HOMOGENEITY ANALYSIS OF A TIME SERIES

Eeke van der Burg
Department of Data Theory
University of Leiden

Paper prepared for the Table Ronde 'Analyse des Données',
Toulouse, January 9-10, 1984

Time series of RE

The data set we analysed consisted of the amounts of computing time (RE) used by the members of the department of Data Theory. We took 52 weeks starting from 22-11-83. The RE have been categorised into 6 categories, ranging from very little to very much. We have chosen the exponents of 3 as the category boundaries. We did HOMALS on the 10x52 matrix of individuals x weeks. Every week is considered as a variable with the same categories 1 to 6 (some does not occur) (Table 1). In addition we did ANACOR on the 10x74 matrix of frequencies/months for each category. A month is taken as four weeks (table 2). The number of columns of this matrix should have been 78 (=13x6), however in four months the highest category did not occur. A second ANACOR analysis has been applied to the matrix of frequencies/semester for each category. This is the 10x24 (24=4x6) matrix of table 3. Finally we did a PRINCALS analysis on the 10x52 matrix of RE in order to restrict the category quantifications.
The object scores or row points of the four analyses are given in figure 1. In all cases John is an outlier. This is because John hardly used computer time, every week he scored in the lowest category. A second outlier is Jacqueline. That is because she is consuming most computer time. There is another factor that plays a role. We have users with a regular pattern: Eeke, Neufléglise, Aad and John, and those with an irregulair pattern: Jan, Ineke, Renée, Willem, Peter, Jacqueline. In the ANACOR semester solution we find the 'irregulars' in the lower part of the plot. They are identified by the second dimension. In the HOMALS solution the 'irregulars' also lie in a consecutive region. PRINCALS does not find the 'irregulars', neither does ANACOR applied to months.
In all cases Willem is nearest to John because he has so many ones in common with him. The same holds to a lesser extend for Jan and John. In the PRINCALS solution the position of Jan differs from his position in the other solutions. This is because PRINCALS finds something different from HOMALS and ANACOR. PRINCALS groups John, Renée, Willem and Jacqueline together, because they all have a zero start.
In the HOMALS and the ANACOR-months solution we find the effect of Paris (Psychometric Society Meeting) back, John, Aad and Neufléglise stayed at home, the rest went to the conference.

We made a plot of the category quantifications of the HOMALS solution. We collected all categories one, two, etc, together and connected them (figure 2, dimension 1; figure 3, dimension 2). Except for the first categories which are really lower in the first dimension and higher in the second dimension, the graphs do not show many systematic features. We also made a plot of the individual 'functions' in figure 4 (dimension 1) and figure 5 (dimension 2). We took the categories of the 10x52 matrix of RE and replaced the category numbers by category quantifications. Before plotting we allready know the functions of John, because he scored only ones so that his functions coincide with those of category one. The other functions look rather random, they do not give us much information. Data without structure will not be structured by HOMALS or ANACOR.

The category quantifications of ANACOR and PRINCALS are given in figure 6 to 11. We plotted all the figures on the same scale. The ANACOR-months solution is a restriction of HOMALS, because we restrict the category quantifications of one month to be the same. The ANACOR-semesters solution is even more restricted because in that case the quantifications for each semester are the same. Finally the PRINCALS solution is also a restriction of HOMALS because in the PRINCALS case the quantifications are restricted to ordered and identical scores for each dimension. The category quantifications of the ANACOR-months solution show very little variation, especially for category 2, 3, 4 and 5 of the first dimension (figure 6 and 7). Obviously the first categories are quantified lower than the other categories on the first dimension and higher on the second dimension. The ANACOR-semesters solution shows even less variation. The only effect that is really clear is the controversy between ones and the other categories (figure 8 and 9). The PRINCALS quantifications show that the categories higher than the lowest are mostly tied (figure 10 and 11). The second dimension also shows an effect in the higher categories of week 38 for category 6 and weeks 42 and 43 of category 5. The only one who has these scores is Jacqueline. We used the ordinal measurement level in PRINCALS in order to find 'more structure'. We expected more of the distinction between high consumers of RE and low consumers. However the equal start of John, Willem, Renéé and Jacqueline is dominating the solution (figure 1) and the other ones are grouped together. Because there is such a large difference between John and Jacqueline in amounts of computer time both they are placed on the extreme of a dimension. Renée shares the neighbourhood of Jacqueline because they share a nonactive period in week 45 to 48.

The fact that one person dominates a separate dimension is also seen at the HOMALS solution (dim 1 John, dim 2 Jacqueline, dim 3 Neufèglise, dim 4 Ineke, dim 5 Peter). Also the eigenvalues of this solution are badly separated from each other (except for the first one) (figure 1). The ANACOR eigenvalues are a little more separated, but the results are not convincing.
In conclusion we can say that we should not expect too much from a time series analysis. We should be critical of the data. There should be at least some obvious structure.

```
4664111413111354452453456342122216531111411153454421 jan
4344321443334445445244334445455434323143243334321123 eeke
3332223322233242444223334332323343232111234433443443 neuf
3333322344444313344443334324455456565542442244241232 aad
5542322456554543452336454552444622115311144335544415 ineke
1111111111111111111111111111111111111111111111111111 john
1111111143243444433434434244534513344344324411115444 renée
1111111111141111113313335365655611111111121466322355 willem
4454554443343344455324555464553213111314323524522212 peter
1111111333355555555424556644566614444611155411111556 jacq
```

Table 1. Data matrix of RE; 10 individuals x 52 weeks.

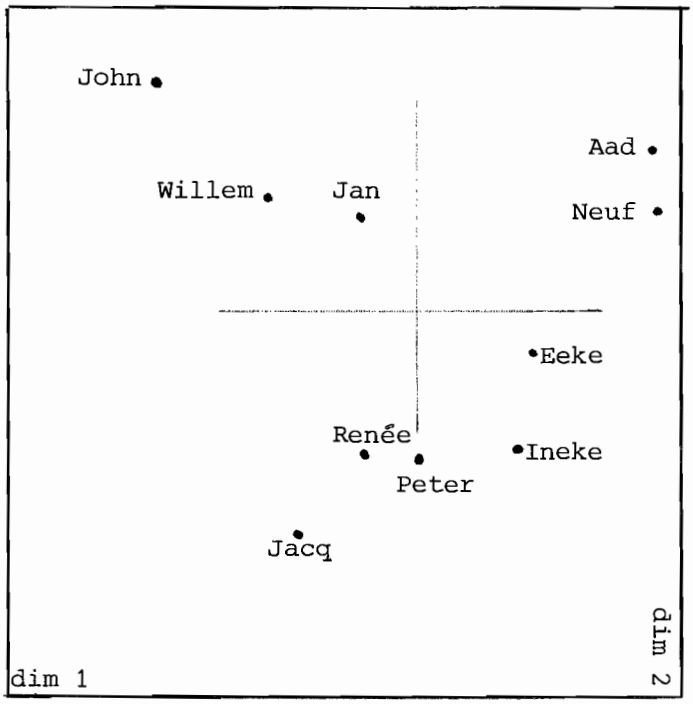
```
000202|300103010001011|01021|00112001110113|00000|101011|4000003001000112011020
001300|1111000310000031|01021|00220000031|00002200121|00|1021000121001210021100
013000022000310000211001030013000001210001300001210031000001120002200002200
004000022000000400102100013000310001120000022000002201012002020010300121000
010120021100000310012101111|00|1111|01012000030122000020101010120001120100210
4000004000040000040000400004000040000040000040000040000040000040000040000
4000004000001120000013000220002200010300001120102100001300011204000000003l0
4000004000030010040000200200103000001021|00002240000040000021010011002011020
0003100002200220000220001120101200002110110203010002011000120102011013000
4000003010000301000004000130101200002020000131003002001011001240000010002l
```

Table 2. Frequencies of RE/month; 10 individuals x 74 categories, months 2,4,5 and 11 have 5 categories, the other months 6.

```
 7 0 1 3 0 2| 0 1 3 4 4 1| 5 4 1 1 1 1| 5 1 1 4 2 0
 1 1 5 6 0 0| 0 1 2 8 2 0| 1 1 3 5 3 0| 2 3 6 2 0 0
 0 6 7 0 0 0| 0 4 4 5 0 0| 2 4 6 1 0 0| 1 1 5 6 0 0
 0 2 6 5 0 0| 1 0 7 5 0 0| 0 1 0 4 6 2| 1 6 1 5 0 0
 0 3 1 3 5 1| 0 1 3 4 4 1| 3 3 1 3 2 1| 3 0 2 5 3 0
13 0 0 0 0 0|13 0 0 0 0 0|13 0 0 0 0 0|13 0 0 0 0 0
 8 1 2 2 0 0| 0 1 4 8 0 0| 1 0 4 6 2 0| 4 1 1 6 1 0
12 0 0 1 0 0| 6 0 6 0 1 0| 7 0 0 0 3 3| 3 3 2 1 2 2
 0 0 3 7 3 0| 0 1 2 5 5 0| 5 1 3 1 2 1| 1 6 2 2 2 0
 7 0 4 0 2 0| 0 1 0 2 8 2| 2 0 0 6 1 4| 7 0 0 1 4 1
```

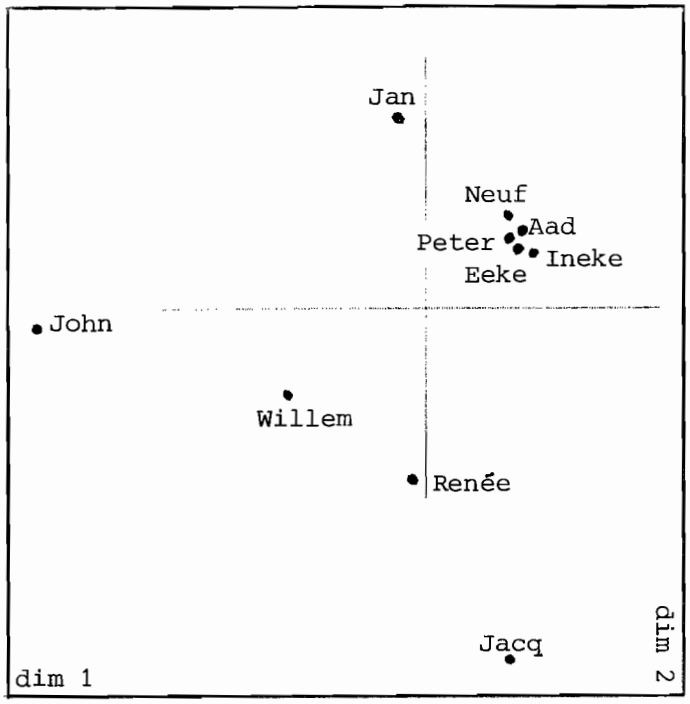
Table 3. Frequencies of RE/semester;
10 individuals x 24 categories.

HOMALS 10 x 52 matrix, weeks



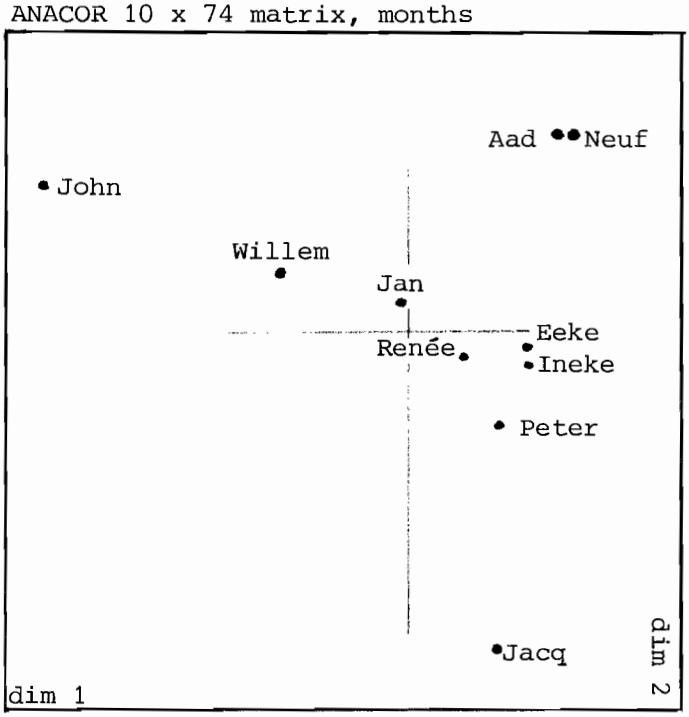
Eigenvalues: .618, .473, .460, .439, .389

PRINCALS (ordinal) 10 x 52 matrix, weeks



Eigenvalues: .459, .242

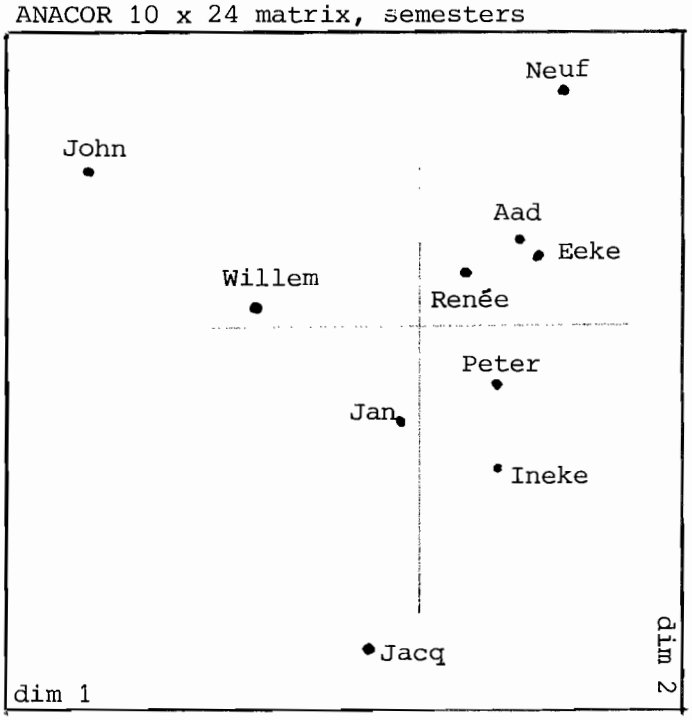
ANACOR 10 x 74 matrix, months



Eigenvalues: .596, .309, .229, .200, .178

dimension 1 reflected

ANACOR 10 x 24 matrix, semesters



Eigenvalues: .459, .189, 135, .086, .068

dimension 1 and 2 reflected

Figure 1. Time series of RE. Object scores of HOMALS and PRINCALS, row points of ANACOR.

RE DATATHEORIE 221182-161183 CAT 1 TO CAT 6

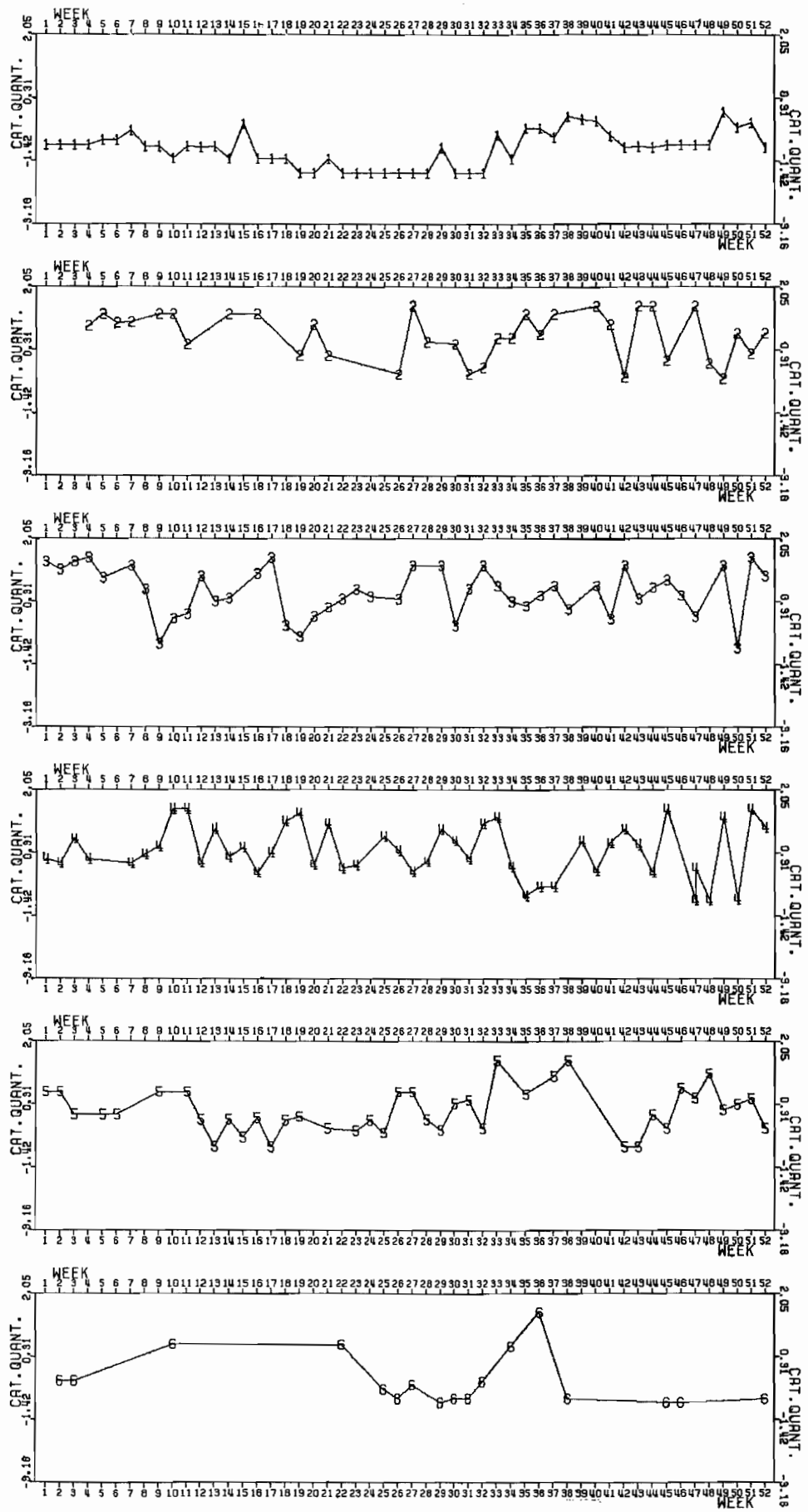


Figure 2. HOMALS 10x52 matrix RE/week.

Category quantifications dimension 1.

RE DATATHEØRIE 221182-161183 CAT 1 TØ CAT 6

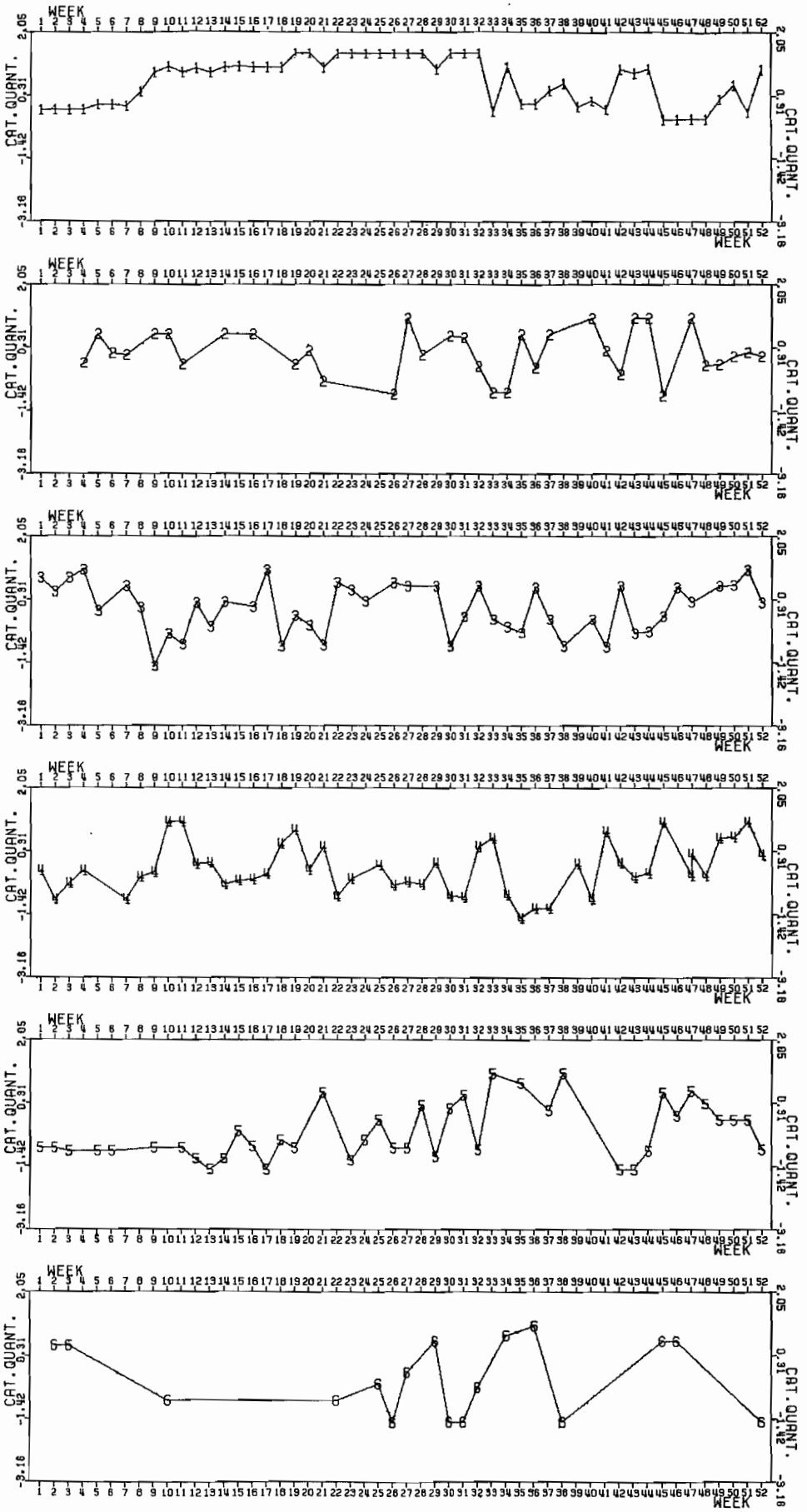


Figure 3. HOMALS 10x52 matrix RE/week.

Category quantifications dimension 2.

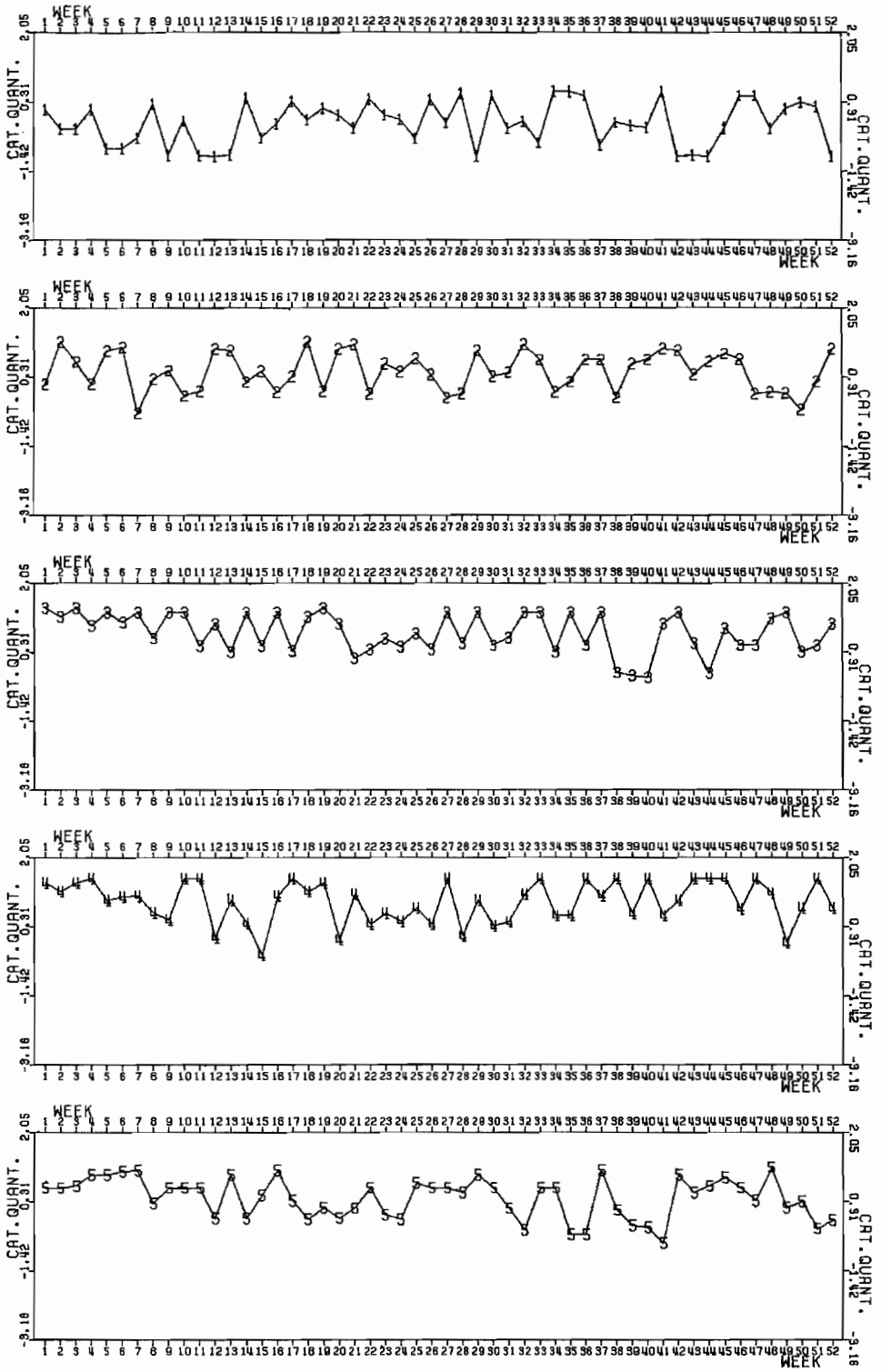

RE DATATHEORIE 221182-161183 INDIVIDUAL 1 TO 10
JAN,EEKE,NEUF,AAD.INEKE,JOHN.RENEE,WILLEM,PETER,JACQ
WEEK
CAT.QUANT.
2.05
0.31
-1.42
-3.16

Figure 4.

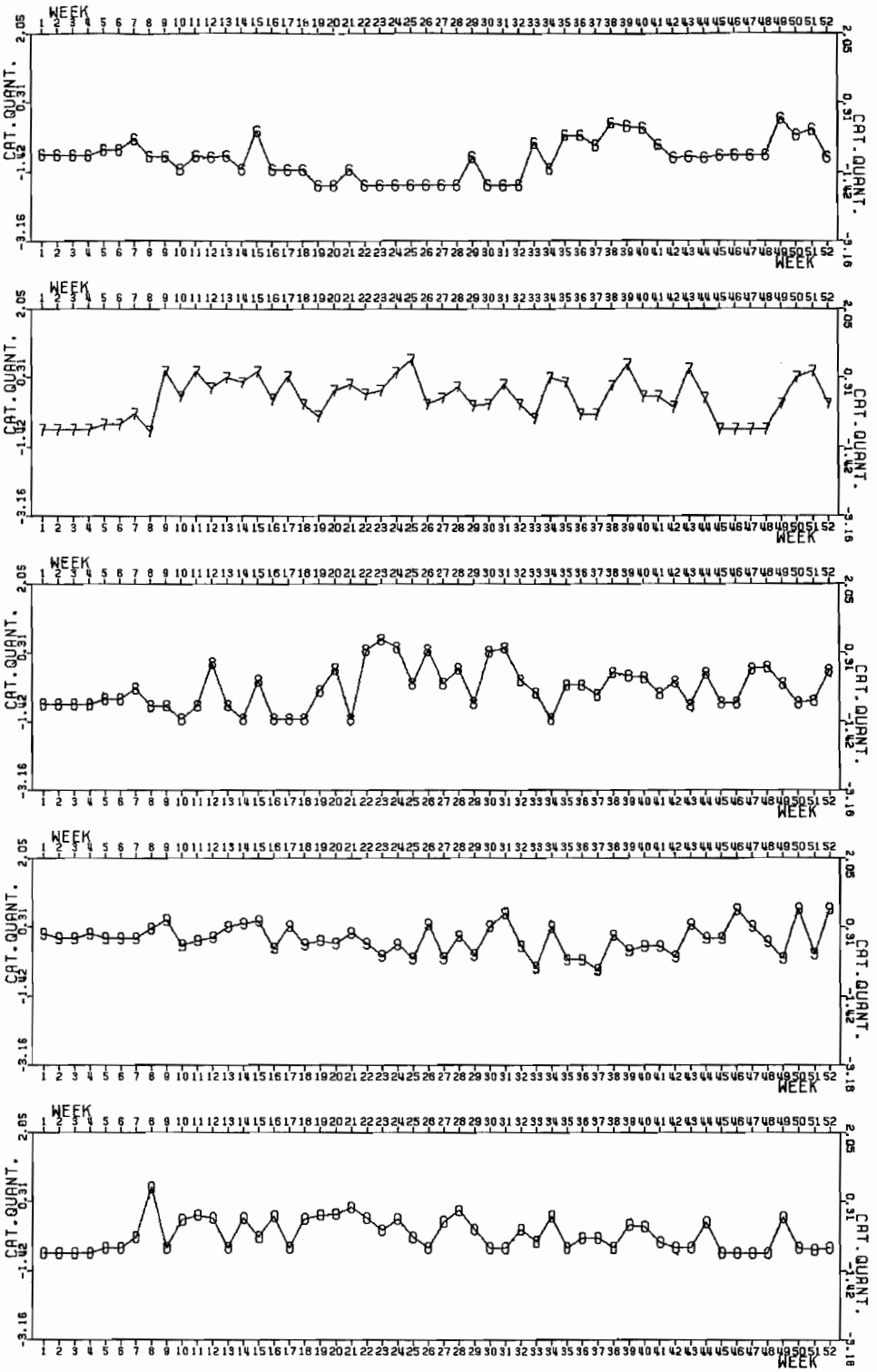


Figure 4. HOMALS 10x52 matrix RE/week.
Individual functions, dimension 1.

RE DATATHEORIE 221182-161183 INDIVIDUAL 1 TO 10
JAN,EEKE,NEUF,AAD,INEKE,JOHN,RENEE,WILLEM,PETER,JACQ

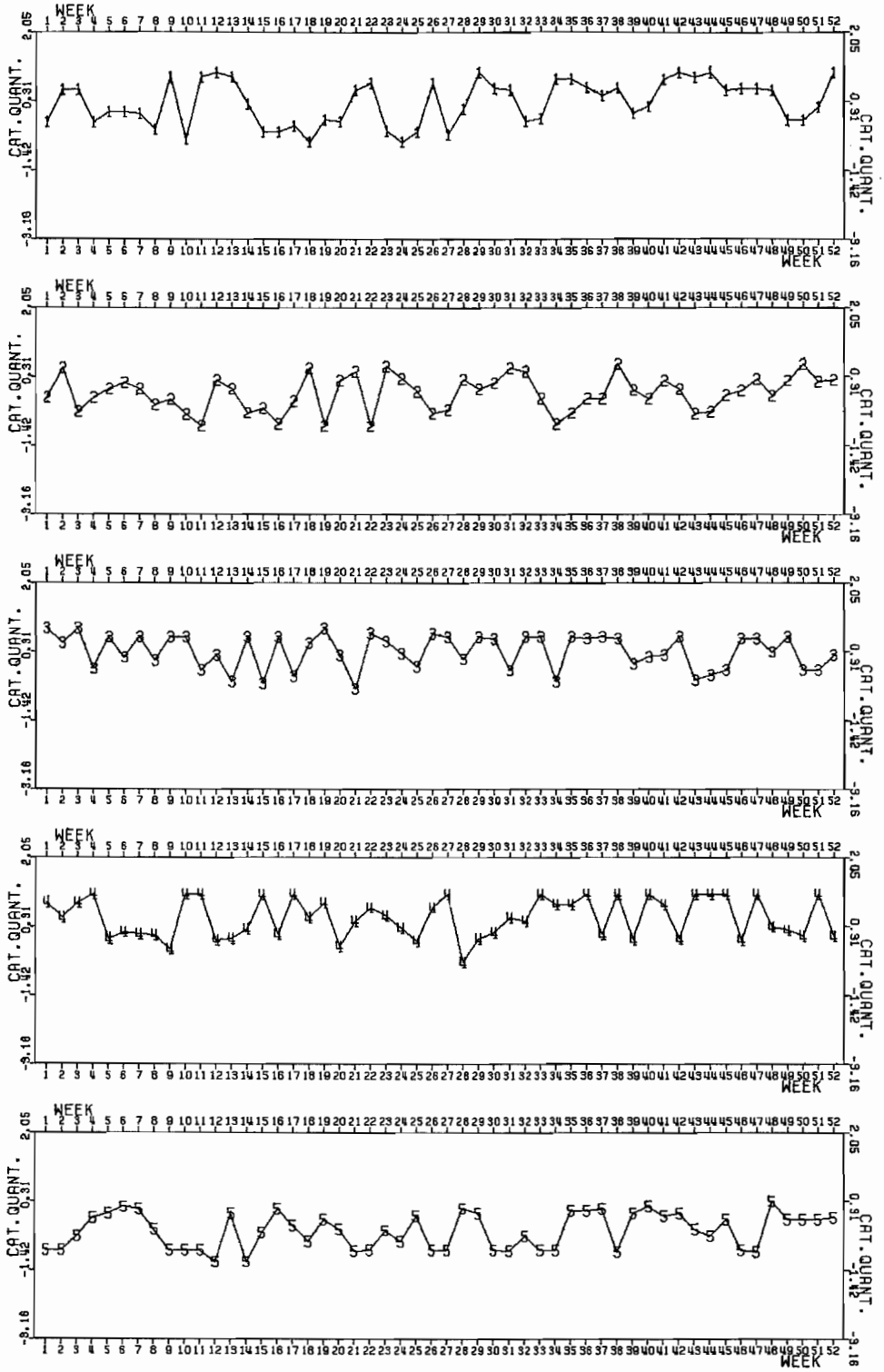


Figure 5.

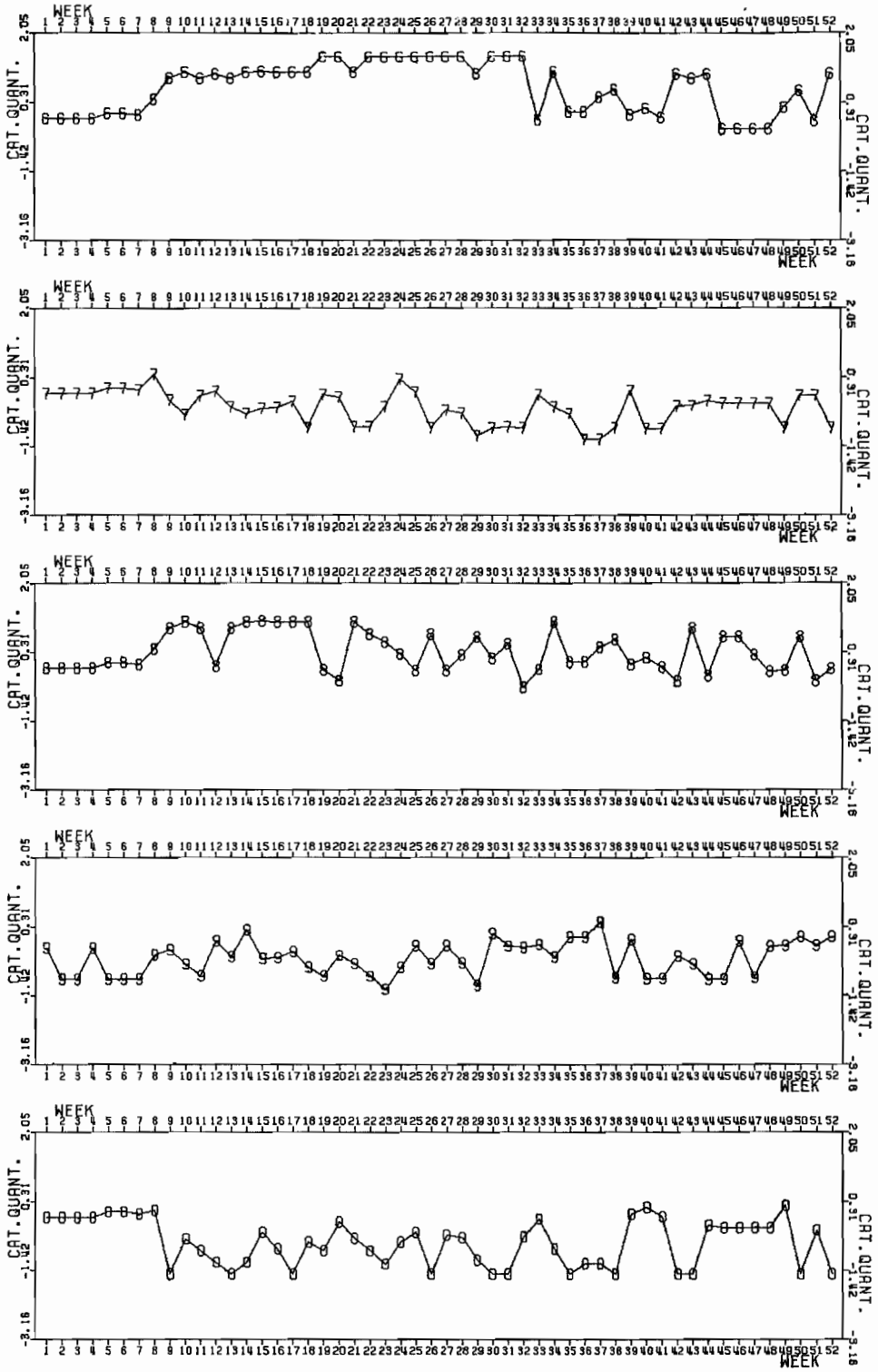


Figure 5. HOMALS 10x52 matrix RE/week.
Individual functions, dimension 2.

RE DATATHEORIE 221182-161183 CAT 1 TO CAT 6

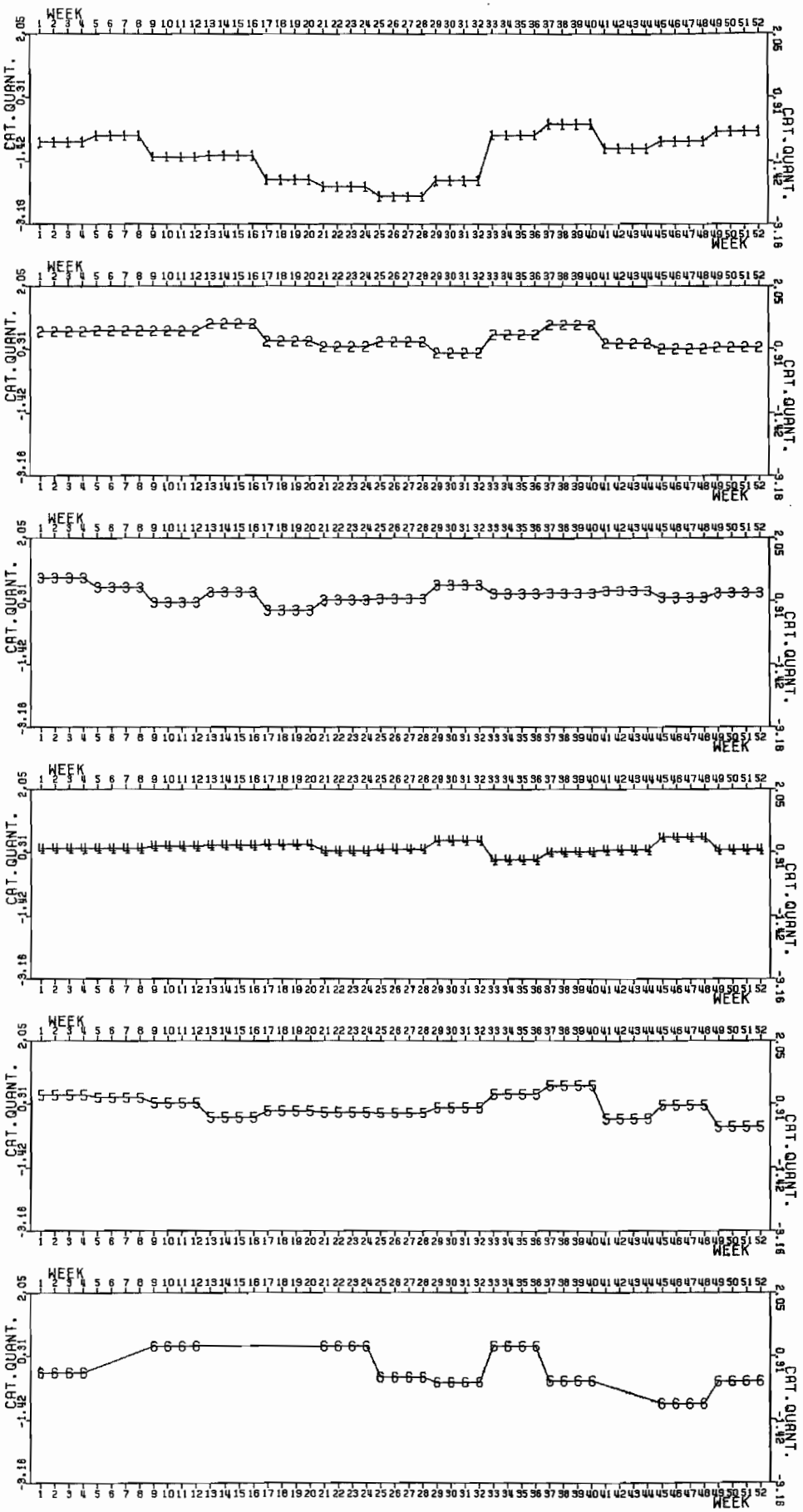


Figure 6. ANACOR 10x74 matrix frequencies/month..
Category quantifications, dimension 1.

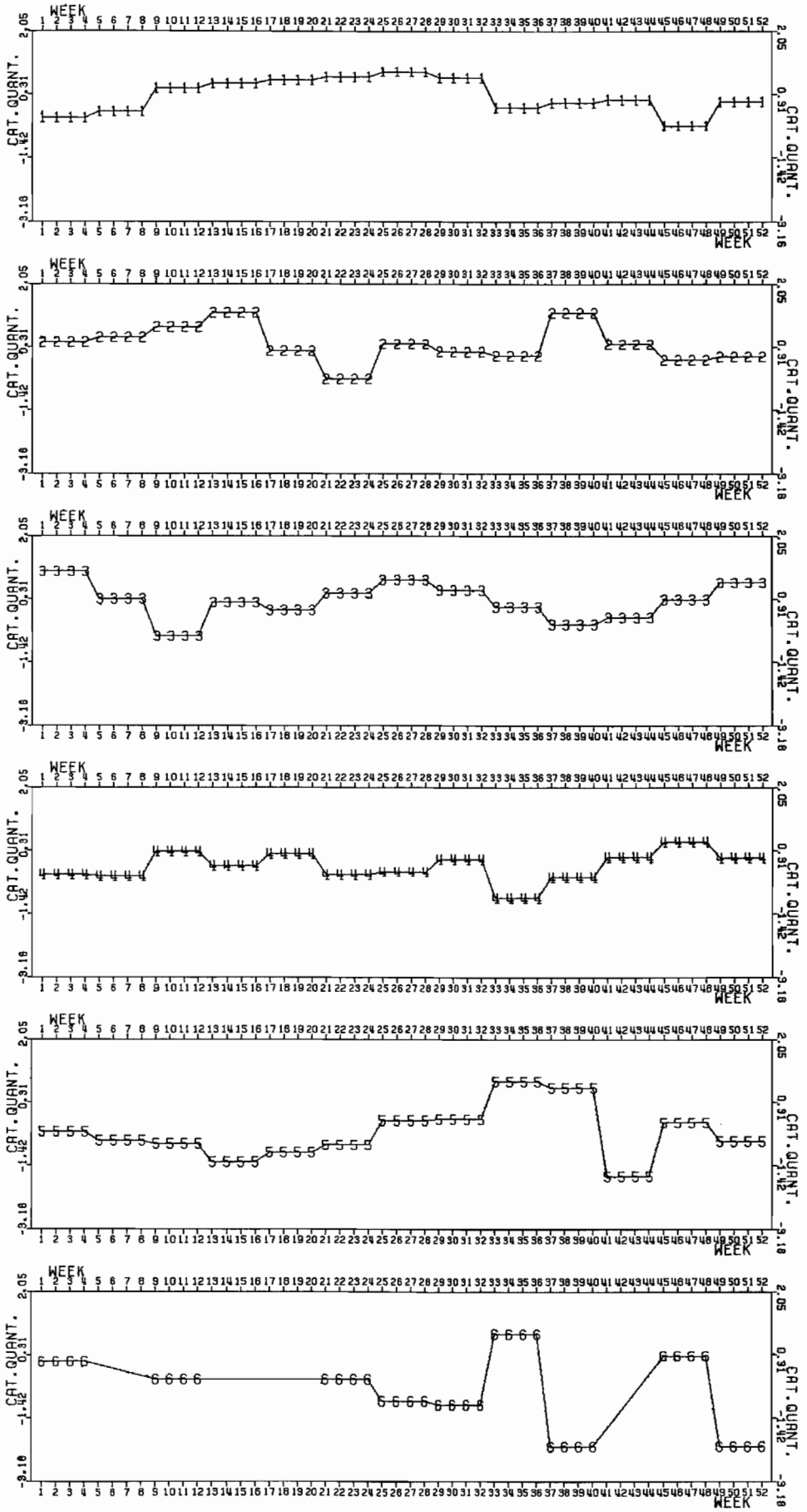


Figure 7. ANACOR 10x74 matrix frequencies/month.
Catgeory quantifications,dimension 2.

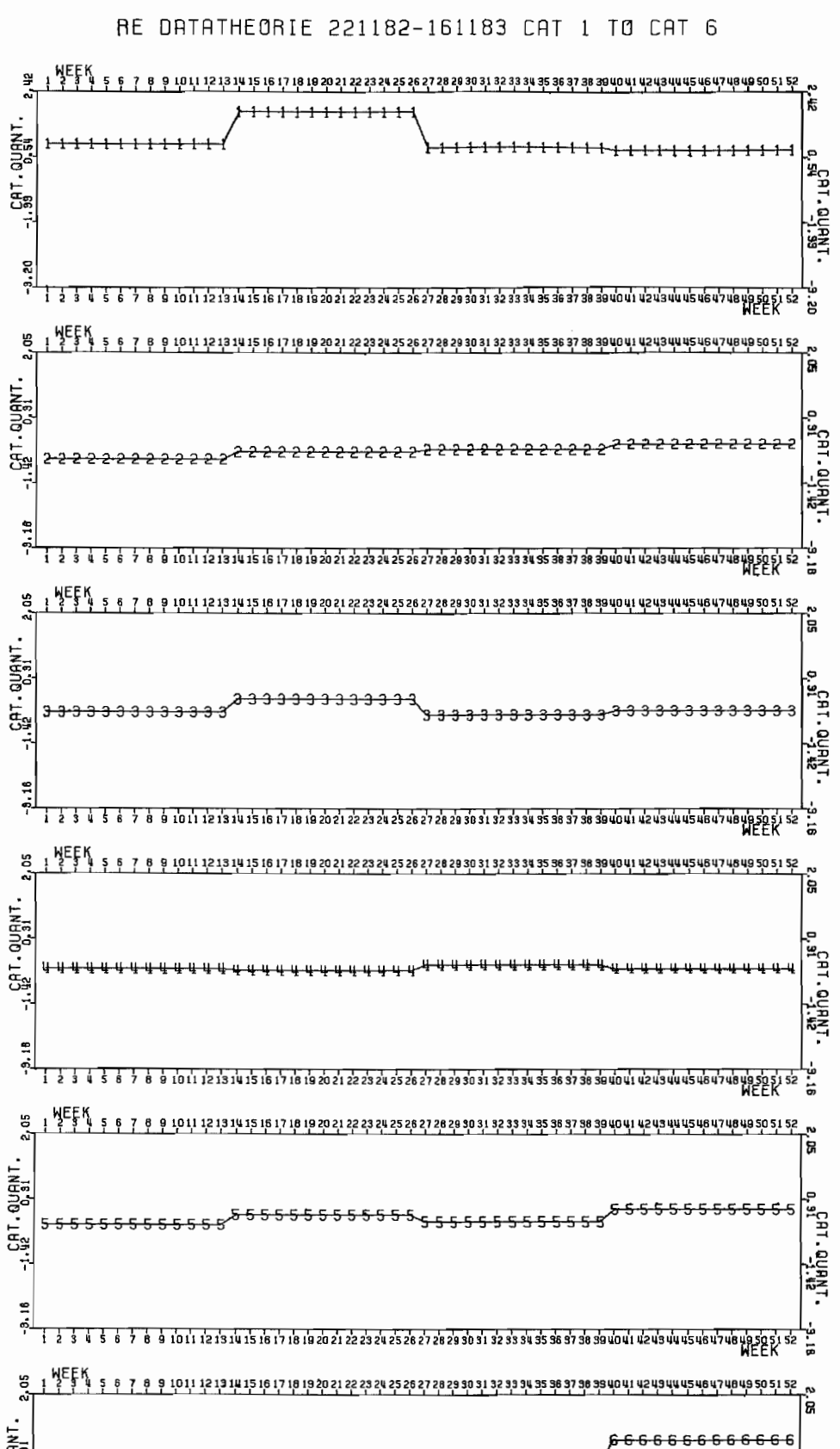


Figure 8. ANACOR 10x24 matrix frequencies/semester.
Category quantifications, dimension 1.

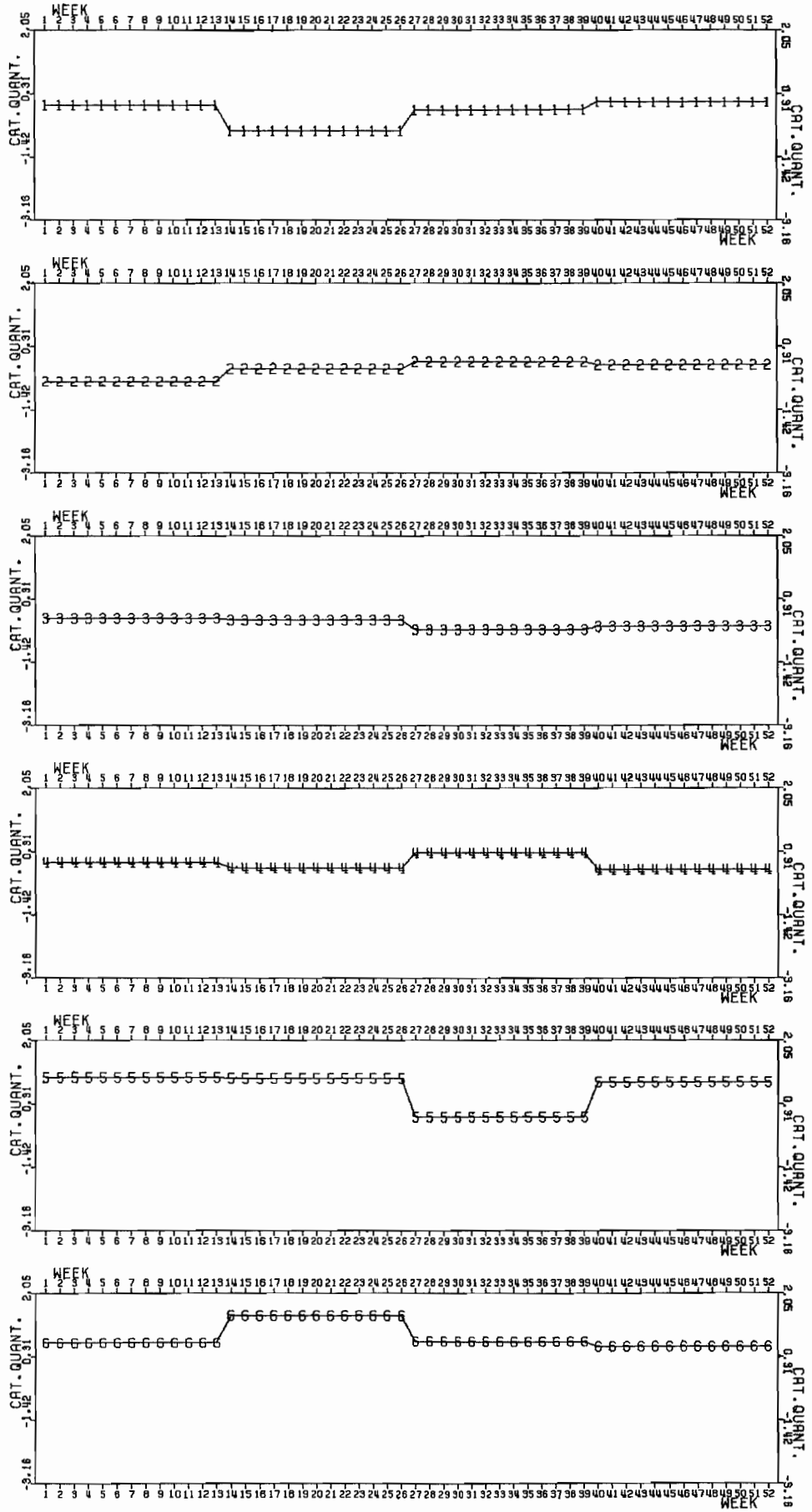


Figure 9. ANACOR 10x24 matrix frequencies/semester.
Category quantifications, dimension 2

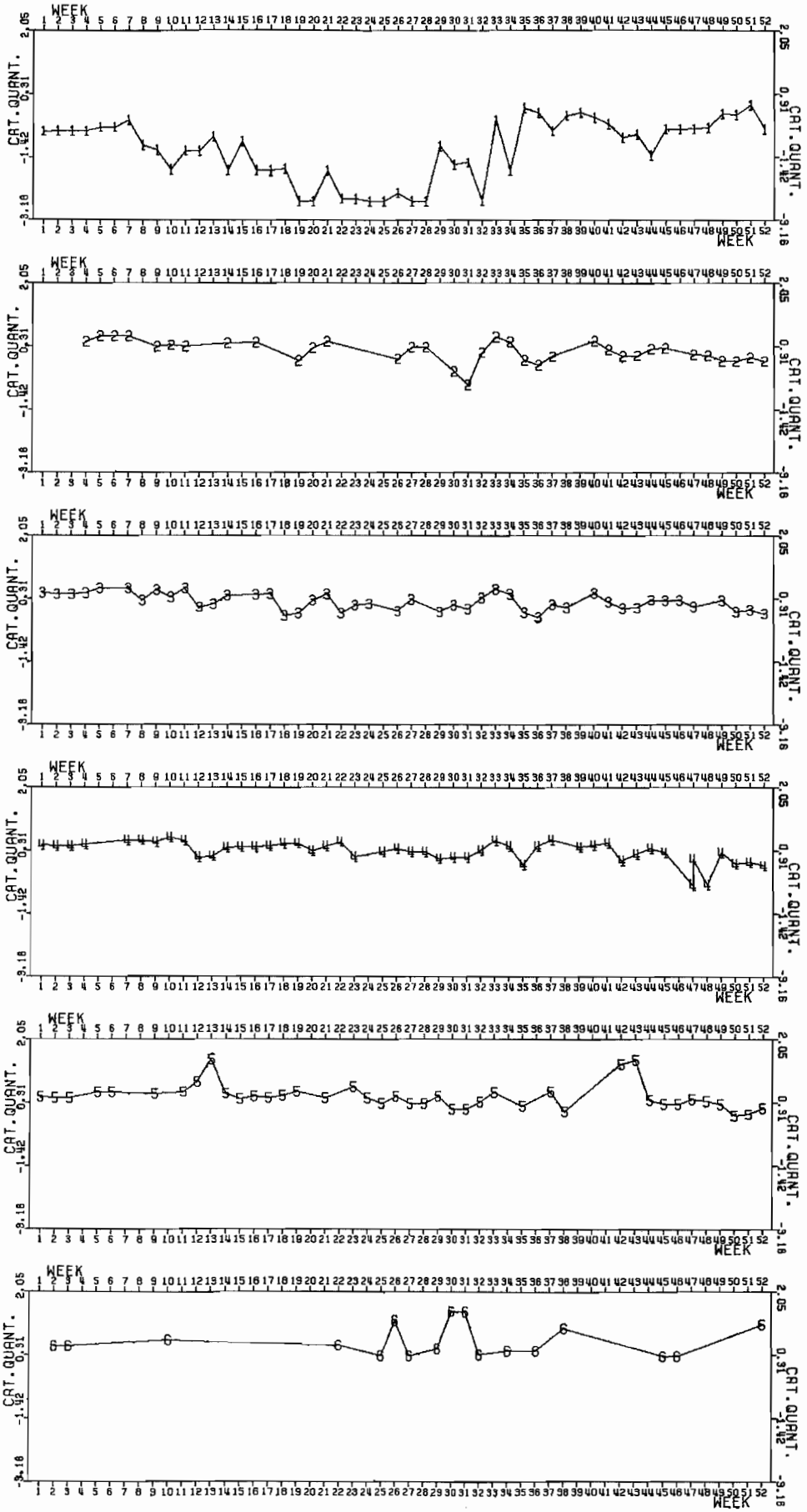


Figure 10. PRINCALS 10x52 matrix RE/week.
Category quantifications, dimension 1.

RE DATATHEORIE 221182-161183 CAT 1 TO CAT 6

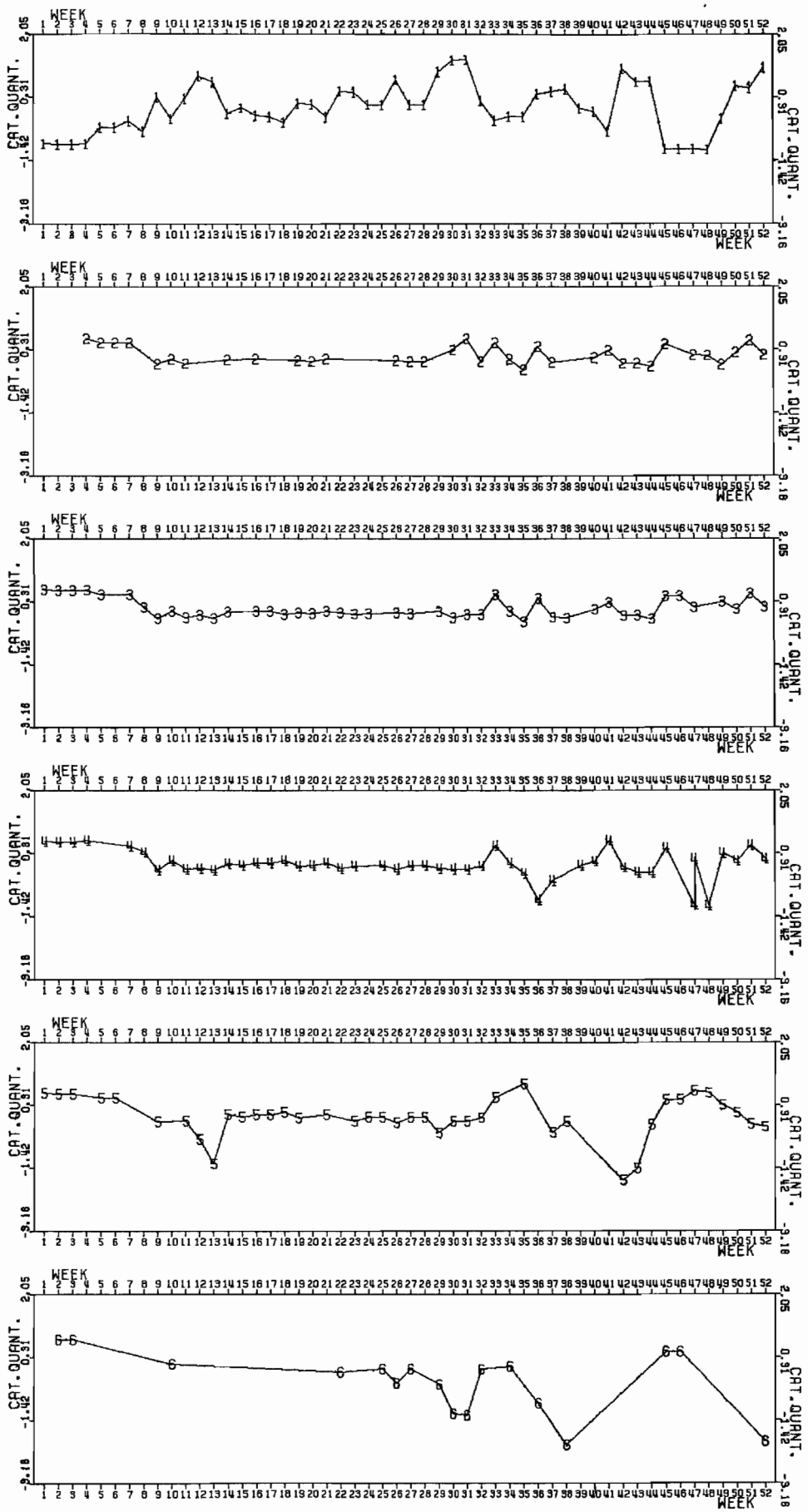


Figure 11. PRINCALS 10x52 matrix RE/week.

Category quantifications, dimension 2.