LINEAR REGRESSION MODELS TO ESTIMATE SOIL LIQUID LIMIT AND PLASTICITY INDEX FROM BASIC SOIL PROPERTIES

Cathy A. Seybold, Moustafa A. Elrashidi, and Robert J. Engel

In Soil Survey, there is a need to estimate liquid limit (LL) and plasticity index (PI) for areas where data are not available. The objectives were to determine if LL and PI prediction equations could be developed from readily available soil properties in Soil Survey, and to test two different data stratification approaches to improve predictability. Measured data in the National Soil Survey Characterization database and multiple linear regression were used for model development. Clay content (<2 μm) and cation exchange capacity were the primary variables used to predict both LL and PI. To predict LL, four equations were developed from 10 taxonomic soil order strata (aggregate of seven soil order strata, Andisols, Spodosols, and Vertisols) that explained between 68% and 81% of the variation in LL, with the Andisols order having the lowest predictability. To predict PI, 10 unique taxonomic soil order equations were developed (Aridisols, Alfisols, Entisols, Inceptisols, Mollisols, Oxisols, Ultisols, Andisols, Spodosols, and Vertisols) that explained between 15% and 77% of the variation in PI, with the Andisols order having the lowest predictability. A few prediction equations were developed from the taxonomic mineralogy strata, which produced models with similar predictability to that of the soil order equations. Validation of the best fitting models with an independent data set showed no significant difference from unit 1 slope and 0 intercept. Predicting LL and PI from readily available soil properties resulted in mostly moderate to strong prediction equations. The most useful equations are those with $R^2 > 0.60$. These prediction equations can be useful in Soil Survey when there are no available data. (Soil Science 2008;173:25-34)

Key words: Liquid limit, plasticity index, prediction, general linear models.

THE Atterberg limits are moisture content limits that divide the states of soil consistency, which is the degree of resistance to deformation. There are three states of soil consistency, the shrinkage limit that separates the solid state from semisolid state; the plastic limit (PL) that separates the semisolid state from plastic state; and the liquid limit (LL) that separates the plastic from liquid state (PCA, 1992). The width of the plastic state (LL minus PL), in terms of moisture content, is the plasticity index (PI). Plasticity is the capability of a soil to undergo unrecoverable deformation at constant volume without cracking or crumbling (McBride, 2002). The LL and PI are used in soil survey for interpreting soils for engineering classifications and other engineering purposes. The Atterberg limits are important for classifying cohesive soil materials and is useful for interpreting soils for shear strength, bearing capacity, compressibility, and swelling potential (McBride, 2002). In addition, LL and PI determinations are cumbersome, time consuming, and are not part of routine soil survey characterization analysis. They are carried out on an ad hoc basis, and such data are not
generally available. A need remains, in soil survey, for a quick and reliable method of estimating the LL and PI that can be applied to a wide range of soils. In addition, the LL and PI need to be predicted from generally available soil properties in soil survey.

There are limited unpublished predictions of LL and PI from soil characterization data that were developed in the 1950s and the early 1970s. These models were developed using least squares estimates and focused on the relationship of clay content (<2 μm) to the Atterberg limits. However, these models can only be applied to a few soil types (e.g., kaolinitic clays) or to specific areas or regions (e.g., Iowa loess). In addition, the A horizons were usually eliminated from the analyses. In these unpublished models, clay was found to explain between 66% and 96% of the variability in the LL, and between 71% and 93% for the PI.

Others have also shown clay to be closely correlated to the Atterberg limits (De Jong et al., 1990; Mbagwa and Abeh, 1998; Odell et al., 1960; Smith et al., 1985). In B and C horizons, total clay content was found to be the most important independent variable for explaining variation in the Atterberg limits (De Jong et al., 1990). In general, the greater the quantity of total clay in a soil, the greater the plasticity and potential shrink and swell (Mitchell, 1993). In addition to the amount of clay, the type of clay and the size and shape of the particles have an effect on the Atterberg limits (Mitchell, 1993; Baver, 1930). In 26 representative soils of Illinois (including A, B, and C horizons), Odell et al. (1960) has shown that the percent montmorillonite in the clay fraction is strongly correlated to the Atterberg limits. In addition, Odell et al. (1960) found that the organic C (OC), total clay, and the quantity of montmorillonite in the clay fraction could explain 86% and 94% of the variability in the LL and PI, respectively. Dumbleton and West (1966) have shown that the plasticity of soils of the same mineralogical type but different origins can show considerable variation in physical properties. As pointed out by Mitchell (1993), the LL and PI values for any one clay mineral can vary over a wide range.

In general, the greater the surface area, the greater the amount of water needed to get to the LL state (Seed et al., 1964; Mitchell, 1993). In 19 British clay soils, total surface area was highly correlated to the LL, and to a lesser extent to the PI (Farrar and Coleman, 1967). Smith et al. (1985) found the LL to be more closely correlated to the specific surface area than to the clay content in 66 soil samples taken from 32 sites across Israel.

To a lesser extent than clay, OC content has been shown to be correlated to the Atterberg limits (De Jong et al., 1990; Mbagwa and Abeh, 1998; Odell et al., 1960; Larney et al., 1988). De Jong et al. (1990) found OC content to be as important as the clay content in explaining the variation in the LL of Ap horizons. In the B and C horizons of the same study, OC was a poor predictor of the Atterberg limits. Mbagwa and Abeh (1998) found organic matter to be best correlated with the PI. In mainly kaolinitic clay soils of Florida, organic matter was found to be weakly correlated to the PI (De La Rosa, 1979). It has been shown that organic matter can increase the PI without increasing the magnitude of the PI (Baver, 1930; Smith et al., 1985). In other words, two soils may have the same PI, but may exhibit plasticity over entirely different moisture ranges.

Cation exchange capacity (CEC) can be an indication of mineral type and has been shown to be highly correlated to the LL (Mbagwa and Abeh, 1998; Odell et al., 1960; Farrar and Coleman, 1967) and to a lesser extent to the PI (Odell et al., 1960; Mbagwa and Abeh, 1998). In 30 samples from Nigeria, 79.9% of the variation in the LL was explained by the CEC alone (Mbagwa and Abeh, 1998). De La Rosa (1979) found clay, CEC, and organic matter, and their interactions to explain 97% of the variation in the PI of 38 soil series (54 samples) from Florida. Exchangeable cations affect the PI by affecting the hydration of the clays (Baver, 1930; Mitchell, 1993). These studies indicate that CEC could be an important variable in predicting LL and PI.

In the National Soil Survey Handbook, there is an LL and PI prediction equation that was developed from a broad range of soil properties (USDA-NRCS, 2005). However, these two prediction equations do not use CEC, and the accuracy of predicting LL and PI could be improved if this diverse data set was stratified into more homogeneous soil groups. The objectives of this study were to determine (i) the relationship between soil properties that are readily obtained in soil survey (e.g., CEC, clay, OC) and LL and PI; and (ii) to determine if useful prediction models could be developed after stratifying by taxonomic order or taxonomic family mineralogy class. These prediction
models will benefit the Natural Resources Conservation Service field soil scientist making entries into the National Soils Information System, which is the United States' national soil survey database. More importantly, these models should improve the accuracy of estimated LL and PI data, which will benefit all users of soil survey data and their interpretations.

MATERIALS AND METHODS

The pre-1999 data in the National Soil Survey Laboratory Characterization Database at Lincoln, Nebraska, were used to develop the LL and PI models. This database contains about 10,000 horizons with measured LL and PI data, representing soils from across the continental United States, Hawaii, and Alaska. Relevant data in the database include taxonomic classifications, morphological descriptions, horizon designations, and analytical data such as OC, exchange properties, particle size separates, pH, and water retention characteristics.

Basic soil properties evaluated as potential predictor variables were pH in water and 0.01 M CaCl₂, total clay, silt, and sand (pipette method); OC (acid-dichromate digestion); water content at —1500 kPa (pressure-membrane extraction using sieved samples), CEC (1.0 N NH₄OAc at pH 7); carbonate clay; bulk density at —33 kPa water content (clod method); and linear extensibility percent (LEP). All methods are described by Burt (2004). All determinations are from air-dried (30 °C–35 °C), crushed, and sieved (<2 mm) soil samples. Data are reported on an oven-dry basis (Burt, 2004). Liquid limit and PL were determined by American Society for Testing and Materials method D 4318 on a less than 0.4-mm base. The PI is the difference between the LL and PL. If either the LL or PL could not be determined, or if PL is greater than the LL, then the soil was eliminated from the data set. Variables were then added and subtracted from the general linear model until the best model was found that contained the best fit regression model (with the highest \( R^2 \) and lowest root mean square error (RMSE)) was developed. Pearson correlations were performed to determine variable colinearity and to help in the selection of predictive variables. Only data elements that contributed significantly \((P = 0.05)\) to predicting the LL or PI and that contributed greater than 5% to the overall improvement of the \( R^2 \) were included in the equations. Scatter plots of the residuals versus the fitted values of each model were used to indicate whether there was nonlinearity, unequal variances, and outliers in the data. All outliers, as identified by the studentized residual in SYSTAT Software (2002), were removed from the data groups. Variables were then added and subtracted from the general linear model until the best model was found that contained statistically significant, intuitively meaningful predictive variables, and variables that are readily obtainable within the National Soils Information System of the Natural Resources Conservation Service. A dummy variable regressor (taxonomic mineralogy/order) was used to evaluate model redundancy between predictive equations (from different data strata) with the same variables (Fox, 1997). The post hoc Tukey test (multiple mean comparison procedure) was used for comparison of equation intercepts (Zar, 1999). When intercepts between two
The range in selected properties of soils used in developing LL and PI regression equations are shown in Table 1. The clay content ranged from 4% to 95%, and OC ranged from 0.01% to 11%. Nonplastic soil layers were not used (PI = 0). Clay content, CEC, and -1500 kPa water were the most highly correlated to LL and also to the PI (Table 2). De Jong et al. (1990) also found water retention to be a good index for the Atterberg limits. In general, the correlation coefficients were lower for the soil property correlations with the PI than with the LL. Others have also reported lower correlations of soil properties with the PI than with the LL (De Jong et al., 1990). Bulk density was significantly and negatively correlated to the LL \((r = -0.55)\). The sand content was significantly and negatively correlated to the PI \((r = -0.45)\) and to a lesser extent to the LL \((r = -0.38)\). Organic C was not significantly correlated to either LL or PI. In contrast, Larney et al. (1988) found organic matter to be highly correlated to the LL and PI. In their study, only Ap horizons were used, and their data set contained a very narrow range in clay contents. In the present study, clay content covers nearly the entire range found in soils (Table 1). Clay content was highly correlated to the -1500-kPa water \((r = 0.93)\). These two variables (clay content and -1500-kPa water) would provide redundant information if both were included in a regression model. Clay content is an easily obtained property in soil survey and is the preferred predictor variable over -1500-kPa water. However, in cases where clay content is not accurately measured,
(because of clay dispersion problems in the particle size analysis) then -1500-kPa water is preferred as a predictor variable over clay content.

Carbonate clay content was not significantly correlated to the LL (r = -0.02), PL (r = -0.01), or PI (r = -0.01). Others have also reported inorganic C to have little effect or no correlation on the LL or PL (Odell et al., 1980; Smith et al., 1985). Conversely, in highly calcareous soils of Egypt, Stakman and Bishay (1976) showed increasing LL with an increasing CaCO₃ content up to about 35% CaCO₃. In their study, the PL showed a slight increase, and thus the PI showed the same tendency as the LL.

Inorganic C generally reduced the Atterberg limits and the P1 of B and C horizons in a study by De Jong et al. (1990). However, the impact of inorganic C was relatively small.

**Prediction of LL**

Clay content and CEC explained 81% of the variation in the LL of the entire data set (n = 6592), excluding data with -1500-kPa water-clay ratios that were greater than 0.6. The regression equation is:

\[
LL = 0.655 \text{ (clay)} + 0.406 \text{ (CEC)} + 12.459 \tag{1}
\]

The equation RMSE (SD about the regression line) is 6.8%. De La Rosa (1979) also used CEC and clay content in addition to OC and their interactions to explain nearly all the variations in the LL (R² = 0.97) of 54 soil samples from Florida. In the present study, there were no interactions that significantly improved the prediction of LL from the entire data set.

Using taxonomic order or taxonomic mineralogy as a dummy variable regressor alone explained 20% and 24% of the variation in the LL, respectively. This indicates that stratifying the data set by mineralogy or taxonomic order and developing a regression equation for each strata could improve the prediction of LL.

The database was stratified by taxonomic order and equations developed for each soil order (excluding Histosols and Gelisols). The resulting regression equations for the orders Entisols, Aridisols, Alfisols, Mollisols, Inceptisols, Oxisols, and Ultisols were not significantly different from each other (data not shown). Therefore, the data from these seven soil orders were combined, and an overall equation was developed (Table 3). Clay content and CEC explained 79% of the variation in LL when the data from the seven soil orders were combined.

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Prediction equation</th>
<th>R²</th>
<th>RMSE</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicting LL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entisols</td>
<td>0.517 (clay) + 0.309 (CEC) - 2.326 (OC) - 1.533</td>
<td>0.77</td>
<td>5.646</td>
<td>210</td>
</tr>
<tr>
<td>Aridisols</td>
<td>0.5 (clay) + 0.181 (CEC) - 1.156</td>
<td>0.64</td>
<td>5.631</td>
<td>635</td>
</tr>
<tr>
<td>Alfisols</td>
<td>0.685 (clay) - 0.003 (clay)² + 0.442 (CEC) - 6.118</td>
<td>0.74</td>
<td>6.526</td>
<td>1022</td>
</tr>
<tr>
<td>Mollisols</td>
<td>0.917 (clay) - 0.005 (clay)² + 0.259 (CEC) - 8.199</td>
<td>0.72</td>
<td>5.632</td>
<td>1459</td>
</tr>
<tr>
<td>Vertisols</td>
<td>0.132 (clay) + 1.53 (LEP) + 16.237</td>
<td>0.51</td>
<td>7.793</td>
<td>319</td>
</tr>
<tr>
<td>Inceptisols</td>
<td>0.741 (clay) - 0.005 (clay)² + 0.246 (CEC) - 4.402</td>
<td>0.61</td>
<td>6.794</td>
<td>295</td>
</tr>
<tr>
<td>Oxisols</td>
<td>0.327 (clay) + 1.428</td>
<td>0.42</td>
<td>6.821</td>
<td>237</td>
</tr>
<tr>
<td>Ultisols</td>
<td>0.68 (clay) - 0.003 (clay)² - 1.292</td>
<td>0.59</td>
<td>5.997</td>
<td>480</td>
</tr>
<tr>
<td>Spodosols</td>
<td>0.355 (CEC) + 1.425</td>
<td>0.69</td>
<td>2.207</td>
<td>40</td>
</tr>
<tr>
<td>Andisols</td>
<td>0.237 (w15bar) + 1.162 (OC) + 5.398</td>
<td>0.15</td>
<td>16.37</td>
<td>140</td>
</tr>
<tr>
<td>Kaolinitic</td>
<td>0.232 (clay) + 0.35 (CEC) - 16.604 (Db) + 33.273</td>
<td>0.50</td>
<td>7.863</td>
<td>246</td>
</tr>
<tr>
<td>Mixed</td>
<td>0.75 (clay) - 0.004 (clay)² + 0.341 (CEC) - 1.884 (OC) - 8.895</td>
<td>0.69</td>
<td>5.429</td>
<td>2797</td>
</tr>
</tbody>
</table>

1. clay: total clay (%); OC: organic carbon (%); w15bar: -1500 kPa water content; CEC: cation exchange capacity (cmol(+) kg⁻¹); LEP: linear extensibility percent (%); Db: bulk density (g cm⁻³).

2. Includes taxonomic soil orders Entisols, Aridisols, Alfisols, Mollisols, Inceptisols, Oxisols, and Ultisols.
Bulk density did not significantly improve the prediction of the combined soil order equation. A unique LL prediction equation was developed for the Vertisols, Andisols, and Spodosols soil orders (Table 3). For the Vertisols order, clay content, CEC, and LEP explained 78% of the variation in the LL. Linear extensibility percent indicates the amount of swelling and shrinkage of the soil. The greater plasticity of the soil indicates a greater potential shrink-swell (Mitchell, 1993). The CEC alone explained 81% of the variation in LL for the Spodosols order. However, a small number of soil samples were used in the development of the equation (n = 41), which may limit its overall application. The LL in the Andisols order was the most difficult to predict and had the largest RMSE. This could be caused by the variation in the amount of andic properties between soil samples or layers within a pedon. Some horizons may have andic properties and some may not, and every combination in between. It may be useful in future analyses to further stratify the Andisols (or the entire data set) by texture modifier—hydrous, medial, and ashy soil layers (Soil Survey Staff, 1999). Spodic horizons could also be grouped with soils with andic soil properties since they respond similarly.

The RMSE of the models developed from the taxonomic order strata, excluding Andisols, ranged from 6.69% to 5.49%, and were lower than the RSME for Eq. 1 (overall LL prediction equation). The lower RMSE indicate that stratifying the large data set and separating out the Vertisols, Andisols, and Spodosols increased the accuracy of predicting the LL.

A comprehensive coverage of prediction equations for the mineralogy classes was not possible because of insufficient data. The mineralogy class strata provided a different way of grouping the soils. Which strata would provide the best fit models and the most accurate predictions cannot be determined from this study. The smectitic and mixed mineralogy models for predicting LL were not significantly different from the "Orders" model in Table 3. The kaolinitic mineralogy model is unique in that bulk density was useful in explaining some of the variations in LL when clay and CEC were used as predictor variables. As with the Andisols, it was equally as difficult to predict the LL of the kaolinitic mineralogy strata.

Plasticity Index

Clay content and CEC explained 71% of the variation in the PI of the entire data set (n = 6592), excluding data with −1500-kPa water-extraction clay ratios that were greater than 0.6. The prediction equation is:

\[ PI = 0.408 (\text{day}) + 0.434 (\text{CEC}) - 1.525 \] (2)

The equation RMSE is 6.72%. The PI was more difficult to predict than the LL. Reasons for this could be that the PI can be the same for two soils, but exhibit plasticity over entirely different moisture ranges (Baver, 1930). In other words, soil properties may vary, but the PI may be the same in some cases, causing PI to be less predictable.

Using taxonomic order or taxonomic mineralogy as a dummy variable regressor alone explained 16% and 17% of the variation in the PI, respectively. Taxonomic mineralogy or soil order explained less variation in PI than in the LL, probably for the same reasons as stated previously. However, stratifying the data set by taxonomic mineralogy or soil order could improve the prediction of PI.

Plasticity index prediction equations were developed for 10 soil orders (Table 3). All of the soil order PI equations were determined to be unique. In general, the PI was more difficult to predict than the LL, as indicated by the lower \( R^2 \) and, in some cases, higher RMSE (Table 3). The CEC and clay contents were the primary predictor variables used in predicting PI. Clay is a major contributor to the plasticity of soils. Organic C was a useful predictor variable of the PI for the Entisols, and LEP was a useful predictor variable for the Vertisols. For the Alfisols, Mollisols, Inceptisols, and Ultisols soil orders, the squared clay content (clay\(^2\)) was found to be significant in predicting PI, which indicates nonlinearity (curvature) in the relationship between clay content and the PI.

Clay, CEC, and OC explained 77% of the variation in PI for the Entisols order. Clay, squared clay, and CEC explained 74%, 72%, and 61% of the variation in PI for the Alfisols, Mollisols, and Inceptisols, respectively. Clay content and CEC explained 64% of the variation in PI for the Andisols. Clay content and LEP explained 51% of the variation in PI for the Vertisols. The CEC alone explained 69% of the variation in PI for the Spodosols, and clay content alone explained 42% of the variation in PI.
PI for the Oxisols. Clay content and squared clay explained 59% of the variation in PI for the Ultisols. The PI of the Andisols order was the most difficult to predict, with ~1500 kPa water and OC explaining only 15% of the variability in the PI. As indicated previously, the magnitude of the LL and PL may vary, whereas the PI may not change, causing the PI to have no relationship to the varying soil property. All soils probably have some degree of this, but it may be more of a factor in the Andisols order. In addition, the factors which affect the plasticity of soils for the most part act simultaneously, and it can be therefore difficult to isolate the effect caused by the individual factors (Dumbleton and West, 1966).

The RMSE for the PI prediction equations ranged from 2.2% to 7.79%, excluding that for the Andisols, which was 16.37%. The RMSE of the overall PI equation (Eq. 2) was at the higher end of this range (6.27%). By stratifying the data set, lower RMSE and/or higher $R^2$ were obtained for the Entisols, Aridisols, Alfisols, Mollisols, Ultisols, and Spodosols soil orders than that for Eq. 2. This suggests that the accuracy of predicting PI will improve if the soil order equations are used instead of Eq. 2.

The prediction of PI for the three mineralogy classes (mixed, smectitic, and kaolinitic) resulted in two unique equations (Table 3). The equation for smectitic mineralogy was not significantly different from the Mollisols model. Clay, CEC, and bulk density explained only 50% of the PI variability in the kaolinitic mineralogy data group. Clay, squared clay, CEC, and OC were able to explain 69% of the PI variation in the mixed mineralogy strata. As indicated in the prediction of LL, there are not enough data to evaluate models from all the mineralogy class strata. However, the three models developed from the mineralogy class strata do not seem to provide any better estimates than stratifying by taxonomic order.

**Validation**

The measured versus predicted LL or PI values for Eq. 1 and 2 regression models are shown in Figs. 1A, B. The 95% confidence intervals about the slope of the regression line for prediction of LL (Eq. 1) indicate no significant difference from unity; the confidence intervals contain one (Table 4). For the same equation, there is no significant difference from a 0 intercept; the confidence intervals contain zero. This means that more than 95% of the time, similarly constructed intervals will contain unit 1 slope and 0 intercept. This indicates that the LL prediction equation validated against the independent data set. For PI (Eq. 2), the 95% confidence intervals about the slope of the regression line indicate a significant difference from unity; the confidence intervals do not contain one (Table 4). The slope of the regression line is significantly less than 1, indicating

---

**Fig. 1.** Scatter plot of measured versus predicted (A) LL and (B) PI. The prediction equations were developed using the entire data set with no data stratification (Eq. 1 and Eq. 2). The one-to-one slope lines are marked on the plots with solid lines.
that the predicted PI for the independent data set will be slightly overestimated. The degree of overestimation increases as the PI gets larger. Conversely, for the same equation, there is no significant difference from a 0 intercept. Overall, the PI prediction equation did not validate against the independent data set. The lack of validation may, in part, be caused by the relatively small validation data set (n = 466) compared with the large data set (n = 6592) that was used to develop the model. The linear models and validation results are optimized for the data set. If the data set compositions are changed (only slightly), a different slope and

<table>
<thead>
<tr>
<th>Prediction Eq.</th>
<th>Slope</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eq. 1 (LL)</td>
<td>0.996 (0.933 to 1.058)</td>
<td>2.607 (-0.070 to 5.284)</td>
</tr>
<tr>
<td>Eq. 2 (PI)</td>
<td>0.846 (0.767 to 0.924)</td>
<td>1.068 (-0.713 to 2.848)</td>
</tr>
<tr>
<td>Orders (LL)</td>
<td>0.996 (0.939 to 1.053)</td>
<td>1.669 (-0.841 to 4.059)</td>
</tr>
<tr>
<td>Mollisols (PI)</td>
<td>0.916 (0.827 to 1.005)</td>
<td>-0.247 (-2.325 to 1.831)</td>
</tr>
<tr>
<td>Alfisols (PI)</td>
<td>0.963 (0.804 to 1.122)</td>
<td>0.766 (-2.393 to 3.925)</td>
</tr>
<tr>
<td>Ultisols (PI)</td>
<td>1.005 (0.757 to 1.254)</td>
<td>-1.837 (-6.767 to 3.093)</td>
</tr>
</tbody>
</table>

Fig. 2. Scatter plot of measured versus predicted LL or PI for the (A) Orders (LL), (B) Alfisols (PI), (C) Mollisols (PI), and (D) Ultisols (PI) in Table 3. The Orders includes the Entisols, Aridisols, Alfisols, Mollisols, Inceptisols, Oxisols, and Ultisols. The one-to-one slope lines are marked on the plots with solid lines.
intercept can result. If a larger validation data set is used, one that more closely represents the original size that was used in model development, a different validation result could occur. However, these results suggest that PI is not easily predicted from readily available soil properties in Soil Survey when one prediction equation is used for all soils.

The measured versus predicted LL or PI values for four of the prediction models in Table 3 are shown in Figs. 2A–D. The 95% confidence intervals about the slope of the regression line for prediction of LL (soil orders equations) and PI for Mollisols, Alfisols, and Ultisols soil orders indicate no significant difference from unity; the confidence intervals contain one (Table 4). There is no significant difference from a 0 intercept for the same four prediction equations. These results suggest that LL and PI can be predicted from readily available soil properties in Soil Survey.

**SUMMARY AND CONCLUSIONS**

Clay content and CEC were the most highly correlated and most important independent variables in predicting both LL and PI. Organic C was not significantly correlated to either LL or PI. However, OC was a useful predictor variable in predicting LL for the Andisols and PI for the Entisols order when in the presence of other variables. The LEP was an independent important variable for predicting LL and PI in the Vertisols soil order. Carbonate clay content was not significantly correlated to either LL or PI. An LL and PI prediction equation was developed from the entire range in soils used in this study. The accuracy of predicting LL and PI was improved by stratifying the data set by taxonomic order. The $R^2$ ranged from 0.68 to 0.80 for prediction of LL and from 0.15 to 0.77 for prediction of PI. The PI of the Andisols was the most difficult to predict ($R^2 = 0.15$). Stratifying the data set by taxonomic family mineralogy did not provide a comprehensive coverage of soils (because of lack of data). Only equations for three of the most common mineralogy classes (mixed, smectitic, and kaolinitic) were developed. Based on the limited data and models, stratifying by mineralogy produced similar levels of prediction accuracy. In conclusion, predicting LL and PI from readily available soil properties (e.g., clay and CEC) resulted in mostly moderate to some strong prediction equations. Weak PI prediction equations resulted for the Andisols strata. Some of the better fit models ($R^2 > 0.60$) will be useful in Soil Survey when no measured data or better means of estimating the LL and PI are available. Other techniques such as the non-parametric nearest neighbor approach (Nemes et al., 2006) or other data stratifications may be needed as the next step in improving the prediction of LL or PI. Stratifying the data set by taxonomic family mineralogy class needs further exploration.

**REFERENCES**


to estimate soil water retention. Vadoze Zone J. 5:1222–1235.