Response Modeling Methodology (RMM)

A New Approach to Empirical Modeling of Variation

Haim Shore, Ben-Gurion University, Israel

Auburn University, Industrial and Systems Engineering Deprt., March 2006

References

BOOK:

[1] Shore, H. (2004). Response Modeling Methodology (RMM)- Empirical Modeling for Engineering and the Sciences. World Scientific Publishing Company. Singapore.

Published Papers:

[1] Shore, H. (2000a). General control charts for variables. *International Journal of Production Research*, 38(8), 1875-1897.

[2] Shore, H. (2001a). Modeling a non-normal response for quality improvement. *International Journal of Production Research*, 39 (17), 4049-4063.

[3] Shore, H. (2001b). Inverse normalizing transformations and a normalizing transformation. In *"Advances in Methodological and Applied Aspects of Probability and Statistics"***,** Book, N. Balakrishnan (Ed.), Gordon and Breach Science Publishers. Canada. V. 2, 131-146,

[4] Shore, H. (2002a). Modeling a response with self-generated and externally-generated sources of variation. *Quality Engineering*, 14(4), 563-578.

[5] Shore, H. (2002b). Response Modeling Methodology (RMM)- Exploring the implied error distribution. *Communications in Statistics (Theory and Methods)*, 31(12), 2225-2249.

[6] Shore, H., Brauner, N., and Shacham, M. (2002). Modeling physical and thermodynamic properties via inverse normalizing transformations. *Industrial and Engineering Chemistry Research*, 41, 651-656.

[7] Shore, H. (2003). Response Modeling Methodology (RMM)- A new approach to model a chemo-response for a monotone convex/concave relationship. *Computers and Chemical Engineering*, 27(5), 715-726.

[8] Shore, H. (2004a). Response Modeling Methodology (RMM)- Validating evidence from engineering and the sciences. *Quality and Reliability Engineering International*, 20, 61-79.

[9] Shore, H. (2004c) Response Modeling Methodology (RMM)- Current distributions, transformations, and approximations as special cases of the RMM error distribution. *Communications in Statistics (Theory and Methods)*, 33(7), 1491-1510.

[10] Shore, H. (2004d). Response Modeling Methodology (RMM)- Maximum likelihood estimation procedures. *Computational Statistics and Data Analysis*. In press.

[11] Shore, H. (2005). Accurate RMM-based approximations for the CDF of the normal distribution. *Communications in Statistics (Theory and Methods)*, 34(3). In press.

Two-Way Distinction
(to model Variation)

- Modeling Systematic Variation vs. Modeling Random Variation
- Theory-based Modeling vs. Empirical Modeling

Two-way Partition of Modeling Variation

Modeling:	**Systematic Variation**	**Random Variation**
Theory-based	Science (relational models)	Statistics (statistical distributions)
Empirical	**?**	**?**

Scientific Relational Models

- **Kinetic Energy:** $E_k(V) = M\,(V^2/2)$
- **Radioactive Decay:** $R(t) = R_0 \exp(-kt)$
- **Antoine Equation:** $\log(P) = A + B / (T+C)$,
- **Arrhenius Formula:** $R_e(T) = A \exp[-E_a/(k_B T)]$
- **Gompertz Growth-Model:** $Y = \beta_1 \exp[-\beta_2 \exp(-\beta_3 x)]$
- **Einstein's:** $E = MC^2 / [1-(v/C)^2]^{1/2}$

Models of Random Variation

- Normal Distribution
- Exponential Distribution
- Cauchy Distribution
- Pearson Family of Distributions
- Johnson Families (S_B, S_U, Log Normal)
- Tukey's g- and h- Systems of Distributions
- Box-Cox transformations

Empirical Modeling of Random Variation (Current)

Modeling:	**Systematic Variation**	**Random Variation**
Theory-based	Science (Relational models)	Statistics (Statistical distributions)
Empirical	**?**	**Heuristics?** **Moment Matching?** **Percentile Matching?**

Empirical Modeling of Systematic Variation (Current)

Modeling:	**Systematic Variation**	**Random Variation**
Theory-based	Science (relational models)	Statistics (Statistical distributions)
Empirical	**?**	Heuristics, Moment Matching, Percentile Matching

Empirical Modeling of Variation (RMM)

Modeling:	**Systematic Variation**	**Random Variation**
Theory-based	Science (relational models)	Statistics (statistical distributions)
Empirical	**?**	**?**

The RMM Model

Two questions:

- - What is the error-structure of the relational models?
- - Are there certain “patterns” that repeatedly appear in **current** relational models? (assuming monotone convex-concave relationships)

The RMM Model (1^{st} question)

- **Antoine Equation (again):**

$$\log(P) = A + B / (T+C),\ B<0$$

[P-pressure, T- temperature (K^0)]

- **What are the random errors associated with this model?**

Suppose that T was “frozen” (namely, constant). Then (one option):

$$\log(P) = A + B / (T+C) + \varepsilon_2,\ B<0$$

ε_2– Random error, representing:

“Self-generated Variation”

The RMM Model (Cont'd)

- Since T is not constant, we can write:

$$T = \mu_T + \varepsilon_1$$

ε_1 – a random additive error, representing

"Externally-generated" random variation

- **Antoine Equation, with the errors:**

$$\log(P) = A + B / (\mu_T + \varepsilon_1 + C) + \varepsilon_2, \ B<0$$

The RMM Model (2nd question)

What are the "repeated" patterns?

- **Kinetic Energy:** $E_k(V) = M\,(V^2/2)$
- **Radioactive Decay:** $R(t) = R_0 \exp(-kt)$
- **Antoine Equation:** $\log(P) = A + B\,/\,(T+C)$,
- **Arrhenius Formula:** $R_e(T) = A \exp[-E_a/(k_B T)]$
- **Gompertz Growth-Model:** $Y = \beta_1 \exp[-\beta_2 \exp(-\beta_3 x)]$

The RMM Model (axiomatic derivation)

- Assume that “Self-generated” and “Externally generated” components of variation **interact**:

$$Y = f_1(\eta,\varepsilon_1;\ \boldsymbol{\theta}_1)\ f_2(\varepsilon_2;\ \boldsymbol{\theta}_2)$$

f_1- Modeling “Exernally-generated” variation

f_2- Modeling “Self-generated” variation

We now wish to model f_1 and f_2

The RMM Model (axiomatic derivation)

Modeling "Self-generated" *random* variation (f_2):

- One option (a proportional model):

$$f_2(\varepsilon_2; \boldsymbol{\theta_2}) = M(1+ \varepsilon_2)= M(1+\sigma_{\varepsilon 2}Z_2)$$

M- A central value, for example: the median

Z_2- A standard normal r.v.

- Alternatively ($\varepsilon_2 <<1$): $f_2(\varepsilon_2; \boldsymbol{\theta_2}) = \exp(\mu_2 + \sigma_{\varepsilon 2}Z_2)$

The RMM Model
(axiomatic derivation)

Modeling "Externally-generated" *random* and *systematic* variation (f_1):

"The Ladder of fundamental monotone convex/concave functions"

The RMM Model (axiomatic derivation)

The "Ladder" ($V=\eta+\varepsilon_1$)

•

•

Exponential-exponential: $\exp[a \exp(V)]$

Exponential-power: $\exp(aV^b)$

Exponential: $\exp(V)$

Power: V^a

Linear: V

The RMM Model (axiomatic derivation)

Modeling "Externally-generated" *random* and *systematic* variation:

$$f_1(\eta,\varepsilon_1;\, \boldsymbol{\theta}_1) = \exp\{(\alpha/\lambda)[(\eta+\varepsilon_1)^{\lambda} - 1]\}$$

Does this represent well the "Ladder"?

The RMM Model (axiomatic derivation)

$$f_1(\eta,\varepsilon_1; \boldsymbol{\theta}_1) = \exp\{(\alpha/\lambda)[(\eta+\varepsilon_1)^{\lambda} - 1]\}$$

Special Cases:

- Linear: $\lambda=0$, $\alpha=1$
- Power: $\lambda=0$, $\alpha\neq1$
- Exponential: $\lambda=1$
- Exponential-power: $\lambda\neq0$, $\lambda\neq1$
- Exponential-exponential: ??

The RMM Model (axiomatic derivation)

$$f_1(\eta,\varepsilon_1;\,\boldsymbol{\theta}_1) = \exp\{(\alpha/\lambda)[(\eta+\varepsilon_1)^{\lambda} - 1]\}$$

Insert: $\exp\{(\beta/\kappa)[(\eta+\varepsilon_1)^{\kappa} - 1]\}$

For $\kappa=0$, $\beta=1$:

$$\exp\{(\beta/\kappa)[(\eta+\varepsilon_1)^{\kappa} - 1]\} = \eta+\varepsilon_1$$

Adding two parameters allows a repeat of the cycle "Linear-power-exponential"

The RMM Model (axiomatic derivation)

Significance of the model for f_1:

The “Ladder” becomes a “conveyer” that takes one, in a **continuous** manner, from one point to the next on the

Spectrum of convexity intensity

The RMM Model (axiomatic derivation)

The complete model:

$$f_1(\eta,\varepsilon_1;\, \theta_1)\, f_2(\varepsilon_2;\, \theta_2) = \exp\{(\alpha/\lambda)[(\eta+\varepsilon_1)^{\lambda} - 1] + \mu_2 + \varepsilon_2\}$$

$\varepsilon_1 = \sigma_{\varepsilon 1} Z_1$, $\varepsilon_2 = \sigma_{\varepsilon 2} Z_2$, $\{Z_1, Z_2\}$ from a bi-variate standard normal distribution with correlation ρ

Three sets of parameters:

- The linear predictor: $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ..$
- "Structural parameters", $\{\alpha, \lambda, \mu_2\}$
- "Error parameters", $\{\rho, \sigma_{\varepsilon 1}, \sigma_{\varepsilon 1}\}$

"Not all parameters were created equal"

The RMM Model (axiomatic derivation)

- Write: $Z_2 \mid (Z_1=z_1) = \rho z_1 + (1-\rho^2)^{1/2} Z$

The model, expressed in terms of *independent* standard normal variables:

$W = \log(Y) =$

$$(\alpha/\lambda)[(\eta+\sigma_{\varepsilon 1} Z_1)^{\lambda} - 1]+\mu_2+\sigma_{\varepsilon 2}\,[\rho Z_1 + (1-\rho^2)^{1/2} Z_2]$$

- For: $\rho = \pm 1$, $\eta = 1$ ($\mu_2=\log(\text{Med.})$):

$W = (\alpha/\lambda)[(1+\sigma_{\varepsilon 1} Z)^{\lambda} - 1]+\mu_2+\sigma_{\varepsilon 2}\rho Z$, an **INT**

The RMM Model (axiomatic derivation)

Eight variations of the RMM model

- Two ways to write each error ($\varepsilon_2 << 1$, $\varepsilon_1 << \eta$):

$$1+\varepsilon_2 \cong \exp(\varepsilon_2),$$

$$(\eta+\varepsilon_1)^\lambda = \eta^\lambda(1+\varepsilon_1/\eta)^\lambda \cong \exp[\lambda \log(\eta)+\lambda\varepsilon_1/\eta]$$

- Two equivalent forms to express the standard normal variables as indpendent (uncorrelated):

$$Z_2 \mid (Z_1=z_1) = \rho z_1 + (1-\rho^2)^{1/2} Z$$

$$Z_1 \mid (Z_2=z_2) = \rho z_2 + (1-\rho^2)^{1/2} Z$$

RMM Estimation

- Phase 1:Estimating the linear predictor (Set I): Canonical Correlation Analysis plus Linear Reg.

- Phase 2: Iteratively alternating between Estimating Set II (the “Structural Parameters”) via W-NL-LS and Set II (the “Error Parameters”) via max. of log-likelihood

RMM Modeling: Random Variation

Modeling via the INT ($\rho=\pm 1$, one variation):

$$Y= \exp\{(A/\lambda)[(1+BZ)^{\lambda} - 1]+\log(M)+DZ\}$$

Z- Standard normal

Examples:

- Approximations for the Poisson quantile and CDF of the Normal distribution
- Approximations for gamma

RMM Modeling (INT)- The Poisson

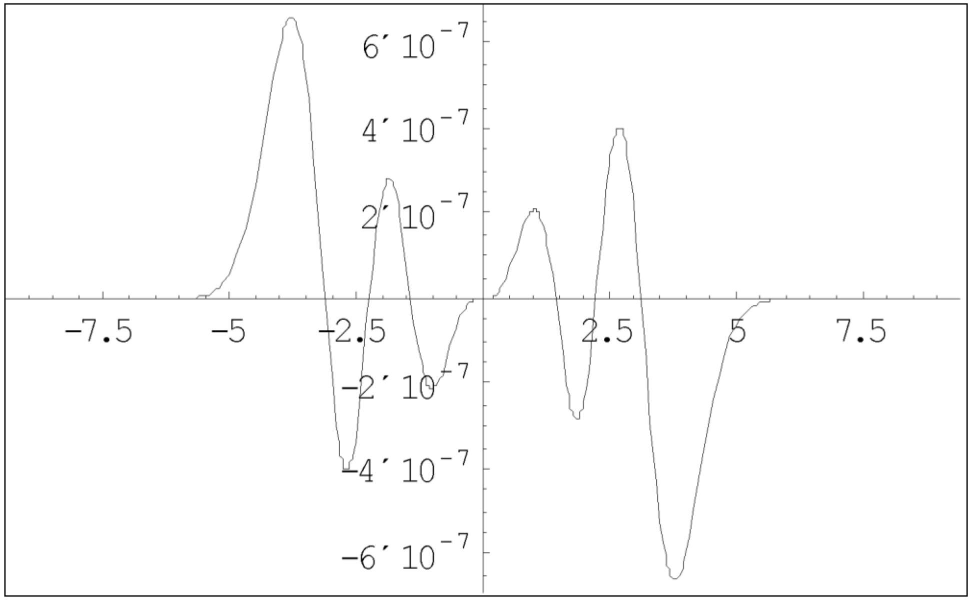
PARAMETERS					QUANTILES						
λ	**M**	**B**	**C**	**D**	**0**	**1**	**2**	**4**	**6**	**8**	**10**
1/2	.1794	-2.786	-.5660	.1925	.281	1.02	1.97	4.01	6.01	7.99	10.0
1	.4785	-1.934	-.6212	.2078	.284	1.02	1.97	4.01	6.01	7.99	10.0
2	1.315	-1.112	-.6903	.2265	.289	1.02	1.97	4.01	6.00	7.99	10.0
5	4.345	-.3190	-.8111	.2581	.296	1.02	1.97	4.01	6.01	7.98	10.0
10	9.325	-.0612	-.9325	.2892	.303	1.02	1.97	4.01	6.01	7.98	10.0
20	22.25	-.0039	-1.089	.3285	.308	1.02	1.97	4.01	6.01	7.98	10.0

Source: *Communications in Statistics (T&M)*, 33(7), 2004

RMM Modeling (INT)- CDF of Standard Normal Distribution

For example: y= -log(1-P) **=**(M) exp{B[exp(Cz)-1] + Dz

(**Source:** *Communications in Statistics (T&M), 34(3), 2005)*



RMM Modeling (INT)- Gamma

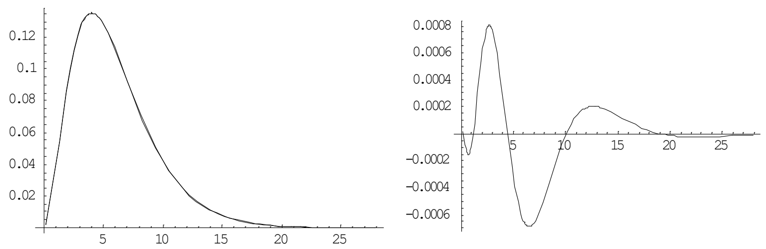


Figure 10.1. Plots of the density function, exact and fitted (left) and of the error for Gamma (3,2) (based on quantile-matching).

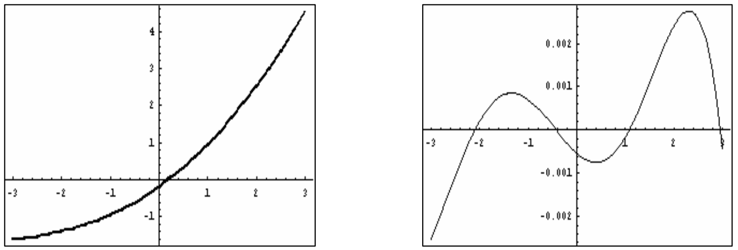


Figure 10.2. Plots of the quantile function (approximate and exact, left) and of the deviation (approximate minus exact, right) for the Gamma (3,2). Horizontal axis is the standard normal quantile (z). Vertical axis (left plot) is the standardized gamma quantile.

RMM Modeling (for sample data)

Fitting and estimating procedures for modeling **random** variation have been developed:

- - Maximum likelihood estimation
- - Estimation by “Percentile-Matching”
- - Estimation by “Moment-Matching” (two-moment, partial and complete)

RMM Modeling (for sample data)

Intra-Galactic Velocities (modeling random var.)

- From Karian and Dudewicz (KD, 2000, p. 208). In astronomy, the cluster named A1775 is believed to consist of two clusters that are in close proximity. KD concluded that the **generalized beta distribution** (GBD) may model well this dataset (therein, p. 208). The four-parameter d.f. of GBD is given by (therein, p. 119):

$$f(y) = (y - \beta_1)^{\beta 3} (\beta_1 + \beta_2 - y)^{\beta 4} / [\ \beta(\beta_3+1, \beta_4+1)\ \beta_2^{(\beta 3+\beta 4+1)}],$$

$$\beta_1 \le y \le \beta_1 + \beta_2$$

Intra-Galactic Velocities (modeling random var.)

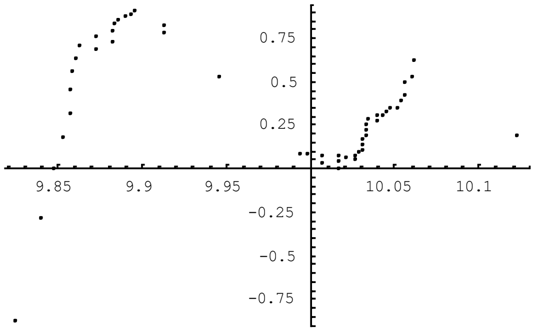


Figure 3: Example 2- Relative errors (in %) of the fitted (50) vs. actual log of the intra-galactic velocities (RMM analysis)

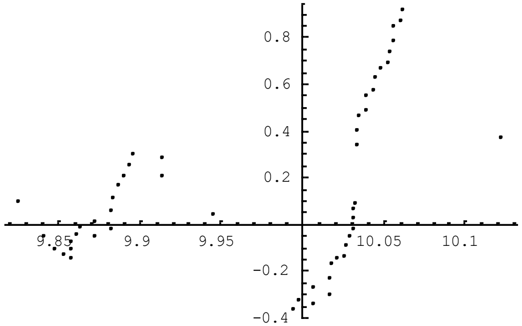


Figure 4: Example 2- Relative errors (in %) of the fitted (51) vs. actual log of the intra-galactic velocities (KD model, based on GBD)

Average relative error: Beta (KD)- 0.28%; RMM- 0.39%

Source: ***Computational Statistics and Data Analysis,*** In press

The Wave-Soldering Process (modeling systematic variation)

- **GLM: Myers, Montgomery, Vining (2002, Book, John Wiley&Sons)**

  $\log(\mu)=3.08 + 0.44C - 0.40G + 0.28AC - 0.31BD$

  $\mu^{1/2} = 5.02 + 0.43A - 0.40B + 1.19C - 0.39E - 1.10G + 0.58AC - 0.96BD$ **(Poisson Model)**

- **RMM: Shore (2004, Book, World Scientific Publishing)**

  $\eta= -0.1786 + 0.16A - 0.14B + 0.53C - 0.16E - 0.49G + 0.22AC - 0.43BD$ **(R^2-adj.=0.848)**

- **Comparison with the square-root link:**

  For A/C: RMM: 0.164/0.526= **0.31**; **GLM:** 0.43/1.19= **0.36**

  For (AC)/(BD): RMM: 0.224/(-0.431)= **-0.52**; GLM: 0.58/(-.96)= **-0.60**.

RMM Modeling in Chemical Engineering

- **DIPPR** (Design Institute for Physical Property Data)- A leading data-base for chemical properties and their modeling
- Notation:

  - **The acceptable model:** Model with best fit relative to the data set in the data-base

  - **Goodness-of-fit:** By various measures for fit (for example: R^2, AIC, MSE)

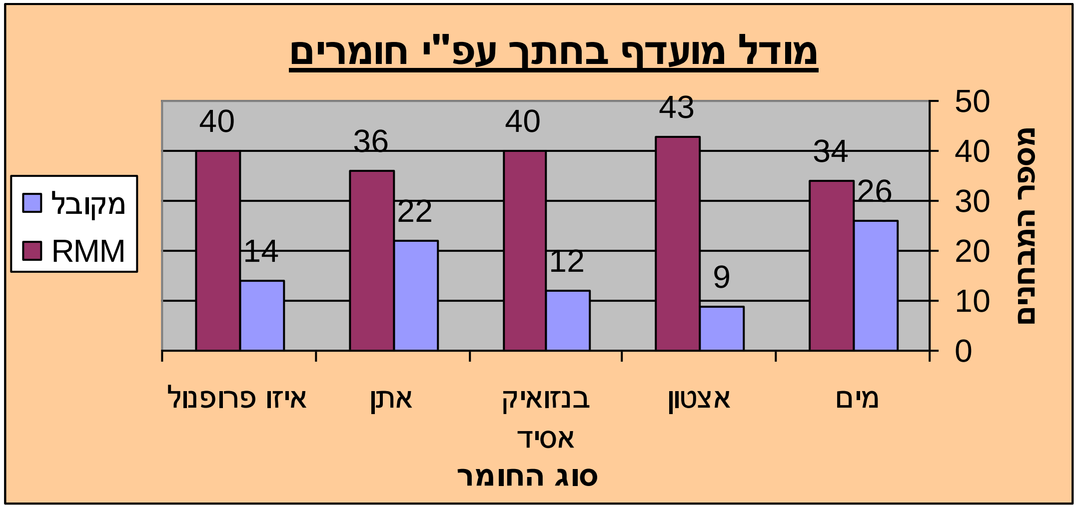
RMM Modeling in Chemical Engineering (Cont'd)

- **Substances examined (5):** Water, Acetone, Benzoic acid, Ethane, Iso-propanol
- **Properties examined (13):** Solid density, Liquid density, Solid vapor pressure, Vapor pressure, Heat of vaporization, Solid heat capacity, Liquid heat capacity, Ideal gas heat capacity, Liquid viscosity, Vapor viscosity, Liquid thermal conductivity, Vapor thermal conductivity, Surface tension

RMM Modeling in Chemical Engineering (Cont'd)

Preferred Model

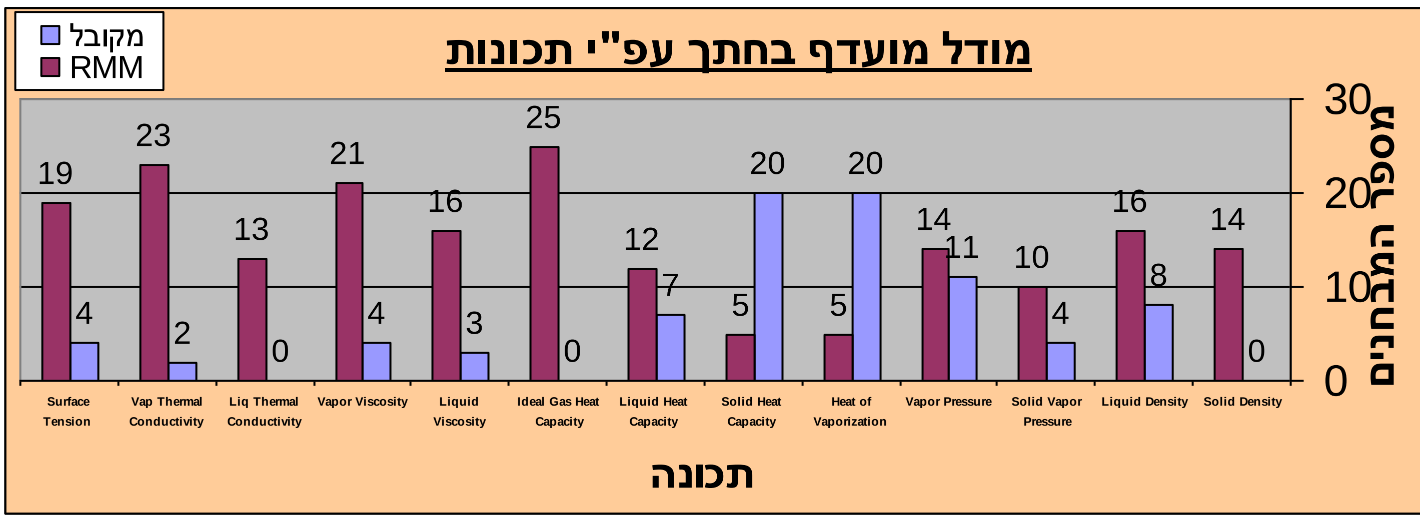
(by substance, across properties and goodness-of-fit measures)



RMM Modeling in Chemical Engineering (Cont'd)

Preferred Model

(by property, across substances and goodness-of-fit measures)



RMM Modeling

- Thank you

(shor@bgu.ac.il)