Chapter 9

EPOLLS Model for Average Horizontal Displacement

9.1. Development of Regional-EPOLLS Component.

The EPOLLS model for horizontal displacements is comprised of the Regional-EPOLLS, Site-EPOLLS, and Geotechnical-EPOLLS model components. With each component of the EPOLLS model, the average and standard deviation of the horizontal surface displacements (*Avg_Horz* and *StD_Horz*) can be predicted for a liquefaction-induced lateral spread. Organization of the EPOLLS model is discussed in Section 5.4 and depicted in Figures 5.3 and 5.4. Statistical regression methods, reviewed in Appendix D, were used to develop all of the EPOLLS model equations.

This section describes development of the *Regional-EPOLLS* (or *R-EPOLLS*) model component for average horizontal displacements (*Avg_Horz*). This component of the EPOLLS model is designed for application in seismic hazard evaluations of large geographic regions (hence, the name "regional" model component). Consistent with the level of information available in these studies, ground deformations can be estimated with the R-EPOLLS model using only *seismological parameters* such as earthquake magnitude, distance to seismic source, and severity of shaking at a site. However, for a given earthquake and site, a separate prediction of soil liquefaction and consequent lateral spreading must be made prior to using the R-EPOLLS model to estimate surface displacements.

Selection of regressor variables.

Candidate regressor variables for inclusion in the Regional-EPOLLS model are defined in Section 6.4 and Table A.2. These variables, compiled in the EPOLLS database, are:

EQ_Mw	*Focal_Depth*	*Hypoc_Dist*	*Accel_max*
EQ_Ms	*Epict_Dist*	*Fault_Dist*	*Duration*

Two measures of earthquake magnitude were compiled in the database, moment magnitude (*EQ_Mw*) and surface-wave magnitude (*EQ_Ms*). A simple correlation analysis showed a slightly stronger relationship between *Avg_Horz* and *EQ_Ms* in the EPOLLS data. However, the surface-wave magnitude is known to become "saturated" for very large earthquakes with M_w greater than about 8 (Kramer 1996). Hence, the surface-wave scale is a poor measure

of the seismic energy released in the 1964 Alaskan earthquake (M_w=9.2). When the fifteen Alaskan case studies are temporarily excluded, a significantly stronger correlation is observed between *Avg_Horz* and *EQ_Mw* than with *EQ_Ms*. Because the moment magnitude scale gives a more reliable measure of seismic energy, as well as a stronger correlation for M_w less than 9, *EQ_Ms* was dropped from further consideration in the Regional-EPOLLS model.

Next, consideration was given to the four compiled measures of distance to the earthquake source: focal depth (*Focal_Depth*), epicentral distance (*Epict_Dist*), hypocentral distance (*Hypoc_Dist*), and fault distance (*Fault_Dist*). Here, a simple analysis showed that *Fault_Dist* and *Focal_Depth* were more strongly correlated to *Avg_Horz* than either *Epict_Dist* or *Hypoc_Dist*. Hence, *Epict_Dist* and *Hypoc_Dist* were dropped from further consideration for the R-EPOLLS model. At this point, model selection procedures (described in Section D.5) suggested that the best R-EPOLLS model would include *EQ_Mw, Fault_Dist, Accel_max,* and *Duration*. Therefore, the regressor *Focal_Depth* was dropped from consideration.

Many empirical attenuation models assume an exponential relationship between ground motion parameters and distance to the earthquake source. That is, empirical attenuation equations are frequently written using logarithms of the source distance (for example, Equations 6.3 and 6.4). Similarly, empirical models given by Youd and Perkins (1987) and Bartlett and Youd (1995) for lateral spreading displacements are also written using logarithms of the source distance. Hence, in the next step of the development, transforms of *Fault_Dist* were considered for the R-EPOLLS model. Transformations considered were $\log_{10}$(*Fault_Dist*), ln(*Fault_Dist*), (*Fault_Dist*)$^{-1}$, (*Fault_Dist*)$^{-0.5}$, (*Fault_Dist*)$^{0.5}$, and (*Fault_Dist*)$^{0.33}$. The transformed source distance variables were all considered in place of *Fault_Dist* in a four-term model as well as an additional variable in a five-term model. The results of this analysis showed that transformations of the source distance were not needed: all transforms had a lower correlation with *Avg_Horz* than the untransformed *Fault_Dist*, and adding a transformed variable to a model already containing *Fault_Dist* produced a lower $\bar{R}^2$. Therefore, the R-EPOLLS model could be written simply in terms of *Fault_Dist*, and logarithmic or other transformations of the source distance were not considered further.

From physics, a body subjected to a constant acceleration will have a velocity equal to the acceleration multiplied by time, and the resulting displacement will be proportional to the acceleration multiplied by the square of time. Hence, the following product terms were considered for the R-EPOLLS model: (*Accel_max*Duration*) and (*Accel_max*Duration*^{2}). However, neither of these product terms were significant when added to the four-variable R-EPOLLS model.

Transformation of the response variable.

The analyses to this point, including formal variable selection procedures based on maximum $\bar{R}^2$, stepwise selection, and C_p plots (see Section D.5), indicated that the best Regional-EPOLLS model would be of the form:

$$Avg\_Horz = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration)$$

However, a residuals plot from this simple model suggested an increasing variance in the errors at larger values of the predicted *Avg_Horz*. That is, a funnel shape was observed in the residuals plot similar to that shown in Figure D.5b. This condition violates a fundamental assumption in the regression analysis and suggests the need to transform the response variable *Avg_Horz*. Following the suggestions of Montgomery and Peck (1992), three variance stabilizing transforms were considered for the R-EPOLLS model: $(Avg\_Horz)^{0.5}$, $\log_{10}(Avg\_Horz)$, and $\ln(Avg\_Horz)$. While each transform reduced the dispersion in the residuals, the square root transformation seemed to yield the most uniform variance of errors. Moreover, the square root transformation produced a slightly greater R^2 than the other options.

Thus, the final form of the model selected for the Regional-EPOLLS model is:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration) \qquad (9.1)$$

Each of the chosen regressor variables in this model are significant to the 99.5% level in partial *F*-tests. Additional variable selection analyses (maximum $\overline{R}^2$, stepwise selection, and C_p plots) confirmed that these four regressors would give the best R-EPOLLS model written in terms of $(Avg\_Horz)^{0.5}$.

Evaluation for influential observations.

Influential observations, as defined in Section D.7, include outliers, high leverage points, and highly influential observations. Data points of this type, which exert a disproportionally large influence on a regression model, can be detected using several diagnostic tests as outlined in Section D.7. Each of these tests were performed for the EPOLLS data and the Regional-EPOLLS model specified in Equation 9.1. Case studies that were consistently marked as influential observations were investigated for possible errors in the compiled data.

While various influence diagnostics indicated possible problems with about ten case studies, only six were consistently indicated by most of the diagnostic tests. Three EPOLLS case studies (Slide Nos. 43, 53, and 54) appear to be highly influential observations, but are fairly well-documented observations and were thus retained in fitting the R-EPOLLS model. Three other EPOLLS case studies (Slide Nos. 13, 101, and 118), which were consistently identified as highly influential observations, are poorly-documented case studies. That is, the average horizontal displacements for these lateral spreads were estimated from very little data. Since the fit of the R-EPOLLS model is unduly influenced by these observations containing unreliable data, Slide Nos. 13, 101, and 118 were not used in fitting the R-EPOLLS model.

Evaluation for multicollinearity.

Multicollinearity arises from linear dependencies among the regressor variables and can produce an unstable model in a multiple linear regression analysis. Multicollinearity and

associated diagnostic tests are discussed in Section D.6. All of these diagnostic tests were examined, including variance inflation factors and condition indices. For the R-EPOLLS model in Equation 9.1, no problems with multicollinearity were indicated.

In the model given by Equation 9.1, linear dependence and multicollinearity might be expected from the four selected regressor variables. Using other empirical equations, *Accel_max* and *Duration* can be estimated from *EQ_Mw* and *Fault_Dist*, meaning that these four variables are not independent of one another. Complete independence of the regressor variables and no multicollinearity would be ensured if magnitude and distance are not used in the same model with acceleration and duration. For example, multicollinearity problems are avoided in regression models for liquefaction or lateral spreading that are expressed either in terms of magnitude and source distance, or acceleration and duration, but not with all four variables (see Liao et al. 1988, and Bartlett and Youd 1992b).

However, multicollinearity results from the *linear* relationship between two or more variables in a regression model. Total independence among the regressor variables is not required. For example, x and x^2 are not linearly related, so using both x and x^2 in a regression model would not necessarily create a problem with multicollinearity. On the other hand, if the range of x is small, the correlation between x and x^2 could be nearly linear and create multicollinearity. Hence, it is important to check any model for possible multicollinearity in a multiple linear regression analysis. While the four regressor variables in Equation 9.1 (*EQ_Mw*, *Fault_Dist*, *Accel_max*, and *Duration*) are not completely independent of one another, the diagnostic tests do not indicate a multicollinearity problem in the regression model. Any linear dependence among these four variables is not sufficient to cause serious problems in the Regional-EPOLLS model.

Regional-EPOLLS component for Avg_Horz.

The selected R-EPOLLS model (Equation 9.1) was fit to 68 case studies (three case studies were discarded as influence data, *Avg_Horz* is unknown in seven cases). The data used to fit this model component are shown in Table 9.1. The final regression equation is:

$$\sqrt{Avg\_Horz} = 0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration - 2.21 \qquad (9.2)$$

In a global *F*-test, this regression is significant to a 99.99% level while each model variable is significant to a 99.95% level in partial *F*-tests. For the transformed model in Equation 9.2, R^2=0.514 and $\bar{R}^2$=0.483. A plot of the R-EPOLLS residuals (in terms of $Avg\_Horz^{0.5}$) are shown in Figure 9.1a. The model residuals appear to be uniformly dispersed about zero, indicating satisfactory model performance. Partial regression plots, described in Section D.8, are shown for each of the R-EPOLLS variables in Figure 9.2. These plots also demonstrate an acceptable model performance and no need for transformation of the regressor variables.

The regression analysis for the Regional-EPOLLS model is in terms of the square root of *Avg_Horz*. However, when the model is actually used, the predicted values must be squared to give the average horizontal displacement. As discussed in Section D.9, de-transformation of the response variable in this manner introduces a bias into the model. To compensate, the mean square error (MSE) can be used as a bias reduction factor as shown in Equation D.25. For the Regional-EPOLLS model, low-bias predictions of the average horizontal displacement are thus given by:

$$Avg\_Horz = (0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration - 2.21)^2 + 0.149 \qquad (9.3)$$

Performance of the Regional-EPOLLS model is examined further in Section 9.5.

9.2. Development of Site-EPOLLS Component.

The *Site-EPOLLS* (or *S-EPOLLS*) component is designed to use in conjunction with the Regional-EPOLLS model to give improved predictions when additional, site-specific data are available. In developing the Site-EPOLLS component, additional information on the topography and geometry of the lateral spread was added to the R-EPOLLS model described in Section 9.1.

Preservation of R-EPOLLS model coefficients.

The Site-EPOLLS model equation is formed by adding regressor variables to those already used in the Regional-EPOLLS model. That is, predictions from the Site-EPOLLS model will require *EQ_Mw, Fault-Dist, Accel_max, Duration*, and the additional variables chosen in this section. From a theoretical standpoint, the model coefficients on the seismological parameters (*EQ_Mw, Fault-Dist, Accel_max, Duration*) should be the same in all model components. In other words, the influence of *EQ_Mw* on *Avg_Horz* should be unchanged throughout the EPOLLS model, with the associated parameter equal to 0.613 as given in Equation 9.3. A complication in fitting the Site-EPOLLS is the absence of topographical and geometrical parameters from many EPOLLS case studies. While the seismological parameters used in the R-EPOLLS component are known for all 78 lateral spreads in the EPOLLS database, information on the topography and geometry of the slide area is not available for about one-fourth of the case studies. The design of the EPOLLS model, using complementary equations, and the significant number of case studies with missing data requires a somewhat unconventional approach to the regression analysis.

In fitting the Site-EPOLLS model, the coefficients on the R-EPOLLS variables are held constant (restricted) while the added model parameters are estimated. Restricting the four model

coefficients in this manner is equivalent to fitting the S-EPOLLS model to the residuals of the R-EPOLLS predictions ($R_{R\text{-}E}$). In equation form, the S-EPOLLS model is fit by regressing $R_{R\text{-}E}$ on the candidate parameters describing the geometry and topography:

$$R_{R\text{-}E} = \sqrt{Avg\_Horz} - (0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration)$$

$$= f(geometrical\ and\ topographical\ parameters) \qquad (9.4)$$

When Equation 9.4 is re-arranged, the fitted Site-EPOLLS model will be expressed in terms of the square root of *Avg_Horz* and retain the same coefficient values as used in the Regional-EPOLLS model component. Note that the intercept term (-2.21) in Equation 9.2 is not included in Equation 9.4 because this value will be incorporated into the intercept term of the fitted Site-EPOLLS model.

The primary advantage of using restricted coefficients to fit the Site-EPOLLS model (equivalent to fitting the residuals in Equation 9.4) is that the maximum possible number of case studies are used in estimating each model coefficient. While the coefficients in the Site-EPOLLS model are regressed on a smaller subset of the EPOLLS database (see Table 9.1), the coefficients on *EQ_Mw, Fault-Dist, Accel_max,* and *Duration* are based on the larger set of 68 lateral spreads. Hence, the completed Site-EPOLLS model is regressed on the maximum possible number of case studies and preserves the model coefficients previously estimated for the Regional-EPOLLS component.

Selection of regressor variables.

Candidate regressor variables for inclusion in the Site-EPOLLS model, representing the geometry and topography of the slide area, are defined in Sections 6.5 and 6.6 as well as Tables A.3 and A.4. These variables, compiled in the EPOLLS database, are:

Slide_Area	*Divergence*	*Face_Height*	*Eff_%Slope*
Slide_Length	*Direct_Slide*	*Top_%Slope*	

In addition, three category variables were recorded for each lateral spread:

Banded_Area?	*Free_Face?*	*Bulkhead?*

Note that the variable *Bot_%Slope* is considered only in the Geotechnical-EPOLLS model (see Section 9.3) because subsurface explorations are required to determine this value.

In the first step of the variable selection process, the variable *Slide_Area* was examined. The surface area of a lateral spread is undefined when *Banded_Area?* = TRUE (that is, sliding occurs in a band with poorly-defined flanks that are much farther apart than the head and toe). Consequently, *Slide_Area* is known in less than half of the EPOLLS case studies and is a poor choice for inclusion in the Site-EPOLLS model. Partial *F*-tests, performed on 27 case studies with the candidate S-EPOLLS parameters listed above, indicated that the area of the slide was

significant to only about a 40% level. Hence, the variable *Slide_Area* was dropped from further consideration for the Site-EPOLLS component.

Next, a variable selection process was used to identify the best parameters to be added in the Site-EPOLLS component. The automated selection procedures described in Section D.5 could not be used in this analysis with the restricted coefficient values carried forward from the R-EPOLLS model. Instead, a manual selection procedure was undertaken (the three category variables were not considered at this stage). Variable selection, following a modified maximum $\bar{R}^2$ procedure, proceeded as follows:

(1) Each of the candidate S-EPOLLS variables were individually added to the R-EPOLLS model and $\bar{R}^2$ was calculated for the resulting five-parameter models. The added regressor variable yielding the highest $\bar{R}^2$ value was added to the model.
(2) Each of the remaining S-EPOLLS variables were then individually added and $\bar{R}^2$ was calculated for the resulting six-parameter models. Again, the variable yielding the highest $\bar{R}^2$ was chosen.
(3) This procedure was repeated until no further improvement in $\bar{R}^2$ was obtained by adding any of the remaining variables.
(4) Partial *F*-tests were then used to check the significance of all variables in the model. Any variable found to be insignificant at this point was dropped and step (2) was repeated.
(5) Finally, the "best" model yields the maximum $\bar{R}^2$ with all variables significant in partial *F*-tests.

One problem with this manual selection procedure is that not all possible combinations of the candidate variables are considered. Hence, using engineering judgement, a few additional models were defined for further consideration. In the end, two good models were identified. The first candidate model was evaluated with 50 case studies:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration, Slide\_Length, Direct\_Slide) \quad (9.5)$$

for which R^2=0.641, $\bar{R}^2$=0.591, and *Slide_Length* and *Direct_Slide* are significant to the 85% level. The second candidate model, evaluated with 58 case studies, was of the form:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration, Slide\_Length, Top\_\%Slope, Face\_Height) \quad (9.6)$$

for which R^2=0.631, $\bar{R}^2$=0.580, and *Slide_Length, Top_%Slope,* and *Face_Height* are all significant to the 80% level. Comparing the models represented by Equations 9.5 and 9.6 is complicated by the different subsets of the database used to evaluate each. Equation 9.5 can be evaluated for only 50 case studies because the variable *Direct_Slide* is unknown for eight of the

58 cases studies used to evaluate Equation 9.6. In general, both models appear to be equally good. However, the surface slope and free face height (*Top_%Slope* and *Face_Height*) are much easier to determine that the direction of sliding (*Direct_Slide*) when evaluating a potential lateral spread. For this reason, Equation 9.6 was chosen for the Site-EPOLLS model component.

Next, a transformation of the variable *Top_%Slope* was considered. Lateral spreading usually occurs on sites with relatively long, fairly uniform, mild slopes that can be approximated as an infinite slope. Hence, the shear stresses in the liquefiable soil underlying a lateral spread will be approximately proportional to the variable *Tau_Slope* defined as:

$$Tau\_Slope = \cos(\theta)\cdot\sin(\theta); \quad \theta = \tan^{-1}(Top\_\%Slope/100) \tag{9.7}$$

Substitution of *Tau_Slope* for *Top_%Slope* in Equation 9.6 produced no discernible improvement in the model performance. This is not unexpected since, for small angles, *Tau_Slope* is approximately equal to the surface slope. A similar transformation of the variable *Eff_%Slope* also failed to produce any improvement.

Finally, the significance of the three category variables *Banded_Area?, Free_Face?*, and *Bulkhead?* were investigated when added to the model of Equation 9.6. The variable *Banded_Area?*, discussed more completely in Section 6.5, was significant to only a 25% level in a partial *F*-test and was therefore dropped. This can be interpreted to mean that the banded-area case studies (*Banded_Area?*=TRUE) are not significantly different than those with a bounded surface area (*Banded_Area?*=FALSE). Similarly, the category variable *Free_Face?* was significant to only a 38% level when added to Equation 9.6. The absence of a free face along the toe of a slide is already represented in the Site-EPOLLS model with *Face_Height* equal to zero. Hence, the lack of significance for the variable *Free_Face?* is not surprising. On the other hand, the category variable *Bulkhead?* appeared to be significant when added to the model. Closer inspection of the database showed that nine of twelve lateral spreads with a flexible bulkhead along the free face were located on the banks of the Shinano River in Niigata, Japan. Therefore, it appears that the category *Bulkhead?* mostly represents the Niigata case studies. Without more data from other sites, *Bulkhead?* is probably a poor indicator of the possible influence of a flexible bulkhead and was not included in the S-EPOLLS model. The possible influence of flexible bulkheads on lateral spreads in the EPOLLS database is discussed further in Section 9.4. Because none of the category variables were included in the final model, the interaction terms suggested in Section D.3 are unnecessary.

Evaluation for influential observations.

Having selected the model form in Equation 9.6, the next step in the regression analysis is to check for influential observations using the statistical tests described in Section D.7. Recall that three case studies (Slide Nos. 13, 101, and 118) were identified earlier as influential observations and not used to fit the Regional-EPOLLS model. All three of these case studies are

missing at least one of the regressor values selected for the Site-EPOLLS model and are therefore not available for fitting this model component.

Diagnostic tests for the S-EPOLLS model consistently indicated that only Slide No. 53 was a high influence observation. Located in San Francisco, this site was subjected to liquefaction in both the 1906 and 1989 earthquakes. While significant lateral spreading occurred in 1906, only very small displacements occurred in 1989 for Slide No. 53. Consequently, it is difficult to delineate the limits of the slide area in 1989 and the value of *Slide_Length* may be questionable for this case study. Therefore, Slide No. 53 was dropped from the fit of the S-EPOLLS model. Subsequent diagnostic tests confirmed that the remaining case studies were not adversely affecting the fit of the Site-EPOLLS model component.

Evaluation for multicollinearity.

The selected Site-EPOLLS model (Equation 9.6) was now checked for possible multicollinearity using the tests described in Section D.6. As was the case for the Regional-EPOLLS model, none of these tests (including variance inflation factors and condition indices) showed any signs of significant problems with multicollinearity.

Site-EPOLLS component for Avg_Horz.

The selected Site-EPOLLS model, as specified in Equation 9.6, was fit to the 57 case studies indicated in Table 9.1. As discussed above, only one case study was discarded from the 58 available cases with no missing regressor values. The model coefficients on the four regressors from the Regional-EPOLLS model were unchanged, and the fitted regression equation is:

$$\sqrt{Avg\_Horz} = 0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration$$

$$+ 0.000523\,Slide\_Length + 0.0423\,Top\_\%Slope + 0.0313\,Face\_Height - 2.44 \quad (9.8)$$

In a global *F*-test, this regression is significant to a 99.99% level while the three additional model variables are each significant to greater than 82% in partial *F*-tests. For the transformed model in Equation 9.8, R^2=0.656 and $\overline{R}^2$=0.606. A plot of the Site-EPOLLS residuals (in terms of $Avg\_Horz^{0.5}$) are shown in Figure 9.1b. The S-EPOLLS residuals, which appear to be uniformly dispersed about zero, indicate satisfactory performance. Partial regression plots are shown for the three S-EPOLLS variables in Figure 9.3. These plots also indicate an acceptable model performance and no need for transformation of the regressor variables. On the other hand, the scatter in the partial residuals of *Top_%Slope* signify the relatively weak significance of this variable, apparently resulting from the narrow range in values for the lateral spreads in the EPOLLS database.

The square root of *Avg_Horz* is again used in the regression analysis for the Site-EPOLLS model. As with the Regional-EPOLLS component, a bias results when de-transforming the model predictions. This bias can be reduced using the mean square error (MSE) of the transformed model. For the Site-EPOLLS model, low-bias predictions of the average horizontal displacement are given by:

$$Avg\_Horz = (0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration$$

$$+ 0.000523\,Slide\_Length + 0.0423\,Top\_\%Slope + 0.0313\,Face\_Height - 2.44)^2 + 0.111 \qquad (9.9)$$

Performance of the Site-EPOLLS model is examined further in Section 9.5.

9.3. Development of Geotechnical-EPOLLS Component.

The *Geotechnical-EPOLLS* (or *G-EPOLLS*) component is designed to give more refined predictions of horizontal displacements in site-specific studies where additional data from soil borings are available. In the Geotechnical-EPOLLS model, parameters describing the site geology are added to the R-EPOLLS and S-EPOLLS components defined in Sections 9.2 and 9.3.

Preservation of R-EPOLLS and S-EPOLLS model coefficients.

Because the G-EPOLLS model is used only in conjunction with the previous two components, the model coefficients from the R-EPOLLS and S-EPOLLS equations are held constant in the G-EPOLLS component. Designing the Geotechnical-EPOLLS model in this way allows for a regression based on the maximum possible amount of data. No subsurface data is available for nineteen of the EPOLLS case studies, and several of the candidate geotechnical parameters are unknown in many others. This reasoning is the same as given in Section 9.2 for the Site-EPOLLS model.

Hence, in fitting the G-EPOLLS model, the coefficients on the R-EPOLLS and S-EPOLLS variables are held constant (restricted) while the added model parameters are estimated. In the same manner as for the Site-EPOLLS model in Section 9.2, restricting the model coefficients in this manner is equivalent to fitting a model to the residuals of the S-EPOLLS predictions ($R_{S\text{-}E}$). In equation form, the G-EPOLLS model is fit by regressing $R_{S\text{-}E}$ on the candidate geotechnical parameters:

$$R_{S-E} = \sqrt{Avg\_Horz} - (0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Durat$$

$$+ 0.000523\,Slide\_Length + 0.0423\,Top\_\%Slope + 0.0313\,Face\_Height)$$

$$= f(geotechnical\ parameters) \qquad (9.10)$$

Selection of regressor variables.

Candidate regressor variables for inclusion in the Geotechnical-EPOLLS model are defined in Section 6.7 and Table A.5. Recall that two values of each geotechnical variable, representing the average and range of subsurface soil conditions, are compiled for each lateral spread in the EPOLLS database. For example, *Avg-Z_GrWater* is the average depth to ground water at the site while *Rng-Z_GrWater* is the range (maximum minus minimum) in depth to ground water. For the model of the average horizontal displacement, only the *Avg-* geotechnical regressors were considered. The potential regressor variables for the G-EPOLLS model component are:

Avg-Z_GrWater	*Avg-FS_Min*	*Avg-D50_Liq*
Avg-Z_TopLiq	*Avg-N160_MnFS*	*Avg-Cu_Liq*
Avg-Z_BotLiq	*Avg-Z_MnFS*	*Avg-Clay_Liq*
Avg-Thick_Liq	*Avg-N160_Min*	*Avg-Fine_Liq*
Avg-Index_Liq	*Avg-Z_MnN160*	*Avg-N160_Cap*
Avg-FS_Liq		*Avg-Fine_Cap*
Avg-N160_Liq		*Avg-Fine_Base*

In addition, the variable *Bot_%Slope*, which can be defined only with subsurface information, is considered a potential regressor for the G-EPOLLS model.

Miyajima and his associates (1991) suggested that displacements in a lateral spread might be predicted using the velocity of movements multiplied by time. Model tests performed by Miyajima et al. (1991) have shown a relationship between the duration of ground movements and the thickness of the liquefied soil beneath. Also, as discussed in Section 3.6, a correlation between velocity of the surface movements and the surface slope was observed in these model tests. Therefore, an additional parameter considered for the G-EPOLLS model is defined as:

$$Slope\_Thick = Top\_\%Slope * Avg\text{-}Thick\_Liq \qquad (9.11)$$

Several of the candidate geotechnical parameters actually represent different ways of measuring the same factor. In the first step of the variable selection for the G-EPOLLS model,

the "best" (statistically most significant) of these variables were identified and the other redundant variables were dropped from consideration. For example, *Avg-Index_Liq*, *Avg-FS_Liq,* and *Avg-N160_Liq* all represent the average shear strength in the liquefied soil deposit. When added separately to the S-EPOLLS model, *Avg-Index_Liq* is a more significant variable than either *Avg-FS_Liq* or *Avg-N160_Liq*, so both were dropped from further consideration. Similarly, *Avg-FS_Min* and *Avg-Z_MnFS* were chosen for the minimum strength and depth to the minimum strength, respectively, in the liquefied soil while the redundant variables *Avg-N160_MnFS, Avg-N160_Min*, and *Avg-Z_MnN160* were dropped.

Unfortunately, the variable *Bot_%Slope* is unknown in about half of the EPOLLS case studies where the other candidate regressors are defined. While it might be an important parameter, *Bot_%Slope* is defined in too few of the EPOLLS case studies to properly judge its relative significance when selecting regressor variables for the G-EPOLLS model. Similarly, *Avg-D50_Liq, Avg-Cu_Liq,* and *Avg-Clay_Liq* were all dropped from consideration because these parameters are known for only about one-fourth of the EPOLLS case studies. Hence, the only variable retained at this point for representing the gradation of the liquefied soil deposit is *Avg-Fine_Liq*.

Variable selection for the Geotechnical-EPOLLS model component now progressed with twelve candidate regressor variables:

Avg-Z_GrWater	*Avg-Index_Liq*	*Avg-N160_Cap*
Avg-Z_TopLiq	*Avg-FS_Min*	*Avg-Fine_Cap*
Avg-Z_BotLiq	*Avg-Z_MnFS*	*Avg-Fine_Base*
Avg-Thick_Liq	*Avg-Fine_Liq*	*Slope_Thick*

The modified maximum $\bar{R}^2$ procedure, described in Section 9.2, and engineering judgment were used to identify good combinations of these regressors for the G-EPOLLS model.

Three candidate models were thus identified for the G-EPOLLS component. The first candidate model was evaluated with 40 case studies:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration, Slide\_Length, Top\_\%Slope,$$

$$Face\_Height, Avg\text{-}Z\_GrWater, Avg\text{-}Z\_MnFS, Avg\text{-}Fine\_Base) \qquad (9.12)$$

for which R^2=0.734, $\bar{R}^2$=0.642, and all three of the added variables are significant to the 90% level in partial *F*-tests. This model gives the highest $\bar{R}^2$ value, but this is somewhat misleading because *Avg-Fine_Base* is unknown in five case studies which could otherwise be used to fit the G-EPOLLS model. Moreover, the average fines content of the soil below the liquefied deposit is more difficult to determine, relative to the other regressors in the candidate models given below. Hence, Equation 9.12 was not selected for the Geotechnical-EPOLLS model.

The second candidate model, evaluated with 45 case studies, fits the data almost as well as the model in Equation 9.12. This second candidate for the G-EPOLLS model is of the form:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration, Slide\_Length, Top\_\%Slope, Face\_Height, Avg\text{-}Z\_MnFS, Avg\text{-}Z\_TopLiq) \quad (9.13)$$

for which R^2=0.700, $\overline{R}^2$=0.623, and the two added variables are significant to the 90% level.

The third candidate model also yields a good fit to 45 case studies with R^2=0.693 and $\overline{R}^2$=0.613. This model is of the form:

$$\sqrt{Avg\_Horz} = f(EQ\_Mw, Fault\_Dist, Accel\_max, Duration, Slide\_Length, Top\_\%Slope, Face\_Height, Avg\text{-}Z\_GrWater, Avg\text{-}Z\_BotLiq) \quad (9.14)$$

However, the two added variables in Equation 9.14 are significant to only the 76% level. Therefore, Equation 9.13 is preferred over the model in Equation 9.14.

The final choice for the G-EPOLLS regression model is represented by Equation 9.13. That is, *Avg-Z_MnFS* and *Avg-Z_TopLiq* are added to the Site-EPOLLS model to form the Geotechnical-EPOLLS component. The two selected regressor variables have an additional advantage in that both are not dependent on any particular in situ test. Since the depths to the top of the liquefied soil and to the minimum factor of safety can be measured by any suitable means, such as a cone penetrometer test, the G-EPOLLS model does not require Standard Penetration Test data.

Evaluation for influential observations.

Having chosen Equation 9.13 as the form of the Geotechnical-EPOLLS model, the next step is to check for possible influence data. Recall that four case studies (Slide Nos. 13, 53, 101, and 118) were dropped as high influence observations in fitting the R-EPOLLS and S-EPOLLS components. Each of these case studies are missing at least one of the required geotechnical variables; therefore, data from these four lateral spreads are not used in fitting the G-EPOLLS model.

Diagnostic tests indicated that two lateral spreads were highly influential. First, Slide No. 43 is a case study with seven soil borings available. Since this is a relatively well-documented case study, Slide No. 43 was retained in fitting the G-EPOLLS model. The second case study identified as a high influence observation was Slide No. 8. Here, the subsurface geology is not

well characterized, so this case study was discarded before fitting the final Geotechnical-EPOLLS model.

Evaluation for multicollinearity.

As done for the other components of the EPOLLS model for average horizontal displacements, the G-EPOLLS model specified in Equation 9.13 was evaluated for possible multicollinearity. Again, diagnostic tests including variance inflation factors and condition indices did not indicate any problems due to multicollinearity.

Geotechnical-EPOLLS component for Avg_Horz.

The selected Geotechnical-EPOLLS model (Equation 9.13) was fit to 44 case studies as indicated in Table 9.1. Only one case study was discarded as a high influence observation from the 45 available cases with no missing regressor values. The model coefficients on the seven regressors from the Regional-EPOLLS and Site-EPOLLS models were unchanged, and the fitted G-EPOLLS regression equation is:

$$\sqrt{Avg\_Horz} = 0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration$$

$$+ 0.000523\,Slide\_Length + 0.0423\,Top\_\%Slope + 0.0313\,Face\_Height$$

$$+ 0.0506\,Avg\text{-}Z\_MnFS - 0.0861\,Avg\text{-}Z\_TopLiq - 2.49 \qquad (9.15)$$

In a global *F*-test, this regression is significant to a 99.99% level while the two additional model variables are significant to a 92% level in partial *F*-tests. For the transformed model in Equation 9.15, R^2=0.707 and $\overline{R}^2$=0.629. A plot of the Geotechnical-EPOLLS residuals (in terms of $Avg\_Horz^{0.5}$) are shown in Figure 9.1c. As with the previous model components, the G-EPOLLS residuals appear to be uniformly dispersed about zero indicating satisfactory performance. Partial regression plots are shown for the two G-EPOLLS variables in Figure 9.4. These plots also indicate an acceptable model performance and no need for transformation of the regressor variables.

When the square root of *Avg_Horz* is de-transformed in the G-EPOLLS model, a known bias results as discussed in Section D.9. As with the Regional-EPOLLS and Site-EPOLLS components, this bias can be reduced using the mean square error (MSE) of the transformed model. For the Geotechnical-EPOLLS model, low-bias predictions of the average horizontal displacement are given by:

$$Avg\_Horz = (0.613\,EQ\_Mw - 0.0139\,Fault\_Dist - 2.42\,Accel\_max - 0.0114\,Duration$$

$$+ 0.000523\,Slide\_Length + 0.0423\,Top\_\%Slope + 0.0313\,Face\_Height$$

$$+ 0.0506\,Avg\text{-}Z\_MnFS - 0.0861\,Avg\text{-}Z\_TopLiq - 2.49)^2 + 0.124 \qquad (9.16)$$

Performance of the Geotechnical-EPOLLS model is examined further in Section 9.5.

9.4. Additional Issues Concerning Model Specification.

Lack of data from less damaging slides.

A fundamental problem with empirical models like the EPOLLS model arises out of the dependence on observational data. In post-earthquake damage reconnaissance and subsequent site studies, greater attention will naturally focus on the most severe ground failures. Sites with very small surface displacements are more likely to go undetected and unreported. For this reason, sites subjected to soil liquefaction but with little or no lateral movements are probably under-represented in the published literature and EPOLLS database. Because the EPOLLS database is probably prejudiced towards more damaging ground failures, the EPOLLS model is likely to conservatively over-predict the displacements in a small-deformation lateral spread. Another way of viewing this problem is illustrated conceptually in Figure 9.5. Here, displacement is plotted against some hypothetical variable ψ that represents a combination of all significant site parameters. Data that is "missing" from the EPOLLS database, because sites with small deformations are disproportionally investigated, would fall in the shaded region in the lower left of this plot. For these marginal lateral spreads, a model regressed on the available data may over-predict the surface displacements.

One possible way to compensate for this problem would be to specify a bi-linear spline model, as shown in Figure 9.5, with a break at ψ^* (called a knot). If the knots can be located, a spline model can be fit with linear regression procedures (Freund and Littell 1991). For the EPOLLS model, knots would probably be needed for each regression variable. However, without the data from the small deformation events that are "missing", appropriate values for the knots are difficult to determine. There appears to be no logical manner of choosing such values without the missing data to provide guidance. Also note that a linear regression model forced to pass through zero (no intercept) would be less accurate for the large-deformation slides. Hence, while the EPOLLS model might be conservatively biased toward more damaging lateral spreads, there seems to be no clear way to quantify and compensate for the deficiency in the available site data.

On the other hand, numerous sites that experienced relatively small deformations are included in the EPOLLS database. At two sites in San Francisco subjected to lateral spreading in 1906 (EPOLLS Slide Nos. 1 and 2), no significant average horizontal movement occurred in the 1989 earthquake (Slide Nos. 53 and 54), although evidence of liquefaction was observed. Histograms of the average horizontal and vertical displacements in the EPOLLS data set are shown in Figure 6.9. These plots indicate that the majority of the EPOLLS case studies experienced less than 1.6 m of average horizontal movement and less than 0.7 m of average settlement (eleven cases had less than 0.2 m of average horizontal movement). Because this data was used to fit the EPOLLS model, concerns over a potential model bias resulting from a lack of data on less damaging slides are somewhat alleviated.

Sites with flexible bulkheads.

In compiling the EPOLLS database, lateral deformations behind relatively large, stiff walls were excluded as not conforming to the definition of lateral spreading adopted in Section 3.1. However, some sites with relatively "flexible" bulkhead walls (usually steel sheet piles) were included in the EPOLLS database. These case studies are identified with *Bulkhead?* = TRUE, but the decision over whether a wall is "stiff" or "flexible" is subjective. Hence, the possibility of significant restraint provided by a bulkhead wall was investigated further.

As discussed previously in Section 9.2, the category variable *Bulkhead?* was not included in the EPOLLS model. Some statistical tests indicated that *Bulkhead?* might be a significant category variable in the Site-EPOLLS component. However, because most EPOLLS case studies with *Bulkhead?* = TRUE were located in Niigata, Japan, these statistical tests are not a reliable indication of influence and *Bulkhead?* was not included in the model. Moreover, for all twelve of the EPOLLS case studies with a flexible bulkhead, the face height is, on average, only 1.8% of the slide length (at the maximum, the face height is 3.6% of the slide length). Given that the slide area is considerably larger than the face height, it is understandable that a flexible wall probably does not impart a significant restraint to the movement of the slide mass.

Influence of gradation in the liquefied soil deposit.

In fitting their empirical model for lateral spreads, Bartlett and Youd (1992a; 1992b; 1995) excluded data from lateral spreads that were underlain by liquefied, gravelly soils (defined as soils with D_{50} greater than 2 mm). In Bartlett and Youd's analysis, data from five lateral spreads in Alaska and Idaho with gravelly soils appeared to be outliers. These same sites are present in the EPOLLS database. However, the influence diagnostics performed during development of all three EPOLLS model components gave no indication that these case studies were outliers or high influence observations. Hence, in the EPOLLS data, the influence of gravel in the liquefiable soil deposit does not appear to have a significant impact on the behavior of a lateral spread.

Overall, the gradation of the liquefiable soil does not appear in any of the EPOLLS model components. This should not be interpreted to mean that fines or gravel play no role in the

ultimate deformation of a lateral spread. Instead, one can only say that the EPOLLS database does not support the inclusion of soil gradation parameters or the exclusion of sites with gravelly soils as outliers. Additional data from more sites with liquefiable soils having different grain size characteristics might lead to different conclusions.

Signs of model coefficients.

Examination of the regressed equation for the Regional-EPOLLS model (Equation 9.3) shows a positive coefficient on *EQ_Mw* and a negative coefficient on *Fault_Dist*. The signs on these coefficients are consistent with expected behavior; namely, displacements should be greater in larger-magnitude earthquakes and smaller at greater distances from the fault rupture. Similar conclusions can be reached when examining the coefficients in the Site-EPOLLS and Geotechnical-EPOLLS model components (Equations 9.9 and 9.16). That is, displacements should be greater with a longer slide area, greater surface slope, higher free face, and greater depth to the minimum FS_{liq} (all associated with positive coefficients) but smaller displacements would be expected with increases in the depth to the top of liquefied soil (negative coefficient).

By a similar argument, positive coefficients on *Accel_max* and *Duration* would be expected, indicating greater displacements with stronger seismic shaking at a site. However, as can be seen in Equation 9.3, both parameters are associated with negative coefficients. Coefficients with the "wrong" sign occur fairly often in regression analyses. Montgomery and Peck (1992) list four common causes:

(1) Computational errors may have occurred in the regression analysis, often due to round-off and truncation errors. Following the recommendations of Freund and Littell (1991), output from the computer code used in the EPOLLS regression analysis was checked to confirm that severe roundoff had not occurred.

(2) Multicollinearity may exist. As discussed in Section 9.1, the R-EPOLLS model was investigated for possible problems with multicollinearity and no evidence was detected.

(3) Range of the regressor data is too small. The uncertainty in an estimated parameter value is greater when the range of the associated regressor data is small. Consequently, when the true value of a model parameter is a small positive number, a regression estimate of that parameter may be a small negative number due to uncertainty in the estimate. Unfortunately, the range of the regressor data can not be increased in the EPOLLS database without more lateral spread case studies.

(4) Important regressors have not been included in the model. This is an obvious problem with the R-EPOLLS model, which is designed to allow approximate predictions with only data on the earthquake source and the resulting ground motions at a site.

The negative coefficients on *Accel_max* and *Duration* in the EPOLLS model probably result from the small range of regressor values in the database and from missing site parameters. Regardless of the underlying cause, the performance of the fitted model does not appear to be adversely affected (see Section 9.5).

Inclusion of Slide_Length in Site-EPOLLS.

Glaser (1993; 1994) points out that a prediction model should not employ the areal extent of a lateral spread because the surface area is difficult to estimate for a future event. Reluctantly, the variable *Slide_Length* was included in the Site-EPOLLS model component. In repeated trials, *Slide_Length* consistently appeared as a significant regressor in models containing many possible combinations of variables. Clearly, the EPOLLS data set strongly supports the inclusion of the variable *Slide_Length* to get the best possible model. As a result, predictions of average horizontal displacements with either the Site-EPOLLS or Geotechnical-EPOLLS components will require predictions of the length of the slide mass.

The *Slide_Length* depends on the geology and topography of a site. The head and toe of the slide may coincide with breaks in the topography. For example, ground displacements will probably progress in opposite directions in separate slide masses to either side of a low ridge. Extensive lateral spreading on both sides of a stream will converge toward the stream, and the toe of both slides can be assumed to meet at the centerline of the stream. In addition, the edge of a liquefiable deposit will roughly define the boundaries of a lateral spread, with consideration for the possibility of marginal slumping (see Figure 3.6). The lateral contact between young, liquefiable alluvium and older, stable sediments can also be used to define the areal extent of a possible lateral spread. In general, geologic and topographic site data should be sufficient to at least bracket estimates of the possible length of sliding in a lateral spread.

9.5. Performance of EPOLLS Model for Average Horizontal Displacement.

Overview of final model equations.

The final three components of the EPOLLS model for predicting average horizontal displacements are given in Equations 9.3, 9.9, and 9.16. Here, to allow for a simpler presentation of these final equations, new notation is introduced to replace the database field names for all nine of the chosen regressor variables. Each of the variables used in the EPOLLS model for predicting average horizontal displacements are defined in Table 9.2. Based on these variables, three new terms are defined:

$$D_R = (613 M_w - 13.9 R_f - 2420 A_{\max} - 11.4 T_d) / 1000. \qquad (9.17)$$

$$D_S = (0.523 L_{slide} + 42.3 S_{top} + 31.3 H_{face}) / 1000. \qquad (9.18)$$

$$D_G = (50.6 Z_{FSmin} - 86.1 Z_{liq}) / 1000. \qquad (9.19)$$

Now, using these definitions, the three components of the EPOLLS model are expressed in the

following equations:

- **Regional-EPOLLS:**

$$Avg\_Horz = (D_R - 2.21)^2 + 0.149 \quad (9.20)$$

- **Site-EPOLLS:**

$$Avg\_Horz = (D_R + D_S - 2.44)^2 + 0.111 \quad (9.21)$$

- **Geotechnical-EPOLLS:**

$$Avg\_Horz = (D_R + D_S + D_G - 2.49)^2 + 0.124 \quad (9.22)$$

Because the final equations are written with squared terms, the EPOLLS model can not predict a zero average displacement. This is consistent with a basic premise of the method: liquefaction and lateral spreading must be expected at a site prior to using the EPOLLS model to estimate surface displacements. Hence, the model should always predict some horizontal deformation greater than zero. From inspection of the above equations for the Regional-, Site-, and Geotechnical-EPOLLS models, the smallest average displacement that can be predicted is, respectively, 0.149, 0.111, and 0.124 meters. The smallest possible predictions correspond to D_R=2.21, (D_R+D_S)=2.44, and (D_R+D_S+D_G)=2.49. At lesser values, minimum EPOLLS model predictions should be assumed. For example, if (D_R+D_S)=2.10, a value of (D_R+D_S)=2.44 should be used in Equation 9.21 to get a minimal predicted *Avg_Horz* = 0.111 m. On the other hand, predictions at these minimum values are not reliable because they involve extrapolation beyond the range of the data used to fit the model. Valid limits for these factors are discussed further in Section 9.6.

Comparison of the quality of fit for the EPOLLS model components.

In Sections 9.1 through 9.3, the quality of fit for the selected model components was expressed in terms of the coefficient of multiple determination (R^2) and the adjusted coefficient of multiple determination ($\overline{R}^2$). These previously reported values, used during model development, were computed without the bias reduction factors and without the high influence data excluded from fitting the model. In this section, R^2 and $\overline{R}^2$ are computed for each model component, specified in Equations 9.20 through 9.22, that incorporates the bias reduction factors. Furthermore, to more rigorously test the performance of the model in predicting the field case studies, the quality of fit is based on all available data including the five high-influence case studies that were not used to fit the model. The total number of EPOLLS case studies, with all values required for a given model component, are indicated in the top of Table 9.3. For example, while only 44 case studies were used to fit the Geotechnical-EPOLLS component, all nine

variables required to use the G-EPOLLS model are known in 45 case studies.

For the 71 case studies with no missing values, the Regional-EPOLLS model (Equation 9.20) yields an R^2=0.537 and $\overline{R}^2$=0.509 as shown in Table 9.3. For the Site-EPOLLS model, R^2=0.710 and $\overline{R}^2$=0.670 for the 58 case studies with no missing data. To compare the performance of the Site-EPOLLS and Regional-EPOLLS models, the coefficient of multiple determination should be calculated on the same subset of the database. Hence, as shown in Table 9.3, for the same 58 case studies, the Regional-EPOLLS model gives R^2=0.622 and $\overline{R}^2$=0.594. Finally, for the 45 case studies with all nine variables known, the R-EPOLLS, S-EPOLLS, and G-EPOLLS models yield, respectively, R^2=0.641, 0.731, 0.752 and $\overline{R}^2$=0.605, 0.680, 0.688. By comparing the values in each row of Table 9.3, the improvement gained with each component of the EPOLLS model can be judged.

For the three EPOLLS models, the best indications of the quality of fit are given in the shaded box at the bottom of Table 9.3. Here, the R^2 and $\overline{R}^2$ values are computed for the largest possible number of case studies with Equations 9.20 through 9.22. Respectively, the Regional-EPOLLS, Site-EPOLLS, and Geotechnical-EPOLLS for average horizontal displacements, yield $\overline{R}^2$=0.509, 0.670, and 0.688 for the case studies in the EPOLLS database.

A graphical comparison among the three EPOLLS model can be made with a scatter plot of the predicted versus observed *Avg_Horz* as discussed in Section D.8. These scatter plots are shown in Figure 9.6 for the R-EPOLLS, S-EPOLLS, and G-EPOLLS models specified in Equations 9.20 through 9.22. Recall that a "perfect" prediction would produce points on a 45° line in these plots and that R^2 is a numerical measure of the scatter about this line. Comparing Figures 9.6a and 9.6b, the improvement in predictions from the Site-EPOLLS model over those from the Regional-EPOLLS model can be seen in the reduced scatter in the plotted points (corresponding to a higher R^2). The improvement in predictions from the Geotechnical-EPOLLS model over those from the Site-EPOLLS model is small, but measurable in the reduced scatter in Figure 9.6c and the slightly higher R^2. Comparing Figures 9.6b and 9.6c, it appears that the G-EPOLLS model produces the most improvement in predictions of *Avg_Horz* greater than about 3.5 m. On the other hand, all three models appear to consistently over-predict displacements when *Avg_Horz* is less than about one meter, although these errors are less in the S-EPOLLS and G-EPOLLS models.

The residuals or errors (observed minus predicted *Avg_Horz*) indicate the precision of the model predictions. For the Regional-EPOLLS model, the residual in the predicted *Avg_Horz* is less than 1.0 m in about three-quarters (55 out of 71) of the lateral spread case studies. The precision in the Site-EPOLLS model is significantly better, with the residual less than 0.75 m in about three-quarters (44 out of 58) of the cases. By the same criteria, however, the residuals from the Geotechnical-EPOLLS model are less than 0.75 m in less than three-quarters (32 out of 45) of the cases. Comparing the model residuals in this manner is complicated by the different case

studies in each subset. Still, in very general terms, predictions of the average horizontal displacement from the R-EPOLLS model appear to be correct to within about 1.0 m in about three-quarters of the lateral spreads studied. Similarly, predictions of *Avg_Horz* from the S-EPOLLS and G-EPOLLS models appear to be correct to within ± 0.75 m in about three-quarters of the case studies.

Double cross-validation using random split of database.

The 78 case studies in the EPOLLS database were randomly divided into two subsets, with 39 cases in each, as indicated in the top half of Table 9.4. A computer algorithm was used to ensure that the individual case studies were randomly assigned to the two subsets. The equally divided subsets were then used to perform a double cross-validation as described in Section D.8. The results of this analysis, expressed in terms of a prediction R^2_p, are given in the lower half of Table 9.4. As expected, the R^2_p values based on half of the available data are less than the corresponding R^2 values for models fit to the full database. More importantly, the values of $R^2_{p,A}$ and $R^2_{p,B}$ are approximately equal for each model component. These positive results signify a good performance of all three model equations in predicting average horizontal displacements.

Comparison with Bartlett and Youd's MLR model.

Other than the EPOLLS model, probably the best empirical model for predicting horizontal displacements on a liquefaction-induced lateral spread is given by Bartlett and Youd (1995). Discussed in Section 4.4, Bartlett and Youd's MLR model consists of Equations 4.8 and 4.9. A logical question to investigate is how well predictions from the EPOLLS model compare with those from Bartlett and Youd's model.

Unfortunately, the two models require different site parameters or similar parameters that are defined differently (for example, the average slope of the slide area versus the surface slope at a specific location). Because of differences in the model parameters, the best way to compare the two models is to predict displacements at actual sites where all of the necessary values are known. In addition, Bartlett and Youd's model predicts displacements at specific locations on a lateral spread while the EPOLLS model predicts an average displacement for the entire slide area. Hence, EPOLLS model predictions of *Avg_Horz* are compared with an average of all displacements, within the same slide boundaries, predicted with Bartlett and Youd's MLR model.

To avoid compiling a new database of parameters for Bartlett and Youd's MLR model, the data used in fitting their model and published in Bartlett and Youd (1992b) were used in this comparison analysis. Data from several of the lateral spreads studied by Bartlett and Youd are also case studies in the EPOLLS database. The first task, then, is to find all of the case study displacement vectors in Bartlett and Youd's database that are located within the boundaries of an EPOLLS lateral spread. Unfortunately, the displacement vectors in Niigata, Japan, are not sufficiently documented by Bartlett and Youd (1992b) to positively locate their position. Consequently, none of the large amount of data from Niigata could be used in this comparison.

In the end, only fifteen lateral spreads were found where locations in Bartlett and Youd's data could be matched with EPOLLS case studies. Furthermore, at nine of these sites, data was available for predicting only one displacement vector with Bartlett and Youd's model; this was deemed insufficient to represent an "average" displacement suitable for comparison with the EPOLLS model.

Consequently, a comparison between these two models was made using data from just six lateral spreads. Because Bartlett and Youd's MLR model requires data from soil borings, these comparisons were made using the Geotechnical-EPOLLS model which also uses data from site soil borings. The results of this analysis are given in Table 9.5, where the case studies are arranged in order of increasing observed deformation. In the four lateral spreads with an average horizontal displacement of about one meter or less, the G-EPOLLS model predicted a larger average displacement than Bartlett and Youd's MLR model. However, in the three cases with the smallest observed movements, this comparison is based on six or fewer predictions from the MLR model. On the other hand, the G-EPOLLS model predicted much smaller displacements for the two lateral spreads with the largest movements.

This comparison with six case studies suggests that the EPOLLS model might predict larger average horizontal displacements for lateral spreads undergoing relatively small deformations, but smaller displacements in more damaging lateral spreads, than Bartlett and Youd's MLR model. However, *there is insufficient data to draw any firm conclusions from the comparison of the two models presented here.*

Performance at very large distances.

Based on data from shallow earthquakes around the world, Ambraseys (1988) proposed a maximum limit for the farthest distance from a fault rupture to an observed liquefaction event. Plotted in the top of Figure 9.7, Ambraseys' equation can be written as:

$$M_w = 4.68 + 0.0092(R_f)_{max} + 0.90\log(R_f)_{max} \qquad (9.23)$$

where M_w is the moment magnitude of a shallow earthquake and $(R_f)_{max}$ is the distance (km) measured from the fault rupture to the most distant liquefaction-induced ground failure.

In deriving their empirical model for lateral spreading, Bartlett and Youd (1992b) used data from sites that were nearly all within 30 km of the fault rupture. To constrain the fit of their regression model, so it would attenuate properly at greater distances, Bartlett and Youd introduced data used by Ambraseys (1988). At these large distances, Bartlett and Youd reasoned that their model should predict very small displacements when average values were used for the other site parameters in their empirical model. Bartlett and Youd were able to show that their final regression model attenuated displacements in a manner consistent with Ambraseys' empirical limit at distances up to 200 km.

In the EPOLLS model, attenuation of the predicted surface displacements at large distances is better constrained by the field data. The moment magnitude and fault distance for each EPOLLS case study are plotted in the top of Figure 9.7. The EPOLLS database contains several lateral spreads that occurred at fault distances greater than 30 km, all of which are within the maximum distance to a liquefaction failure as predicted by Equation 9.23.

If liquefaction near these maximum distances caused a lateral spread, very small displacements would be expected. To investigate the performance of the Regional-EPOLLS model at very large distances, the average horizontal displacement was predicted at distances corresponding to Equation 9.23. That is, for a given earthquake magnitude, the average deformation was predicted for a lateral spread located at the farthest limits of liquefaction failures as indicated by Ambraseys' bounding equation. The peak surface acceleration at this hypothetical site was estimated using Equation 6.3 and the duration of shaking was computed with Equation 6.5 (a focal depth of 15 km was assumed). The R-EPOLLS model was thus used to estimate displacements for M_w, R_f, A_{max}, and T_d that represent the level of shaking at the greatest distance from the fault rupture where a lateral spread could theoretically result. Note that these predictions involve extrapolation beyond the limits of the EPOLLS database, especially at distances greater than about 70 km. Moreover, the EPOLLS model only predicts surface movements at sites where liquefaction and lateral spreading occur. Therefore, this analysis involves predicted displacements that are greater than zero for sites at the extreme limits of distance where lateral spreads could theoretically develop.

The average horizontal displacements predicted with the Regional-EPOLLS model, at the limiting distances indicated by Equation 9.23, are plotted in the bottom of Figure 9.7. For an earthquake magnitude less than 7.4, the R-EPOLLS model predicts significant displacements (*Avg_Horz* up to 0.86 m) at lateral spreads at these maximum distances. The model is better overall in this regime, because most of the EPOLLS case studies are within 70 km of the fault rupture, but these predictions involve extrapolation beyond the data used to fit the model. Of more concern is the model performance at even greater distances where there is relatively little field data. As can be seen in Figure 9.7, the R-EPOLLS model predicts, for M_w greater than 7.4, a minimal average horizontal displacement of 0.149 m at sites located more than 100 km from the fault rupture. Hence, it appears that the EPOLLS model behaves satisfactorily when extrapolated out to very large distances.

9.6. Criteria for Model Predictions.

The reliability of the EPOLLS model is unknown outside the range of the field data used in fitting the equations. Hence, predictions of lateral spread displacements should not involve extrapolations beyond the limits of the parameter values in the EPOLLS database. The general maximum limits of the nine parameters used in the Regional-, Site-, and Geotechnical-EPOLLS

components are listed in Table 9.6. A better picture of how well a given set of site parameters compares with the data used to fit the EPOLLS model is obtained by comparison with the histograms in Figure 9.8. For example, the range of L_{slide} used in fitting the model extends from 20 to 1360 m. However, as can be seen in the histogram of L_{slide} in Figure 9.8, most of the case studies used to fit the model have an L_{slide} less than 600 m. Hence, relatively more confidence could be placed in an S-EPOLLS prediction at an L_{slide} of 200 m as opposed to 1000 m.

In addition to checks for simple extrapolation beyond the limits of the parameters in Table 9.6, model predictions should also be checked for possible *hidden extrapolation* involving combinations of the regressor variables. Evidence of possible simple or hidden extrapolation can be obtained by comparing, for a given site, values of D_R, (D_R+D_S), and $(D_R+D_S+D_G)$ with the limits given in Table 9.7. The minimum and maximum values for these factors in Table 9.7 represent the range in values for the data used to fit the three model components. Extrapolation is indicated if the criteria in Table 9.7 is exceeded for any model component and the preceding components. For example, a calculation with any EPOLLS model component for D_R = 4.10 should be considered unreliable because the prediction requires extrapolation beyond the limits of the data used to fit the R-EPOLLS model. If D_S = 0.15 for the same site, a prediction with the Site-EPOLLS model for (D_R+D_S) = 4.25 would still be questionable because D_R is outside the range of the field data. Hence, when using the Geotechnical-EPOLLS model, extrapolation is indicated if any of the limits in Table 9.7 are exceeded.

Hidden extrapolation is discussed further in Section D.9 where a more rigorous diagnostic test is given. This test requires a multiplication involving the matrix of the model parameters for a given site [x] and the matrix $([X]^T[X])^{-1}$. For each of the three EPOLLS model components, [x] and $([X]^T[X])^{-1}$ are defined in Figure 9.9. The result of this multiplication is then compared to the value of h_{max} for each model component in Table 9.8. Refer to Section D.9 for further details on testing for hidden extrapolation.

As also discussed in Section D.9, a *prediction interval* can be used to express the uncertainty in a model prediction arising out of the imperfect fit of the model to the database. Calculation of the prediction interval is done in accordance with Equation D.27 and also requires [x] and $([X]^T[X])^{-1}$ from Figure 9.9. In addition, the mean square error (MSE), given for each component in Table 9.9, is also needed. Finally, the prediction interval is computed with the *t*-distribution for a certain level of significance. In Table 9.9, values of the *t*-distribution, for five levels of significance and the appropriate degrees-of-freedom (n-p), are given for each EPOLLS model component.

Recall from Sections 9.2 and 9.3 that the Site-EPOLLS and Geotechnical-EPOLLS models are regressed on the residuals of the preceding model component. In these regression analyses, the design matrix [X] is modified to restrict the parameter values to be unchanged from the preceding model component. The resulting, modified $([X]^T[X])^{-1}$ matrices are thus inappropriate

for detecting hidden extrapolation or calculating a prediction interval. Hence, the $([X]^T[X])^{-1}$ matrices for the S-EPOLLS and G-EPOLLS models in Figure 9.9 were computed in a separate regression analysis where the model parameters were not restricted. This resulting $([X]^T[X])^{-1}$ yields conservative prediction intervals and tests for hidden extrapolation. That is, a test for hidden extrapolation is more stringent because it is based on a smaller subset of the database than actually used to fit the model. For the same reason, the prediction interval will be somewhat larger than the true value.

Table 9.1. Data used to fit and evaluate the EPOLLS model for average horizontal displacements.

Slide_ID	Earthquake	Avg_Horz	EQ_Mw	Fault_Dist	Accel_max	Duration	Slide_Length	Face_Height	Top_%Slope	Avg-Z_TopLiq	Avg-Z_MnFS
1	1906: San Francisco, California	1.00	7.7	13.0	0.44	45.	1360.	0.0	0.6	1.7	3.3
2	1906: San Francisco, California	1.20	7.7	12.0	0.46	45.	790.	0.0	1.3	2.0	4.0
4	1906: San Francisco, California	0.50	7.7	23.5	0.29	72.	530.	7.5	0.1	3.4	3.1
5	1923: Kanto, Japan	0.90	7.9	20.0	0.24	87.	480.	2.4	0.7	1.0	2.4
8	1948: Fukui, Japan	1.69	7.0	0.0	0.25	4.	250.	0.0	0.4	12.2	12.2
9	1948: Fukui, Japan	1.56	7.0	0.0	0.25	4.	290.	0.0	0.3	6.9	11.0
11	1964: Prince William Sound, Alaska	0.29	9.2	31.0	0.52	88.	220.	1.8	0.1	2.0	4.4
12	1964: Prince William Sound, Alaska	0.47	9.2	31.0	0.52	88.	370.	1.8	0.2	2.5	6.3
14	1964: Prince William Sound, Alaska	0.30	9.2	60.0	0.31	75.	90.	1.8	0.05	5.1	9.1
15	1964: Prince William Sound, Alaska	1.20	9.2	60.0	0.31	75.	50.	1.8	-0.2	2.5	8.2
16	1964: Prince William Sound, Alaska	1.22	9.2	60.0	0.31	75.	125.	3.0	0.2	1.4	4.2
17	1964: Prince William Sound, Alaska	0.09	9.2	60.0	0.31	75.	110.	3.0	0.2	1.2	4.6
18	1964: Prince William Sound, Alaska	1.60	9.2	60.0	0.31	75.	235.	4.9	0.1	1.7	4.5
21	1964: Prince William Sound, Alaska	0.30	9.2	100.0	0.21	77.	67.	7.3	0.1	7.3	10.0
22	1964: Prince William Sound, Alaska	0.24	9.2	100.0	0.21	77.	25.	3.7	0.1	3.7	5.5
23	1964: Prince William Sound, Alaska	0.20	9.2	100.0	0.21	77.	180.	3.7	0.1	3.7	5.2
24	1964: Prince William Sound, Alaska	1.40	9.2	100.0	0.21	77.	150.	3.0	0.7	3.0	6.4
26	1964: Niigata, Japan	3.94	7.6	16.0	0.16	19.	370.	5.5	0.0	3.0	7.0
27	1964: Niigata, Japan	3.76	7.6	15.0	0.16	19.	470.	4.6	0.0	1.0	4.5
28	1964: Niigata, Japan	2.08	7.6	15.0	0.16	19.	330.	4.2	0.0	1.8	2.8
29	1964: Niigata, Japan	4.21	7.6	13.0	0.17	19.	520.	4.6	-0.7	3.2	6.9
30	1964: Niigata, Japan	4.78	7.6	14.0	0.17	19.	320.	5.0	0.0	3.0	8.7
31	1964: Niigata, Japan	1.22	7.6	14.0	0.17	19.	320.	0.0	0.0	2.2	6.4
32	1964: Niigata, Japan	2.34	7.6	14.0	0.17	19.	500.	0.0	0.3	1.9	5.5
33	1964: Niigata, Japan	4.71	7.6	13.0	0.17	19.	480.	3.5	-0.3	3.4	4.3
35	1964: Niigata, Japan	4.59	7.6	11.0	0.17	19.	450.	3.3	1.0	2.4	9.4
37	1964: Niigata, Japan	3.23	7.6	12.0	0.17	19.	910.	0.0	0.2	2.2	6.7
38	1964: Niigata, Japan	4.74	7.6	12.0	0.17	19.	370.	0.0	0.6	2.6	5.4
39	1964: Niigata, Japan	2.76	7.6	11.0	0.18	19.	740.	0.0	0.3	2.0	2.9
40	1971: San Fernando, California	1.02	6.7	0.5	0.50	15.	930.	9.0	1.0	7.0	12.4
41	1971: San Fernando, California	0.90	6.7	1.0	0.50	14.	1340.	0.0	1.8	5.7	6.3
43	1979: Imperial Valley, California	1.40	6.5	1.6	0.46	14.	110.	1.5	0.3	1.8	3.4
44	1979: Imperial Valley, California	0.10	6.5	4.4	0.40	12.	320.	1.7	0.4	1.2	3.6
45	1983: Nihonkai-Chubu, Japan	1.47	7.9	60.0	0.25	30.	245.	0.0	5.2	1.3	6.7
46	1983: Nihonkai-Chubu, Japan	1.46	7.9	60.0	0.25	30.	325.	0.0	2.1	0.9	10.5
47	1983: Nihonkai-Chubu, Japan	1.58	7.9	60.0	0.25	30.	535.	0.0	1.0	2.5	5.9
48	1983: Nihonkai-Chubu, Japan	1.26	7.9	60.0	0.25	30.	600.	0.0	0.8	2.9	4.0
49	1983: Nihonkai-Chubu, Japan	1.55	7.9	60.0	0.25	30.	870.	0.0	0.4	2.6	5.9
50	1983: Borah Peak, Idaho	0.30	6.9	8.0	0.30	12.	40.	0.0	4.0	1.5	2.6
51	1983: Borah Peak, Idaho	1.20	6.9	0.0	0.40	11.	105.	0.0	9.5	6.1	9.6
52	1987: Superstition Hills, California	0.17	6.5	24.0	0.21	26.	20.	2.4	-0.47	2.2	3.5
54	1989: Loma Prieta, California	0.00	7.0	65.0	0.17	8.	350.	0.0	1.0	2.9	4.0
56	1989: Loma Prieta, California	0.06	7.0	1.0	0.39	11.	150.	4.3	0.15	5.3	7.6
113	1993: Hokkaido Nansei-oki, Japan	0.93	7.7	15.0	0.25	60.	70.	1.1	0.9	1.9	2.6
114	1993: Hokkaido Nansei-oki, Japan	0.81	7.7	15.0	0.25	60.	90.	0.0	1.0	1.5	3.8
6	1948: Fukui, Japan	1.96	7.0	0.0	0.25	4.	270.	0.0	0.4		
7	1948: Fukui, Japan	1.89	7.0	0.0	0.25	4.	190.	0.0	0.5		
10	1964: Prince William Sound, Alaska	0.30	9.2	31.0	0.52	88.	87.	1.8	0.07		
25	1964: Niigata, Japan	3.75	7.6	16.0	0.16	19.	290.	6.0	0.7		
34	1964: Niigata, Japan	0.98	7.6	13.0	0.17	19.	140.	3.0	0.5		
53	1989: Loma Prieta, California	0.00	7.0	67.0	0.17	8.	1000.	0.0	0.6		
109	1993: Hokkaido Nansei-oki, Japan	1.38	7.7	12.0	0.25	60.	350.	0.0	0.37		
110	1993: Hokkaido Nansei-oki, Japan	0.68	7.7	13.0	0.25	60.	380.	0.0	0.08		
111	1993: Hokkaido Nansei-oki, Japan	0.57	7.7	13.0	0.25	60.	100.	1.0	1.0		
112	1993: Hokkaido Nansei-oki, Japan	1.38	7.7	13.0	0.25	60.	140.	0.0	1.0		
115	1993: Hokkaido Nansei-oki, Japan	1.48	7.7	15.0	0.25	60.	160.	1.3	0.54		
116	1993: Hokkaido Nansei-oki, Japan	0.67	7.7	15.0	0.25	60.	160.	0.7	-0.1		
117	1993: Hokkaido Nansei-oki, Japan	1.36	7.7	15.0	0.25	60.	140.	0.8	1.33		
13	1964: Prince William Sound, Alaska	2.40	9.2	35.0	0.47	87.					
19	1964: Prince William Sound, Alaska	0.17	9.2	119.0	0.18	86.					
55	1989: Loma Prieta, California	0.30	7.0	12.0	0.25	14.					
101	1990: Luzon, Philippines	3.00	7.6	65.0	0.20	16.					
102	1990: Luzon, Philippines	2.00	7.6	65.0	0.20	16.					
103	1990: Luzon, Philippines	1.50	7.6	65.0	0.20	16.					
104	1991: Telire-Limon, Costa Rica	1.50	7.7	25.7	0.27	12.					
105	1991: Telire-Limon, Costa Rica	1.30	7.7	24.1	0.28	12.					
106	1991: Telire-Limon, Costa Rica	2.20	7.7	24.1	0.28	12.					
107	1991: Telire-Limon, Costa Rica	1.00	7.7	25.7	0.27	14.					
108	1991: Telire-Limon, Costa Rica	0.20	7.7	25.7	0.27	14.					
118	1994: Northridge, California	0.08	6.7	2.8	0.83	9.					
121	1994: Northridge, California	0.05	6.7	12.1	0.46	14.					

Data used to fit Regional-EPOLLS component

Data used to fit Site-EPOLLS component

Data used to fit Geotechnical-EPOLLS component

High influence data not used in fitting model

Not used in fitting model, values unknown

Table 9.2. Definition of variables used in the EPOLLS model for horizontal displacements.

Variable	Units	Field Name in Database	Definition
M_w	--	*EQ_Mw*	moment magnitude of earthquake
R_f	km	*Fault_Dist*	shortest horizontal distance from site to the surface projection of the fault rupture or zone of seismic energy release
A_{max}	g	*Accel_max*	peak horizontal acceleration at the ground surface of the site that would occur in the absence of excess pore pressures or liquefaction generated by the earthquake
T_d	sec	*Duration*	duration of strong earthquake motions at the site, defined as time between the first and last occurrence of a surface acceleration ≥ 0.05 g
L_{slide}	m	*Slide_Length*	maximum horizontal length from head to toe of the lateral spread in prevailing direction of movement
S_{top}	%	*Top_%Slope*	average slope across the surface of the lateral spread, measured as the change in elevation over the distance from the head to the toe • when a free face is present, the surface slope is measured from the head of the slide to the crest of the free face • negative S_{top} indicates a surface that slopes in a direction opposite to the prevailing direction of movement
H_{face}	m	*Face_Height*	height of free face, measured vertically from toe to crest of the free face • H_{face}= 0 when no free face present • when free face is a stream bank, measure H_{face} from the bottom of stream and do not include height of narrow levees along top
Z_{FSmin}	m	*Avg-Z_MnFS*	average depth to the minimum factor of safety in potentially liquefiable soil
Z_{liq}	m	*Avg-Z_TopLiq*	average depth to the top of liquefied soil

Table 9.3. Quality of fit statistics for the EPOLLS model components for average horizontal displacement.

Component of EPOLLS model:		**Regional** (Eq. 9.20)	**Site** (Eq. 9.21)	**Geotechnical** (Eq. 9.22)
Number of case studies with no missing values:		71	58	45
Number of case studies used to fit component:		68	57	44
Based on ***71*** case studies with all four variables required for the Regional-EPOLLS model	R^2 =	0.537		
	$\overline{R}^2$ =	0.509		
Based on ***58*** case studies with all seven variables required for the Site-EPOLLS model	R^2 =	0.622	0.710	
	$\overline{R}^2$ =	0.594	0.670	
Based on ***45*** case studies with all nine variables required for the Geotechnical-EPOLLS model	R^2 =	0.641	0.731	0.752
	$\overline{R}^2$ =	0.605	0.680	0.688
Evaluation of model components using all case studies with no missing data	*# Cases* =	71	58	45
	R^2=	0.537	0.710	0.752
	$\overline{R}^2$ =	0.509	0.670	0.688

Table 9.4. Results of double cross-validation analysis for three EPOLLS models.

	Subset "A"				Subset "B"			
Slide Nos. of EPOLLS case studies randomly assigned to each subset	2	17	40	100	1	22	44	107
	3	25	41	101	6	23	45	110
	4	26	50	106	8	24	46	111
	5	27	51	108	10	29	47	112
	7	28	52	109	12	30	48	113
	9	31	53	116	14	32	49	114
	11	33	54	117	18	37	102	115
	13	34	55	119	19	38	103	118
	15	35	56	121	20	42	104	120
	16	39	57		21	43	105	

	$R^2_{p,A}$	$R^2_{p,B}$
Regional-EPOLLS model	0.406	0.383
Site-EPOLLS model	0.623	0.582
Geotechnical-EPOLLS model	0.693	0.696

Note: The prediction $R^2_{p,A}$ is based on a model fit to Subset "A" used to predict the data in Subset "B"

Table 9.5. Comparison between predictions from the Geotechnical-EPOLLS model and Bartlett and Youd's MLR model.

Slide No. of lateral spread in EPOLLS database	Average of observed horizontal displacements	Bartlett and Youd's (1995) MLR model		Average horizontal displacement predicted with Geotechnical-EPOLLS model
		Number of predicted displacement vectors	Average of predicted horizontal displacements	
44	0.10 m	4	0.03 m	0.54 m
52	0.17 m	6	0.16 m	0.29 m
41	0.90 m	5	0.15 m	0.83 m
40	1.02 m	23	0.66 m	1.26 m
43	1.40 m	28	1.43 m	0.24 m
49	1.55 m	69	3.01 m	1.37 m

Table 9.6. Limiting range of EPOLLS model parameters for predicting average horizontal displacement.

Variable	Units	Minimum Value	Maximum Value
M_w	--	6.5	9.2
R_f	km	0.0	119
A_{max}	g	0.16	0.52
T_d	sec	4	88
L_{slide}	m	20	1360
S_{top}	%	-0.7	5.2
H_{face}	m	0.0	9.0
Z_{FSmin}	m	2.4	12.4
Z_{liq}	m	0.9	7.3

Table 9.7. Limiting range of EPOLLS model factors for predicting average horizontal displacement.

Component	Factor	Minimum Value	Maximum Value
Regional-EPOLLS	D_R	2.57	3.88
Site-EPOLLS	$D_R + D_S$	2.81	4.35
Geotechnical-EPOLLS	$D_R + D_S + D_G$	2.82	4.53

Table 9.8. Values of h_{max} for testing for hidden extrapolation in predicting average horizontal displacement.

Model Component:	Regional-EPOLLS	Site-EPOLLS	Geotechnical-EPOLLS
h_{max} =	0.17	0.41	0.72

Table 9.9. Parameters for computing prediction intervals on the average horizontal displacement.

Model Component:	Regional-EPOLLS	Site-EPOLLS	Geotechnical-EPOLLS
MSE =	0.149	0.111	0.124
(n - p) =	63	49	34
(1-α)	*$t_{\alpha/2,n-p}$ for (1-α)% prediction interval*		
50%	0.678	0.680	1.170
80%	1.295	1.299	1.691
90%	1.669	1.677	2.032
95%	1.998	2.010	2.345
99%	2.656	2.680	3.002

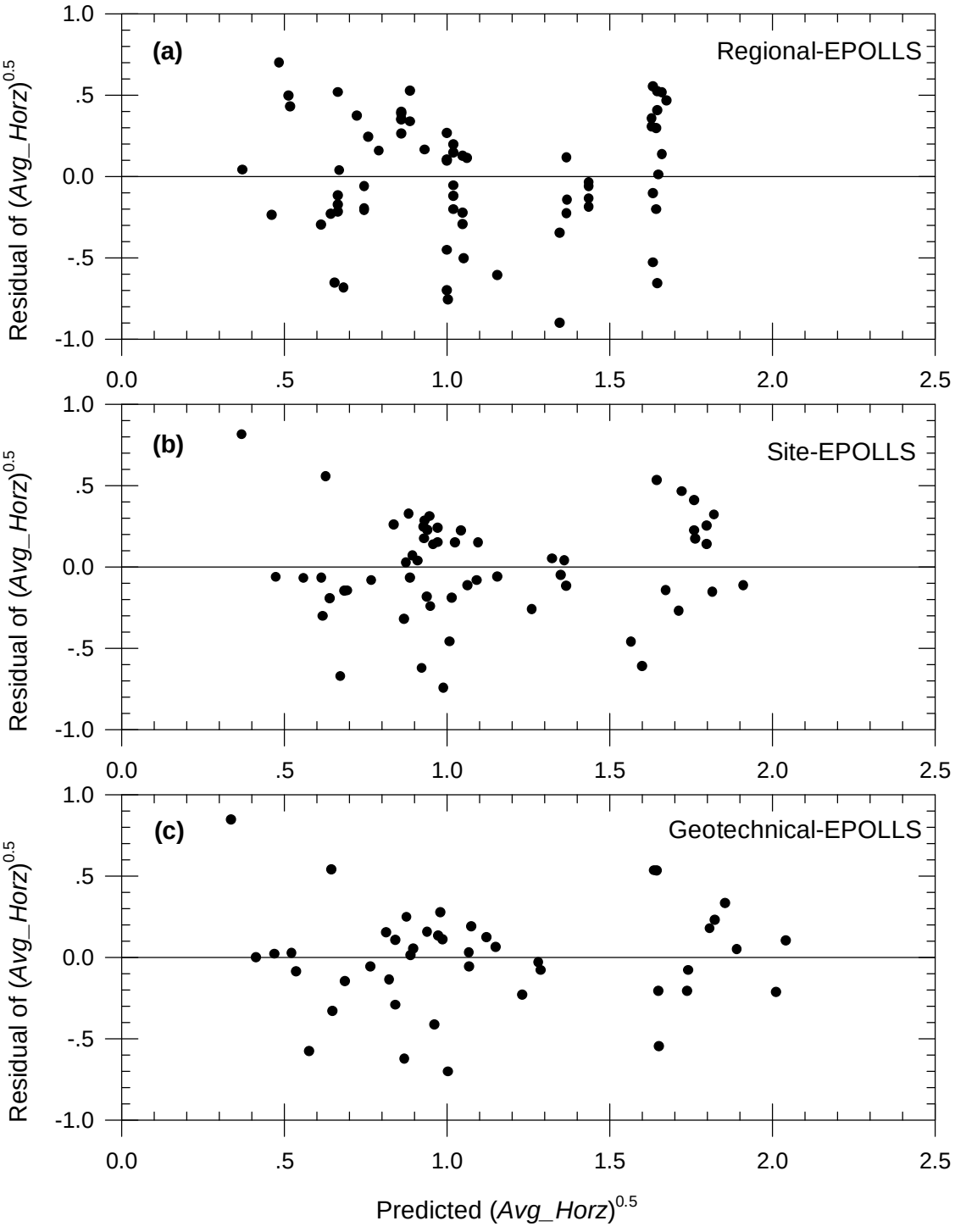

Figure 9.1. Residuals of the fitted (a) Regional-EPOLLS, (b) Site-EPOLLS, and (c) Geotechnical-EPOLLS model components.

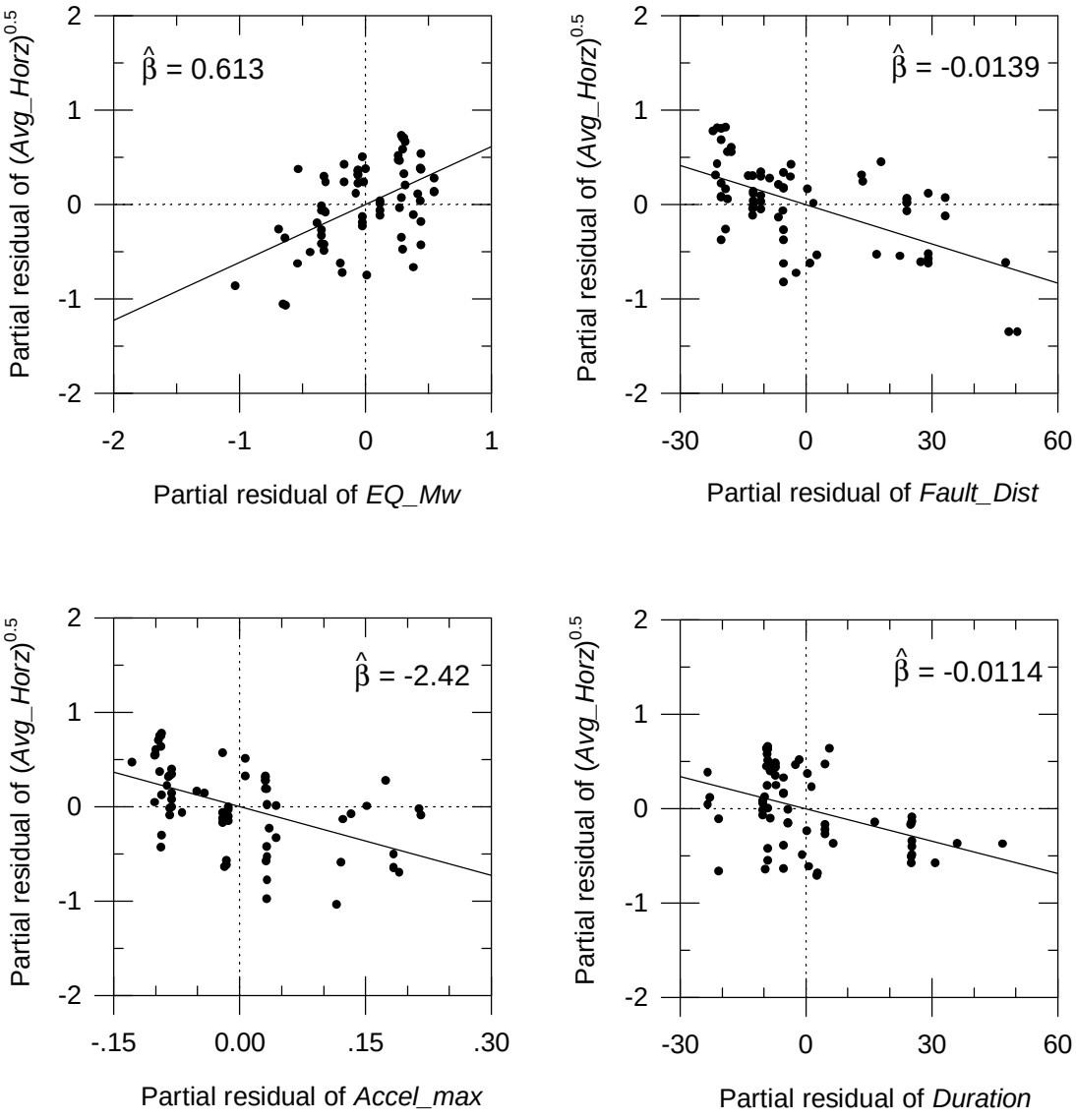

Figure 9.2. Partial regression plots for regressor variables in Regional-EPOLLS component.

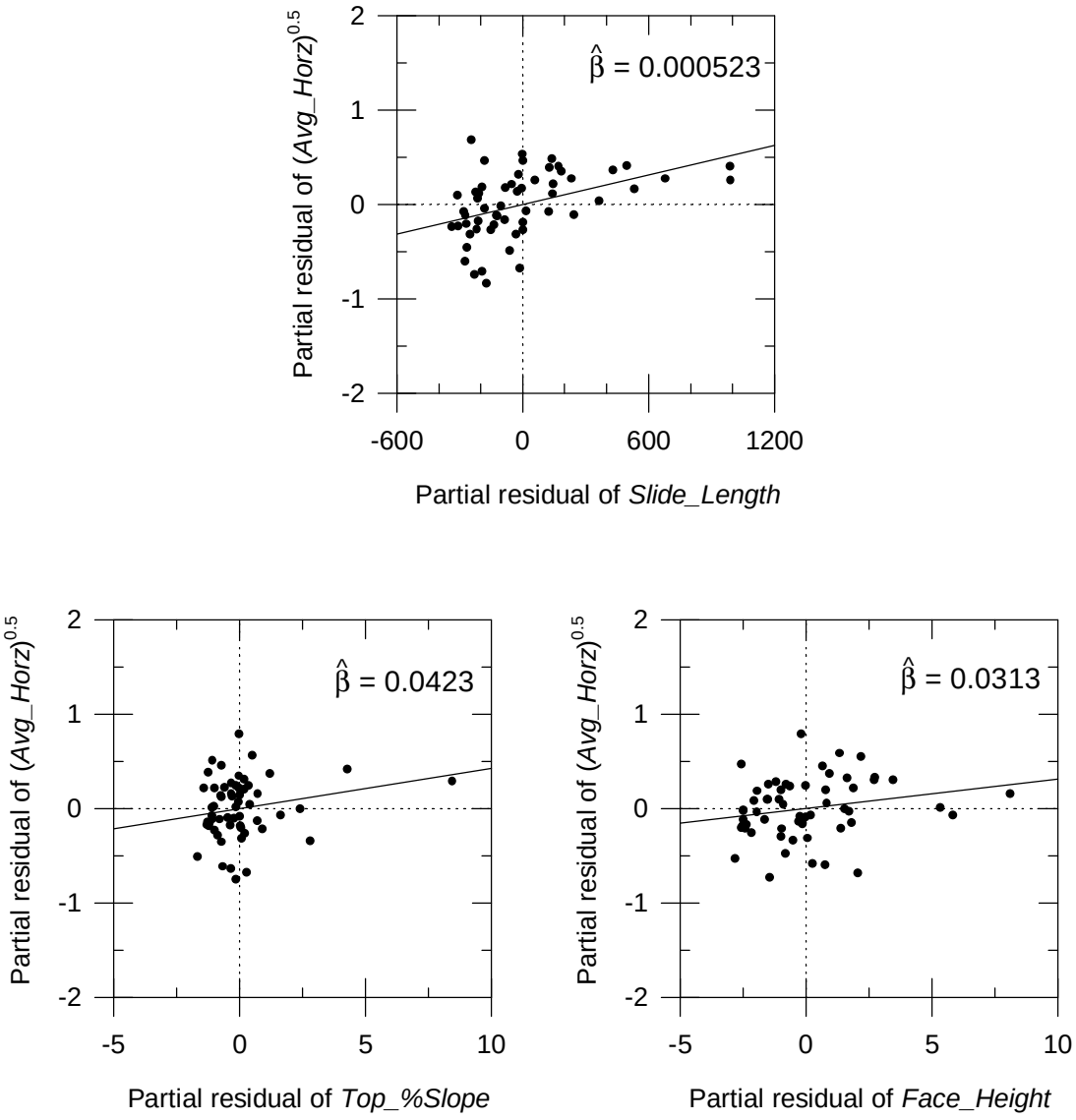

Figure 9.3. Partial regression plots for regressor variables in Site-EPOLLS component.

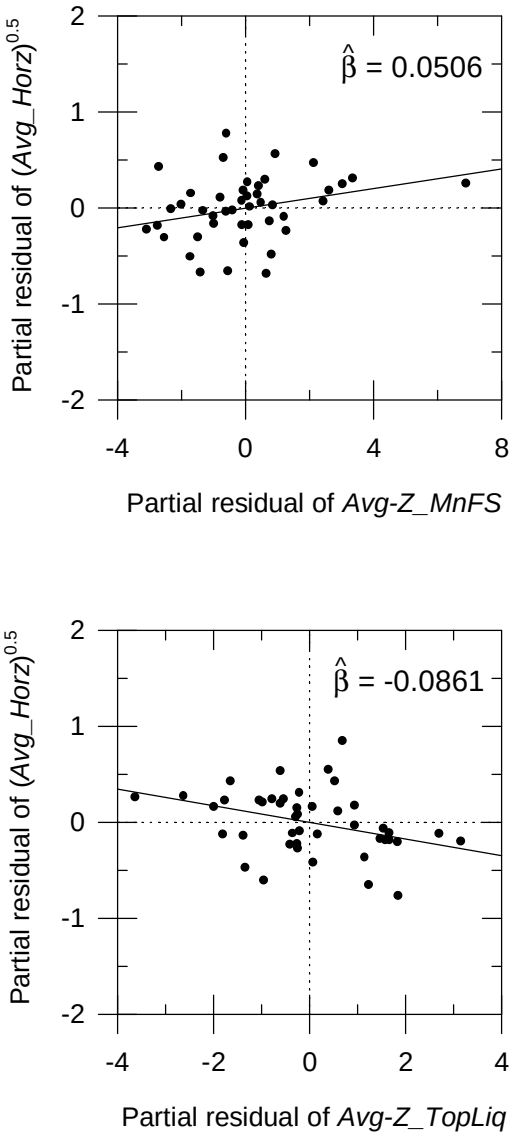

Figure 9.4. Partial regression plots for regressor variables in Geotechnical-EPOLLS component.

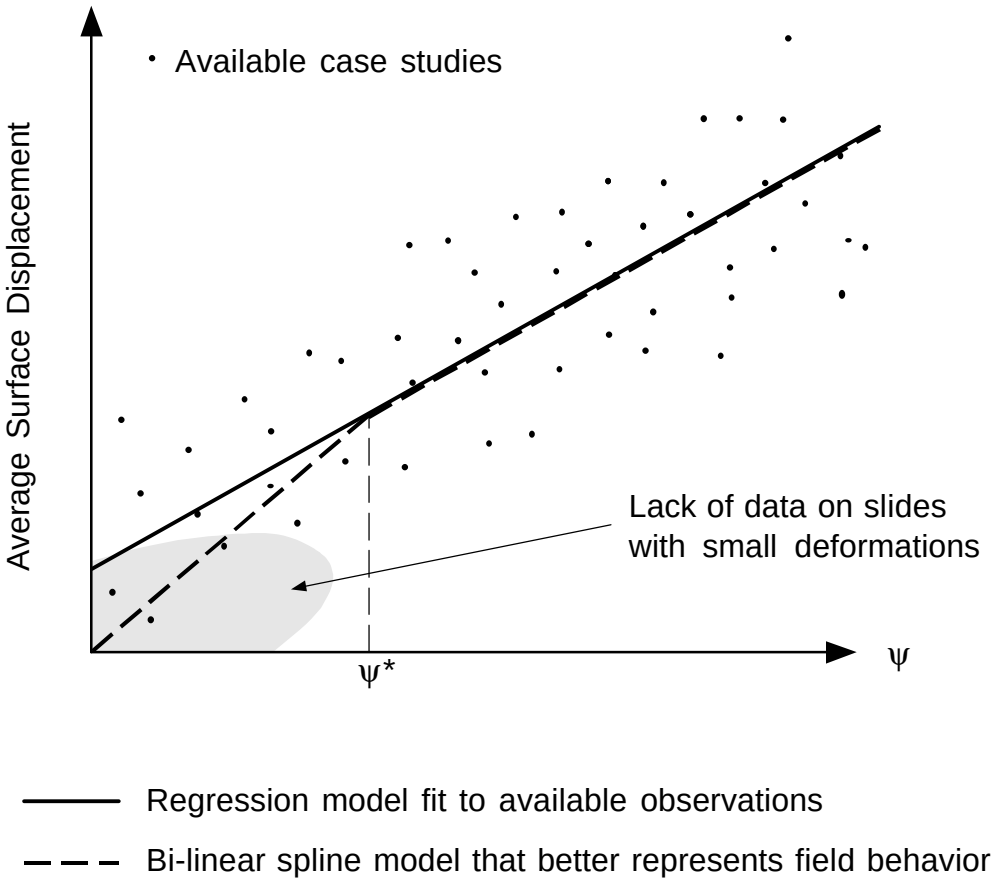

Figure 9.5. Conceptual illustration of possible model bias resulting from lack of data on less damaging lateral spreads.

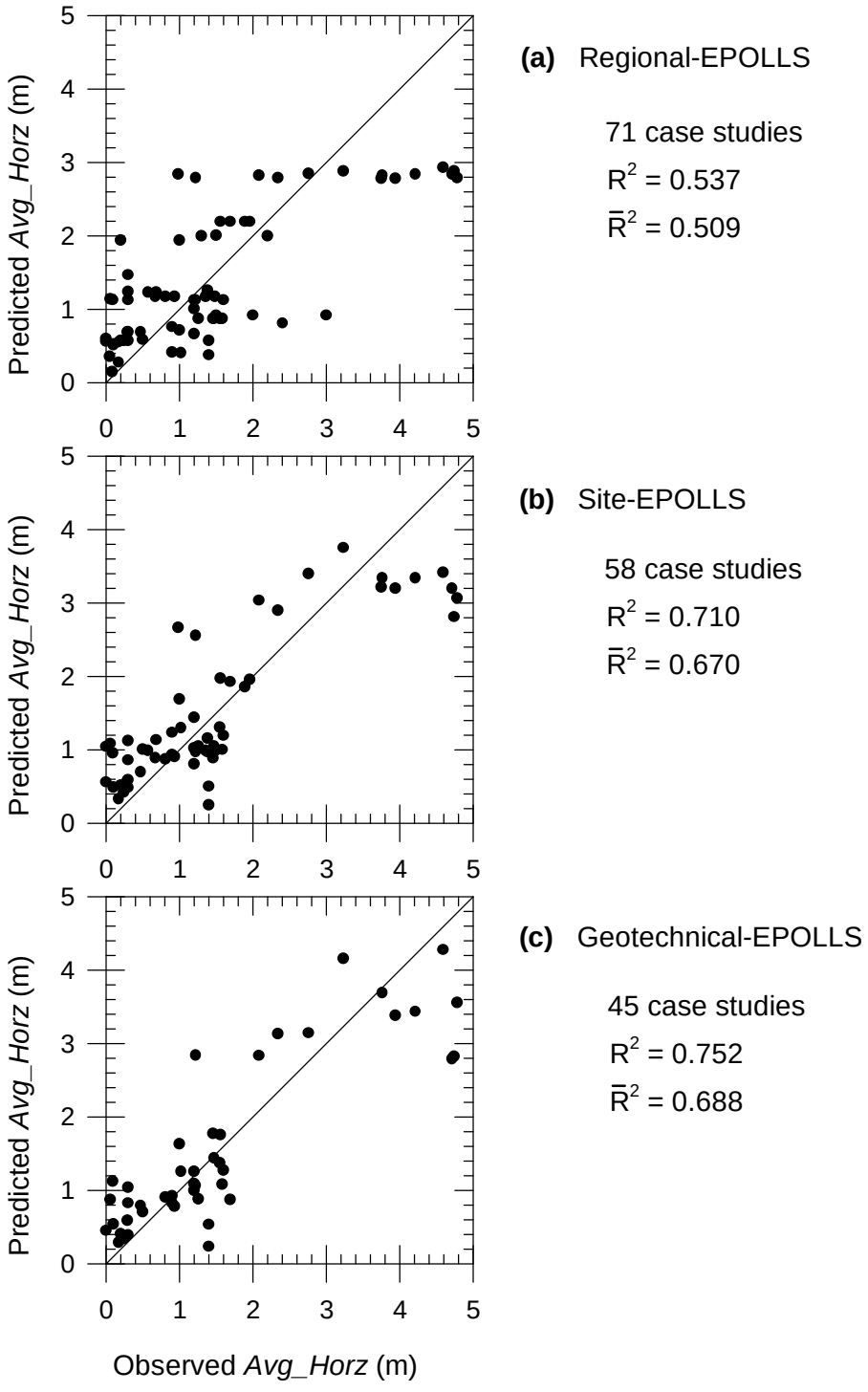

Figure 9.6. Performance of the (a) Regional-EPOLLS, (b) Site-EPOLLS, and (c) Geotechnical-EPOLLS model components.

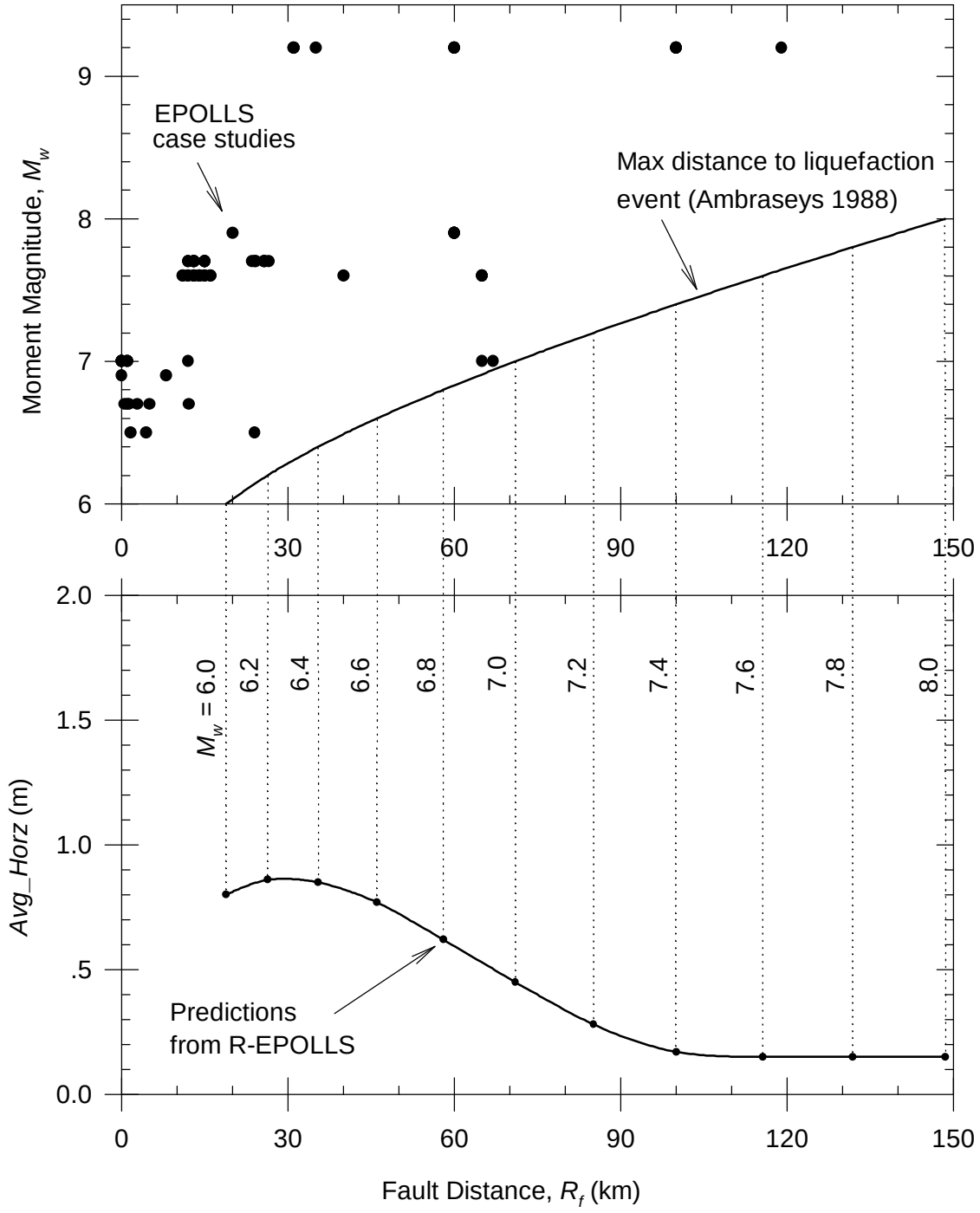

Figure 9.7. Performance of R-EPOLLS model at maximum distance to liquefaction events.

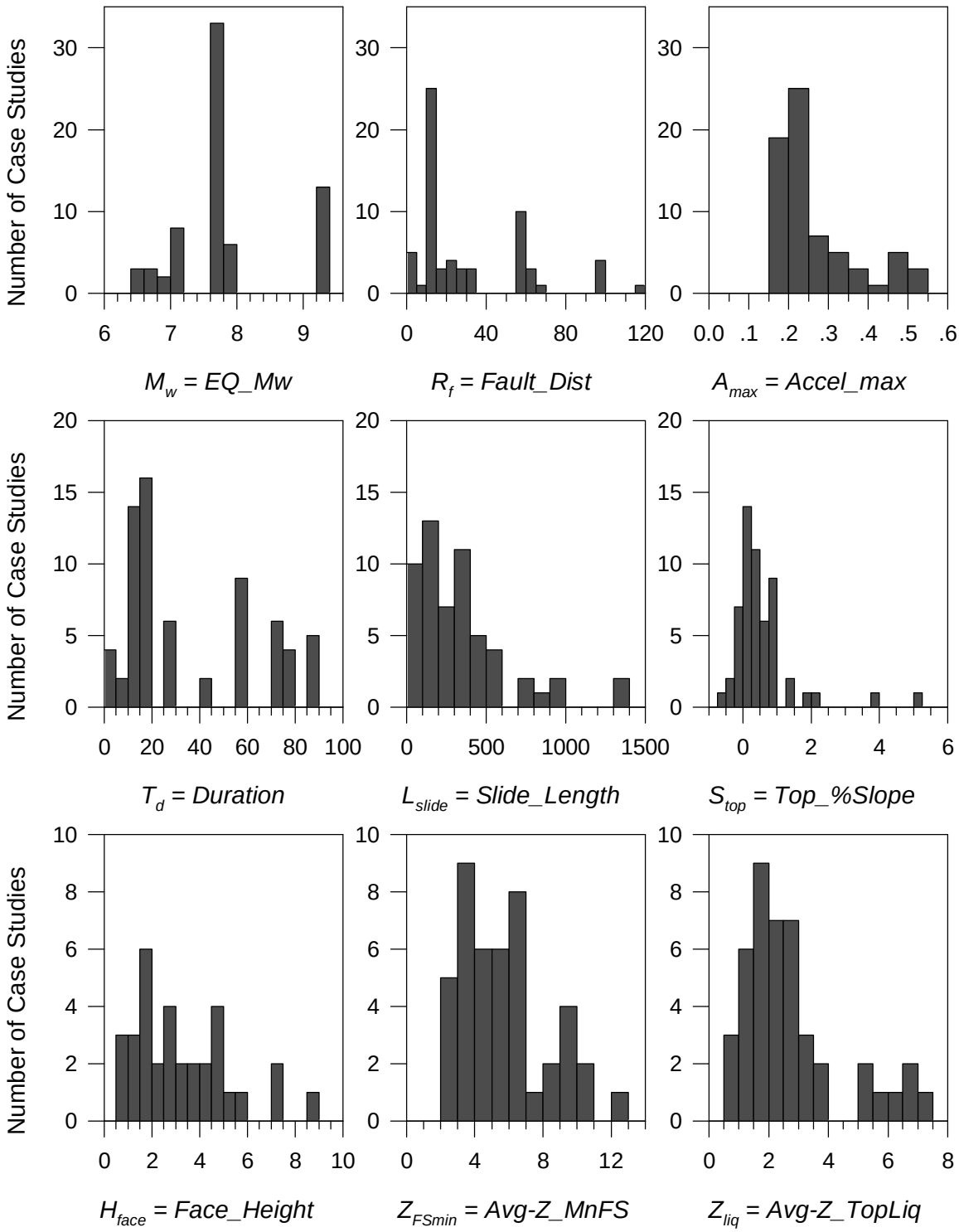

Figure 9.8. Histograms of data used to fit the EPOLLS models for average horizontal displacement.

(*a*) ***Regional-EPOLLS; for*** $[x] = [1.0\ M_w\ R_f\ A_{\max}\ T_d]$

$$([X]^T[X])^{-1} = \begin{bmatrix} 5.39 & -.761 & .00681 & -.624 & .0138 \\ & .111 & -.00112 & .0258 & -.00200 \\ & & 3.72\cdot10^{-5} & .00218 & 5.18\cdot10^{-6} \\ & & & 1.73 & -.00299 \\ & \textit{symmetric} & & & 6.48\cdot10^{-5} \end{bmatrix}$$

(*b*) ***Site-EPOLLS; for*** $[x] = [1.0\ M_w\ R_f\ A_{\max}\ T_d\ L_{slide}\ S_{top}\ H_{face}]$

$$([X]^T[X])^{-1} = \begin{bmatrix} 6.85 & -.973 & .0109 & -.441 & .0161 & -8.22\cdot10^{-5} & -.0441 & 3.26\cdot10^{-4} \\ & .143 & -.00171 & .00921 & -.00242 & 8.62\cdot10^{-7} & .00536 & -.00112 \\ & & 5.63\cdot10^{-5} & .00297 & 5.71\cdot10^{-6} & -1.04\cdot10^{-7} & -1.57\cdot10^{-4} & -3.57\cdot10^{-5} \\ & & & 2.27 & -.00521 & -2.38\cdot10^{-4} & -.0483 & -.00358 \\ & & & & 8.74\cdot10^{-5} & 1.13\cdot10^{-6} & 1.31\cdot10^{-4} & 2.59\cdot10^{-5} \\ & & & & & 2.50\cdot10^{-7} & 1.09\cdot10^{-5} & 1.92\cdot10^{-6} \\ & & & & & & .0103 & .00165 \\ & \textit{symmetric} & & & & & & .00386 \end{bmatrix}$$

(*c*) ***Geotechnical-EPOLLS; for*** $[x] = [1.0\ M_w\ R_f\ A_{\max}\ T_d\ L_{slide}\ S_{top}\ H_{face}\ Z_{FSmin}\ Z_{liq}]$

$$([X]^T[X])^{-1} = \begin{bmatrix} 10.2 & -1.48 & .0105 & -1.60 & .0334 & -2.15\cdot10^{-5} & -.0648 & -.0393 & .110 & -.0884 \\ & .223 & -.00173 & 0.174 & -.00515 & -1.43\cdot10^{-5} & .00909 & .00507 & -.0199 & .0144 \\ & & 6.72\cdot10^{-5} & .00438 & -3.79\cdot10^{-6} & 3.52\cdot10^{-7} & -1.15\cdot10^{-4} & 5.21\cdot10^{-5} & 5.34\cdot10^{-5} & -2.27\cdot10^{-4} \\ & & & 3.19 & -.0124 & -2.63\cdot10^{-4} & -.0457 & .0189 & -.0160 & -.0411 \\ & & & & 1.97\cdot10^{-4} & 1.47\cdot10^{-6} & -6.90\cdot10^{-6} & -2.79\cdot10^{-4} & 5.87\cdot10^{-4} & -1.51\cdot10^{-4} \\ & & & & & 2.91\cdot10^{-7} & 1.48\cdot10^{-5} & 5.55\cdot10^{-6} & 3.09\cdot10^{-6} & -5.80\cdot10^{-6} \\ & & & & & & .0119 & .00321 & -.00190 & -4.52\cdot10^{-4} \\ & & & & & & & .00629 & -.00111 & -.00298 \\ & & & & & & & & .00865 & -.00741 \\ & \textit{symmetric} & & & & & & & & .0182 \end{bmatrix}$$

Figure 9.9. $([X]^T[X])^{-1}$ matrices, used to check for hidden extrapolation and compute prediction intervals.