STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, March 2024, pp 310–324.
Published online in International Academic Press (www.IAPress.org)

Solving a Typical Small Sample Size MRSM Dataset Problem Using a Flexible Hybrid Ensemble Approach for Credibility

Delson Chikobvu , Domingo Pavolo*

Department of Mathematical Statistics and Actuartial Sciences, University of the Free State, Republic of South Africa

Abstract Multiresponse surface methodology often involves small data analytics which, statistically, have regression modelling credibility problems. This is worsened by dataset, model selection and solution methodology uncertainties.It is difficult for solution methodologies which select and use single best models per response at simultaneous optimisation to effectively deal with these problems.This paper exploited the fact that model selection criteria choose differently, in a flexible hybrid ensemble system, to generate several solutions for integration and comparison.Mean square prediction error, with bias-variance-covariance decomposition values, were computed and analysed at simultaneous optimisation. Results suggest that the credibility of the final solution is enhanced when working with multiple models, solution methodologies and results.However, the results do not show any significance of small sample size correction to model selection criteria and analysis of bias-variance-covariance decompositions at simultaneous optimisation does not encourage dependence on theoretical optimality for best results.

Keywords Credibility of Results, Flexible Hybrid Ensembling, Multiresponse Surface Methodology, Simultaneous Optimisation, Small Sample Size Problems, Solution Uncertainty.

AMS 2010 subject classifications 90C05

DOI: 10.19139/soic-2310-5070-1111

1. Introduction

Multiple response surface methodology (MRSM) involves experimental designing and execution, regression modelling, model selection, response surface analysis, and optimisation [1, 2, 3, 4]. MRSM often involves small data analytics. The credibility of working with small sample size datasets is a big issue in statistics [5, 6]. Practitioners face the problem of model uncertainty in MRSM work and classical model selection (MS) and model averaging (MA) are the common techniques used for attempting to solve this problem [7], though MS is more common. Over the years, many MS criteria have been developed resulting in the problem of MS criterion uncertainty as the criteria often make different choices [8]. Moreover, MS has been criticised for its failure to account for dataset and model uncertainty, model bias and loss of information contained in discarded response models resulting in overly specified prediction accuracy, and understatement of variance [3, 9, 10]. These problems compound to solution uncertainty which leaves a question on the credibility of the final result.

Optimisation in MRSM is multi-objective in nature, and is performed after regression modelling and model selection of single "best" models for each response [3, 4]. Once single "best" models for each response are available, optimisation is performed to determine optimum or desired working process parameters. The existence of so many MRSM solution methodologies that often give different results implies the existence of solution methodology uncertainty which, again, leads to solution uncertainty and lack of solution credibility. A typical

*Correspondence to: Domingo Pavolo (Email: dpavolo68@gmail.com). Department of Mathematical Statistics and Actuarial Sciences, University of the Free State. Bloemfontein, Republic of South Africa.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Table 1. Averaged results for two-factor CCD MRSM conveyor belting experiment

Run	$T_c(min.)$	$R_t(mm)$	Ave. Hardness	Ave. Adhesion
1	16	7.2	60	10.60
2	30	7.2	63	13.34
3	16	22.8	53	6.20
4	30	22.8	61	12.10
5	23	15	58	11.80
6	23	15	58	12.10
7	13	15	44	6.50
8	33	15	63	13.30
9	23	4	63	13.30
10	23	26	56	3.50
11	23	15	58	12.20
12	23	15	57	12.30
13	23	15	58	12.10

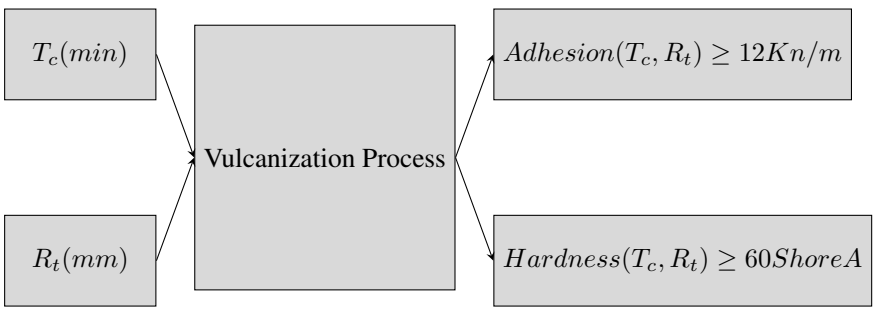

Figure 1. Conveyor belting vulcanisation process and the desired minimum quality requirements.

Table 2. Showing the tabular form of equation 1 required in work instruction

$R_t(mm)$	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$T_c(min.)$	T_7	T_8	T_9	.	.	.	.	.	.	.	.	.	.	T_{20}

small sample size MRSM dataset problem borrowed from [11] of an experiment done to determine rubber covered mining conveyor belting optimum cure times ($T_c(min.)$), for different rubber thicknesses ($R_t(mm)$), is shown in Table 1. The intention was to derive the relationship

$$CureTime(T_c)(min.) = f(R_t(mm)) + error \tag{1}$$

Given the small sample size MRSM experiment dataset in Table 1 and that the component adhesion and cover hardness are parametric functions of cure time and rubber thickness (see Figure 1), the company wanted to submit a work instruction to the shop floor with a table of the form in Table 2. Rawlings et al. in [5] defined as small sample size anything satisfying the relationship $n/k < 40$ and stated that no MS exercise should be taken seriously when the sample sizes are as small as $(n-k) <= 10$, where n is sample size. In this case $n/k = 13/6 < 40$ and $(n-k) = (13-6) = 7 < 10$ [11]. MRSM literature has many such small sample size examples [3]. To guarantee

solution credibility, there was obvious need for rigour in the solution methodology working with such a small sample size MRSM dataset.

The rest of this paper is organised as follows: Section 2 describes the solution methodology, while Section 3 presents the empirical example investigation results, Section 4 is discussion and conclusion.

2. Solution Methodology

Zhang in [12] suggested that for the practitioner who is less concerned with theoretical optimality, it is more important to have available methods that are simple but flexible enough to be used in a variety of practical situations. Burnham and Anderson in [9] proposed multi-model inference (M-MI). An ensemble is defined as a system that is constructed with several individual models/algorithms working in parallel and whose outputs are combined with a suitable decision fusion strategy to produce a single answer for a given problem [14] as shown in Figure 2.

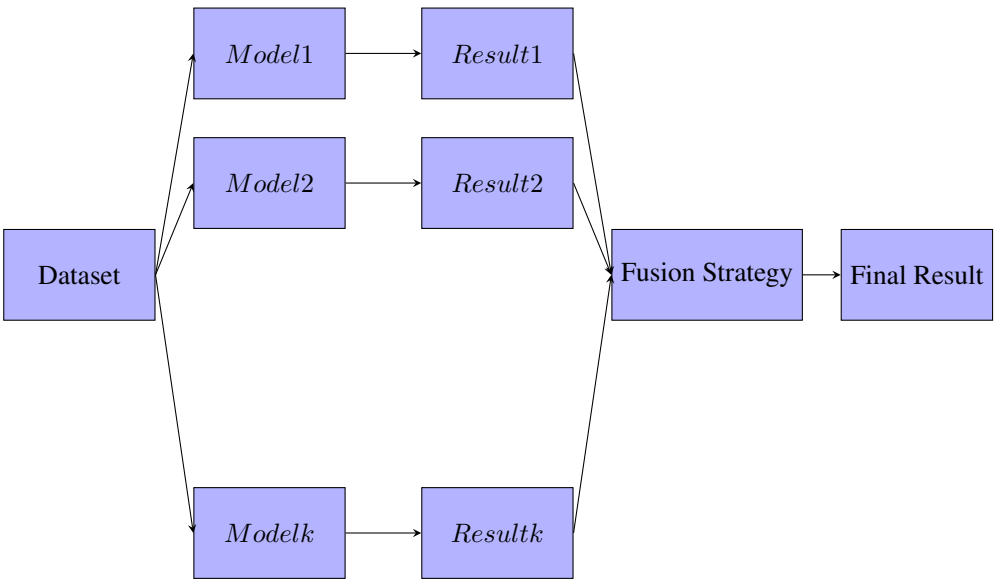

Figure 2. A general ensemble system architecture.

A bias-variance decomposition of a single regression model is shown in equation (2):

$$E[F_i - E(f_i)]^2 = E(F_i) - E(f)^2 + E[F_i - E(F_i)]^2 \tag{2}$$

where F_i is the fitted value and $E(f)$ is the expected value of the function f [13].

This decomposition can be reduced to

$$MSE(f) = (bias(f))^2 + var(f) \tag{3}$$

Krogh and Vedelsby in [15] derived that

$$E_{ens} = E_{ave} - D, \tag{4}$$

where E_{ave} is the average error of the base models, and D is the diversity term estimating the amount of variability among the base models. $E_{ens} \leq E_{ave}$ since D is non-negative. This implies that

$$MSE_{ens} \leq \frac{1}{M}(\sum_{i=1}^{M} MSE_i), \tag{5}$$

where MSE_i is the mean square error of each base model in the ensemble.

Ueda and Nakano in [16] then produced the bias-variance-covariance decomposition of the generalisation of ensemble estimators. In this decomposition it is assumed that:

$$f_{ens} = \frac{1}{k}\sum_{i=1}^{k}[F_i(X)], \tag{6}$$

then

$$E[f_{ens} - E(f)]^2 = Bias^2 + \frac{Var.}{k} + (1 - \frac{1}{k})Covar. \tag{7}$$

$$Bias = \frac{1}{k}\sum_{i=1}^{k}(F_i - E(f)) \tag{8}$$

and

$$Covar. = \frac{1}{(k(k-1))}\sum_{i=1}^{k}\sum_{j=1,i\neq j}^{k} E[F_i - E(F_i)][F_j - E(F_j)]. \tag{9}$$

This shows that if we are able to design an ensemble with both low bias, variance and uncorrelated individual solution methodologies, we can expect improved generalisation performance.

Unlike the MRSM approach which is based on selecting single "best" models for each response before simultaneous optimisation, the ensemble system integrates several accurate results after simultaneous optimisation and is thus less risky, more credible and can be more accurate than any of its base members if they are both accurate and diverse [17, 18, 19, 20, 21, 22]. The ensemble approach is popular for solving complex problems including small sample size dataset problems [23, 24, 25, 26, 27]. The following flexible hybrid ensemble solution methodology was, therefore, employed:

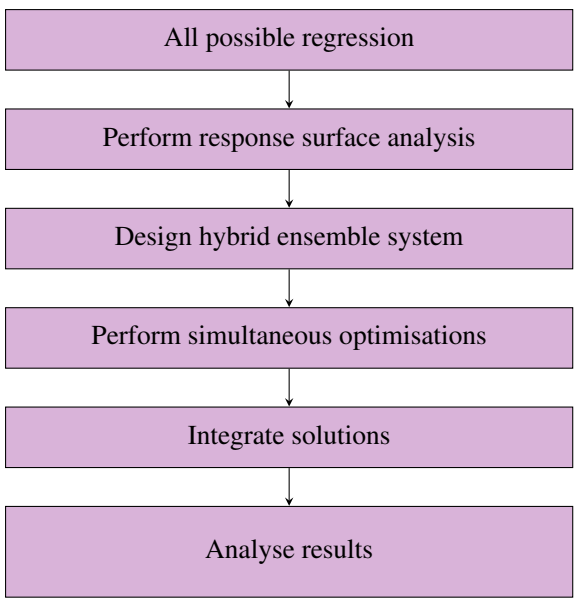

Figure 3. The flexible hybrid ensemble solution methodology.

3. Application of Solution Methodology

3.1. Application of the hybrid ensemble methodology

Hybrid ensembles are heterogeneous mixtures of base models, base methodologies and ensemble systems [28, 29]. Hybrid ensemble systems have been demonstrated to perform better than homogeneous systems [30, 31]. Figure 3 is a hybrid ensemble of five solution methodologies used to obtain five solutions which were then integrated into two before a final credible solution was selected.Figure 4: Presents the diagrammatic flow of the hybrid ensemble.

All possible regression modelling was carried out on the small sample size MRSM dataset of Table 1 for both the hardness and adhesion responses. Response surface conformance analysis (RSCA) was done on the two sets of twenty-five OLS adhesion and hardness response models. The hardness response had only one model with a conforming response surface, hardness response model $[T_c, R_t, T_c.R_t, T_c2]$ [11]. The hybrid ensemble base methodologies, therefore, focused on selecting candidate sets of adhesion response models for simultaneous optimisation with the single hardness response model to generate accurate and diverse results for integration.

Table A1, in Appendix, shows the fifteen MS criteria used to assess the twenty-five adhesion response models. Table A2, in Appendix, shows the adhesion response models selected as best in red by each MS criterion. The fifteen MS criteria were then split into three groups of five which are (i) MS criteria not corrected for small sample size inefficiency (Table 3), (ii) MS criteria corrected for small sample size inefficiency (Table 4), and (iii) prediction MS criteria (Table 5). Tables 3 to 5 show the MS criteria, their values and the models they selected. The Votes column shows the number of times a particular response model was selected, for example model $[T_c.R_t, R_t^2]$ was selected four times by MS criteria AIC, BIC, KIC and TIC. These three groups were used to construct the first three base ensembles in the hybrid ensemble.Table 7 shows the implications of comparisons of results from various base ensembles of Figure 4.

Table 3. Showing adhesion response models selected as best by five small sample size uncorrected MS criteria

Models	AIC	BIC	HQ	KIC	TIC	Votes
$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	11.7	15.1	0.5	20.7	13.7	1
$[T_c, T_c.R_t, T_c^2, R_t^2]$	9.8	12.6	0.5	17.8	11.8	3
$[T_c.R_t, R_t^2]$	9.8	11.5	4.2	15.8	11.8	4

Table 4. Showing adhesion response models selected as best by five small sample size corrected MS criteria

Models	AICc	SBC	HQc	KICc	MKIC	VOTES
$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	20.2	1.9	1.1	87.3	18.2	1
$[T_c, T_c.R_t, T_c^2, R_t^2]$	14.8	1.7	1.1	64.4	12.0	2
$[T_c.R_t, R_t^2]$	11.0	4.8	4.5	38.1	5.4	3

The fourth base ensemble methodology utilised the first five "best" AIC_c adhesion models as base models for simultaneous optimisation as proposed by Burnham and Anderson in [9] in their multi-model inference system (M-MI). Table 6 shows the five adhesion response models and their AIC_c values and the computed smoothed AIC_c $(S - AIC_c)$ weights.The formula for $S - AIC_c$ is,

$$\frac{e^{((-AICc_i)/2)}}{\sum_{i=1}^{k} e^{((-AICc_i)/2)}}, \tag{10}$$

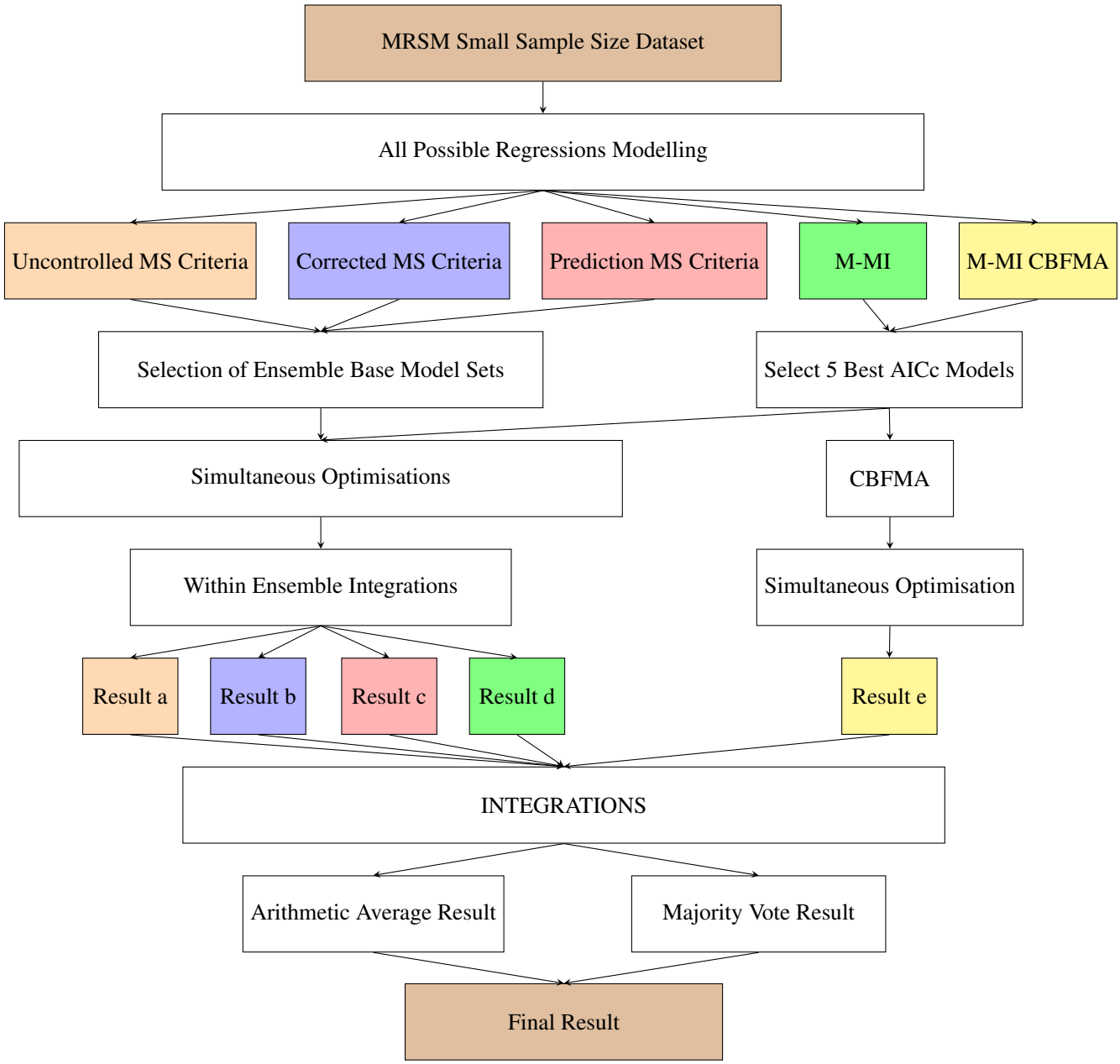

Figure 4. The diagrammatic flow of the hybrid ensemble.

where one calculates the weighted average $B = \sum_{i=1}^{k} u_i B_i$ from the i estimators of B, a regression parameter in the response model, and $\sum_{i=1}^{k} u_i = 1$.

The fifth and last methodology combined the adhesion response models of the M-MI ensemble system using their S-AICc weights, as shown in Table 7, to come up with an averaged model estimator for simultaneous

Table 5. Showing adhesion response models selected as best by Prediction MS criteria.

Model	$R^2_{pred.}$	Adeq. Prec.	$C_p - k$	PRESS	APC_p	Votes
$[T_c, T_c.R_t, T_c^2, R_t^2]$	51.5	4.1	0.1	59.3	2.4	1
$[R_t, T_c.R_t, R_t^2]$	49.9	5.7	0.0	81.3	2.6	1
$[T_c.R_t, R_t^2]$	65.6	1.8	0.0	42.1	2.9	3
$[T_c, R_t, R_t^2]$	49.4	11.9	2.1	62.2	2.9	1
$[R_t, T_c.R_t]$	52.0	5.4	0.0	58.6	4.0	1

Table 6. Showing adhesion response models selected as best by the AICc MS criterion.

Model	AIC_c Value	$S - AIC_c$ Weight
$[T_c.R_t, R_t^2]$	11.0	0.566078
$[T_c, T_c.R_t, R_t^2]$	13.9	0.132785
$[R_t, T_c.R_t, R_t^2]$	14.3	0.108715
$[T_c.R_t, T_c^2, R_t^2]$	14.4	0.103413
$[T_c, R_t^2]$	14.7	0.089009

Table 7. Showing the computation of the frequentist model averaged estimator (M-MI CBFMA).

Model	AICc Wt.	Intercept	T_c	R_t	$T_c.R_t$	T_c^2	R_t^2
$[T_c.R_t, R_t^2]$	0.566	10.497	0	0	0.020	0	-0.026
$[T_c.R_t, T_c^2, R_t^2]$	0.133	10.390	0	0	0.019	0.001	-0.025
$[T_c, T_c.R_t, R_t^2]$	0.109	9.140	0.091	0	0.016	0	-0.022
$[T_c, R_t^2]$	0.103	6.180	0.324	0	0	0	-0.011
$[R_t, T_c.R_t, R_t^2]$	0.089	11.080	0	-0.098	0.021	0	-0.023
M-MI	CBFMA	9.941	0.043	-0.009	0.018	0.000	-0.024

optimisation with the hardness response model. The difference between the M-MI ensemble and the M-MI CBFMA methodology was that, whilst M-MI combined results after simultaneous optimisation, M-MI CBFMA combined the five models using MA before simultaneous optimisation. After simultaneous optimisation of the adhesion response base models with the single hardness response model with a conforming response surface, results from the individual solution methodologies were compared and integrated using arithmetic averaging (Ave.) and majority vote (M. Vote). Majority vote meant selecting the cure time with the highest frequency of occurrence for a given rubber thickness. Equation (5) of Krogh and Vedelsby in [15] suggested that the mean square prediction error (MSPE) of the hybrid ensemble would be better than or equal to the average of the base methodologies. Table 8 shows the expected implications of analysing results from different base ensembles.

3.2. Bias-variance-covariance decomposition

Computations of the bias-variance decomposition of base models and bias-variance-covariance decomposition of base ensemble estimators at simultaneous optimisation were done using equations (2) to (7). The formula for MSPE for a sample size n with Y_i the ith estimated response value at simultaneous optimisation and Y_T the targeted

Table 8. Showing the implications of comparisons of results from different base ensembles.

No.	COMPARISONS/ANALYSIS	IMPLICATIONS
1	Sample size inefficiency uncorrected vs. corrected MS criteria	Effect of inefficiency correction
2	Uncorrected + corrected criteria vs. Prediction MS criteria	Error of optimism
3	M-MI ensemble vs. M-MI CBFMA ensemble	Effect of model averaging
4	All eight results	Solution uncertainty/credibility

response value is given as:

$$MSPE = \frac{1}{n}(\sum_{i=1}^{n}(Y_i - Y_T)^2) = \frac{1}{n}(\sum_{i=1}^{n} e\_i^2 \tag{11}$$

and

$$Bias = \frac{1}{n}(\sum_{i=1}^{n} Y_i) - Y_T \tag{12}$$

Table 9 shows the simultaneous optimisation and determination of the various components of the bias-variance decomposition of the adhesion-hardness response model pair $[T_c.R_t, R_t^2] - [T_c, R_t, T_c.R_t, T_c^2]$ using Excel. The $T_c(min.)$ column is the estimated cure time for each rubber thickness $R_t(mm)$. The value e^2 is the square of the residual, the difference between the targeted value and the fitted value. Var. is the variance (ref. equation (3)).

Table 9. The simultaneous optimisation of adhesion-hardness response model pair $[T_c.R_t, R_t^2] - [T_c, R_t, T_c.R_t, T_c^2]$ using Excel.

T_c (min)	R_t (mm)	Adhesion Y_T	$T_c.R_t, R_t^2$ Estimate	. e^2	Hardness Y_T	$T_c, R_t, T_c.R_t, T_c^2$ Estimate	. e^2
21	7	12	12.217	0.047	60	60.13	0.017
22	8	12	12.419	0.175	60	60.36	0.131
22	9	12	12.427	0.182	60	60.15	0.022
23	10	12	12.586	0.343	60	60.35	0.125
23	11	12	12.511	0.261	60	60.17	0.030
24	12	12	12..628	0.395	60	60.35	0.125
24	13	12	12.470	0.221	60	60.21	0.042
24	14	12	12.261	0.068	60	60.06	0.003
25	15	12	12.305	0.093	60	60.24	0.059
25	16	12	12.012	0.000	60	60.13	0.016
26	17	12	12.013	0.000	60	60.29	0.082
27	18	12	12.004	0.000	60	60.39	0.150
29	19	12	12.369	0.136	60	60.40	0.163
30	20	12	12.357	0.127	60	60.30	0.090
.	.	.	12.327	MSPE:0.146	.	60.25	MSPE:0.076
.	.	.	Bias: 0.327	Var.:0.039	.	Bias:0.250	Var.:0.012

Following adhesion-hardness model pair $[T_c.R_t, Rt^2] - [T_c, R_t, T_c.R_t, T_c^2]$, simultaneous optimisations and computations of bias-variance decomposition for the rest of the base models of all base methodologies were done and contributed to the bias-variance-covariance decomposition of the ensembles in Tables 10, 11, 12 and 13, respectively.

Table 10. Bias-variance-covariance decomposition of the Uncorrected MS ensemble.

Model	MSPE	Bias	Var.	Covar.	MSPE	Bias	Var.	Covar.
$[T_c.R_t, R_t^2]$	0.146	0.327	0.039	.	0.076	0.250	0.012	.
$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	0.516	0.616	0.137	.	0.065	0.220	0.018	.
$[T_c, T_c.R_t, T_c^2, R_t^2]$	0.495	0.619	0.113	.	0.076	0.250	0.012	.
AVE.	0.323	0.472	0.079	0.112	0.074	0.246	0.013	0.014

Table 11. Bias-variance-covariance decomposition of the Corrected MS ensemble.

Model	MSPE	Bias	Var.	Covar.	MSPE	Bias	Var.	Covar.
$[T_c.R_t, R_t^2]$	0.1464	0.3270	0.0394	.	0.0755	0.2500	0.0117	.
$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	0.5161	0.6155	0.1373	.	0.0651	0.2200	0.0177	.
$[T_c, T_c.R_t, T_c^2, R_t^2]$	0.4951	0.6186	0.1125	.	0.0755	0.2500	0.0117	.
AVE.	0.3234	0.4724	0.0791	0.1108	0.0742	0.2463	0.0125	0.0141

Table 12. Bias-variance-covariance decomposition of the the Prediction MS ensemble.

Model	MSPE	Bias	Var.	Covar.	MSPE	Bias	Var.	Covar.
$[T_c.R_t, R_t^2]$	0.146	0.327	0.039	.	0.076	0.250	0.012	.
$[R_t, T_c.R_t, R_t^2]$	0.151	0.370	0.014	.	0.096	0.285	0.014	.
$[T_c, T_c.R_t, T_c^2, R_t^2]$	0.495	0.617	0.113	.	0.076	0.250	0.012	.
$[T_c, R_t, R_t^2]$	0.112	0.288	0.029	.	0.105	0.289	0.020	.
$[R_t, T_c.R_t]$	0.142	0.285	0.061	.	0.184	0.402	0.023	.
AVE.	0.191	0.363	0.048	0.062	0.098	0.282	0.015	0.019

Table 13. Bias-variance-covariance decomposition of the M-MI ensemble.

Model	MSPE	Bias	Var.	Covar.	MSPE	Bias	Var.	Covar.
$[T_c.R_t, R_t^2]$	0.146	0.327	0.039	.	0.076	0.250	0.012	.
$[T_c, T_c.R_t, R_t^2]$	0.142	0.354	0.016	.	0.096	0.285	0.014	.
$[R_t, T_c.R_t, R_t^2]$	0.151	0.370	0.014	.	0.096	0.285	0.014	.
$[T_c.R_t, T_c^2, R_t^2]$	0.105	0.301	0.015	.	0.096	0.285	0.014	.
$[T_c, R_t^2]$	0.103	0.237	0.047	.	0.097	0.279	0.019	.
AVE.	0.138	0.325	0.032	0.033	0.084	0.265	0.013	0.015

Tables 10, 11, 12 and 13 show that a response model with the best MSPE and bias on the adhesion side normally has the worst MSPE and bias at the hardness response model side at simultaneous optimisation, and the reverse is true. This suggests an MSPE (adhesion) – MSPE (hardness) trade-off that could be reminiscent of the popular bias-variance trade-off in statistics and machine learning. Khuri and Conlon in [32] called this simultaneous optimisation compromise.

3.3. Integration of ensembles

Tables 14, 15, 16 and 17 show the integrations of individual base model cure time estimates for each of the four ensemble systems with bias-variance-covariance computations shown in Tables 10, 11, 12 and 13, respectively. The weights used to compute the weighted averages (W. AVE) in Tables 14, 15 and 16 were from the VOTES column of Tables 3, 4 and 5. Table 16 used the S-AICc weights of Table 6. The weights of response models in Tables 14 - 16 were computed from the votes shown in Tables 3, 4 and 5. The yellow shading in Tables 14 -18 shows where the cure time estimates of each base model agreed with the weighted average (W. AVE) values in the last column. The formula for relative accuracy of base models and methodologies relative to the weighted average or hybrid ensemble results is shown in equation (15).

$$RelativeAccuracy = \frac{NumberofBaseModelsorEnsembleCorrectPredictions}{TotalNumberofInstances} \tag{13}$$

Table 14. Results integration for Uncorrected MS criteria ensemble cure time estimates.

Weight	0.125	0.500	0.375	Wtd.
$R_t, (mm)$	$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	$[T_c.R_t, R_t^2]$	$[T_c, T_c.R_t, T_c^2, R_t^2]$	Ave.
7	21	21	21	21
8	22	22	22	22
9	22	22	22	22
10	23	23	23	23
11	23	23	23	23
12	24	24	24	24
13	24	24	24	24
14	24	24	24	24
15	25	25	25	25
16	25	25	26	25
17	25	26	27	26
18	27	27	28	27
19	28	29	29	29
20	31	30	30	30
Acc. Count	11	14	11	.
Rel. Acc.	0.79	1.00	0.79	.

The adhesion base response models selected in both the small sample size inefficiency uncorrected and corrected criteria ensemble systems were the same. The only difference in Tables 14 and 15 is in the weights as shown in the VOTES of Tables 3 and 4.

3.4. Integration of the results of the five solutions methodologies

Both the Ave. and the M. Vote fusion methods had the same cure time estimates. The three base methodologies of Prediction MS, M-MI and M-MI CBFMA agreed on all the cure time estimates. The uncorrected and the corrected MS information criteria ensembles had the same results, though different from the other three on three rubber thicknesses.

The bias-variance-covariance decomposition of the hybrid ensemble with the five different solution methodologies is shown in Table 18.

The small sample size uncorrected and corrected MS information criteria ensembles comparatively had the worst MSPE and bias values on the adhesion response side, but the best on the hardness side. M-MI and M-MI CBFMA,

Table 15. Results integration for Corrected MS ensemble.

Weight	0.1667	0.5000	0.3333	Wtd.
$R_t, (mm)$	$[T_c, R_t, T_c.R_t, T_c^2, R_t^2]$	$[T_c.R_t, R_t^2]$	$[T_c, T_c.R_t, T_c^2, R_t^2]$	Ave.
7	21	21	21	21
8	22	22	22	22
9	22	22	22	22
10	23	23	23	23
11	23	23	23	23
12	24	24	24	24
13	24	24	24	24
14	24	24	24	24
15	25	25	25	25
16	25	25	26	25
17	25	26	27	26
18	27	27	28	27
19	28	29	29	29
20	31	30	30	30
Acc. Count	11	14	11	.
Rel. Acc.	0.79	1.00	0.79	

Table 16. Results integration for Prediction MS criteria ensemble.

Weight	0.4286	0.1429	0.1429	0.1429	0.1429	Wtd.
$R_t, (mm)$	$[T_c.R_t, R_t^2]$	$[T_c, T_c.R_t, T_c^2, R_t^2]$	$[T_c, R_t, R_t^2]$	$[R_t, T_c.R_t, R_t^2]$	$[R_t, T_c.R_t]$	Ave.
7	21	21	21	21	21	21
8	22	22	22	22	22	22
9	22	22	22	22	22	22
10	23	23	23	23	23	23
11	23	23	23	23	24	23
12	24	24	24	24	25	24
13	24	24	24	24	26	24
14	24	24	24	24	27	24
15	25	25	25	25	28	25
16	25	25	26	26	28	26
17	25	26	27	27	29	27
18	27	27	28	28	29	28
19	28	29	29	29	30	29
20	31	30	30	30	31	30
Acc. Count	11	14	11	14	4	.
Rel. Acc.	0.79	1.00	0.79	1.00	0.29	

on the other hand, had the best MSPE and bias values on the adhesion side. The Prediction MS criteria ensemble had the worst MSPE and bias values on the hardness side but was middle-of-the-road on the adhesion side. It was therefore, difficult to select a “best” methodology in this case.

Table 17. Results integration for M-MI ensemble.

Weight	0.566608	0.13279	0.10872	0.10341	0.08901	Wtd.
$R_t, (mm)$	$[T_c.R_t, R_t^2]$	$[T_c, T_c.R_t, R_t^2]$	$[R_t, T_c.R_t, R_t^2]$	$[T_c.R_t, T_c^2, R_t^2]$	$[T_c, R_t^2]$	Ave.
7	21	21	21	21	21	21
8	22	22	22	22	22	22
9	22	22	22	22	22	22
10	23	23	23	23	23	23
11	23	23	23	23	23	23
12	24	24	24	24	24	24
13	24	24	24	24	24	24
14	24	24	24	24	25	24
15	25	25	25	25	26	25
16	25	26	26	26	27	26
17	26	27	27	27	28	27
18	27	28	28	28	30	28
19	29	29	29	29	30.5	29
20	30	30	30	30	30.5	30
Acc. Count	11	14	14	14	7	
Rel. Acc.	0.79	1.00	1.00	1.00	0.50	

Table 18. Showing the cure time estimates of the various methods and their integration.

$R_t, (mm)$	Uncor. MS	Cor. MS	Pred. MS	M-MI	M-MI CBFMA	Average	M. Vote
7	21	21	21	21	21	21	21
8	22	22	22	22	22	22	22
9	22	22	22	22	22	22	22
10	23	23	23	23	23	23	23
11	23	23	23	23	23	23	23
12	24	24	24	24	24	24	24
13	24	24	24	24	24	24	24
14	24	24	24	24	24	24	24
15	25	25	25	25	25	25	25
16	25	25	26	26	26	26	26
17	26	26	27	27	27	27	27
18	27	27	28	28	28	28	28
19	28	29	29	29	29	29	29
20	30	30	30	30	30	30	30
Acc. Count	11	11	14	14	14	.	.
Rel. Acc.	0.79	0.79	1.00	1.00	1.00	.	.

4. Discussion and Conclusion

The discussion of the results is structured around the comparisons/analysis and implications presented in Table 8.

Table 19. Comparing bias-variance-covariance decomposition results of all methods.

.	.	Adhesion	.	.	.	Hardness	.	.
Model	MSPE	Bias	Var.	Covar.	MSPE	Bias	Var.	Covar.
Uncorrected MS	0.323	0.472	0.079	0.111	0.074	0.246	0.013	0.014
Corrected MS	0.323	0.472	0.079	0.111	0.074	0.246	0.013	0.014
Prediction MS	0.236	0.371	0.043	0.112	0.113	0.304	0.018	0.021
M-MI	0.148	0.344	0.024	0.130	0.093	0.278	0.016	0.016
M-MI CBFMA	0.134	0.343	0.017	.	0.096	0.290	0.014	.
AVE.	0.233	0.400	0.049	0.079	0.090	0.273	0.015	0.018

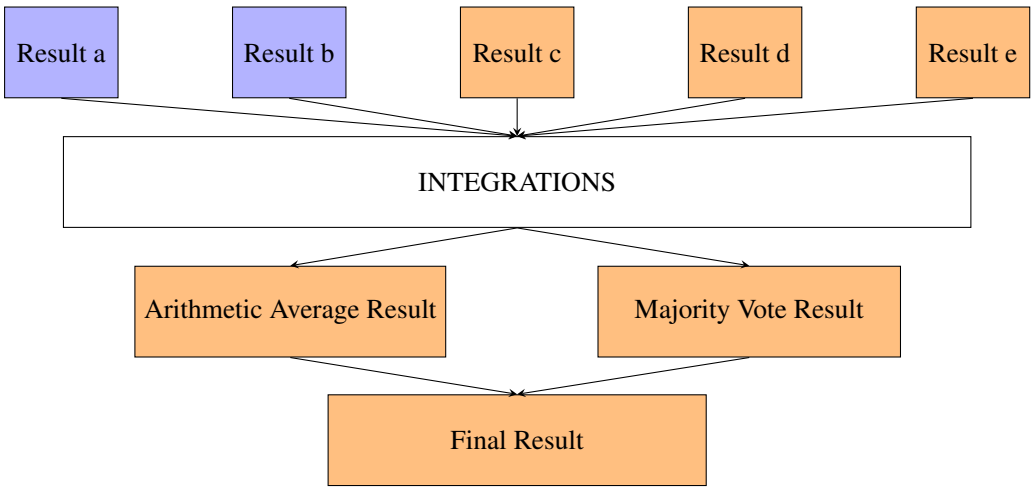

Figure 5. Comparison of Results

4.1. Effect of small sample size inefficiency correction

The adhesion models selected by the two sets of five small sample size uncorrected and corrected MS information criteria were the same. The cure time estimates of the small sample size uncorrected and corrected MS information criteria ensembles were thus the same. The differences on the MSPE and biases of the two ensembles were insignificant. This suggests that the effect of small sample size correction on model selection criteria may be insignificant at multiple model selection criteria level.

4.2. Error of optimism

The cure time estimates of the uncorrected and corrected MS criteria ensembles are different on three rubber thicknesses from, and the theoretical accuracy values of the ensembles are worse than those of, the prediction MS criteria and M-MI ensembles. This demonstrates the error of optimism. Response models that have very good fit to the dataset are not necessarily good at generalisation.

4.3. Effect of CBFMA

The M-MI and M-MI CBFMA ensembles had the same cure time estimates but different theoretical accuracies. The M-MI CBFMA had a better theoretical accuracy than the M-MI. This agrees with literature.

4.4. Solution uncertainty/credibility

Six out of eight results (0.75) were in agreement as shown in Figure 5. Based on the principle that the greater the uncertainty, the lower the credibility, this result had the least uncertainty and hence the highest credibility. In this particular MRSM example, the hybrid ensemble did not produce a final result that was more accurate than all the base methodologies, but one more credible than individual base methodologies. The fact that different answers were obtained by different solution methodologies exposed the risk of depending on a result from a single methodology. This suggests that a multiple solution methodology approach exposes the problem of solution methodology uncertainty which is not exposed by the use of a single solution methodology and supports ensembling of results from different methodologies. Multiple methodologies allow for comparison and justifications on selection of final result. The accuracy values in the bias-variance-covariance decomposition at simultaneous optimisation do not give a clear indication of the "best" methodology or "best" response model pair.

Future research areas would look at:

• A simple software that can easily generate solutions from different methodologies for integration using multiple fusion methodologies and the most logical solution selected.

• A further study on simultaneous optimisation compromise and it's relationship to prediction accuracy

REFERENCES

1. A. I. Khuri, *The Analysis of Multiresponse Experiments: A Review, Tecnical Report 302*, Department of Statistics, University of Florida, Gainesville, FL32611.Khuri, 1988.
2. S. Mukhopadhyay, and A.I. Khuri, *Response Surface Methodology*, Wires Computational Statistics, 2, March/April, 128-149, 2010.
3. R.H. Myers, D.C. Montgomery, and C.M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimisation Using Designed Experiments*, 4th Edition, ISBN: 978-1-118-91601-8, 2016.
4. A.I. Khuri, *A General Overview of Response Siurface Methodology*, Biometrics Biostatistics International Journal, 5(3): 00133. DOI: 10.15406/bbij.2017.05.00133, 2017.
5. J.O. Rawlings, S.G. Pantula, and A.D. Dickey, *Applied Regression Analysis: A Research Tool*, Second Edition, Springer-Verlag New York Inc., 1998.
6. D.G. Jenkins, and P.F. Quintana-Ascencio, *A Solution to Minimum Sample Size for Regressions*, PLoS ONE 15(2): e0229345. https:// doi.org/10.1371/journal.pone.0229345, 2020..
7. T. Xie, *Prediction Model Averaging Estimator*, Economics Letters 131, 5–8, 2015.
8. M. Schomaker, and C. Heumann, *When and when not to use optimal model averaging*, Statistical Papers, in press. DOI: 10.1007/s00362-018-1048-3, 2018
9. K. Burnham, and D. Anderson, *Model selection and multi-model inference*, A practical information-theoretic approach, Springer, New York, 2002.
10. P. Mitra, H. Lian, R. Mitra, H. Liang, and M. Xie, *A general framewor for frequentist model averaging*, Science China Mathematics, 62(2), 205-226, 2019.
11. D. Pavolo, and D. Chikobvu, *Optimising Vulcanisation Time of Rubber Covered Mining Conveyor Belts Using Multiresponse Surface Methodology*, International Journal of Operations and Quantitative Management (IJOQM), Vol. 26, Issue # 1, pages 29-48, 2020.
12. P. Zhang, *On the choice of penalty term in generalised FPE criterion*, In Selecting Models from Data (P. Cheeseman and R. W. Oldford, eds.), 41-49, Springer-Verlag, New York, 1994.
13. W. Wang, *Some Fundamental Issues in Ensemble Methods*, International Joint Conference on Neural Networks (IJCNN 2008) pp: 2243 – 2250, 2008.
14. S. Geman, E. Bienenstock, and R. Doursat , *Neural networks and the bias/variance dilemma*, Neural Computation, vol. 4, no. 1, pp. 1–58, 1992.
15. A. Krogh,and J. Vedelsby, *Neural Network Ensembles, Cross Validation and Active Learning*, NIPS, 7, 231–238, 1995.
16. N. Ueda, and R. Nakano, *Generalisation error of ensemble estimators*, IEEE International Conference on Neural Networks, 1996, vol. 1, pp. 90–95, 1996.
17. L. K. Hansen, and P. Salamon, *Neural networ ensembles*, IEEE transactions on pattern analysis and machine intelligence, 12(10):993–1001, 1990.
18. D. Opitz, and R. Maclin, *Popular Ensemble Methods: An Empirical Study*, Journal of Artificial Intelligence Research, 11:169–198, 1999.
19. R. Polikar, *Ensemble based systems in decision making*, IEEE Circuits Syst. Mag 6(3):21–45, 2006.
20. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, Springer Series in Statistics, USA, 2nd Edition, 2016.
21. C. Ju, A. Bibaut, and M. Van de Laan, *The Relative Performance of Ensemble Methods with Deep Convulational Neural Networks for Image Classification*, J. Appl. Stat. Vol. 45 (15), pp: 2800-2818, 2018.
22. Y. Brain, H. Chen, *When Does Diversity Help Generalisation in Classification Ensembles?*, IEEE Transactions in Cybernetics, arXiv:1910.13631v1[cs. LG], 30 Oct. 2019.
23. T.G. Dietterich, *Ensemble methods in machine learning*, In Multiple classifier systems, pages 1–15. Springer, 2000.

24. R. Polikar, A. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios, and C. Clark, *An ensemble based data fusion approach for early diagnosis of alzheimer disease*, Information Fusion, 9:83–95, 2008.
25. P. Yang, Y.H. Yang, B.B. Zhou, and A.Y. Zomaya,, *Current Bioinformatics, A review of ensemble methods in bioinformatics: Including stabilioty of feature selection and ensemble feature selection*, Vol. 5, (4):296-308, 2010 (updated on 28 Sep. 2016).
26. A. Ahangi, F.A. Langrondi, F. Yazdanpanah, and S. Mirroshandel, *A Novel Fusion Mixture of Active Experts Algorithm for Traffic Signs Recognition Multimedia Tools and Applications*, DOI: 10.1007/s11042-019-0, 2019..
27. S. Zhou, R. Hu, Y. Liu, and Z. Tang, *Margin Based Pareto Ensemble Pruning: An Ensemble Pruning Algorithm That Learns to Search Optimised Ensembles*, Computational Intelligence and Neuroscience. Hindawi, Vol. 2019, Article ID 7560872, 12 pages, 2019.
28. Z-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, CSC Press, Tylor and Francis, USA. ISBN 13:978-1-4398-3005-5. (eBook- PDF), 2012.
29. P. Kazienko, E. Lughofer, and B. Trawonsk, *Editorial on the Special Issue "Hybrid and Ensemble Techniques in Soft Computing: Recent Advances and Emerging Trends*, Soft Computing (2015) 19: 3353 – 3355. DOI 10.1007/s00500-015-1916-x, 2015.
30. Y. Chali, H. Sadid, and M. Mojahid, *Complex question answering:homogeneous or heterogeneous, which ensemble is better?*, 19th International Conference on Application of Natural Language to Information Systems (NLDB 2014), Jun 2014, Montpellier, France. pp. 160-163. ffhal-01402563f.
31. A. Petrokova, M. Affenzeller, and G. Merkurjeva, *Heterogeneous versus Homogeneous Machine Learning Ensembles*, Information Technology and Management Science, 18(1) pp. 135-140. Doi: 10.1515/itms-2015-0021, 2015.
32. A.I. Khuri, and M. Conlon, *Simultaneous Optimisation of Multiple Responses Represented b Polynomial Regression Functions*, Technometrics, 23, 363-375, 1981. http://dx.doi.org/10.2307/1268226.