**Attributes in the** [**3dREMLFIT**](https://docs.google.com/document/d/1wYOqsYpovM44xn8axNFKaXsqTXHS1O0KJ-pj-rBalog/edit?usp=sharing) **.xmat.1D format**

**RW Cox – August 2019 A.D.**

* Attributes are stored in an XML-ish header before the actual matrix numbers.
* Attributes are of the form **name = "quoted string"** - the quotes can be single or double.
* Below is a sample header, followed by the first row of the matrix (there are 444 rows in the actual matrix, each with 20 numbers):

# <matrix

# ni\_type = "20\*double"

# ni\_dimen = "444"

# ColumnLabels = "Run#1Pol#0 ; Run#1Pol#1 ; Run#1Pol#2 ; Run#1Pol#3 ; Run#2Pol#0 ; Run#2Pol#1 ; Run#2Pol#2 ; Run#2Pol#3 ; Run#3Pol#0 ; Run#3Pol#1 ; Run#3Pol#2 ; Run#3Pol#3 ; vis#0 ; aud#0 ; roll#0 ; pitch#0 ; yaw#0 ; dS#0 ; dL#0 ; dP#0"

# ColumnGroups = "12@-1,1,2,6@0"

# RowTR = "2"

# GoodList = "0..40,45..264,267..449"

# NRowFull = "450"

# RunStart = "0,150,300"

# Nstim = "2"

# StimBots = "12,13"

# StimTops = "12,13"

# StimLabels = "vis ; aud"

# Nglt = "1"

# GltLabels = "V-A"

# GltMatrix\_000000 = "1,20,12@0,1,-1,6@0"

# BasisNstim = "8"

# BasisOption\_000001 = "-stim\_times"

# BasisName\_000001 = "vis"

# BasisFormula\_000001 = "BLOCK(20,1)"

# BasisColumns\_000001 = "12:12"

# BasisOption\_000002 = "-stim\_times"

# BasisName\_000002 = "aud"

# BasisFormula\_000002 = "BLOCK(20,1)"

# BasisColumns\_000002 = "13:13"

# CommandLine = "3dDeconvolve -input pb05.FT.surf.rh.r01.scale.niml.dset pb05.FT.surf.rh.r02.scale.niml.dset pb05.FT.surf.rh.r03.scale.niml.dset -censor motion\_FT.surf\_censor.1D -polort 3 -num\_stimts 8 -stim\_times 1 stimuli/AV1\_vis.txt &apos;BLOCK(20,1)&apos; -stim\_label 1 vis -stim\_times 2 stimuli/AV2\_aud.txt &apos;BLOCK(20,1)&apos; -stim\_label 2 aud -stim\_file 3 &apos;motion\_demean.1D[0]&apos; -stim\_base 3 -stim\_label 3 roll -stim\_file 4 &apos;motion\_demean.1D[1]&apos; -stim\_base 4 -stim\_label 4 pitch -stim\_file 5 &apos;motion\_demean.1D[2]&apos; -stim\_base 5 -stim\_label 5 yaw -stim\_file 6 &apos;motion\_demean.1D[3]&apos; -stim\_base 6 -stim\_label 6 dS -stim\_file 7 &apos;motion\_demean.1D[4]&apos; -stim\_base 7 -stim\_label 7 dL -stim\_file 8 &apos;motion\_demean.1D[5]&apos; -stim\_base 8 -stim\_label 8 dP -jobs 2 -gltsym &apos;SYM: vis -aud&apos; -glt\_label 1 V-A -fout -tout -x1D X.xmat.1D -xjpeg X.jpg -x1D\_uncensored X.nocensor.xmat.1D -fitts fitts.FT.surf.rh.niml.dset -errts errts.FT.surf.rh.niml.dset -bucket stats.FT.surf.rh.niml.dset"

# >

1 -0.99999999284744 0.9932885915041 -1.0000000007947 0 0 0 0 0 0 0 0 0 0 -0.056317329311536 0.1472171255615 -0.030924689328919 -0.14155002441671 -0.0522833100934 -0.081843944456843

For computational details, idly peruse this scan of my handwritten notes about 3dREMLfit’s algorithms and models: <https://afni.nimh.nih.gov/pub/dist/doc/misc/3dREMLfit/3dREMLfit_mathnotes.pdf>

Some attributes are necessary for 3dREMLfit to operate, and some are optional. The leading ‘#’ character on each line is not necessary, and is there for peculiar historical/hysterical reasons and also for compatibility with some other AFNI software (e.g., 1dplot).

Attributes can be in any order inside the **<matrix ... >** header.

Note that index counting (e.g., for rows and columns, mentioned below) starts at 0, not 1, as decreed by the Almighty: <http://mathworld.wolfram.com/PeanosAxioms.html>

* ni\_type = "20\*double" [REQUIRED]
  + This indicates there are 20 numerical values per row in the data section (past the header), and they are to be interpreted as doubles (64 bit floating point values) when read in.
  + In this example, the matrix has 20 columns (regressors) – numbered from 0..19, as mentioned above.
  + In the code, this numeric value (20) is called **nreg** = number of regressors; that is how I will refer to it below, as needed.
  + The “\*double” is needed, since the parser for this format allows data columns of various types, but in this case all the data columns are numeric.
* ni\_dimen = "444" [REQUIRED]
  + This value indicates there are 444 rows in the data section.
  + In this example, the matrix corresponds to 444 time points (TRs).
  + Also see NRowFull below.
* ColumnLabels = "Run#1Pol#0 ; Run#1Pol#1 ; Run#1Pol#2 ; Run#1Pol#3 ; Run#2Pol#0 ; Run#2Pol#1 ; Run#2Pol#2 ; Run#2Pol#3 ; Run#3Pol#0 ; Run#3Pol#1 ; Run#3Pol#2 ; Run#3Pol#3 ; vis#0 ; aud#0 ; roll#0 ; pitch#0 ; yaw#0 ; dS#0 ; dL#0 ; dP#0" [OPTIONAL but highly recommended]
  + Defines the string label for each column in the matrix.
  + If this attribute is present, there *must* be as many labels as columns (nreg).
  + Labels cannot contain whitespace characters unless ’in quotes’.
    - In this example, single quotes would have to be used, to distinguish from the double quotes used to delineate the attribute itself.
  + Labels must be separated as shown above, with a semicolon (labels can contain commas, if you insist).
  + In this example, columns 0..11 and 14..19 are regressors of no interest, and columns 12 and 13 (vis#0 and aud#0) are the regressors of interest (response models for stimuli).
    - Which regressors correspond to stimuli and which do not will be marked out in the ’Stim’ attributes described later.
  + Labels are attached to output volumes in the results datasets, to make it easy for the AFNI user to see which volume corresponds to the statistical estimates for which stimulus.
* ColumnGroups = "12@-1,1,2,6@0" [NOT USED]
  + This attribute is not actually used by 3dREMLfit for anything at this time [Aug 2019].
  + Its intended function is to mark matrix columns as being in different groups.
    - In this example, the first 12 columns are “baseline and drift model” (group -1), the next 2 columns belong to distinct stimuli, and the last 6 columns belong to the motion regressors (and other dataset-based) regressors of no interest.
* RowTR = "2" [OPTIONAL]
  + This attribute is not actually used by 3dREMLfit now [Aug 2019].
  + It defines the inter-scan time interval (TR) in seconds.
  + The TR is needed for construction of the matrix from the stimulus response model, but that has already been done, so this attribute is really just for documentation and completeness.
* GoodList = "0..40,45..264,267..449" [HIGHLY REQUIRED]
  + The matrix provided to 3dREMLfit is the censored matrix; that is, the time points (TRs) to be censored have had the corresponding rows removed from the full matrix.
    - The data volumes to be censored will be removed from the input dataset during processing by 3dREMLfit.
  + The GoodList attribute lists the TR indexes from the original (uncensored) time series dataset that *are present* in the matrix file; that is, it is the opposite of the “censor list”.
  + There must be the same number of integers specified here as the number of time points specified by the ni\_dimen attribute (here, 444).
  + The brute force approach would be just to list all the integers, comma separated, in one long string.
  + For the sake of compactness, contiguous sequences of integers can be given, as in the example, where "0..40" means the same as listing all the integers 0, 1, 2, …, 40.
  + In this example, there were 450 time points in the original EPI dataset, and clearly 6 of them have been censored, since the matrix has only 444 rows.
    - This attribute is required so that the temporal autocorrelation ARMA(1,1) voxelwise model doesn’t falsely assume that the data to be processed occurs with constant TR.
    - The RunStart attribute (below) subserves this purpose also, marking the temporal discontinuities between multiple EPI imaging runs.
    - If there were no censoring, then GoodList = "0..449" would work fine (but still would be required by 3dREMLfit).
* NRowFull = "450" [REQUIRED]
  + This attribute gives the number of TRs in the full (uncensored matrix).
  + It is needed for creating the “fitts” and “errts” output datasets, and also for consistency checking to make sure that the user is inputting data that matches the matrix.
* RunStart = "0,150,300" [OPTIONAL]
  + If there is more than one imaging run – that is, there is a long temporal discontinuity between some time points in the dataset to be processed – then this attribute gives the list of the starting TR indexes for each run.
  + In this example, there were 3 runs of 150 TRs each: 0..149, 150..299, and 300..499.
    - The ARMA(1,1) model for the noise temporal correlation is built to have zero correlations for time point pairs from different runs; see the math notes for details on how this is implemented.
  + As with GoodList, this attribute is needed for correct temporal autocorrelation model fitting.
  + If RunStart is not present, then the input EPI dataset is presumed to contain only one imaging run.
* The “Stim” group of attributes mark off some columns as being “of interest” for statistics – presumably from task stimuli. These are [OPTIONAL] as a group, but if Nstim is present, then the others must be present as well.
  + Statistics (betas and *t*-statistics) will be computed only for columns marked as belonging to stimuli, since no one is ever interested in the statistics for the drift and motion parameters (e.g.). If the “Stim” attributes are not present, statistics will not be calculated unless GLTs are used.
  + Nstim = "2"
    - This attribute indicates how many distinct stimuli present.
    - Each stimulus will correspond to 1 or more contiguous columns in the matrix.
  + StimBots = "12,13"
    - This attribute should have Nstim integer entries
    - It indicates the column indexes (remember, counting starts at 0) that correspond to the start of each stimulus's column group
  + StimTops = "12,13"
    - This attribute should have Nstim integer entries.
    - It indicates the column indexes that correspond to the end of each stimulus's column group.
    - In this example, the model for each stimulus has just one column, so the StimBots and StimTops attributes are identical.
    - In deconvolution type models (e.g., AFNI TENTS, FIR models) or in parametric regression, a single stimulus will have multiple regression columns in its response model.
  + StimLabels = "vis ; aud"
    - This attribute should have Nstim string entries, separated by semicolons.
    - These are used (among other things) to process symbolic general linear tests (GLTs) among beta coefficients, given on the 3dREMLfit command line via the “-gltsym” option.
* The “GLT” group is used to specify one or more general linear tests among the beta coefficients, directly in the matrix file. These are completely [OPTIONAL].
  + As mentioned above, GLTs can also be specified outside the matrix file, on the 3dREMLfit command line.
    - GLTs in the matrix file are specified as sets of coefficients to be applied to the beta estimates.
    - GLTs on the 3dREMLfit command line can use symbolic names for the stimuli to specify the coefficients to be attached to the betas.
  + Nglt = "1"
    - If present, this attribute specifies the number of GLTs in the matrix file. It should be an integer from 1 to 1000000.
  + GltLabels = "V-A"
    - This attribute contains Nglt string labels, one for each GLT specified
    - The labels are attached to the output data volumes to make it easy for the user to see which volume corresponds to what statistical test.
  + GltMatrix\_000000 = "1,20,12@0,1,-1,6@0"
    - There should be Nglt of these attributes, with a six digit suffix starting at \_000000, then \_000001, and so forth. (If you want more than 1 million GLTs, you are legally insane and should be confined for your own safety.)
    - Each GltMatrix\_xxxxxx attribute has *r*＊nreg+2 numeric values, which are used to define an *r* ✕ nreg matrix for some *r* ≥ 1.
      * The first value in the attribute is the number of rows *r* in the GLT matrix.
        + *r* = 1 corresponds to a *t*-test of the weighted sum of betas against the null hypothesis that the sum is 0.
        + *r* > 1 corresponds to an *F-*test of the *r* weighted beta sums defined by the individual rows against the null hypothesis that these sums are all zero.
      * The second value in the attribute is the number of columns in the GLT matrix.
        + This value *must* be the same as nreg, or 3dREMLfit will not like the matrix file (i.e., it will exit with an error message). It is present here to make the matrix definition self-contained, and as a check that the creator of the matrix file is not deranged.
      * The remaining values are the rows of the GLT matrix, nreg numbers per row, *r* rows, row after row.
        + In the example, there are only 2 nonzero numbers in the single row, corresponding (naturally) to the test vis-aud≟0.
        + There is no requirement that a GLT be a “contrast”; that is, the sum of the weights in the rows do not need to be 0.
* The “Basis” group of attributes is [NOT USED] by 3dREMLfit at this time.
  + I won’t describe them now, since this exercise is really getting dull.
    - Their function is to describe the response model used to construct the stimulus columns, and the example above is from AFNI program 3dDeconvolve.
  + I don’t even recall why I put this stuff in here (for Rick Reynolds, maybe?).
* CommandLine = "3dDeconvolve -input ......" [OPTIONAL]
  + This option is used to write the command that generated the matrix file into the output dataset(s) history note, for the potential elucidation of any user of the data. Otherwise, it is not needed or used.

Things such as volume labels and history notes are specific to AFNI-formatted datasets, including NIFTI files containing an AFNI-formatted extension which holds such piffle. Output datasets with this auxiliary information are written by 3dREMLfit, but such AFNI extensions will not be inherited by descendants created in non-AFNI software.

Note that the attributes and matrix are stored in an XML-ish structure of the form

<matrix ... *lots of attributes ...* >

Lots of numbers in text format

</matrix>

The software tries to be reasonably forgiving of formatting errors, but is not infinitely forgiving. The final **</matrix>** can be omitted, for example, if no other data follows the “**Lots of numbers**”.

Later Notes

* 3dREMLfit attribute names are now no longer case sensitive. The data element name **<matrix>...</matrix>** is still case sensitive.
* You can extract the matrix from a .xmat.1D file generated in 3dDeconvolve using a Unix command along this line:
  + grep -v '^#' Fred.xmat.1D > Fred.rawmat.1D
  + This works since all the non-matrix lines start with the shell comment character ‘#’.

**Why Use 3dREMLfit?**

Features on the road less traveled include:

* Voxelwise estimation of the temporal autocorrelation ARMA(1,1) model, rather than using the same temporal autocorrelation model for all voxels.
  + Is somewhat better than a global autocorrelation model at the usual TRs (> 1s) current in FMRI data acquisition:
    - <http://dx.doi.org/10.1038/s41467-019-09230-w>
    - <https://www.biorxiv.org/content/10.1101/323154v2>
  + Output of the pre-whitened residuals as well as the “ordinary” residuals.
  + REML estimation of the ARMA(1,1) parameters does not require that the time points be on a pure regularly spaced (fixed TR=Δ*t*) grid – time breaks from censoring or imaging runs are allowed for.
  + Voxelwise REML estimation is parallelized across multiple CPU cores, for speed.
* Allows for voxelwise regressors of no interest to be appended to the design matrix.
* Symbolic specification of linear tests among the beta estimates.
* Can do both GLSQ and OLSQ in the same program run, if you want to compare.
* Also see <https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dREMLfit.html>