Hello All,

Ahead of our telecon tomorrow I wanted to follow up on Ray's and Norbert's comments in response to the questions I had posed earlier.

Firstly, I think there are some points on which we can all agree, and that it would be useful to state these so that we can concentrate on trying to resolve or discuss issues in areas where views differ. We can probably agree that:

1) the residual error standard deviations from **back-calculation** of the calibration data used to define the matrix, should be viewed only as an indicator of the 'goodness of fit' and never as a guide to the predictive abilities of the math model when it is applied to truly independent data. Above all, these standard deviations should not be used as the sole criterion for evaluating a math model.

2) proper evaluation of the predictive capabilities of the math model requires data that are truly independent of the 'base' calibration data. The residual error standard deviations of such independent data are one metric that can be used in evaluation of the math model's predictive capability.

3) any technique that brings objectivity to the development of balance calibration load designs, or can be used to objectively assess the suitability of a design prior to performing a calibration, should be of great benefit to the balance community.

In regard to (3), when I first learned of Norbert's work from Dennis back in late 2007, I was excited at the prospect of there now being an objective way of identifying terms that might be inadequately defined by the calibration data. In a message to Norbert at that time I wrote - "*What is NEW to me in your approach is this idea of 'screening' the load design to reject terms that would not be properly defined by it. The additional 'screening' to remove terms that are not of real significance seems like a very logical thing to do, but for me the real 'breakthrough' is being able to define the terms that MUST be removed because leaving them in the model will corrupt the result.*" So I'm definitely a supporter of the idea of objective, quantitative, screening of balance calibration load designs, to eliminate terms from the math model that would be corrupt because they were inadequately defined by the data. But I am concerned by the Canard balance result, where rigid adherence to the criteria and threshold levels used within BALFIT resulted in at least two key interaction terms being dropped from the model. The resulting 'optimized' math model was certainly not that, and in fact it does not even come close to providing adequate modeling of the balance. Perhaps the currently used thresholds are not appropriate, or perhaps other additional criteria are required? I don't know the answer to that, but I don't think the poor result for the Canard balance should be shrugged off as some anomaly. Without an explanation of why 'things went wrong' in that particular case, it's impossible to have confidence in the process.

I would like to make some additional comments, (supported by results and plots), to add to those already provided by Ray and Norbert, and it seems easiest to do this by including the E-mail message below, with my comments (in blue), plus results and plots, interlaced.

Hello Ray,
Thanks for you Q&A email. Below please find some additional comments in red that I wanted to make. I have also attached two result files in order to demonstrate the advantage of using VIFs for the definition of a regression model. Talk to you during next weeks telecon ...
Best regards,
Norbert

On 3/4/09 10:36 AM, "Rhew, Ray D. (LARC-D201)" <ray.d.rhew@nasa.gov> wrote:

All,

I have addressed some of the questions brought up by this discussion below.
Thanks to Robin and Norbert for their diligence in providing additional information on this subject balance.

Ray

Q1. Can anyone clearly demonstrate any circumstance in which any of the several matrix models I assessed, or the single model that Ray assessed (which produced an essentially identical result), would fail to predict the forces and moments correctly, i.e. with acceptable accuracy? And I submit that merely citing large VIF values does not demonstrate this, especially in light of the fact that Ray proposed a moment transfer transformation that yielded low VIF magnitudes, yet resulted in performance essentially identical to that of the same matrix model derived with unshifted moments and large VIF magnitudes.

A1: The VIF effects the prediction variance and the uncertainty in the coefficient estimates. These quantities are not typically used in our community. Norbert has added these to BalFIT and they are available in commercial regression software. I learned about them by taking courses in statistics and modeling. As I mentioned in my original email, the back calculated residuals may not be effected by large VIF and therefore may not be the correct or the only quantity to use for evaluating calibration models. The case where the test or use of the instrument acquires data VERY close to the data applied during calibration, will show smaller impact to poor VIF math models. However, when the application of the math model is to test data that is not VERY close to the calibration data, the effects of the large VIF will begin to result in large uncertainties. One example is to put in the case of full normal force and zero bending moment. While this is extreme, the two models will give very different answers for normal force because the sensitivities for each are quite different. Again the objective is to provide the "best" model for a test and looking at different characteristics is a good thing. (references: Introduction to Linear Regression Analysis, Third Edition, Montgomery, Peck and Vining, Section 3.8 has a good discussion on VIF; Response Surface Methodology Second Edition, Myers and Montgomery, Appendix 2; Wikipedia definition may help, http://en.wikipedia.org/wiki/Variance_inflation_factor;)

Norbert's comments: - I agree 100% with Ray's statements. It is my experience that many people forget that the back-computed residuals only tell the user how "good" the regression model is at points that were used to develop the model in the first place. This is, however, not the way regression models are used in the field. It is more important to know how the regression model is expected to perform at points that were not a part of the original data. The VIF threshold essentially leads to regression models that perform better (have smaller residuals) at these "off-design" points.

How can I support this assertion? - I have attached two pdf files in order to give an example with "real-world" data. The data is from the FMS calibration of the MC60-D balance. The first file contains back-computed residuals for data set ABC09 (cal data obtained in ABCS machine) and check load residuals for data sets MAN15 and MAN19 (manual loads) if a 96 term math model is used for the analysis. The second file contains the same set of residuals if BALFIT's optimized regression model is used for the analysis. As expected, the 96 term math model has smaller "back-computed" residuals for data set ABC09. However, the 96 term math model has large VIFs that cannot be ignored. Therefore, it is expected that the 96 term regression model will not perform as well as BALFIT's optimized regression model at the check load points (MAN 15 and MAN 19) as the optimized math model does not violate the VIF threshold. This is exactly what you will see if you have the two pdf files open side-by-side! Again, to summarize, I believe it is very important to know how well the regression models will perform at data points that are NOT used to determine the regression model. This knowledge is as important as knowing how good the back-computed residuals look.

Robin's comments (3/10/09):

My question was really in reference to the Canard balance results, specifically those for the #1 balance where the models Ray and I had used gave results that were in good agreement, despite the apparently large differences in VIF values for Ray's model compared to mine. This seems to me to negate any claim that a 'correct' result can't be obtained without employing the BMC-shift that Ray proposed, (which reduced the VIF values to roughly 1). I was asking if anyone could show that the model using un-shifted moments (large VIFs) was incorrect?

Ray's response was general in nature and, like Norbert, I completely agree with what he has written. But since Ray advanced the case of NF = maximum combined with BM = 0 as an extreme example of an independent test case that would expose differences between models exhibiting large and small VIFs, I thought to make a test of that. I decided to run some tests for the conditions NF = +/-max = +/-9.983 lb. and BM-a = 0. For the 'shifted BMC' case, MEASURED DATA exist for the condition NFmax/BM-b=0, where BM-b = BM-a - 1.666*NF. However data for the hypothetical case of NFmax/BM-a=0 do not exist, and thus the {R1,R2,R3} bridge response vector must be estimated by extrapolation from the measured data at spanwise loading stations y = 1.966", 1.666", and 1.366". Using these data I made both linear and parabolic extrapolations to obtain {R1,R2,R3} data appropriate to the condition NFmax/BM-a=0. Personally I don't think it's very wise to make a quadratic extrapolation at any time, (and especially not when there are only 3 original points), but as Norbert had used parabolic extrapolation to obtain data at spanwise location y = 0.76" for one of the models he investigated, I decided to include both the linear and quadratic extrapolated results for comparison. The test case consisted of 10 data points as follows:

1) For use with the matrix model appropriate to the 'unshifted' BM-a data and {R1,R2,R3} millivolt vector:

NF	BM-a	HM	R1mV	R2mV	R3mV	
9.983	19.626578	0	4.131524	4.686478	-0.047506	Canard #1 Pt 18 Cen Out (y = 1.966")
9.983	16.631678	0	3.541662	4.086818	-0.035199	Canard #1 Pt 15 Cen Mid (y = 1.666")
9.983	13.636778	0	2.95264	3.489024	-0.023237	Canard #1 Pt 12 Cen In (y = 1.366")
9.983	0	0	0.268574	0.762508	0.0320729	Canard #1 LINEAR EXTRAPOLATION to BMC at y=0"
9.983	0	0	0.281249	0.790654	0.0268682	Canard #1 PARABOLIC EXTRAPOLATION to BMC at y=0"
-9.983	-19.636561	0	-4.128768	-4.685217	0.044978	Canard #1 Pt 8 CEN OUT (y = 1.967")
-9.983	-16.631678	0	-3.54229	-4.088866	0.033503	Canard #1 Pt 5 CEN MID (y = 1.666")
-9.983	-13.656744	0	-2.953898	-3.490614	0.022473	Canard #1 Pt 2 Cen IN (y = 1.368")
-9.983	0	0	-0.272058	-0.763723	-0.028981	Canard #1 LINEAR EXTRAPOLATION to BMC at y=0"
-9.983	0	0	-0.153752	-0.644106	-0.0239399	Canard #1 PARABOLIC EXTRAPOLATION to BMC at y=0"

2) For use with the matrix model appropriate to the 'shifted' BM-b data and {(R2-R1),R2+R1),R3} millivolt vector:

NF	BM-b	HM	(R2-R1)mV	(R2+R1)mV	R3mV	
9.983	2.9949	0	0.554954	8.818002	-0.047506	Canard #1 Pt 18 Cen Out (y = 1.966")
9.983	0	0	0.545156	7.62848	-0.035199	Canard #1 Pt 15 Cen Mid (y = 1.666")
9.983	-2.9949	0	0.536384	6.441664	-0.023237	Canard #1 Pt 12 Cen In (y = 1.366")
9.983	-16.631678	0	0.493934	1.031082	0.0320729	Canard #1 LINEAR EXTRAPOLATION to BMC at y=0"
9.983	-16.631678	0	0.509405	1.071903	0.0268682	Canard #1 PARABOLIC EXTRAPOLATION to BMC at y=0"
-9.983	-3.004883	0	-0.556449	-8.813985	0.044978	Canard #1 Pt 8 CEN OUT (y = 1.967")
-9.983	0	0	-0.546576	-7.631156	0.033503	Canard #1 Pt 5 CEN MID (y = 1.666")
-9.983	2.974934	0	-0.536716	-6.444512	0.022473	Canard #1 Pt 2 Cen IN (y = 1.368")
-9.983	16.631678	0	-0.491665	-1.035781	-0.028981	Canard #1 LINEAR EXTRAPOLATION to BMC at y=0"
-9.983	16.631678	0	-0.490354	-0.797858	-0.0239399	Canard #1 PARABOLIC EXTRAPOLATION to BMC at y=0"

After running the load calculation with both matrix models, I processed the residual error data in Excel to transform the BM-b result to BM-a, and calculate average errors and standard deviations of the residual loads for point groupings as below:

Ray' matrix {NF,BM-b,HM}					Description	# of	Robin's matrix {NF,BM-a,HM}			
	dNF%	dBM-a%	dHM%	dBM-b%		points		dNF%	dBM-a%	dHM%
Average:	0.2547	0.3585	-0.4075	0.9758	ALL points, measured	10	Average:	0.4023	0.2701	0.3569
Std. Devn:	0.9743	0.9646	0.8753	8.5858	and extrapolated		Std. Devn:	1.0318	0.9346	0.7827
Average:	-0.0754	0.0134	-0.0861	0.5087	measured data used in	6	Average:	-0.0122	-0.0222	0.2150
Std. Devn:	0.0568	0.0415	0.4457	0.4579	the extrapolation		Std. Devn:	0.0721	0.0404	0.4259
Average:	0.0068	0.0949	-0.9746	0.5942	NFmax/BM-a=0 LINEAR	2	Average:	0.2714	-0.0722	0.4505
Std. Devn:	0.0094	0.0012	1.3550	0.0442	EXTRAPOLATION		Std. Devn:	0.0092	0.0051	1.3001
Average:	1.4929	1.6574	-0.8049	2.7588	NFmax/BM-a=0 PARABOLIC	2	Average:	1.7766	1.4893	0.6890
Std. Devn:	2.1644	2.0344	1.5771	25.5820	EXTRAPOLATION		Std. Devn:	2.1709	2.0329	1.5993
Average:	-0.1085	-0.0211	-0.0889	0.4623	NFmax/BM-b=0	2	Average:	-0.0454	-0.0566	0.2114
Std. Devn:	0.1010	0.0650	0.4134	0.9948	measured points		Std. Devn:	0.1015	0.0653	0.4134

From these results two things are pretty clear:

1) there is essentially no difference between the results obtained with the {NF, BM-a, HM} and {NF, BM-b, HM} matrix models when the latter is transformed back to BM-a. (Note that both models used exactly the same terms.) The {NF, BM-a, HM} model and load schedule showed large VIF values while those for the {NF, BM-b, HM} model and load schedule were roughly 1.0, but the two models DID NOT "... *give very different answers for normal force because the sensitivities for each are quite different.*", as Ray had predicted?

2) the data obtained with parabolic extrapolation show much larger errors than the linearly extrapolated data.

<u>MC60-D balance</u>:

I also ran some tests using this MC60-D balance since it offered a chance to compare the process I use to evaluate a matrix with Norbert's evaluations using BALFIT. My procedure is to first derive a matrix for the math model having the greatest number of terms that I deem reasonable considering the nature of the loading design. In this case, with the ABCS loading schedule, I started with a 6x96 model, with the intercepts being included in the regression.

Having determined a matrix I perform the load back-calculation to assess the quality of the curve fit. If that is satisfactory I then assess the individual matrix terms in a 'normalized' format in which I'm screening for any suspiciously large term contributions that might indicate bad terms resulting from lack of definition in the data. This is, unfortunately, a very subjective process but that's all I've had to offer – which is why I was excited to learn of Norbert's objective approach. If my assessment reveals any suspicious terms, I re-derive the matrix with those terms eliminated, and repeat the assessment. In the case of the MC60-D balance, my assessment resulted in me excluding the "F.|F|", F^3 , and $|F^3|$ term groups, as well as all 2-component product terms involving absolute values of either component. This reduced the 6x96 model I had started with, (6x97 including intercept), to just 6x33, (6x34 including intercept).

In addition to the above 6x97 and 6x34 matrices, I also derived a 6x96 matrix (FMS, without intercept), plus Norbert's 'optimized' matrix, (6x {19,22,13,20,18,21}). Then when Norbert added information for a 6x39 matrix I thought I would add that too, and had derived what I usually assume for a 6x39 model, (basic 6x27 + |F| + F^3 terms), before realizing the Norbert's 6x39 model was different, (basic 6x27 + |F| + F.|F| terms)! Finally I ended up with matrices for each of 6 math models, and used each of them to make load calculations for each of the files that Norbert had cited, (A09 calibration data, and A15 & A19 MANUAL check loads). I also added file A12; this contains the 6-component 'W/T test' loads that FMS routinely use with all ABCS calibrations, which are probably better 'independent' loads than the manual loads, which often repeat load conditions already contained in the ABCS calibration data.

The standard deviations of the residual load errors that I obtained are shown in plotted form below. The results confirm the expected result that, for the back-calculations of the calibration data, the standard deviations of the residual errors increase as the number of terms included in the math model decreases. However, beyond that the results really do not support Norbert's statement -- "*Therefore, it is expected that the 96 term regression model will not perform as well as BALFIT's optimized regression model at the check load points (MAN 15 and MAN 19) as the optimized math model does not violate the VIF threshold. This is exactly what you will see if you have the two pdf files open side-by-side!* ".

Q2. Given that the forces and moments predicted by Norbert's recommended' matrix model have, for all cases assessed, resulted in significantly larger errors (relative to the known applied loads and moments) than obtained with the other matrix models, what is the rationale for using that model when it demonstrably degrades the achievable accuracy of the measured forces and moments? In the 35 years that I worked in W/T testing, I never once met a customer test engineer who would have knowingly selected an option that sacrificed measurement accuracy when other options existed.

A2: Again, we need to answer a fundamental question, " What is the correct quantity (ies) to use for evaluating the models ?" If we only use the back calculated residuals, then it is not an issue. However, each case needs to be evaluated properly to determine the "best" model. This is discussed in the first answer.

Robin's comments (3/10/09):

The results I gave previously compared residual errors for the Canard #1 balance when using Ray's matrix *{NF, BM-b, HM} / {(R2-R1), (R2+R1), R3)*, my matrix *{NF, BM-a, HM} / {R1, R2, R3)*, and Norbert's 'optimum' matrix *{BM1, BM2, HM} / {R1, R2, R3)*. Those comparisons <u>were not based solely on back-calculations of the calibration data</u>, but also included cases to test the predictive capabilities of the three matrices. Since there were no real independent check load data available, I resorted to simulating 12 check cases equivalent to loading the balance with its maximum NF substantially outside the area covered by the grid that had been used to apply the calibration loads. The *{R1, R2, R3}* bridge response values appropriate to these loads were determined from linear extrapolation of the measured values acquired at the three spanwise stations. For both the basic calibration data and for these simulated check load data, the residual errors were very similar for the math models Ray and I had proposed, while those for Norbert's model were <u>much</u> larger. Previously I distributed the results with some error plots, but I'm including the standard deviations summary again here. Note that "BM-a" is the moment about the original BMC, "BM-b" is the moment about the 'shifted' BMC, and the two are related by the equation "BM-a = BM-b + NF*1.666". Both the "NF / BM-b" computed by Ray's model, and the "BM1 / BM2" computed by Norbert's model, were transformed back to {NF, BM-a} and the error statistics computed using Excel.

Error statistics summary for the 199-point calibration dataset:

	Error Standard Deviation			
	NF	BM-a	BM-b	HM
Reference value for %:	10	20	3	3
Matrix 1 {NF, BM-a, HM} :	0.052	0.037	---	0.262
Matrix 2 {NF, BM-b, HM} :	0.054	0.036	0.437	0.279
Norbert's {NF, BM-a, HM}:	1.254	0.161	---	0.823

```
Error statistics summary for the 12-point 'pseudo' check loads dataset:
                                   Error Standard Deviation
                                  NF     BM-a     BM-b      HM
Reference value for %:            10      20        3        3
Matrix 1 {NF, BM-a, HM} :      3.368   0.597     ---     1.295
Matrix 2 {NF, BM-b, HM} :      3.369   0.632   22.638    1.117
Norbert's {NF, BM-a, HM}:     12.680   1.645     ---     7.648
```

In the case of the results for the back-calculation of the calibration data, the residual errors with Norbert's model were highly linear in nature, a direct consequence of the BM2 on R1 and BM1 on R2 linear interaction terms having been eliminated. With those two terms eliminated the model cannot possibly characterize the balance's behaviour accurately, unless the balance geometry is defined by the 'electrical' locations of the gages. I can't understand why anyone would choose to sacrifice accuracy like this, which is why I asked the question.

Q3. QUESTION: Did you ever compute VIFs for Dennis' original data set for canard balance 1 using DESIGN-EXPERT 7.0? How do your VIFs for Dennis' direct compute regression model compare with the numbers that I list in the enclosed PDF file?

A3: No. I only have one set of data for the balance (the set I analyzed).

Q4. Can the VIF be determined ahead of time?

A4: Yes. The VIF is computed based on the calibration load schedule or design and therefore the design can be evaluated prior to execution to ensure the proposed model can be supported. A very powerful tool. Most of my career has been developing load schedules based on historical practice. However, the statistical tools available make the job easier, more quantitative and therefore more defendable.

Robin's comments (3/10/09):

Yes, when load and moment are the independent variables I know that the VIF is only a function of the calibration load schedule and the regression model terms. But Ray had written that "*Using VIF only to evaluate/reduce a model can be dangerous in certain applications ...*", and my question was whether those **'certain applications'** could be predicted in advance?

Q5. Give an example of this being used previously.

A5: I first applied the moment shifting technique to reduce linear dependencies to a sonic boom model/sting system. The model was too small for a balance. Therefore, we strain-gaged the sting to measure Normal, Pitch, Yaw and Side at a distance away from the models moment center. The moments could be positive or negative as well as the forces. Therefore, the best model was developed from the moment shift to remove the co-linearity.

Norbert's comments: - I am right now in the process of preparing a paper for the AIAA summer conference in Denver that will discuss the application of BALFIT's optimization algorithm to a similar sting balance. In my analysis I did NOT have to shift the moment center because the gage outputs R1+R2 and R2-R1 were used for the analysis. This is also the gage output combination that I would suggest for a canard balance.

Robin's comments (3/10/09):

In such a case, as with the Canard balance, shifting the BMC to roughly the model's moment centre will result in both positive and negative moment values, but it doesn't change the variation in those values as the location of the applied load is varied. Yes, the VIFs are made smaller but the range of variation of the calibrating forces and moments, and the corresponding bridge responses, are identical. In the case of the Canard balance I think it has been demonstrated that the BMC-shifting had no effect on the final results when the moments were compared at the same BMC. Perhaps the balance geometry, in the form of the magnitude of the ratio of the moment transfer distance to the separation of the two moment-sensing bridges, may have a bearing on the ability (or inability) to solve for the matrix without having to shift the BMC? However, for the Canard balance, BMC-shifting was not necessary to obtaining a 'good' solution.

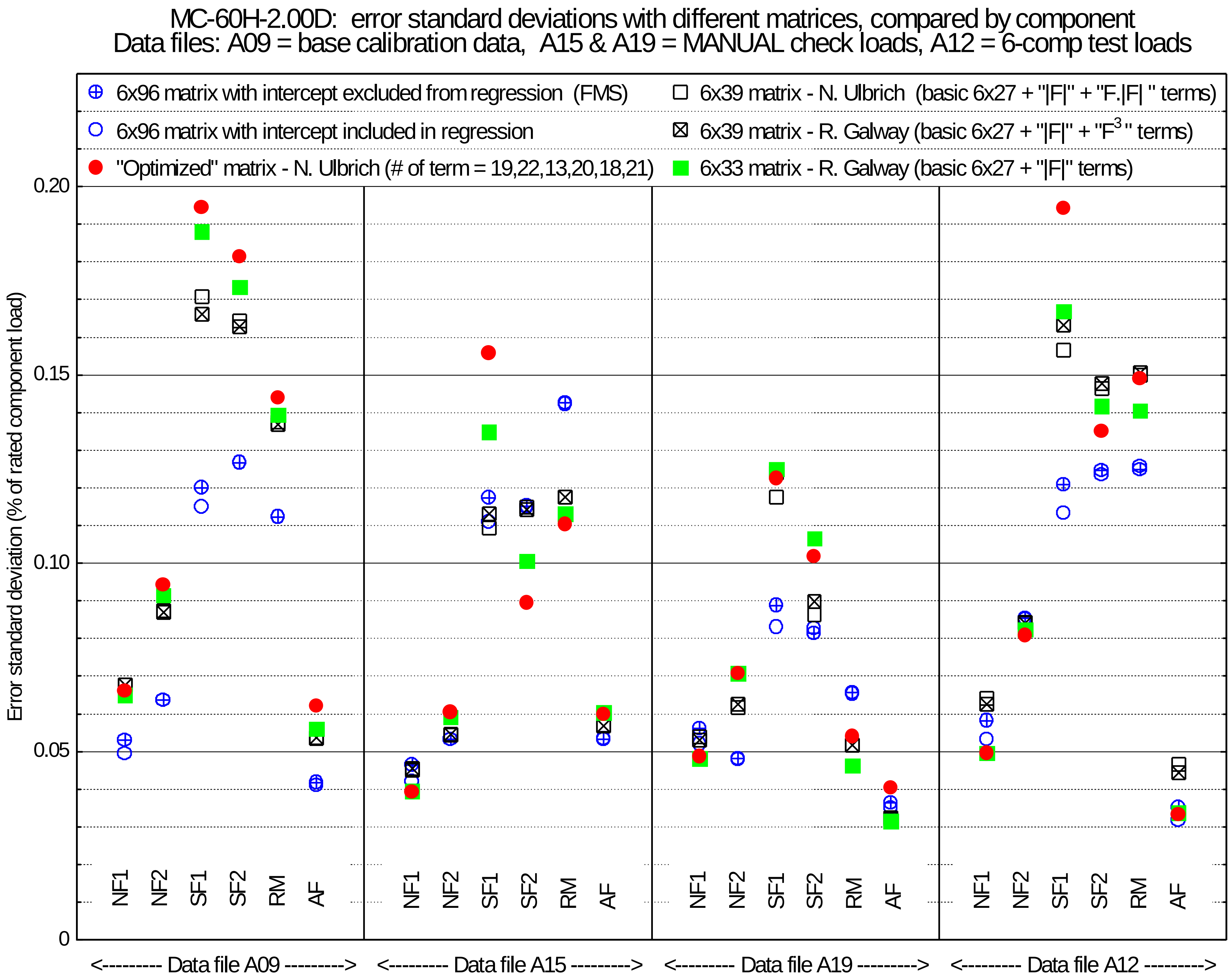
Norbert: With your 'sting balance' you say you did NOT have to shift the BMC because you were using R1+R2 and R2-R1 in your analysis. I presume then that you are using the so-called 'non-iterative' math model in which the responses become the independent variables? Otherwise surely you would see the same large VIFs that you and Ray have confirmed for the Canard balance geometry?

This has become a rather long document but a fair part of it is information that you have all already received. Hopefully it can provide some points for discussion during our telecon tomorrow.

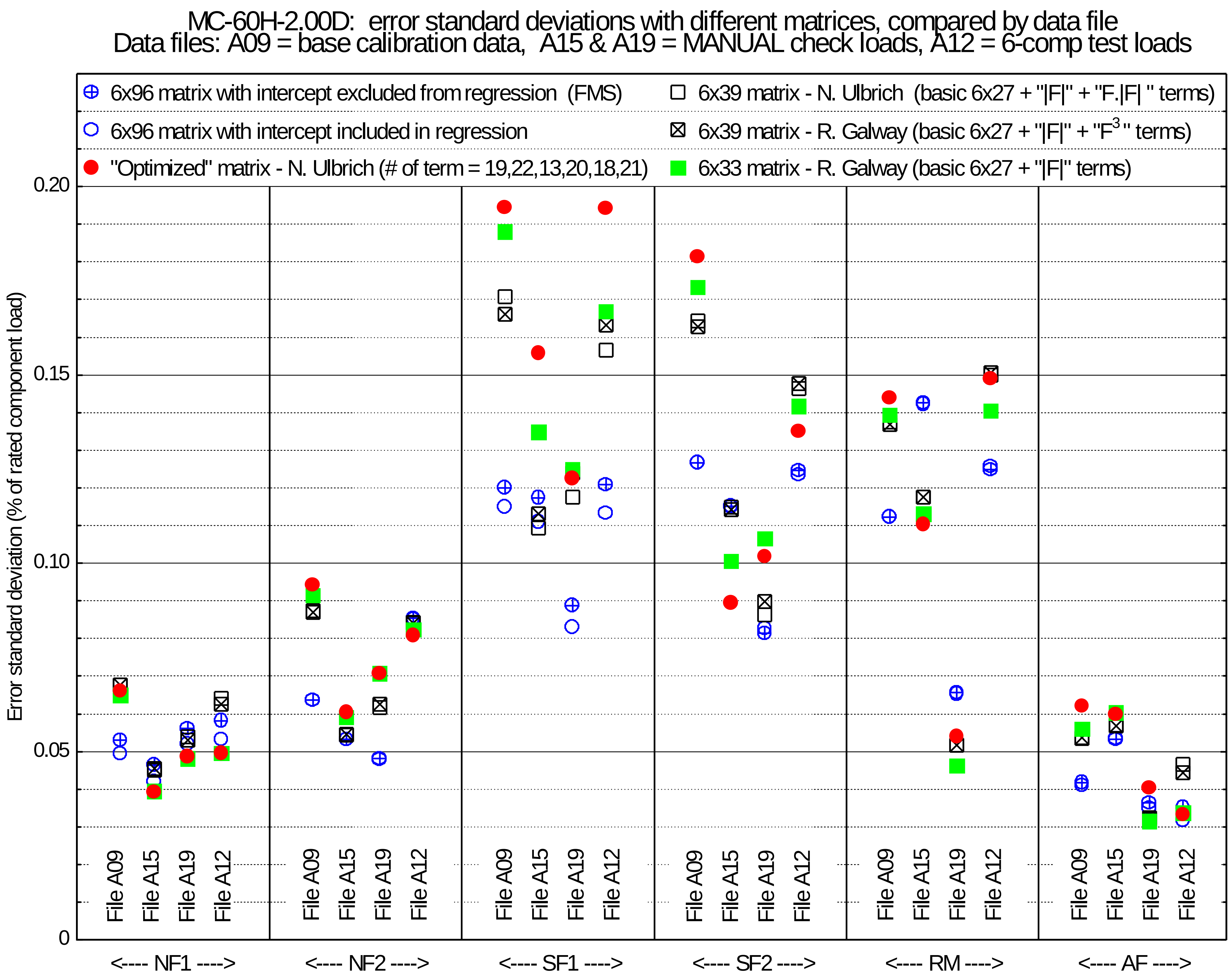
Regards, Robin

10 March, 2009

The plots of the MC60-D balance residual error standard deviations for several math models are shown below.



MC-60H-2.00D: error standard deviations with different matrices, compared by component
Data files: A09 = base calibration data, A15 & A19 = MANUAL check loads, A12 = 6-comp test loads
6x96 matrix with intercept excluded from regression (FMS)
6x96 matrix with intercept included in regression
"Optimized" matrix - N. Ulbrich (# of term = 19,22,13,20,18,21)
6x39 matrix - N. Ulbrich (basic 6x27 + "|F|" + "F.|F| " terms)
6x39 matrix - R. Galway (basic 6x27 + "|F|" + "F3" terms)
6x33 matrix - R. Galway (basic 6x27 + "|F|" terms)
Error standard deviation (% of rated component load)
0.20
0.15
0.10
0.05
0
NF1
NF2
SF1
SF2
RM
AF
<--------- Data file A09 --------->
<--------- Data file A15 --------->
<--------- Data file A19 --------->
<--------- Data file A12 --------->



MC-60H-2.00D: error standard deviations with different matrices, compared by data file
Data files: A09 = base calibration data, A15 & A19 = MANUAL check loads, A12 = 6-comp test loads
6x96 matrix with intercept excluded from regression (FMS)
6x96 matrix with intercept included in regression
"Optimized" matrix - N. Ulbrich (# of term = 19,22,13,20,18,21)
6x39 matrix - N. Ulbrich (basic 6x27 + "|F|" + "F.|F| " terms)
6x39 matrix - R. Galway (basic 6x27 + "|F|" + "F3" terms)
6x33 matrix - R. Galway (basic 6x27 + "|F|" terms)
Error standard deviation (% of rated component load)
0.20
0.15
0.10
0.05
0
File A09
File A15
File A19
File A12
<---- NF1 ---->
<---- NF2 ---->
<---- SF1 ---->
<---- SF2 ---->
<---- RM ---->
<---- AF ---->