questions. One of these I want to address here, viz. the potential relationship between the notations, and by implication of the manuscripts.

*Table 1:* % isolated pressus in 11 manuscripts

| ms | Bv33 | VL | Bv34 | E | B | G | L | Ch | MR | An | Mp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| % | 47 | 45 | 38 | 21 | 15 | 13 | 9 | 9 | 8 | 5 | 70 |

The theoretical approaches to (textual) variants analysis that I found in the literature[1] prove unable to cope with the present distribution of notations. In short, the theoretical stance on how to deal with contamination determines what relationships one finds. This observation reinforces the conclusion of the monks of Solesmes, who saw contamination as their biggest problem in their attempt towards a critical edition of the Graduale Romanum,[2] and it is another illustration of Grier's observation that "… in virtually all cases, the stemma does not represent absolute, objective truth …".[3] While keeping this warning in mind, a reappraisal of existing methods from a fresh viewpoint might be helpful.

[1] As discussed in Vinton Dearing, *Principles and practice of textual analysis*, University of California Press (1974). See also Antonín Hruby, Statistical methods in textual criticism, in: *General Linguistics* 5, no. 3, supplement 1962, pp. 77–138. In the discussion following the presentation of the paper Charles Atkinson brought to my attention the work of Joseph Diamond, *A tradition of 3 tropes* (Institute of Medieval Music, Ottawa 1991). The first chapters of this work present a thorough critique of existing theoretical approaches. However, Diamond does not criticise the mathematics of the methods of the Solesmes monks.

[2] Les moines de Solesmes, *Le Graduel Romain*, édition critique, tome II, les sources; tome IV, le texte neumatique (Solesmes, 1957 and 1960–62).

[3] James Grier, *The Critical Editing of Music* (Cambridge University Press, 1996), 65.

## Properties of variants in western chant

Indeed, one thread that runs through the work of David Hughes[4] is the observation that most of the variants found in chant manuscripts:
– are of a trivial nature,
– appear largely at random, and
– allow only a few alternative notations.

Then it is obvious that we cannot neglect the possibility that scribes, though working independently, arrive at the same notation, at the same place, purely by accident. In the face of few alternatives, if one scribe disagrees with a second one, he cannot fail to agree with some third. It is an example of a familiar puzzle: if I have a bag of socks in two colours, how many do I need to grab in darkness, to make sure that I have at least one matching pair. We must *expect*, rather than *conclude*, that contamination is an important phenomenon with such variants, and find a way to deal with it.

There is yet a very different, but again fundamental reason why existing theories of variants analysis might not apply to plainchant. These methods essentially assume that all sources are in manuscript form. But it is evident that for a long time chant was transmitted orally and that the oral transmission remained important for a while even after notation became possible. In other words, the aural/oral tradition continued to influence the written tradition. First of all, a chant manuscript might be seen as a witness of the local contemporary oral tradition, which, apart from reform movements, is slowly evolving and subject to influences from elsewhere. In our age of computer games and virtual reality we might see the oral tradition at a particular time and place as a virtual source with its own transmission history.

So, if we consider actual manuscripts as copies from virtual sources, and if there is a non-negligible probability of agreement by chance, is there anything that we can say about the coherence of the written tradition? The answer is, yes we should still be able to.

We might accept the new starting point that initially all manuscripts are assumed to be unrelated, unless the opposite can be proven. How do we then provide a proof?

## Chance agreement

First, let us develop the idea of accidental agreement. I use the pressus notations in Mp and Bv34 as an example. There is only one alternative, so there are

[4] David Hughes has worked a lifetime with chant variants. His latest published paper of a general nature is: The implications of variants for chant transmission, in: Cahn and Heimer, *De musica et cantu* (Olms Verlag, 1993) 65–73. His latest publication is: The alleluias Dies sanctificatus and Vidimus stellam as examples of late chant transmission, in: *Plainsong and Medieval Music* 7 (1998) 101–128.

two possible notations. If the manuscripts are independent of one another, we should not be able to guess from the notations in Mp what Bv34 shows and vice-versa. If in Mp we find 70% pressus, or whichever actual value, we should find the same proportion if we look only at the places where Bv34 writes a pressus and again the same proportion for those places where Bv34 does not write a pressus.

*Table 2a* shows the calculation and a comparison with the actual agreement found. We see that the agreement between actual value and estimated random value is good: the difference is 1% or 1 case for each type of variation. We must conclude that we do not see a difference between the actual distribution and a chance distribution for the pressus notations in Mp and Bv34.

*Table 2a:* Estimated level of agreement for unrelated manuscripts and actual agreement, example of Mp and Bv34, 81 cases

| | **Mp** | **Bv34** | | |
|---|---|---|---|---|
| % pressus | 70 | 38 | | |
| % other | 30 | 62 | | |
| variations: | | | | |
| Mp | Bv34 | estimated random % | actual % | difference % |
| pressus | pressus | 70*38=27 | 28 | 1 |
| pressus | other | 70*62=43 | 42 | –1 |
| other | pressus | 30*38=11 | 10 | –1 |
| other | other | 30*62=19 | 20 | 1 |

When we look at E and L, the situation is different, as *Table 2b* shows. The difference between actual value and estimated random value is 7%, 5 or 6 cases. Is this difference large enough to conclude that the notations in E and L are related? What we need is a method to interpret such differences. We need statistical theory. In the following I give a brief description in plain language, in which I nevertheless have tried to catch the essentials as I understand them.

*Table 2b:* Estimated level of agreement for unrelated manuscripts and actual agreement, example of E and L, 79 cases

| | **E** | **L** | | |
|---|---|---|---|---|
| % pressus | 21 | 9 | | |
| % other | 79 | 91 | | |
| variations: | | | | |
| E | L | estimated random % | actual % | difference % |
| pressus | pressus | 21* 9 = 2 | 9 | 7 |
| pressus | other | 21*91 = 19 | 12 | –7 |
| other | pressus | 79* 9 = 7 | 2 | –7 |
| other | other | 79*91 = 72 | 79 | 7 |

**Kappa, the measure of association**

In the seventies, statisticians[5] introduced a concept known as the measure of association between two sets of data, usually referred to as kappa.

Kappa is defined as the difference between the actual agreement and the estimated accidental agreement, expressed as a fraction of the greatest possible difference between the two. The greatest possible difference is defined as between full agreement and (the estimated) accidental agreement. And of course, full agreement is equal to the number of cases in the data.

Kappa is defined as a fraction, and therefore it is independent of the number of cases. For full agreement, kappa = 1, by definition; if the agreement is at random, kappa = 0, also by definition. The calculation of kappa can be implemented in such a way that kappa has the value –1 when there is no agreement at all, that is, if the manuscripts agree to disagree.

Therefore, in the case of two possible notations, as for the pressus, three parameters enter into the calculation of the kappa value for a pair of manuscripts: the percentage of one notation in each manuscript, needed for the estimate of the agreement at random, and the % actual agreement between them.

**The relevance z of a particular kappa value**

A second and an even more important bit of information that statistical theory gives us, is what my textbook calls the z of kappa, and I call it the relevance of a particular kappa value. We need this parameter because we want to know when we must start to worry about a divergence between actual agreement and estimated random agreement. The interpretation of z is carefully worded in the theory: the value of z allows us to determine the probability that a particular non-zero value of kappa might have arisen by chance, even if the sources are independent (so that the value of kappa should have been 0). The relevance z depends on the number of cases.

The two things to remember about z are: 1. as z increases, the likelihood increases that the agreement is not by chance; 2. we must choose a threshold for the absolute value of z. For a small z we attribute the agreement between manuscripts to chance, with a large (positive or negative) z we may suspect a possible relationship.

Furthermore, in my implementation of the algorithms I calculate the correlation of each manuscript with all the manuscripts in the database, including itself. In the latter case we know that kappa = 1 by definition, but the value of z is useful information as it gives the z between a source and its perfect copy.

[5] I used Siegel and Castellan, *Nonparametric statistics for the behavioral sciences*, 2nd international edition (McGraw-Hill, 1988), a standard textbook in The Netherlands. In particular I refer to 284 ff. for their description of the kappa statistic.

This is the largest z obtainable from the available data for this source, and if this value happens to be smaller than the threshold we adopt, we cannot draw conclusions for this source.

**Choice of a threshold value for the relevance**

The appropriate value of the threshold is a matter of our own choice. The conclusion of the theory, simplified, is the following:

At z = 2 (or –2), there is a 5% probability of deciding that two mss are related when in fact they are not.

At z = 3 (or –3), this probability of a wrong decision has decreased to 0.3 %. At larger absolute z values, the probability continues to decrease.

Statisticians therefore often choose z = 3 as a threshold of relevance, but more could be said about which value to choose. For the purpose of the present demonstration I shall accept z = 3 as the threshold value.

**Relationship of kappa with the percentage agreement count**

In what circumstances is it important to calculate kappa and z, rather than the raw % of agreement?

It can be shown that for closely related manuscripts, where the agreement is high, the plain variants % and kappa, though numerically different, will yield the same relationships, but we expect that kappa will give a crisper result over a wider range of disagreement. Moreover, in the relevance z we have a measure that will tell us when the variants distribution becomes indistinguishable from a random distribution.

Thus, the main grouping presented by the monks of Solesmes in their *Le Graduel Romain* study is not challenged in principle by the statistical method. But it is now an open question what levels of contamination may or may not be found by an application of proper statistical theory. The data of the monks should be re-evaluated. We may expect that such a new evaluation will lead to the removal of large amounts of statistically insignificant clutter from their diagrams. We will obtain a better view of the main branches of the manuscript tradition and their relationships. Maybe we will be able to disentangle the apparent Gordian knot and see which manuscripts, if any, show the influence of more than one local tradition. Maybe we will be able to penetrate a little into that nebulous region that remained inaccessible until now.

**Application to the isolated pressus**

The case of the pressus is a useful example to illustrate the possibilities of the approach. After having removed entries with z values smaller than 3, we find positive and significant values of z and kappa shown in *Tables 3a* and *3b*.

*Table 3a:* Relevance: z values better than 3 for the isolated pressus

| | **E** | **G** | **B** | **Ch** | **L** | **MR** | **Bv33** | **Bv34** | **VL** | **An** | **Mp** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **E** | 7.1 | 4.0 | 3.5 | | | | | | | | –4.1 |
| **G** | 4.0 | 5.4 | 3.4 | | | | | | | | –4.8 |
| **B** | 3.5 | 3.4 | 5.6 | 3.2 | 3.1 | | | | | | –5.5 |
| **Ch** | | | 3.2 | 3.9 | 3.3 | | | | | | –5.1 |
| **L** | | | 3.1 | 3.3 | 3.9 | | | | | | –5.0 |
| **MR** | | | | | | 3.6 | | | | | –3.1 |
| **Bv33** | | | | | | | 8.0 | | | | |
| **Bv34** | | | | | | | | 9.2 | | | |
| **VL** | | | | | | | | | 5.7 | | |
| **An** | | | | | | | | | | | –4.7 |
| **Mp** | –4.1 | –4.8 | –5.5 | –5.1 | –5.0 | –3.1 | | | | –4.7 | 7.8 |

*Table 3b:* Kappa values at z >3

| | **E** | **G** | **B** | **Ch** | **L** | **MR** | **Bv33** | **Bv34** | **VL** | **An** | **Mp** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **E** | 1.00 | 0.64 | 0.54 | | | | | | | | –0.47 |
| **G** | 0.64 | 1.00 | 0.61 | | | | | | | | –0.55 |
| **B** | 0.54 | 0.61 | 1.00 | 0.70 | 0.66 | | | | | | –0.63 |
| **Ch** | | | 0.70 | 1.00 | 0.84 | | | | | | –0.64 |
| **L** | | | 0.66 | 0.84 | 1.00 | | | | | | –0.65 |
| **MR** | | | | | | 1.00 | | | | | –0.43 |
| **Bv33** | | | | | | | 1.00 | | | | |
| **Bv34** | | | | | | | | 1.00 | | | |
| **VL** | | | | | | | | | 1.00 | | |
| **An** | | | | | | | | | | | –0.60 |
| **Mp** | –0.47 | –0.55 | –0.63 | –0.64 | –0.65 | –0.43 | | | | –0.60 | 1.00 |

We are left with only a few manuscripts whose notations are closer than expected by chance: the 3 Sankt-Gallen sources E, B and G, and 2 of the frankish sources, Ch and L. The manuscript B appears to connect the two regions. The present data do not indicate a relationship between the 3 Beneventan manuscripts, though there are sufficient data as their z values on the diagonal show. An has too few pressus to make a placement possible, MR is a borderline case. Mp turns out to require a separate treatment, since it has significant negative correlations not to be discussed here.

*Figure 1* shows a Venn-diagram of positive relationships suggested by Tables 3a and 3b.

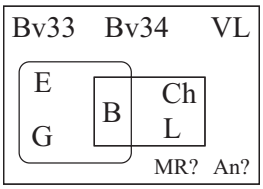


*Figure 1*

**How not to lie with statistics**

Statistics provide a link between a body of data and an inference from those data. Such an inference is suggested when we decide that it is improbable that the particular distribution of the data would have arisen by chance. The key word here is improbable, not: impossible. Statistical analysis indeed yields no more than a suggestion, not a proof. If we want to accept a result as real, we should have an explanation. We need a model.

Therefore, the model is possibly the most important part of the analysis. It operates in two problem areas, each of which needs much more thinking than I have been able to do so far. The first area is in the definition of what constitutes a variant. The second area is in the interpretation of the result.

**The choice of data**

For a proper application of statistics the data must meet several requirements. Two of these are immediately relevant, as they are not necessarily fulfilled in chant.

1. Individual data should be independent. But chant makes extensive use of formulae or stock phrases, and different formulae may well follow different patterns of agreement. If we mix any, our estimate of accidental agreement will be wrong. In one approach, the one taken by the monks of Solesmes, one aims at a wide cross-section of types of variant, taking care that no particular one predominates. Alternatively, one might limit oneself to the study of just one phenomenon, be it a formula or an aspect of style or whatever property of chant that shows regional variation in its notation. To my mind, the latter case would be preferable, as it would tell us something about the development of chant itself as reflected by the manuscript tradition.

2. All data should be equally probable. It is not enough to look for places of difference. If data are gathered by looking primarily for disagreement, the requirement of equal representation is violated. We overestimate the differences between manuscripts and our estimate of statistical relevance is wrong. The decision to include a data point should be independent of the agreement or disagreement found at that point. This again favours the study of a particular aspect for which agreement as well as disagreement can be defined.

3. The role of lacunae. The monks of Solesmes were well aware of the possible distortion of manuscript relationships by the influence of lacunae. Late into the present work it became clear that effects due to lacunae may be more pernicious than the monks were prepared to admit. Lacunae obscure potential variants and may bring in apparent relationships between otherwise unrelated manuscripts. A full treatment of this subject is required for a re-evaluation of the *Le Graduel Romain* study.

## Interpretation of the final result

In order to accept a relationship suggested by statistical analysis one should verify that such a link is not introduced by lacunae, by limited data or by a few exceptional data. One should go back to the raw data and find a common determinant before one decides to accept a relationship as real. Only then are we ready to draw the usual contour maps of manuscript relationships, on the basis of kappa and z. I use z values to define contour levels, but place manuscripts according to their kappa values.

In the case of the pressus, I have been able to trace some of the habits of the scribes who wrote the manuscripts. Briefly, in the 3 Sankt-Gallen sources we do not find a pressus on a stressed syllable; Beneventan scribes do not make that distinction. If L or Ch show a pressus, the SG manuscripts do so too. I have not found an explanation in properties of the text that will explain the differences between the Sankt-Gallen manuscripts. The negative correlations of Mp with other manuscripts are a consequence of the fact that Mp has no pressus in common with L and Ch.

## Conclusion

In this paper I sketch the potential power of standard non-parametric statistical theory in order to overcome the problem of contamination that until now has hampered the study of relationships between manuscripts of the Mass Proprium. I have not seen a method that is able to deal with contamination in such a simple and straightforward way.

The notation of the pressus in isolation is used as an illustration. Whilst not treated in full detail, which must wait for another occasion, the example serves to show the course of an actual analysis.

The method identifies problem areas, as yet largely untouched, in the definition of variants and the role of lacunae. While the theoretical underpinning must be confirmed, practical work should proceed to re-examine the data of the monks of Solesmes and possibly expand on their results.