PreQual (dtiQA_v7.1.3 Multi) User Guide

Leon Cai and Qi Yang
MASI Lab
July 30, 2020

COMMAND:

```
singularity run
-e
-B /path/to/inputs/directory/:/INPUTS
-B /path/to/outputs/directory/:/OUTPUTS
-B /path/to/freesurfer/license.txt:/APPS/freesurfer/license.txt (Optional, needed only for SynbO-DisCo)
--nv (Optional, see options --eddy_cuda and --eddy_extra_args; GPU support currently experimental)
/path/to/dtiQA_v7.1.3.simg
pe_axis
[options]
```

ARGUMENTS:

Input Directory: The dtiQA_config.csv file and at least one diffusion weighted image must be provided.

- dtiQA config.csv (see below for format, must be named exactly)
- <image1>.nii.gz (diffusion weighted image)
- <image1>.bval (units of s/mm2, in the FSL format)
- <image1>.bvec (normalized unit vectors in the FSL format)
- <imageN>.nii.gz (diffusion weighted image)
- <imageN>.bval (units of s/mm2, in the FSL format)
- <imageN>.bvec (normalized unit vectors in the FSL format)
- t1.nii.gz (Optional, used for Synb0-DisCo, must be named exactly)
- Other files as needed (see --extra eddy args for more information)

Output Directory: Outputs listed at the end of this document

pe_axis: Phase encoding axis of all the input images. We do NOT support different phase encoding axes between different input images at this time. The options are i and j and correspond to the first and second dimension of the input images, respectively. Note that FSL does not currently support phase encoding in the third dimension (i.e. k, the dimension in which the image slices were acquired, commonly axial for RAS and LAS oriented images). This parameter is direction AGNOSTIC. The phase encoding directions of the input images along this axis are specified in the dtiQA_config.csv file. See "dtiQA_config.csv Format" and "Example Phase Encoding Schemes" for more information.

dtiQA_config.csv FORMAT:

```
The format for the configuration CSV file is as follows:

<imagel>,pe_dir,readout_time

:
 <imageN>,pe_dir,readout_time
```

- <image> is the shared file PREFIX between the corresponding NIFTI, BVAL, and BVEC files for that particular image (i.e. HARDI iso.nii.gz -> HARDI iso)
- pe_dir is either + or -, corresponding to the direction along the phase encoding axis (as defined by the parameter pe axis) on which the image is phase encoded.
 - Note that a combination of phase encoding axis and direction map to specific anatomical (i.e. APA, APP, etc.) directions based on the orientation of the image. So, for instance in a RAS image, an axis of j and direction of + map to APP. We infer the orientation of the image from the header of the NIFTI using nibabel tools and output the best anatomical phase encoding direction interpretation of the input direction in the PDF.
- readout_time is a non-negative number, the readout_time parameter required by FSL's eddy. The absolute value of this parameter is used to scale the estimated b0 field. Note a value of 0 indicates that the images are infinite bandwidth (i.e. no susceptibility distortion). See "Example Phase Encoding Schemes" below for more information.

EXAMPLE PHASE ENCODING SCHEMES:

Here are some different example combinations of pe_axis, pe_dir, and readout_time parameters and the corresponding FSL acquisition parameters lines:

pe_axis	pe_dir	${ t readout_timeacqparams}$	line			
i	+	0.05	1,	0,	Ο,	0.05
j	_	0.1	Ο,	-1,	0,	0.1

OPTIONS:

--bval threshold N

A non-negative integer threshold under which to consider a b-value to be zero. Useful when some MRI machines do not allow for more than one b0 volume to be acquired so some users acquire scans with extremely low b-values to be treated like b0 volumes. Setting this value to 0 results in no thresholding. Units = s/mm2.

Default = 50

--denoise on/off

Denoise images prior to preprocessing using Marchenko-Pastur PCA <u>implemented in MRTrix3</u>. The SNR of the b0s of the final preprocessed images are reported in the PDF output regardless of whether this option is on or off.

Default = on

--prenormalize on/off

Intensity normalize images prior to preprocessing by maximizing the intra-mask intensity-histogram intersections between the averaged b0s of the scans. If this option is on, these histograms before and after prenormalization will be reported in the output PDF. This is done to avoid gain differences between different diffusion scans. If this option is off, we assume that the various input images all have the same gain. That being said, we still estimate and report the gain factors and

intensity histograms in a gain QA page and report warnings if estimated gains greater than 5% are found.

Default = on

--symb0 on/off

Run topup with a synthetic b0 generated with the Synb0-DisCo deep-learning framework if no reverse phase encoded images are supplied and a T1 image is supplied. Synb0-DisCo requires at least 24GB of RAM.

Default = on

--extra_topup_args="string"

Extra arguments to pass to FSL's topup. Topup will always run with the following:

```
--subsamp=1,1,1,1,1,1,1,1,1

--miter=10,10,10,10,10,20,20,30,30

--lambda=3.3e-4,6.7e-5,6.7e-6,1e-6,3.3e-7,3.3e-8,3.3e-9,3.3e-11,6.7e-13
```

For topup options that require additional inputs, place the file in the inputs directory and use the following syntax: "--<myinputoption> /INPUTS/<file.ext>". For topup options that produce additional outputs, the file will save in the output directory under the "TOPUP" folder by using the following syntax: "--<myoutputoption> /OUTPUTS/TOPUP/<file.ext>". Note that in this case "/INPUTS" and "/OUTPUTS" should be named exactly as is and are NOT the path to the input and output directory on your file system.

Default = none

--eddy_cuda 8.0/9.1/off

Run FSL's eddy with NVIDIA GPU acceleration. If this parameter is 8.0 or 9.1, either CUDA 8.0 or 9.1 must be installed and properly configured on your system, respectively, and the --nv flag must be run in the singularity command. If this parameter is off, eddy is run with OPENMP CPU multithreading. See --num_threads for more information. CUDA is required to run eddy with --mporder (intra-volume slice-wise motion correction). See --extra_eddy_args for more information.

Default = off

--eddy_mask on/off

Run eddy with or without a brain mask. If on, FSL's brain extraction tool (bet) is used with a low threshold to create a rough brain mask for eddy. This can sometimes produce poor results. If off, no mask is used and produces empirically minor differences in results than when a mask is used. If this option is on, the contour of this mask is drawn in the PDF.

Default = on

--extra eddy args="string"

Extra arguments to pass to FSL's eddy. Eddy will always run with the following:

--repol

Note that if --mporder is passed here, --eddy_cuda must be 8.0 or 9.1 and the singularity option --nv must be passed into the container, as intra-volume slice-wise motion correction requires GPU acceleration.

For eddy options that require additional inputs, place the file in the inputs directory and use the following syntax: "--<myinputoption> /INPUTS/<file.ext>". For eddy options that produce additional outputs, the file will save in the output directory under the "EDDY" folder by using the following syntax: "--<myoutputoption> /OUTPUTS/EDDY/<file.ext>". Note that in this case "/INPUTS" and "/OUTPUTS" should be named exactly as is and are NOT the path to the input and output directory on your file system.

Default = none

--postnormalize on/off

Intensity normalize images after preprocessing by maximizing the intra-mask intensity-histogram intersections between the averaged b0s of the scans. If this option is on, these histograms before and after postnormalization will be reported in the output PDF.

Note: This option was intended for testing and is left for posterity. It is not recommended at this time.

Default = off

--correct bias on/off

Perform <u>ANTs N4 bias field correction</u> as <u>called in MRTrix3</u>. If this option is on, the calculated bias field will be visualized in the output PDF.

Default = off

--split_outputs

Split the fully preprocessed output (a concatenation of the input images) back into their component parts and do NOT keep the concatenated preprocessed output.

Default = Do NOT split and return only the concatenated output

--keep intermediates

Keep intermediate copies of diffusion data (i.e. denoised, prenormalized, bias-corrected, etc.) used to generate final preprocessed data. Using this flag may result in a large consumption of hard disk space.

Note: Due to space concerns, special permission needed to use this option on XNAT.

--num threads N

A positive integer indicating the number of threads to use when running portions of the pipeline that can be multithreaded (i.e. MRTrix3, ANTs, and FSL's eddy without GPU acceleration).

Note: Due to resource concerns, special permission needed to multi-thread on XNAT.

Default = 1 (do NOT multithread)

--project string

String describing project in which the input data belong to label PDF output

Default = proj

--subject string

String describing subject from which the input data were acquired to label PDF output

Default = subj

--session string

String describing session in which the input data were acquired to label PDF output

Default = sess

--help, -h

PIPELINE ASSUMPTIONS:

- All NIFTI images are consistent with a conversion from a DICOM using dcm2niix (at least v1.0.20180622) by Chris Rorden and are raw NIFTIs without distortion correction. We require this as dcm2niix exports b-value/b-vector files in FSL format and removes ADC or trace images auto-generated in some Philips DICOMs. In addition dcm2niix correctly moves the gradients from scanner to subject space, whereas some Philips PARREC converters do not, which may result in spurious results or pipeline failure.
 - We expect raw volumes only, no ADC or trace volumes. ADC volumes are sometimes encoded as having a b-value greater than 0 with a corresponding b-vector of (0,0,0) and trace volumes are sometimes encoded as having a b-value of 0 with a corresponding non-unit normalized b-vector, as in the case of some Philips PARREC converters. We check for these cases, remove the affected volumes, and report a warning in the console and in the PDF.
 - We cannot, unfortunately, account for failure of reorientation of gradients into subject space.
- Images will be processed in the order they are listed in dtiQA config.csv.
- The size of all the volumes across all images must all be the same.
- The location of b0 images inside the input images do not matter.
- As per the FSL format, we do not support non-unit normalized gradients. We also do not support gradient directions of 0,0,0 when the corresponding b-value is non-zero.

Gradients with the latter configurations may cause pipeline failure. We report warnings in the output PDF if we identify these.

- The phase encoding axis of all volumes across all images is the same.
- The phase encoding direction along the axis is the same for all volumes inside an image and is specified in the dtiQA config.csv file.
- Unless --prenormalize is on, we assume all input images have the same gain.
- dtiQA_v7 will preferentially preprocess images with FSL's topup using available images with complementary phase encoding directions (i.e. + and -, "reverse phase encodings"). If none are available and a T1 is available, dtiQA_v7 will synthesize a susceptibility-corrected b0 from the first image listed in dtiQA_config.csv with Synb0-DisCo for use with topup, unless the user turns the --synb0 parameter off. The readout time of this synthetic b0 will be zero and the phase encoding direction will be equal to that of the first image in dtiQA_config.csv. Otherwise, dtiQA_v7 will preprocess without topup and move straight to FSL's eddy.
- dtiQA_v7 uses topup and eddy for preprocessing, both of which at the present moment do NOT officially support DSI acquisitions but only single- and multi-shell. dtiQA_v7 will force topup and eddy to run on DSI data, but may not produce quality results. Please carefully check the PDF output as we report a warning if eddy detected non-shelled data and thus required the use of the force flag.
 - Note that eddy may erroneously detect data as non-shelled if there are fewer directions in one of the shells than others. Because we merge the images for preprocessing, a notable example of this is when a reverse-phase encoded image uses a different shell than the forward images and has significantly fewer directions.
- For preprocessing, eddy will motion correct to the first b0 of each image.
- MRTrix3 by default preferentially uses the qform for understanding NIFTI orientations. Nibabel uses the sform. We set MRTrix3 to use the sform in our pipeline, and thus we preferentially use the sform when the two don't match.
- No b0 drift correction is performed.

PIPELINE PROCESSING STEPS:

- 1. Threshold all b-values such that values less than the --bval_threshold parameter are 0.
- 2. Check that all b-vectors are unit normalized and all b-values greater than zero have associated non-zero b-vectors. For any volumes where this is not the case, we remove them, flag a warning for the output PDF, and continue the pipeline.
- 3. If applicable, denoise all diffusion scans with dwidenoise (Marchenko-Pastur PCA) from MrTrix3.
- 4. If applicable, prenormalize all diffusion scans. To accomplish this, extract all b0 images from each diffusion scan and average them. Then find a rough brain-mask with FSL's bet and calculate an intensity scale factor such that the histogram intersection between the intra-mask histogram of the different scans' averaged b0s to that of the first scan is maximized. Apply this scale factor to the entire diffusion weighted scan. This is done to avoid gain differences between different diffusion scans.
 - a. If prenormalization is not indicated, we still run the prenormalization algorithms to calculate rough gain differences and report the gain factors and intensity histograms in a gain QA page. The outputs of the algorithms, however, are NOT propagated through to the rest of the pipeline.
- 5. Prepare data for and run preprocessing with topup and eddy
 - a. Topup:
 - i. Extract all b0s from all scans, maintaining their relative order.

- ii. (Optional) If a T1 is supplied and no complementary (i.e. reverse) phase encoded images are provided, use Synb0-DisCo to convert the first b0 of the first scan to a susceptibility-corrected b0.
- iii. Build the acquisition parameters file required by both topup and eddy
 - 1. For the number of b0s from each image, add the same phase encoding and readout time line to the acquisition parameters file, as outlined in "Example Phase Encoding Schemes".
 - a. Example: In the case where we have a phase encoding axis of j and two images, one with 7 b0s, + direction, and 0.05 readout time and one with 3 b0s, direction, and 0.02 readout time, this file will have 10 lines. The first 7 lines are identical and equal to [0, 1, 0, 0.05]. The last three lines are also identical and equal to [0, -1, 0, 0.02].
 - 2. (Optional) If Symb0-DisCo is run because no complementary phase encoding directions are supplied and --symb0 is not off, we add an additional line to the end of the file. This line is the same as the first line of the file except that the readout time is 0 instead.
 - a. Example: In the case where we have a phase encoding axis of j and two images, one with 7 b0s, + direction, and 0.05 readout time and one with 3 b0s, + direction, and 0.02 readout time, this file will have 11 lines. The first 7 lines are identical and equal to [0, 1, 0, 0.05]. The next three lines are also identical and equal to [0, 1, 0, 0.02]. Finally, the last line is equal to [0, 1, 0, 0].
- iv. We then concatenate all the b0s maintaining their order and run topup with the acquisition parameters file if images with complementary phase encoding directions are supplied or if a T1 was supplied. Otherwise, we move on to the next step, eddy.

b. Eddy

- i. Using the acquisition parameters file from the topup step, regardless of whether topup was performed, we build the eddy index file such that each volume in each image corresponds to the line in the acquisition parameters file associated with the first b0 of each scan. This is done to tell eddy that each volume in a given scan has the same underlying phase encoding scheme as the first b0 of that scan.
 - 1. Example: In the case where we have two images, one with 7 b0s and 100 total volumes and one with 3 b0s and 10 total volumes, the eddy index file has 100 1's followed by 10 8's.
- ii. Eddy is then run with either a mask generated with bet and the -f 0.25 and -R options or without a mask (aka with a mask of all 1's), depending on user input (see the --eddy_mask option) and with the output of topup if topup was run. Eddy also runs with the --repol option for outlier slice replacement. We also first run eddy with a check looking for shelled data. If the check fails, eddy is then run with the --data_is_shelled flag to force eddy to run on all scans, DSI included. Note that DSI data is not officially supported by FSL… yet?
 - 1. If eddy detects data is not shelled, we report this as a warning
 - 2. As noted in the assumptions section above, eddy may erroneously detect data as non-shelled if there are fewer directions in one of the shells than others. Because we merge the images for preprocessing, a notable example of this is when a reverse-phase

encoded image uses a different shell than the forward images and has significantly fewer directions.

- iii. Eddy also performs byec rotation correction and calculates the voxel-wise signal-to-noise ratios of the b0 images and the voxel-wise contrast-to-noise ratios for the diffusion weighted images. SNR is defined as the mean value divided by the standard deviation. CNR is defined as the standard deviation of the Gaussian Process predictions (GP) divided by the standard deviation of the residuals between the measured data and the GP predictions.
- 6. If the user chooses to, we then perform post-normalization in the same fashion as pre-normalization.
- 7. If the user choose to, we then wrap up preprocessing with an N4 bias field correction as implemented in ANTs via MRTrix3's dwibiascorrect.
- 8. We generate a brain mask using FSL's bet2 with the following options:
 - a. -f 0.25
 - b. -R
- 9. We then apply the mask to the preprocessed images while we calculate tensors using MRTrix3's dwi2tensor function. For visualization we discard tensors that have diagonal elements greater than 3 times the apparent diffusion coefficient of water at 37° C (~ 0.01).
 - a. We also reconstruct the preprocessed image from the tensor fit for further analysis later. dwi2tensor does this for us.
- 10. We then convert the tensor to FA and MD images (and visualize them later too) as well as AD, RD, and V1 eigenvector images for the user. The latter 3 are not visualized.

PIPELINE QUALITY ASSURANCE STEPS:

- 1. We start with the brain mask generated above to generate a mask used for the following quantification of tensor fit using a chi-squared statistic.
 - a. First, we calculate the mean image for each unique b-value (0 not included). Then we run FSL's FAST to isolate the CSF on each meaned image. We then take the average probability of a voxel being CSF across all unique b-values and assign >15% probability to be a positive CSF voxel.
 - b. Then we call the final chi-squared mask to be the intersection of the inverted CSF mask and a 1-pixel eroded version of the brain mask.
- 2. On the voxels inside the chi-squared mask, we perform the following quality assurance:
 - a. We perform a chi-squared analysis for each slice for each volume in the main image by calculating the ratio between the sum-squared error of the fit and the sum-squared intensities of the slice.
 - b. We extract the average FA for a number of white matter ROIs defined by the Hopkins atlas. We do this by non-rigidly registering the atlas to our FA output and extracting the FA values contained in each ROI.
 - c. We check the gradients output by eddy (i.e. the preprocessed gradients) with dwigradcheck from MRTrix3. This performs tractography and finds the optimal sign and order permutation of the b-vectors such that the average tract length in the brain is most physiological.
 - i. These optimized gradients are saved in the OPTIMIZED_BVECS output folder, and the gradients output by eddy in the PREPROCESSED folder are NOT overwritten.
 - ii. The original, preprocessed, and preprocessed + optimized gradients are visualized as outlined below.
- 3. We then visualize the entire pipeline.

- a. On the first page we describe the methods used for that run of the pipeline (what inputs were provided, what sort of preprocessing happened, etc.).
- b. We then visualize the raw images with the interpreted phase encoding schemes.
- c. If Symb0-DisCo was run, we then visualize the distorted b0 (first b0 of first scan) and T1 used as inputs as well as the output susceptibility corrected b0 in their native space.
- d. If pre- or post-normalization was performed, we then visualize the intra-mask histograms before and after these steps as well as the calculated scaling factors. If pre-normalization is not performed, we visualize the histograms that would have been generated with pre-normalization ONLY as a check for gain differences.
- e. We then visualize the first b0 of the images before and after preprocessing with the contours of the brain and stats masks overlaid as well as the contours of the eddy mask overlaid if it is used.
- f. We plot the motion and angle correction done by eddy as well as the RMS displacement and median intensity for each volume and the volume's associated b-value. These values are read in from an eddy output text file and we also compute and save the average of these values. In addition, we plot the outlier slices removed and then imputed by eddy as well as the chi-squared fit, with maximal bounds 0 to 0.2. The median chi-squared values are shown across volumes and slices.
- g. We then plot the original raw b-vectors scaled by their b-values, the preprocessed ones output by eddy, and the optimized ones determined by dwigradcheck applied to the preprocessed ones.
- h. If bias field correction was performed, we then visualize the calculated fields.
- i. We then visualize some central slices of the average volumes for all unique b-values, including b = 0 and report the median intra-mask SNR or CNR calculated by eddy as appropriate.
- j. We visualize the tensors using MRTrix3's mrview, omitting the tensors with negative eigenvalues or eigenvalues greater than 3 times the ADC of water at
- k. We then visualize some central slices of the FA map clipped from 0 to 1 as well as the average FA for the Hopkins ROIs and the quality of the atlas
- 1. Lastly, we visualize some central slices of the MD map clipped from 0 to 0.003 (ADC of water at 37° C).

Output Directory

<imageN %> denotes the original prefix of imageN with the preceding preprocessing step descriptors tacked on the end. For example, in the case of the PRENORMALIZED directory, the prefix for imageJ may or may not include "denoised" depending on whether the denoising step was run.

Folders and files in **bold** are always included. Folders and files in italics are removed if --keep intermediates is NOT indicated

1. THRESHOLDED BVALS

o <image1>.bval

o <imageN>.bval

2. CHECKED (these contain the volumes that have passed the bval/bvec checks)

```
o <image1> checked.nii.gz
      o <image1> checked.bval
      o <image1> checked.bvec
      o <imageN> checked.nii.gz
      o <imageN> checked.bval
      o <imageN> checked.bvec
3. DENOISED (these files are only created if --denoise is on)
      o <image1> checked denoised.nii.gz
      o <imageN> checked denoised.nii.gz
4. PRENORMALIZED (these files are only created if --prenormalize is on)
      o <image1 %> norm.nii.gz
      o <imageN %> norm.nii.gz
5. GAIN CHECK (these files are only created if --prenormalize is off)
      o <image1 %> norm.nii.gz
      o <imageN %> norm.nii.gz
6. TOPUP (these files are only created if topup was run)
      o acqparams.txt (same as OUTPUTS/EDDY/acqparams.txt)
      o preproc input b0 first.nii.gz (only if Synb0-DisCo is run)
      o b0 syn.nii.gz (only if Synb0-DisCo is run)
      o preproc input b0 all.nii.gz or preproc input b0 all smooth with b0 syn.nii.gz
      o preproc input b0 all topped up.nii.gz or
        preproc input b0 all smooth with b0 syn topped up.nii.gz
      o preproc input b0 all.topup log or
        preproc input b0 all smooth with b0 syn.topup log
      o topup field.nii.gz
      o topup results fieldcoef.nii.gz
      o topup results movpar.txt
7. EDDY
      o acqparams.txt (same as OUTPUTS/TOPUP/acqparams.txt)
      o index.txt
      o preproc input.nii.gz
      o preproc input.bval
      o preproc input.bvec
      o preproc input eddyed.nii.gz (renamed from "eddy results.nii.gz")
      o preproc input eddyed.bval
      o preproc input eddyed.bvec
      o eddy mask.nii.gz (only included if --eddy mask is on)

    eddy results.eddy command txt

    eddy results.eddy movement rms (describes volume-wise RMS displacement)

      o eddy results.eddy outlier free data.nii.gz

    eddy results.eddy outlier map (describes which slices were deemed outliers)

      o eddy results.eddy outlier n sqr stdev map

    eddy results.eddy outlier n stdev map

    eddy results.eddy outlier report

    eddy results.eddy parameters (describes volume-wise rotation and translation)

      o eddy results.eddy post eddy shell alignment parameters

    eddy results.eddy post eddy shell PE translation parameters

    eddy results.eddy restricted movement rms

    eddy results.eddy rotated bvecs (describes properly rotated b-vectors)
```

```
eddy_results.eddy_values_of_all_input_parameters
      o eddy results.eddy cnr maps.nii.gz
8. POSTNORMALIZED (these files are only created if --postnormalize is on)
      o <image1 %> topup eddy norm.nii.gz (" topup" only applies if topup was run)
      o <imageN %> topup eddy norm.nii.gz
9. UNBIASED (these files are only created if --correct bias is on; this folder is
   removed if --correct bias is off)
      o normed unbiased.nii.gz (if postnormalization is run) or
        preproc input eddyed unbiased.nii.gz (if postnormalization is not run)
      o bias field.nii.gz
10.PREPROCESSED (these represent the final output of the pipeline)
      o dwmri.nii.gz (dwmri* files deleted only if --split outputs is also set)
      o dwmri.bval
      o dwmri.bvec
      o <imagel> preproc.nii.gz (* preproc files exist only if --split outputs is set)
      o <image1> preproc.bval
      o <image1> preproc.bvec
      o <imageN> preproc.nii.gz
      o <imageN> preproc.bval
      o <imageN> preproc.bvec
      o mask.nii.gz
11.TENSOR
      o dwmri tensor.nii.gz
      o dwmri recon.nii.gz
12.SCALARS
      o dwmri tensor fa.nii.gz
      o dwmri tensor md.nii.gz
      dwmri tensor ad.nii.gz
      dwmri tensor rd.nii.gz
      o dwmri tensor v1.nii.gz
13.STATS
      o atlas ants fa.nii.gz
      o chisq mask.nii.gz
      o chisq matrix.txt

    eddy avg abs displacement.txt

      o eddy median cnr.txt
      o eddy_avg_rel_displacement.txt
      o eddy avg rotations.txt
      o eddy_avg_translations.txt
      o roi avg fa.txt

    stats.csv (contains summary of all motion, SNR/CNR, and average FA stats)

14.OPTIMIZED BVECS
      o dwmri.bval
      o dwmri.bvec
15.PDF

    dtiQA.pdf (final QA document)
```