A METHOD OF LINEAR COMBINATION OF MULTIPLE MODELS FOR EPISTEMIC UNCERTAINTY MINIMIZATION

D. Y. Kwak\textsuperscript{1}, E. Seyhan\textsuperscript{2}, and T. Kishida\textsuperscript{3}

ABSTRACT

In this study, we are evaluating methods to minimize the epistemic uncertainty that is usually caused due to the model selection and logic tree schemes. One of the methods is to simply select the best prediction model resulting in the least variation of errors if the variation is known, and the other one is to seek a multi-model weighting scheme through logic trees. For the latter case, selection of weights varies depending on the availability of distribution and correlation information among the models. Equal weights are often used when only mean predictions are available. Inverse-variance weighting methods are generally used when distributions for each model are available. Optimized weights can be determined by minimizing the prediction variance when distributions and correlations among the multiple models are available. In this study, we describe such methods with mathematical derivation and numerical solutions and apply to example cases with varying combination of variances and correlation levels to compare results. We find that the variance of combined model error is reduced when the models are less correlated, and the optimized weight scheme is more effective when the variance of each model is changing. Although such findings may not be claimed as new, we believe that this study can be considered as a benchmark. We also apply the linear combination method to proxy-based $V_{530}$ estimations using two regional data sets (California and Japan) considering three proxies for $V_{530}$: slope, terrain, and geology. As these three proxy-based models are highly correlated, which results in two options for the best practice supposing that exact correlation is unknown: 1) use of the proxy with the least variance if variance is lower than others; 2) use of equal weight if variances are comparable.

\textsuperscript{1}Assist. Prof., Dept. of Civil and Env. Eng., Hanyang Univ., Ansan, South Korea (email: dkwak@hanyang.ac.kr)
\textsuperscript{2}Lead Earthquake Modeler, Risk Management Solutions, Inc., 7575 Gateway Blvd, Newark, CA 94560, USA
\textsuperscript{3}Assist. Prof., Dept. of Civil Infra. and Env. Eng., Khalifa Univ., Abu Dhabi, United Arab Emirates

A Method of Linear Combination of Multiple Models for Epistemic Uncertainty Minimization

D. Y. Kwak¹, E. Seyhan², and T. Kishida³

ABSTRACT

In this study, we are evaluating methods to minimize the epistemic uncertainty that is usually caused due to the model selection and logic tree schemes. One of the methods is to simply select the best prediction model resulting in the least variation of errors if the variation is known, and the other one is to seek a multi-model weighting scheme through logic trees. For the latter case, selection of weights varies depending on the availability of distribution and correlation information among the models. Equal weights are often used when only mean predictions are available. Inverse-variance weighting methods are generally used when distributions for each model are available. Optimized weights can be determined by minimizing the prediction variance when distributions and correlations among the multiple models are available. In this study, we describe such methods with mathematical derivation and numerical solutions and apply to example cases with varying combination of variances and correlation levels to compare results. We find that the variance of combined model error is reduced when the models are less correlated, and the optimized weight scheme is more effective when the variance of each model is changing. Although such findings may not be claimed as new, we believe that this study can be taken as a benchmark. We also apply the linear combination method to proxy-based $V_{30}$ estimations using two regional data sets (California and Japan) considering three proxies for $V_{30}$: slope, terrain, and geology. As these three proxy-based models are highly correlated, which results in two options for the best practice supposing that exact correlation is unknown: 1) use of the proxy with the least variance if variance is lower than others; 2) use of equal weight if variances are comparable.

Introduction

Forecasting models can vary due to several reasons such as the use of different data types and samples; the use of the same data sets, but different model building approaches. For example, when

---

¹ Assis. Prof., Dept. of Civil and Env. Eng., Hanyang Univ., Ansan, South Korea (email: dkwak@hanyang.ac.kr)
² Lead Earthquake Modeler, Risk Management Solutions, Inc., 7575 Gateway Blvd, Newark, CA 94560, USA
³ Assis. Prof., Dept. of Civil Infra. and Env. Eng., Khalifa Univ., Abu Dhabi, United Arab Emirates

the time-averaged shear wave velocity of the upper 30 m of a site \(V_{S30}\) is unknown, it is usually estimated using a weighted approach of different proxies such as geology, slope, terrain, or combination of them [1-4]. Each proxy uses different amount and quality of data types but predicts a common variable which is \(V_{S30}\). The accuracy of such variables is also important in development of ground motion models (GMPEs or GMMs). For instance, five GMMs were developed from the NGA West2 project [5], where each GMM provides intensity measures (IMs) of PGA, PGV, and various spectral accelerations. All GMMs used a common well-recorded strong ground motion data set called “flatfile” [6], while each GMM was developed independently.

For better accuracy, it is essential to select a model or combine models using logic trees from a suite of model that minimizes the large epistemic uncertainty for the region of the interest. One of the existing methods is to simply select the best prediction model resulting in the least variation of errors if variation is known, and the other one is to seek a multi-model weighting scheme through logic trees. The former case is simple, but even with the model with least variation, the variation of prediction can be reduced by combining with other models. For the latter case, selection of weights varies depending on the availability of error distributions and correlation information among models. Generally, there are three ways to develop weighting schemes:

1. Equal weight: Equal weights are often used when the models are ranked closely, and/or only mean predictions are available coupled with inadequate knowledge.
2. Inverse-variance weight: Inverse-variance weighting method is usually used when error distributions for each model are available.
3. Optimized weight: Optimized weights can be determined by minimizing the prediction variance when both distributions and correlations among the multiple models are available.

Above methods are called as “linear combination method” because each model with certain weight is summed for final model prediction. Seyhan et al. [7] provides inverse-variance weights to proxy-based models predicting \(V_{S30}\) for each region (California, Japan, Taiwan) applied to NGA-West 2 site database. Kishida et al. [8] presents optimized weights to existing local magnitude models to predict the moment magnitude using Iranian data sets. Other than the linear combination method, there is Bayesian Model Averaging (BMA) method of combining models based on the likelihood of model estimations [9]. However, BMA is not discussed in this paper.

In this paper, we describe three aforementioned linear combination methods with mathematical derivations and numerical solutions. We compare the results by applying these methods to example cases with hypothetical setup, and an illustrative model used in earthquake engineering practices such as combination of proxy-based \(V_{S30}\) estimations. Evaluation of prediction improvements across the methods are provided by highlighting the importance of correlations among the multiple models.

**Linear Combination Method**

Linear combination method represents linearly combining multiple models with different weights:

\[
\hat{y} = w_1 \hat{y}_1 + w_2 \hat{y}_2 + w_3 \hat{y}_3 + \cdots \text{ where } \sum_i w_i = 1
\]

where \(\hat{y}\) is the combined prediction, \(\hat{y}_i\) is the prediction from a model \(i\), and \(w_i\) is a weight to the \(i\). The sum of weights is equal to one. A preferred weights scheme should be the one resulting in the
least mean square error (MSE) of $\hat{y}$.

Herein we describe the variance estimation of $\hat{y}$ starting with two models. If we suppose that the true estimate is $y$, the residual $e_i$ for model $i$ can be calculated as:

$$
e_1 = y - \hat{y}_1$$
$$
e_2 = y - \hat{y}_2$$

(2)

We define the variance of $e_1$ and $e_2$ as $\sigma_1^2$ and $\sigma_2^2$, respectively, and covariance as $\sigma_{12}$. The linearly weighted model is then expressed as:

$$\hat{y} = w_1\hat{y}_1 + w_2\hat{y}_2$$

(3)

where $w_1+w_2$ is unity. Supposing the residual of $\hat{y}$ as $e$, it becomes:

$$e = w_1e_1 + w_2e_2$$

(4)

Then, the variance of $e$, $\sigma^2$ will be as follows:

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}$$

(5)

For multiple models more than two, the $\sigma^2$ can be shown using matrix notation:

$$\sigma^2 = [w_1 \ldots w_n] \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

(6)

The preferred weight vector is the one with the least $\sigma^2$. However, to solve Eq. (6) for the least $\sigma^2$, both variance and covariance information are required. Unfortunately, such information is generally limited and sparse. Thus, in this study, we investigate these three weighting schemes: 1) equal weights, 2) inverse-variance weights, and 3) optimized weights. At the end of this section, we compare $\sigma^2$ from the three schemes with example cases.

**Equal Weight Approach**

Equal weight scheme is commonly chosen when the mean prediction from a model is only available without any preference on each model. In this case, the weight for each model will be simply the inverse of the total number of models:

$$w_i = \frac{1}{\text{number of models}}$$

(7)

Substituting $w_i$ into Eq. (6), the variance combining $N$ models becomes as follows:

$$\sigma^2 = \frac{1}{N^2} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

(8)
Depending on the correlation coefficient ($\rho_{ij}$), the $\sigma^2$ in Eq. (8) ranges within two extremes:

- **Minimum - no correlation** ($\rho_{ij}=0$): $\sigma^2 = \frac{1}{N^2} \sum_i^N \sigma_i^2$ 
- **Maximum - perfect correlation** ($\rho_{ij}=1$): $\sigma^2 = \frac{1}{N^2} (\sum_i^N \sigma_i)^2$

**Inverse-Variance Weight Approach**

Inverse-variance weight scheme can be used when adequate information on the variance is available for each model in addition to the mean predictions. Usually prediction models include the variance information, so this method is widely used [7]. The weight is inversely proportional to the variance $\sigma_i^2$ of the model:

$$w_i = \frac{1}{\sigma_i^2} / \sum_i \frac{1}{\sigma_i^2}$$

where sum of weights is unity. In addition to $\sigma^2$, mean misfit of a model ($\mu_i$) can be considered for the proportion of the weight [7]. The weights based on the inverse of variance and square of mean misfit can be expressed as follows:

$$w_i = \frac{1}{\mu_i^2 + \sigma_i^2} / \sum_i \frac{1}{\mu_i^2 + \sigma_i^2}$$

(10)

Note that if $\sigma_i^2$ and $\mu_i$ are equal to all models, the weights become equal (Eq. 7). Substituting $w_i$ in Eq. (10) to Eq. (6), the variance of combined models can be shown below:

$$\sigma^2 = \left(\sum_i \frac{1}{\mu_i^2 + \sigma_i^2}\right)^2 \left[\frac{1}{\mu_1^2 + \sigma_1^2} \cdots \frac{1}{\mu_n^2 + \sigma_n^2}\right] \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_n^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\mu_1^2 + \sigma_1^2} \\ \vdots \\ \frac{1}{\mu_n^2 + \sigma_n^2} \end{bmatrix}$$

(11)

Similar to the equal weight approach, the $\sigma^2$ in Eq. (11) ranges within two extremes depending on the $\rho_{ij}$. The minimum will be given if models are uncorrelated ($\rho_{ij}=0$; $\sigma_{ij}=0$), and the maximum will be obtained if models are perfectly correlated ($\rho_{ij}=1$; $\sigma_{ij}^2=\sigma_i^2 \sigma_j^2$).

**Optimized Weight Approach**

If both variance and covariance information are available, we can choose optimized weights to obtain the least $\sigma^2$. We can derive the following solutions for two models by solving Eq. (5):

$$w_1 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

$$w_2 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

(12)

For multiple models, it is not trivial solving Eq. (5). Rather, we can use Monte-Carlo simulation...
approach as follows to find the weight vector to minimize $\sigma^2$:

1. Generate N number of random vectors for which each vector has sum of unity and each element is within the range of [0 1].
2. Calculate variance using Eq. (6).
3. Find a vector resulting in minimized $\sigma^2$.

Figure 1 shows the number of generations ($n$) and difference of $\sigma_N$ with $\sigma_N$ (i.e., $(\sigma_N - \sigma_N)/\sigma_N$). The $\sigma_N$ is the standard deviation resulting from the simulation with sufficient $n$ (i.e., $10^6$). The difference is shown as mean (Figure 1a) and standard deviation (Figure 1b), which are resulting from 25 simulations per $n$. We assume the minimal sufficient $n$ is the point that the mean of difference is less than $10^{-2}$, and found that 1) having more models increases the mean difference of the standard deviations and 2) $n > 100,000$ provides a sufficient result throughout the number of models considered. With the same $n$, the mean of difference increases with the number of models up to six, and decreases afterward.

![Figure 1](image_url)

Figure 1. Difference of the standard deviation relative to the standard deviation resulting from one million generation. (a) mean and (b) standard deviation of the difference from 25 simulations per $n$. Two to eight number of models are used to compare.

**Sensitivity of Performance Evaluation for Example Cases**

In this section, the performances of three aforementioned weighting schemes are evaluated using example cases. Four cases are created for testing by varying standard deviation ($\sigma$) and the correlation coefficient ($\rho$) as shown below:

- Case 1: Equal $\sigma$ with no $\rho$
- Case 2: Varying $\sigma$ with no $\rho$
- Case 3: Equal $\sigma$ with varying $\rho$
- Case 4: Varying $\sigma$ with varying $\rho$

where $\sigma$ and $\rho$ for each case are shown in Table 1.
Table 1. Standard deviation $\sigma$ and correlation coefficient $\rho$ for each test case.

<table>
<thead>
<tr>
<th>Case</th>
<th>Model 1, $\sigma_1$</th>
<th>Model 2, $\sigma_2$</th>
<th>Model 3, $\sigma_3$</th>
<th>Model 1-2, $\rho_{12}$</th>
<th>Model 2-3, $\rho_{23}$</th>
<th>Model 1-3, $\rho_{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Figure 2 compares the weights assigned for each method and case, and Figure 3 shows the standard deviation of the combined estimate ($\sigma_{comb}$). Our findings are listed below:

- **Case 1**: All three schemes provide evenly distributed weights. Applying those equal weights, $\sigma_{comb} (=0.115)$ resulted in a reduction of 42% from $\sigma_i (=0.2)$.
- **Case 2**: Inverse-variance and optimized weight schemes provide the least $\sigma_{comb} (=0.154)$ which is 23% reduction from $\sigma_1$ that is the least $\sigma$ among other models. Both methods result in identical weights. For equal weight scheme, $\sigma_{comb}$ reduction is only 10%.
- **Case 3**: The weights from equal and inverse-variance schemes are evenly assigned to each model, and the resulting $\sigma_{comb} (=0.163)$ provides 18% reduction from $\sigma_i$. The optimized weight scheme gives different weights, but the reduction on $\sigma_{comb}$ is not as significant as from equal and inverse-variance weight schemes ($\sigma_{comb}=0.161$, 19% reduction). This indicates that if the variances are comparable among models, equal weight scheme approach is enough.
- **Case 4**: The use of equal and inverse-variance weights results in $\sigma_{comb}$ of 0.253 and 0.207, respectively, which are higher than $\sigma_1$. Using the optimized weight scheme approach, Model 3 in Table 1, which has the higher variance and higher covariance than other models, results in zero weight. The $\sigma_{comb} (=0.187)$ is 7% less from $\sigma_1$. This observation indicates that the model with the least $\sigma$ can be adequate to use if the models are highly correlated with each other and equal weight scheme approach results in the higher $\sigma_{comb}$ than the least $\sigma$.

Figure 2. The weights are determined for model 1 to 3 for each method (Eq.: Equal weight, Inver.: Inverse-variance weight, Optim.: Optimized weight) for (a) Case 1, (b) Case 2, (c) Case 3, (d) Case 4.
Application

Three linear combination methods illustrated in this study are applied to proxy-based $V_{S30}$ estimation models. Generally, the prediction models provide variance of errors, but the covariance between models are not available except that they are evaluated using observation data. In that case, we use exact correlation coefficients calculated from measured $V_{S30}$ values as well as three levels of correlation levels supposed: low ($\rho$=0.1), medium ($\rho$=0.4), and high ($\rho$=0.8) to see the impact of correlations.

Performance of Proxy-based $V_{S30}$ Estimations

Two regional data sets are used to test the linear combination methods on proxy-based $V_{S30}$ model:

- **California**: NGA-West 2 site database provides 4149 sites with measured and inferred $V_{S30}$ values [7] (Accessed on March, 2018: http://peer.berkeley.edu/ngawest2/databases/). For partial measurements when the depth of the profile ($z_p$) is less than 30 m, extrapolation methods are preferred [10-11]. In this database, the ‘$V_{S30}$ Code’ column indicates the ranking of the preferred $V_{S30}$, where ‘0’ and ‘1’ represents the fully measured ($z_p \geq 30$m) and partially-measured $V_{S30}$ values ($10 \leq z_p \leq 20$m), respectively. We subset the data set with the $V_{S30}$ code of 0 or 1 conditioned on the available slope, geology, and terrain-based $V_{S30}$ values for the sites [1-3]. This screening provides us with 264 California sites.

- **Japan**: As a part of NGA-Sub project, seismic stations in Japan were characterized from NIED K-NET & KiK-Net and PARI networks, and $V_{S30}$ values were populated [12-13]. Among the total of 2231 sites, 1675 sites had measured or inferred $V_{S30}$ as well as slope, geology, and terrain-based $V_{S30}$ values [2-3,14].

For both data sets, the following proxy-based $V_{S30}$ models are considered: slope, terrain, and geology. Table 2 shows the standard deviation and correlation coefficient of each of these models. These standard deviations and correlation coefficients are calculated using the subset of the NGA-West 2 site database as mentioned above. Since measured $V_{S30}$ values are available, we could estimate $\rho$ among three models directly which are in range of 0.65 – 0.87. In addition to the
exact $\rho$, we added three levels of correlation levels: low ($\rho_{ij} = 0.1$), medium ($\rho_{ij} = 0.4$), high ($\rho_{ij} = 0.8$), and mix ($[\rho_{st}, \rho_{tg}, \rho_{sg}] = [0.1, 0.4, 0.8]$) for linear combination method to evaluate the impact of correlation.

Table 2. Standard deviation ($\sigma$) of a model and correlation coefficient ($\rho$) between models.

<table>
<thead>
<tr>
<th>Region</th>
<th>Slope, $\sigma_s$</th>
<th>Terrain, $\sigma_t$</th>
<th>Geology, $\sigma_g$</th>
<th>$\rho_{st}$</th>
<th>$\rho_{tg}$</th>
<th>$\rho_{sg}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>0.443</td>
<td>0.426</td>
<td>0.312</td>
<td>0.862</td>
<td>0.776</td>
<td>0.726</td>
</tr>
<tr>
<td>Japan</td>
<td>0.432</td>
<td>0.41</td>
<td>0.422</td>
<td>0.865</td>
<td>0.662</td>
<td>0.666</td>
</tr>
</tbody>
</table>

Figures 4 shows weights for each correlation levels for California and Japan data sets. For California (top-row in Table 2), the geology proxy has a very high weight of more than 97%, showing a high correlation with $V_{S30}$ data, and other two proxies have negligible weights for the case of exact correlation (Figure 4a). This trend is also shown for the case of high-correlation level (Figure 4d). As presented in Figure 5a, the $\sigma_{comb}$ for the exact correlation with optimized method is similar to $\sigma_g$ which has the least $\sigma$ among models, and equal and inverse-variance weight schemes result in higher $\sigma_{comb}$ than $\sigma_g$. If the correlation level is low, equal and inverse-variance weights provide lower $\sigma_{comb}$ than $\sigma_g$, and the weights between inverse-variance and optimized schemes are comparable. For the cases of mid-correlation level and mixed correlation, the $\sigma_{comb}$ is similar with $\sigma_g$ for equal weight scheme, and it decreases for inverse-variance and optimized weights.

The weights for each correlation level for Japan data set are shown in the lower panels of Figure 4. For the case of exact correlation with optimized weights, the slope proxy has a low weight, and terrain and geology proxies have similar weights. However, the improvement by adopting the optimized weight scheme is not significant from equal and inverse-variance weight schemes. Figure 5b compares $\sigma_{comb}$ per method for different level of correlation, and low to high correlation levels provide similar $\sigma_{comb}$. For the case of mixed correlation, the optimized weight scheme provides approximately 6% reduced $\sigma_{comb}$ relative to other weight schemes.

Concluding Remarks

In this paper, we described and presented the performance of three weighting schemes (equal weights, inverse-variance weight, optimized weight) of linear combination of models to reduce the existing epistemic uncertainty. We also applied these schemes to various prediction models for select cases and proxy-based $V_{S30}$ models. We found that the optimized weight scheme always provides the minimal variance of combined models, but the covariance between models required for the optimized weight scheme is usually not available. Supposing that the covariance information is unknown, we found that the inverse-variance weight scheme is the most effective if the correlation level is expected to be low, whereas selecting a model with the least standard deviation is more effective if correlation level is expected to be high. If the variances among models are similar, simple equal weights scheme performs reasonably well if correlation levels are expected to be similar among models. Although such findings may not be claimed as new, we believe that this study can be considered as a benchmark.

For the lack of covariance information, it is important to note that the resulting standard
deviation from a combined model could be highly biased regardless of combination method. The difference of standard deviation among difference correlation levels is even more significant than the difference from the selection of combination method (Figure 5). Hence, recognizing the correlation levels among models is recommended when combining models.

Figure 4. Weights determined for slope, terrain, and geology-based models for California (top-row) and Japan (bottom-row) data sets. Each combination method (Eq.: Equal weight, Inver.: Inverse-variance weight, Optim.: Optimized weight) is shown for (a) exact correlation, (b) low-level, (c) mid-level, (d) high-level, (e) and mixed correlation.

Figure 5. Standard deviation of combined estimate for each method and correlation level for (a)
California and (b) Japan data sets.

Acknowledgments

We acknowledge the Pacific Earthquake Research Center for providing Next Generation Attenuation (NGA) West 2 flatfile and site database (http://peer.berkeley.edu/ngawest2/databases/) and NGA-Subduction site database. We also thank our reviewers for their constructive comments.

References


