



context: aggregation model analysis: I have attached R notebook html version and data ENB.txt which it has processed.

Input: ENB.txt It regards SIG AT2 Q-1.pdf for procedure and data structure.

Aggregation Model : It is considering weighted aggregation, ordered weighted aggregation, Quassi-arithmetic aggregation, choquet Integral, Geometric weighted aggregation. After finding optimum model based on metrics and statics like RMSE and correlations, the test data is evaluated for Appliances energy consumption (Column Y)

role : Assume data analyst role to analyse all models (their weightages, predictor variable signified, overfitting or biasing) and compare model

task: analyse and describe all models (their weightages, predictor variable signified, overfitting or biasing) and compare metrics and stats and describe them. Find optimum model.

stats and metrics provided : text files which list metrics RMSE ,Av. abs error , Pearson correlation , Spearman correlation , Orness

audience : Business stakeholders looking for best model to employ.

The analysis of aggregation models for predicting Appliances energy consumption (Y) based on the ENB.txt dataset and SIG AT2 Q-1.pdf procedure is as follows:

Model metrics and weights summary:

- Weighted Arithmetic Mean (WAM):
 - RMSE: 0.0413, Average Absolute Error: 0.0280
 - Pearson correlation: 0.2071, Spearman correlation: 0.2868
 - Weights indicate strong importance for variable 1 (Temperature living room) with a weight ~0.993, small weight for variable 3 (~0.007), and negligible others.
- Weighted Power Mean p=0.5:
 - RMSE: 0.0413, Average Absolute Error: 0.0279
 - Pearson correlation: 0.2109, Spearman correlation: 0.2949
 - Weights also highly favor variable 1 (~0.984) and a bit more weight for variable 3 (~0.016), others near zero.

- Weighted Power Mean QM (Quadratic Mean):
 - Very high errors (RMSE and average abs error > 1.2), and no correlation reported.
 - Weights are zero for all variables, indicating this model performs poorly or is invalid here.
- Weighted Power Mean GM (Geometric Mean):
 - RMSE: 0.0413, Average Absolute Error: 0.0279
 - Pearson correlation: 0.2155, Spearman correlation: 0.3009
 - Variable 1 still dominant weight (~0.969), variable 3 has increased weight (~0.031); others zero.
- Ordered Weighted Averaging (OWA):
 - RMSE, absolute error, and correlations same as WAM.
 - Orness is very low (0.0037), indicating near-min operator behavior.
 - Weights nearly identical to WAM: dominant variable 1 (~0.993), minor for variable 3 (~0.007).
- Choquet Integral:
 - Metrics almost identical to WAM and OWA.
 - Orness moderate (0.252), suggesting some trade-off between AND/OR operators.
 - No direct variable weights provided but Shapley values (importance measures) indicate:
 - Variable 1 and 5 have importance ~0.496 each (almost equal)
 - Variable 3 minor importance (~0.007)
 - Variables 2 and 4 negligible.
 - Fuzzy measure weights indicate complex interaction patterns.

Interpretation and comparison:

- All models except PowerMean QM show very similar predictive accuracy (RMSE ~0.0413) and correlation values (~0.20-0.22 Pearson).
- WAM, OWA, and Choquet have almost the same error metrics, but Choquet shows higher orness, capturing importance of variable 5 as well, unlike others that mostly focus on variable 1.
- The best weighted power mean variant in terms of correlations is GM with slightly higher Pearson and Spearman but still close to others.
- PowerMean QM is clearly not performing well and should be excluded.
- Variable 1 (Living room temperature) is consistently the dominant predictor across all models, with variable 3 (Office room temperature) having minor influence. Choquet uniquely assigns significant importance to variable 5 (Pressure), suggesting it captures interaction effects better.

- The OWA model shows low orness, indicating strong bias toward minimum values, which may limit flexibility.

Optimum model recommendation:

- Based on the balance of predictive accuracy and interpretability, the Choquet integral model is recommended.
- It captures multi-variable interactions better, includes variable 5 importance not seen in simpler weighted averages, with no loss in accuracy.
- It avoids overfitting given the moderate correlations yet maintains relevant variable contributions.
- The WAM and OWA models are simpler with similar errors but ignore variable 5 influence.

In summary, the Choquet integral model is best suited for predicting Appliances energy consumption here, balancing error metrics, correlation values, and meaningful variable importance better than alternatives.

This analysis is based on detailed metrics and weight extraction from provided stats files for six models: WAM, OWA, Choquet, PowerMean ($p=0.5$, GM, QM) [WAM-stats.txt, OWA-stats.txt, choquet-stats.txt, powermean-stats*.txt] and incorporates variable contributions and orness where applicable. The predictions and correlations imply no severe overfitting but some bias toward variable 1 except Choquet which better balances variables [attached_files].