QDET update
Round 5 project

Daniel Barboriak MD
Neuroradiology
January 19, 2017

Acknowledgement

- Thanks to QIBA and NIBIB for funding this project

Background

- DROs simulating T1 mapping and dynamic images obtained with DCE-MRI experiments are helpful, but how we can use these to measure performance is not completely clear

Background

- Software differ in the amount of bias

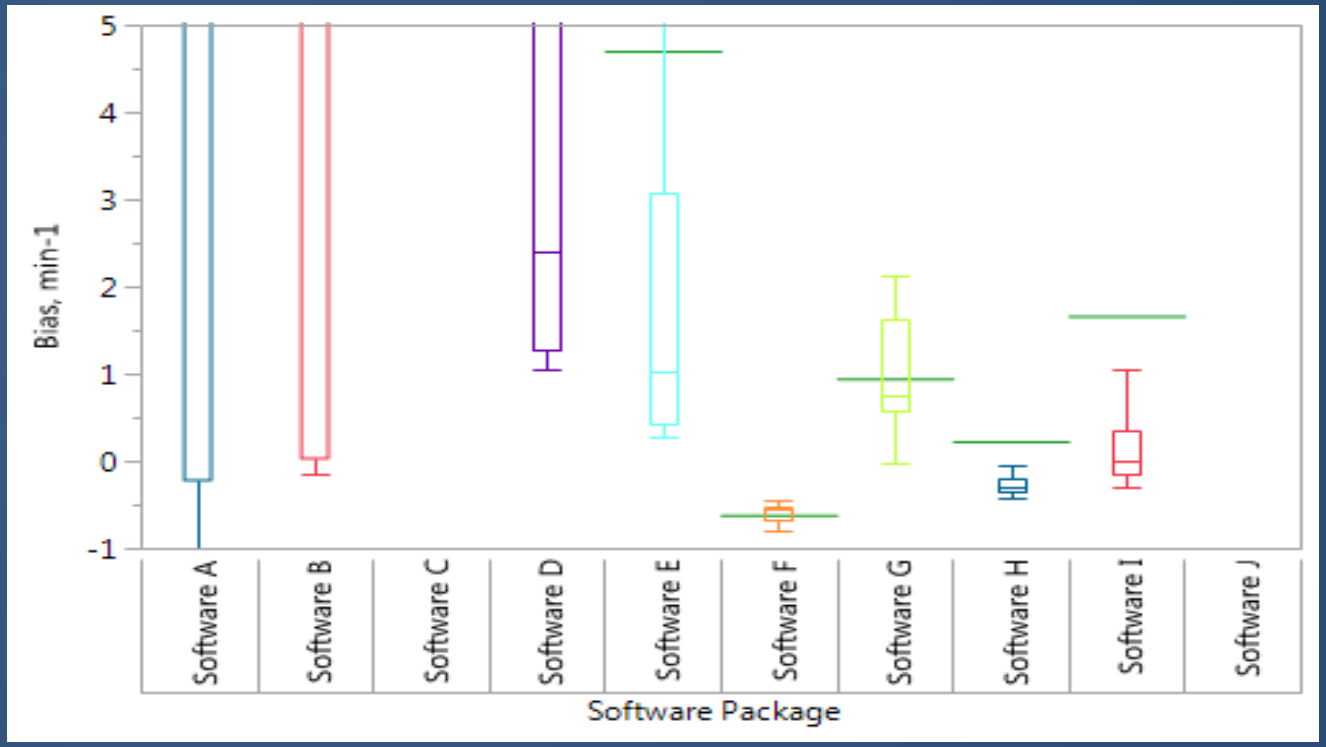


Figure above shows the mean bias for all software packages over all noisy conditions in DRO version 9. Standard box and whisker plots with data range, first and third quartile, median (within box) and mean (green line) are show. Optimal performance is zero bias.

Background

- Software differ in the amount of precision

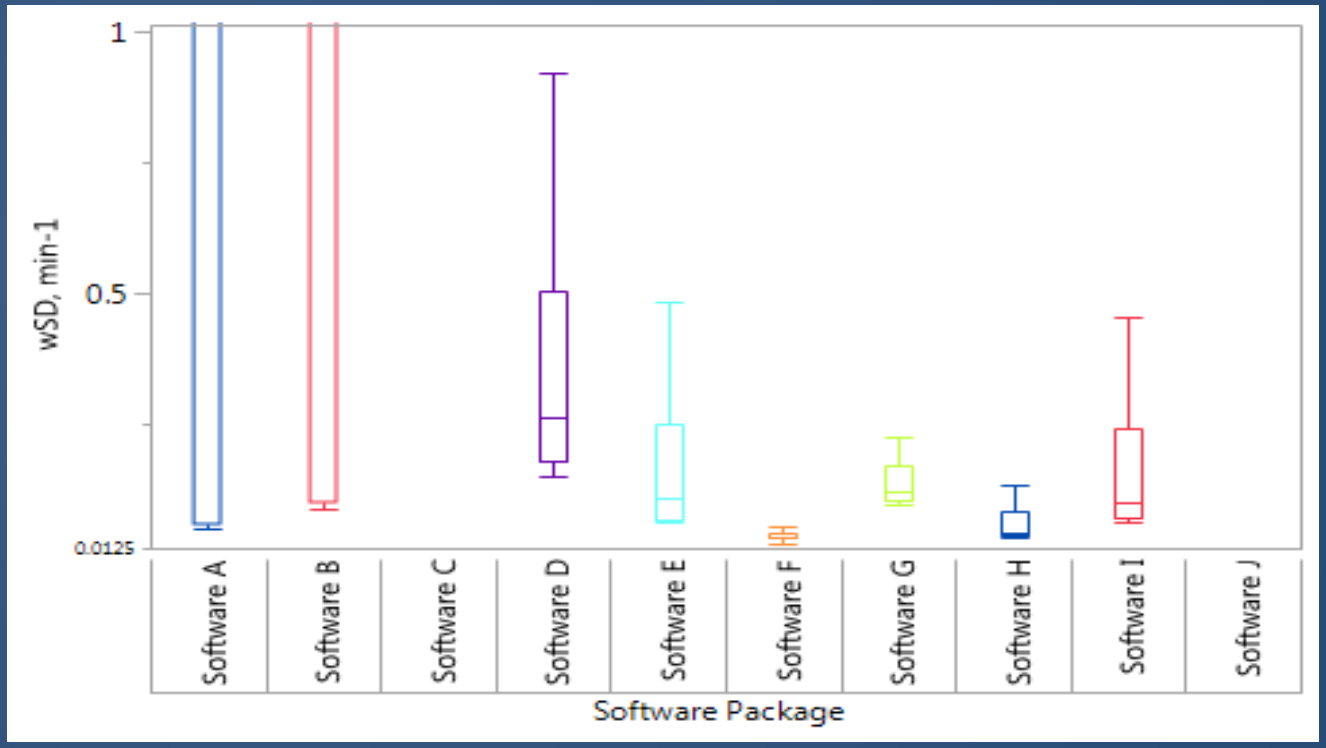


Figure above shows the wSD for all software packages over all noisy conditions in DRO version 9. Standard box and whisker plots with data range, first and third quartile, median (within box) and mean (green line) are shown. Lower wSD indicates better precision.

Background

- Not clear to combine this data. Is software that is more biased but also more precise "better" than software that is less precise but less biased?
- Concept of aggregate vs. disaggregated metrics
  - Total Deviation Index (TDI) has been suggested as a metric with intuitive meaning
  - Different aggregated metrics may give different results

Background

- Opportunity to provide open source aggregated metrics to evaluate DROs and other objects related to imaging
- Used MEVIS / Dr. Laue's QDET as a starting point

Goal 1: Provide open source access to statistical analyses

- Root Mean Square Deviation (RMSD)
- Concordance Correlation Coefficient (CCC)
- Total Deviation Index (TDI)
- Sigma metric
- Bland-Altman repeatability coefficient

Example: T1 mapping

- Image data: V:\QIBA\QIBA Project Round 5\Internal_QDET_Validation\v3_DCEMRIS 4_Sigma_5\T10_DCEMRIS4v046.img

- Intermediate Use Case Mask: V:\QIBA\QIBA Project Round 5\Phantom_Intermed_Use_Case_Masks\In termediate Use Case Sigma 5.tif

RMSD

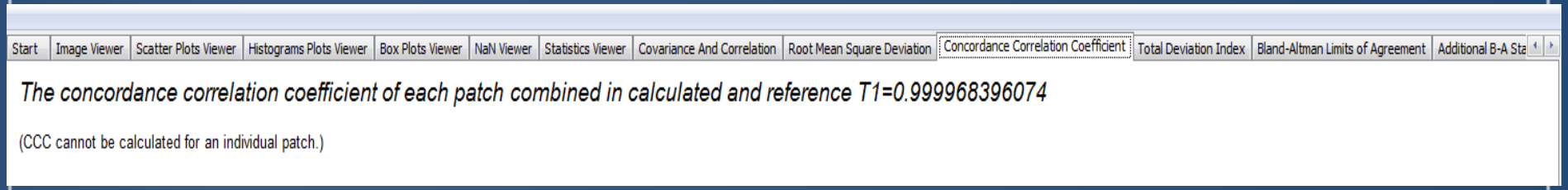
Start | Image Viewer | Scatter Plots Viewer | Histograms Plots Viewer | Box Plots Viewer | NaN Viewer | Statistics Viewer | Covariance And Correlation | Root Mean Square Deviation | Concordance Correlation Coefficient | Total Deviation Index | Bland-Altman Limits of Agreement | Additional B-A Sta

The root mean square deviation of each patch in calculated and reference T1:

	R1 = 0.0004	R1 = 0.0005	R1 = 0.0007	R1 = 0.0010	R1 = 0.0014	R1 = 0.0020	R1 = 0.0028	R1 = 0.0040	R1 = 0.0057	R1 = 0.0080	R1 = 0.0113	R1 = 0.0160	R1 = 0.0226	R1 = 0.0320	R1 = 0.0453
S0 = 500	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 1000	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd = 27.4210137	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 2000	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd = 38.0797260	rmsd = 23.4997692	rmsd = 13.2818925	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 5000	rmsd =	rmsd =	rmsd = 82.4138520	rmsd = 43.3075566	rmsd = 24.9931319	rmsd = 16.4172163	rmsd = 8.6568181	rmsd = 5.0118449	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 10000	rmsd = 169.8408968	rmsd = 69.2654233	rmsd = 46.2454721	rmsd = 24.5654443	rmsd = 13.2437242	rmsd = 6.9080210	rmsd = 3.7666067	rmsd = 2.5854839	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 20000	rmsd = 68.0013888	rmsd = 40.6092620	rmsd = 24.0803259	rmsd = 11.7818384	rmsd = 7.1985755	rmsd = 3.7366117	rmsd = 2.5016229	rmsd = 1.2279136	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =
S0 = 50000	rmsd = 29.0176051	rmsd = 17.0406534	rmsd = 8.4449354	rmsd = 4.5504861	rmsd = 2.6468072	rmsd = 1.3848149	rmsd = 0.7930722	rmsd = 0.5284498	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =	rmsd =

The root mean square deviation of all patches combined in calculated and reference T1=6.21782507301

CCC


Start
Image Viewer
Scatter Plots Viewer
Histograms Plots Viewer
Box Plots Viewer
NaN Viewer
Statistics Viewer
Covariance And Correlation
Root Mean Square Deviation
Concordance Correlation Coefficient
Total Deviation Index
Bland-Altman Limits of Agreement
Additional B-A Sta
The concordance correlation coefficient of each patch combined in calculated and reference T1=0.999968396074
(CCC cannot be calculated for an individual patch.)

TDI

Start | Image Viewer | Scatter Plots Viewer | Histograms Plots Viewer | Box Plots Viewer | NaN Viewer | Statistics Viewer | Covariance And Correlation | Root Mean Square Deviation | Concordance Correlation Coefficient | Total Deviation Index | Bland-Altman Limits of Agre

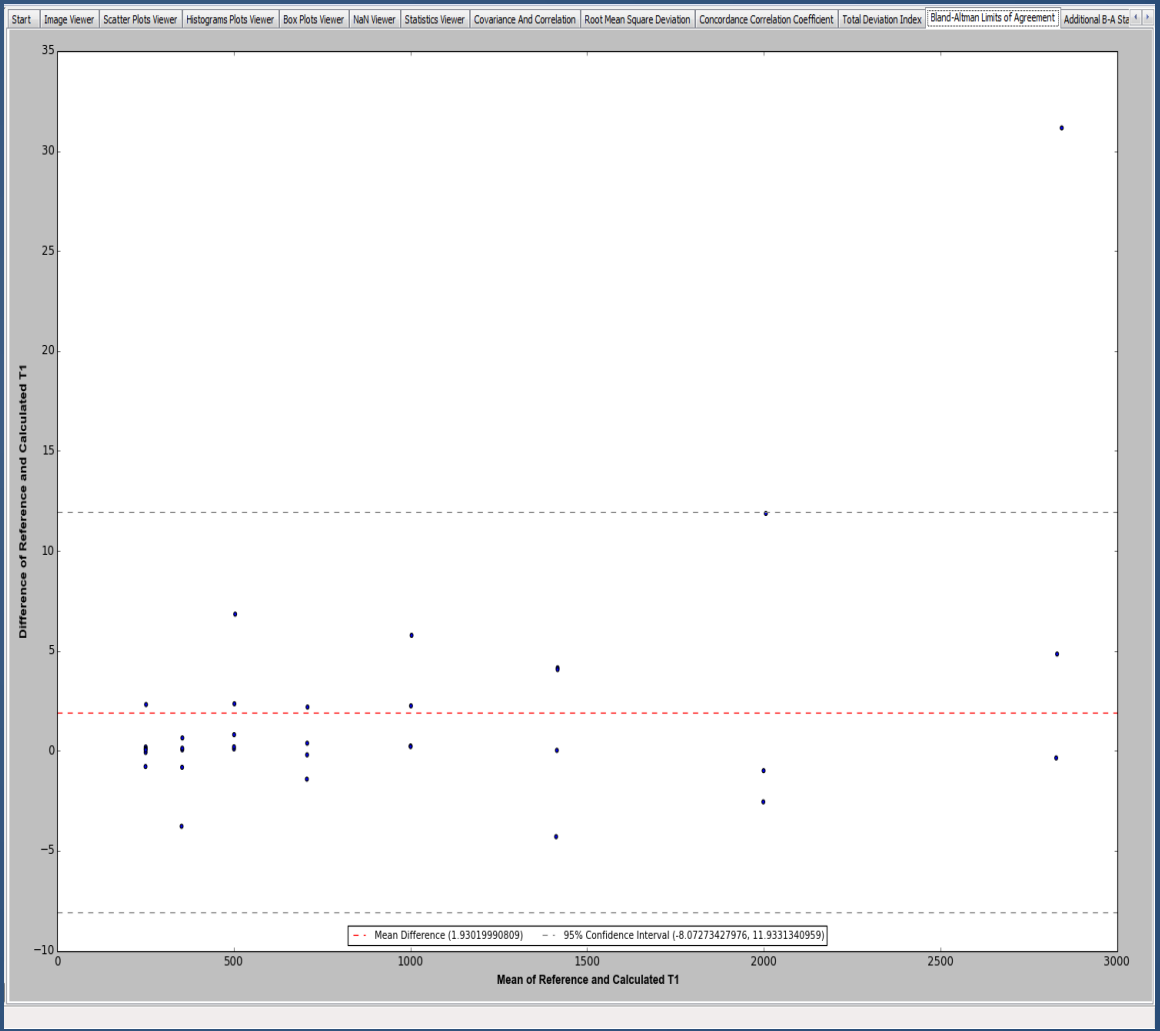
The total deviation indexes of each patch in calculated and reference T1:

	R1 = 0.0004	R1 = 0.0005	R1 = 0.0007	R1 = 0.0010	R1 = 0.0014	R1 = 0.0020	R1 = 0.0028	R1 = 0.0040	R1 = 0.0057	R1 = 0.0080	R1 = 0.0113	R1 = 0.0160	R1 = 0.0226	R1 = 0.0320	R1 = 0.0453
S0 = 500	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 1000	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi = 54.0130311	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 2000	tdi =	tdi =	tdi =	tdi =	tdi =	tdi = 74.9988135	tdi = 46.2847611	tdi = 26.1627185	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 5000	tdi =	tdi =	tdi = 162.3397317	tdi = 85.3012902	tdi = 49.2305366	tdi = 32.3359319	tdi = 17.0517529	tdi = 9.8725096	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 10000	tdi = 334.5026795	tdi = 136.4407345	tdi = 91.0924236	tdi = 48.3879172	tdi = 26.0879598	tdi = 13.6067510	tdi = 7.4185200	tdi = 5.0928483	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 20000	tdi = 133.9516990	tdi = 79.9594738	tdi = 47.4275646	tdi = 23.2082650	tdi = 14.1772741	tdi = 7.3604941	tdi = 4.9277813	tdi = 2.4187095	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =
S0 = 50000	tdi = 57.1520101	tdi = 33.5667580	tdi = 16.6351534	tdi = 8.9636189	tdi = 5.2132455	tdi = 2.7275790	tdi = 1.5620174	tdi = 1.0409234	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =	tdi =

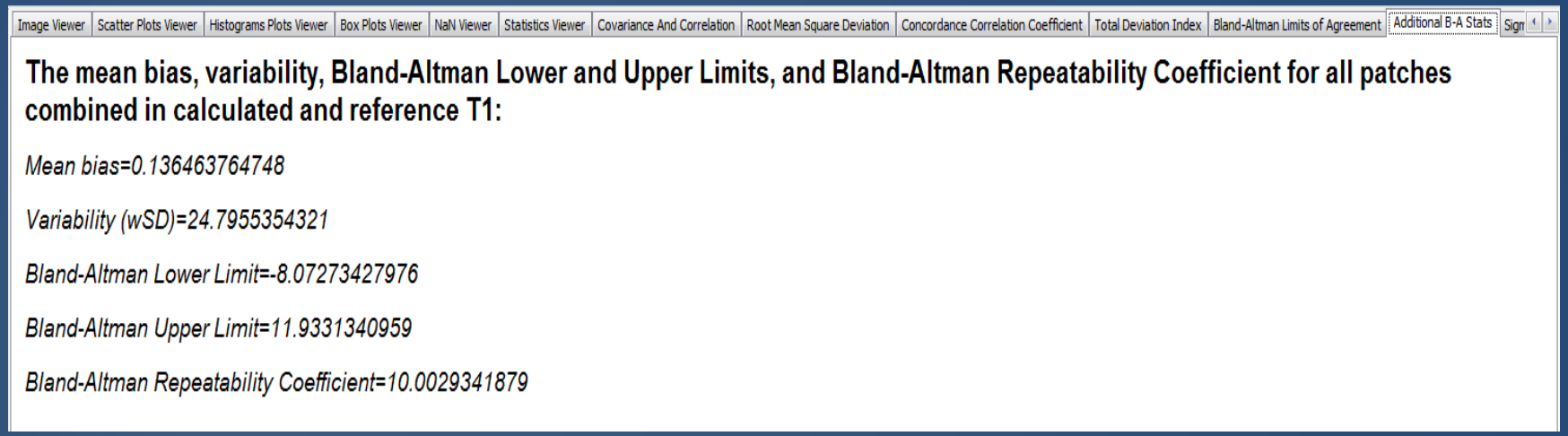
The estimated total deviation index of each patch combined in calculated and reference T1=31.1528125

The total deviation index of each patch combined in calculated and reference T1=12.3322691691

Bland-Altman Limits of Agreement



Bland-Altman statistics



Goal 1: Provide open source access to statistical analyses

- QDET works because it "knows" the structure of the T1 and dynamic DROs
- Opportunity to extend QDET by allowing text input:

100	100	1	sigma	5	1	0.0003536	0.001143440	0.000461835
100	99	1	sigma	5	1	0.0005	0.000853097	0.000043275
100	100	1	sigma	5	1	0.0007071	0.001023908	0.000046254
100	100	1	sigma	5	1	0.001	0.001194814	0.000030588
100	100	1	sigma	5	1	0.001414	0.001628560	0.000037071

Goal 1: Provide open source access to statistical analyses

- Verify that table input works:

Data series: QIBA v3, Sigma 5		
Statistic	QDET Image Result	QDET Table Result
RMSD	72.3410122204	72.3410134604
CCC	0.996101080147	0.99610108011
TDI	142.454159807	142.427606459
Bland-Altman Lower Limit	-131.085300578	-131.085322711
Bland-Altman Upper Limit	150.045842821	150.045833691
Bland-Altman Repeatability Coefficient	140.565571699	140.565578201

Table 1. QDET image result vs. table result for QIBA v3, Sigma 5

Goal 1: Provide open source access to statistical analyses

- Verify against other software programs that the aggregated metrics look correct:

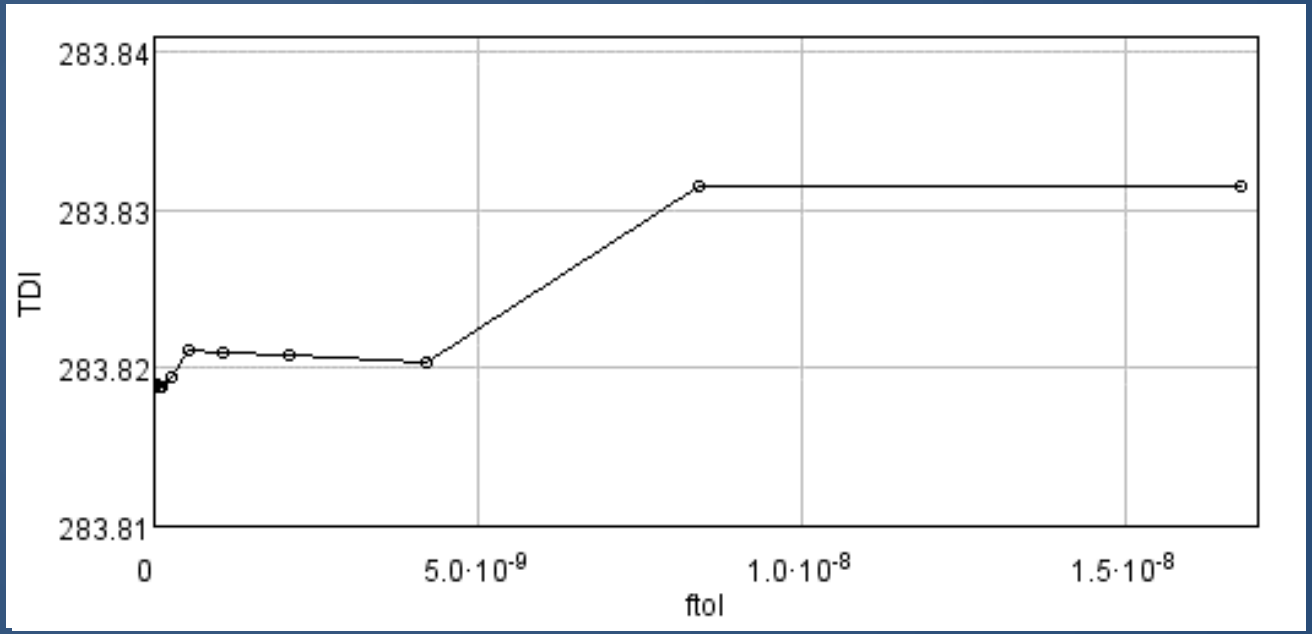
Data series: QIBA v3, Sigma 5		
Statistic	QDET Result	R Result
RMSD	72.3410134604	72.34101
CCC	0.99610108011	0.9961011 (“epiR”) 0.9961 (“Agreement”)
TDI	142.427606459	142.45416
Bland-Altman Lower Limit	-131.085322711	-131.759502
Bland-Altman Upper Limit	150.045833691	150.720013
Bland-Altman Repeatability Coefficient	140.565578201	141.2398

Table 3. QDET text result vs. result from R for QIBA v3, Sigma 5

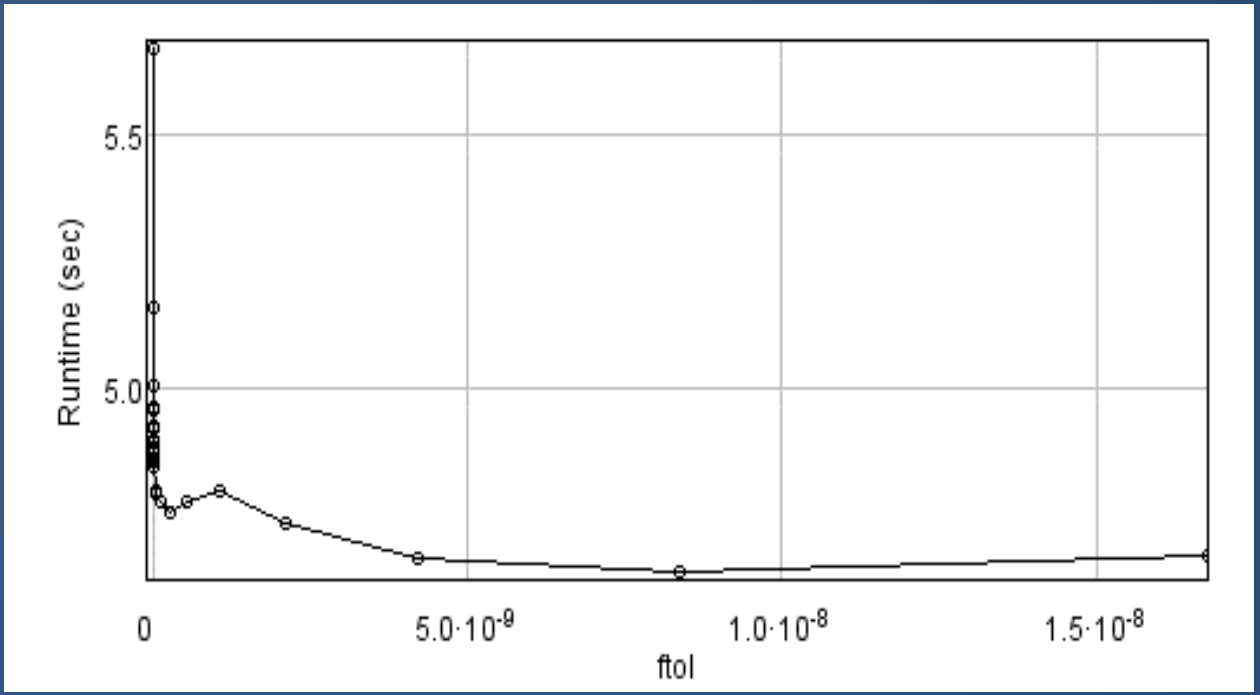
Goal 2: Use scripts to tune parameters

- Non-linear fitting routines (e.g. Levenberg-Marquardt) use "hidden" parameters to determine step size and stopping points. These may vary from software package to software package
- Can use QDET and the software iteratively to find the best performing parameters
- This process was run on the QIBA v3 Sigma 10 DRO. The first step was to run DCEMRIS4 (through an R script), doubling the ftol value with each iteration, starting at $1x10^{-15}$ until ftol reached a maximum value of $3x10^{-8}$. The second step used QDET on all output from the R script to compute RMSD, CCC, and TDI for each ftol variation

Goal 2: Use scripts to tune parameters



Goal 2: Use scripts to tune parameters



Goal 2: Use scripts to tune parameters

- To make this work, you need to run these iteratively from command line
  - For example, to run QDET with T1 calculated and reference images and apply Intermediate Use Case mask:
    - QIBA_evaluate_tool --mode T1 --cfile "V:\QIBA\QIBA Project Round 5\Internal_QDET_Validation\v3_DCEMRIS4_Sigma_5\T10_DCEMRIS4v046.img" --rfile "V:\QIBA\QIBA Project Round 5\H. Laue\QIBA evaluate tool - Original Executable Application\QIBA evaluate tool\reference_data\T1.dcm" --destination "V:\QIBA\QIBA Project Round 5\Internal_QDET_Validation\v3_DCEMRIS4_Sigma_5\v5_Demo_Output.img" --mask "V:\QIBA\QIBA Project Round 5\Phantom_Intermed_Use_Case_Masks\Intermediate Use Case Sigma 5.tif

Goal 3: Demonstration of use of aggregated measures of agreement to rank performance of competing image analysis algorithms

- QDET was used to calculate aggregate measures of agreement (RMSD, CCC, and TDI) for each software package submitted
- A Python script was then used to create spreadsheets of all statistics for every QIBA data set

T1	Software A	Software B	Software C	Software D	Software E
Mean	634.0553571	1147186481	164074.1249	653.6831177	651.8798416
RMSD	144.8120592	7457215516	1191151.712	198.0463201	302.8686174
RMSD Rank	2	15	14	4	11
CCC	0.983786081	6.35348E-08	0.00035044	0.97021142	0.92688791
CCC Rank	2	15	14	5	11
TDI (Nonparametric)	285.1088827	14684319376	2345601.415	389.8624031	595.9541133
TDI (Nonparametric) Rank	2	15	14	4	11
TDI (Parametric)	213.4406506	1712.69279	1712.41457	312.8584171	852.2532603
TDI (Parametric) Rank	1	14	13	6	11
Mean bias	0.42877646	55391190.17	7842.357107	4.395376344	51.80138393
Mean bias Rank	2	15	14	7	13
Variability (wSD)	253.1823559	8838702707	1085499.084	412.6976236	438.7223699
Variability Rank	4	15	14	9	11
Bland-Altman Lower Limit	-246.273405	-1.1082E+10	-1794749.61	-314.450982	-490.609454
Bland-Altman Upper Limit	233.9632418	13376138214	2121617.442	341.3963399	513.9482593
Bland-Altman Repeatability Coefficient	240.1183235	12228952373	1958183.528	327.923661	502.2788565
Bland-Altman Repeatability Coefficient Rank	2	15	14	4	11

Goal 4: Provide guides to interpretation

- Thanks to Dr. Obuchowski
- Example:
  - TDI describes the absolute difference between QIB measurements and their reference values. 95% of differences will be smaller than the TDI95. The TDI95_p is an estimate of TDI determined parametrically, assuming the underlying distribution of differences is Gaussian. The TDI95_np is determined nonparametrically, from the actual absolute difference found in the submitted data.

Next steps

- Publication
- Registration