The Natural Resources Inventory (NRI) is conducted annually. However, data preparation and estimation takes a considerable period of time; the 2012 NRI was released in 2015.

- There is therefore interest in obtaining NRI based forecasts in a more timely fashion, particularly state level estimates of some (combinations of) broad uses.

The overall NRI forecasting project involves a number of different models for different combinations of broad uses, which this talk will refer to as categories.

- The models detailed here are focused on just three categories.
As noted above, there are three categories of interest for this project. These are combinations of NRI broad uses:

1. **Cropland** (Combines broad uses Cultivated Cropland(1), Noncultivated Cropland(2) and CRP(12))
2. **Pastureland** (Combines broad uses Pastureland(3) and Rangeland(4))
3. **Forestland** (identical to broad use Forestland(5))
These three categories are combinations of broad uses that are mostly represented by real points in the NRI. Because the models used here are based on observed transitions of points, pseudopoints cannot be used.

▶ Recall that a pseudopoint is created to reflect (change in) certain broad uses that are collected at a segment level. A pseudopoint is given one imputed broad use for all years prior to its creation, and receives the broad use of the area it is representing going forwards.

▶ Pseudopoints thus can only ever have a single change in their history, and cannot themselves change again.

Because pseudopoints do not work well with this model, broad uses that are primarily represented by pseudopoints can’t be modeled successfully.
Basic Model Structure

This talk considers two models. However both share a basic structure:

▶ Both are discrete space Markov models

Additionally, both models can be written using a mostly common notation. This is elaborated on below.
Notation

Let \( i = 1, \ldots, I \) index the points within a region (one of the 48 continental states). Let \( t = 1, \ldots, T \) index years. Let \( k = 1, \ldots, K \) denote the categories.

Then let \( X_{i,t} \) be a random variable corresponding to the category of point \( i \) in year \( t \).

For purposes of this model, \( X_{i,t} \in \{1, 2, 3, 4\} \), where 4 is a catch-all ’Otherland’ category.

Category 4 allows points to have values other than Cropland, Pastureland and Forestland. There is no actual interest in predicting the amount of ’Otherland’.
Assume that the category of each point $i$ follows a Markov process, i.e. $X_{i,t} | X_{i,t-1}$ is independent of $\{X_{i,0}, \ldots X_{i,t-2}\}$.

**Figure:** Markov Assumption for NRI Points

\[
\begin{align*}
X_{i,t-1} & \rightarrow X_{i,t} & \rightarrow X_{i,t+1}
\end{align*}
\]
Under the Markov assumption, we only need to know the point’s state at time $t$ to predict $t+1$. Let

$$\pi_{kl} = P(X_{i,t+1} = k \mid X_{i,t} = l) \ \forall \ i, t$$

(1)

This assumes that points in the same category have identical transition probabilities.
Model 1 is based solely on the NRI. It is easily expressed via a transition matrix.

- Represent the population at each timepoint by a column vector, where each entry corresponds to the total in a category. In this case the population vector is

\[ V_t = \begin{pmatrix} C_{t+1} & R_{t+1} & F_{t+1} & O_{t+1} \end{pmatrix}^T, \]

where \( C_t, N_t, F_T, O_t \) are the counts of points in Cropland, Pasture+Rangeland, Forestland and Otherland, respectively.

- Define a matrix \( G \) to project the population forwards, so

\[ V_{t+1} = GV_t. \]
The *rows* of $G$ give the expected rate at which members of the different categories transition to the category corresponding to that row.

- The first row thus gives the mean rates at which Cropland, Pastureland, Forestland and Otherland become/remain Cropland.

The *columns* of $G$ give the fates of each category.

- The entries in the first column of $G$ thus gives the probability that a point in Cropland is in Cropland, Pastureland, Forestland and Otherland next year.
  - In the context of the NRI, the number of real points is fixed (ignoring the occasional point that is killed during estimation), so each column of $G$ sums to one.
Using the notation defined already, it is easy to represent Model 1 in matrix form.

\[
\begin{pmatrix}
C_{t+1} \\
R_{t+1} \\
F_{t+1} \\
O_{t+1}
\end{pmatrix} = \begin{pmatrix}
\pi_{1,1} & \pi_{1,2} & \pi_{1,3} & \pi_{1,4} \\
\pi_{2,1} & \pi_{2,2} & \pi_{2,3} & \pi_{2,4} \\
\pi_{3,1} & \pi_{3,2} & \pi_{3,3} & \pi_{3,4} \\
\pi_{1,1} & \pi_{4,2} & \pi_{4,3} & \pi_{4,4}
\end{pmatrix} \begin{pmatrix}
C_t \\
R_t \\
F_t \\
O_t
\end{pmatrix}
\]

All that remains is to estimate the elements of $G$. 
Two possible methods for estimating the elements of $A$ are as follows:

$$\hat{p}_{i,k,l}^{(1)} = \frac{n_{k,l,t}}{n_{l,t}}$$

(2)

Where $n_{k,l,t}$ is the number of real core points in category $l$ in year $t$, and category $k$ in year $t+1$, and $n_{l,t}$ is the number of points in $l$ in year $t$. Note that this estimate only uses the data from two previous years. A second method is

$$\hat{p}_{i,k,l}^{(2)} = T^{-1} \sum_{t=1}^{T-1} \frac{n_{k,l,t}}{n_{l,t}}$$

(3)

Which is simply the average of the estimators defined in (2). Note that (3) may need renormalized so the columns of $\hat{G}$ sum to one.

- For this presentation, all estimation for Model 1 is done using (2).
Model 2 attempts to use the Cropland Data Layer (CDL) as a source of external information.

- The CDL uses computer classification of satellite imagery. It provides classification of all land in the continental United States at a resolution of $30 \times 30$ meters.
- The CDL for a year is usually available by Spring of the following year. Since this is before the NRI for that year becomes available, the CDL may help predict future NRIs.
- The CDL is available for all states starting in 2009.

This model uses the classification of the CDL pixel closest to each NRI point as external information.

- The CDL uses a different set of land use categories than the NRI. For this model, the CDL categories have been mapped to the four NRI categories defined previously.
Model 2 incorporates CDL information by using two assumptions:

1. NRI classification is error-free; what the NRI says is at a point is exactly what is at that point.
2. The CDL classification for the pixel closest to point $i$ depends only on the land use for point $i$.

Combined, these assumptions imply that the CDL for point $i$ at year $t$ depends only on the NRI classification for point $i$ in year $t$.

- Let $Y_{i,t}$ be a random variable corresponding to the CDL classification of the pixel closest to point $i$ in year $t$. 
Figure: Markov Assumption for NRI and CDL

\[ X_{i,t-1} \rightarrow X_{i,t} \rightarrow X_{i,t+1} \]

\[ Y_{i,t-1} \rightarrow Y_{i,t} \rightarrow Y_{i,t+1} \]
By the Markov assumptions above, $Y_{i,t} | X_{i,t}$ is independent of all $\{Y_{i,0} \ldots Y_{i,t-1}, X_{i,0} \ldots, X_{i,t-1}\}$. Let

$$p_{hk} = P(Y_{i,t} = h | X_{i,t} = k) \forall i, t$$  \hspace{1cm} (4)

Under the previous assumptions, the logical choice for predicting the NRI one timepoint ahead is the distribution

$$P(X_{i,t+1} | X_{i,t}, Y_{i,t+1})$$  \hspace{1cm} (5)

- Recall that the CDL $Y$ is available before the NRI $X$, so this distribution make sense.
(5) may then be derived in parametric form.

\[
P(X_{i,t+1} | Y_{i,t+1}, X_{i,t}) = \frac{P(X_{i,t+1}, Y_{i,t+1} | X_{i,t})}{P(Y_{i,t+1} | X_{i,t})}
\]

where

\[
P(X_{i,t+1}, Y_{i,t+1} | X_{i,t}) = P(X_{i,t+1} | X_{i,t})P(Y_{i,t+1} | X_{i,t+1}, X_{i,t})
= P(X_{i,t+1} | X_{i,t})P(Y_{i,t+1} | X_{i,t+1})
\]

and

\[
P(Y_{i,t+1} | X_{i,t}) = \int P(X_{i,t+1}, Y_{i,t+1} | X_t) dX_{i,t+1}
\]

\[
\Rightarrow P(X_{i,t+1} = k | Y_{i,t+1} = h, X_{i,t} = l) = \frac{\pi_{kl}p_{hk}}{\sum_{k=1}^{K} \pi_{kl}p_{hk}}
\]
Model 2 may be estimated via maximum likelihood. One method is to set

\[
\pi_{kl} = \frac{\exp\{w_k^T \beta_l\}}{1 + \sum_{j=1}^{K-1} \exp\{w_j^T \beta_l\}}, \quad k = 1, \ldots, K - 1
\]

\[
\pi_{Kl} = \frac{1}{1 + \sum_{j=1}^{K-1} \exp\{w_j^T \beta_l\}}
\]

(7)

Where \( \beta_l^T = (\beta_{1l}, \ldots, \beta_{Kl}) \) is a vector of \( K \) regression parameters, and \( w_k^T \) is a vector of length \( K \) with a one in position \( k \) and zeroes elsewhere.

- A similar link function is used for the \( p \) parameters, but with \( \gamma \) regression parameters, and \( z \) indicator vectors for the sake of notational distinction.
Using (7) avoids having to maximize on a simplex, since the $\beta$ and $\gamma$ parameters can take values on the entire real line. Given ML estimates $\hat{\beta}$ and $\hat{\gamma}$, invariance of the MLE gives estimates of $\hat{\pi}_{kl}, \hat{p}_{kl}$. From these it is easy to calculate the estimated predictive distribution from (6);

$$
P(\hat{X}_{i,t+1} = k \mid Y_{i,t+1} = h, X_{i,t} = l) = \frac{\hat{\pi}_{kl}\hat{p}_{hk}}{\sum_{k=1}^{K} \hat{\pi}_{kl}\hat{p}_{hk}}$$
Weights in the NRI

NRI points each have a weight $w_i$, which may be (roughly) thought of as how much area that point represents.

- NRI weights are in units of 100 acres, so a point with a weight of 3 represents 300 acres, or about half a square mile.

Interest is therefore in predicting not points, but the total weight in a category for a state, i.e., how many hundreds of acres will be in Cropland in Iowa in 2015?
How to Incorporate Weights?

Two possibilities present themselves for incorporating weights in Model 2:

1. Classify each point to the category with the highest estimated probability, allocate all weight for that point to that category. The predictor is then
   \[ \hat{A}_k = \sum_{i=1}^{I} W_i \times I(k \text{ has maximal estimated probability for point } i) \]

2. Treat each point as representing a multinomial distribution with size \( W_i \) and \( K = 4 \) categories, use the models to estimate the vector of probabilities for each category and point, then predict the expected value of the multinomial. The predictor is then
   \[ \hat{A}_k = \sum_{i=1}^{I} W_i \hat{P}(X_{i,t} = k | \circ) \]

In practice, predictor (1) is equivalent to simply predicting no change. Predictor (2) is more consistent with the matrix projection model setup for Model 1 as well. We therefore consider predictor (2) here.
As noted, these methods are meant to be used for real core points. Since this excludes points in the rotation panels, and pseudopoints, the total weight used by these models is less than the total weight for each state.

- NRI weights are calibrated so that the sum of all weights for a state equals the total area of that state.

A simple fix is to ratio adjust the weights for real core points to sum to the state total. Define

\[ W_i^* = R \times W_I, \]

\[ R = \frac{\sum \text{all points } j W_j}{\sum \text{real core points } i W_i} \]  

(8)
Models 1 and 2 were both fit to real core points from Iowa. Model 1 was fit using (2) with data from 2010 and 2011. Model 2 was fit using data from 2009 - 2011.

- Both models were used to predict 2012.
The following table gives estimates for the 2012 NRI under Model 1 and Model 2. The ‘NRI 2012’ column gives the sum of weights in real core points, and estimates for both models use unadjusted ‘raw’ weights.

<table>
<thead>
<tr>
<th>Category</th>
<th>Model 2 Real Core</th>
<th>Model 1 Real Core</th>
<th>NRI 2012 Real Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cropland</td>
<td>37411.977</td>
<td>37340.119</td>
<td>37803</td>
</tr>
<tr>
<td>Pastureland</td>
<td>4992.639</td>
<td>5034.980</td>
<td>4757</td>
</tr>
<tr>
<td>Forestland</td>
<td>3782.692</td>
<td>3800.528</td>
<td>3599</td>
</tr>
</tbody>
</table>

The root sum of squared errors (RSSE) for Model 1 is 576.32, and 492.11 for Model 2.
Results for Illinois

So everything is as expected right? Model 2 outperforms Model 1 by using an external source of information.

Table: Raw Weighted Estimates for Illinois

<table>
<thead>
<tr>
<th>Category</th>
<th>Model 2 Real Core</th>
<th>Model 1 Real Core</th>
<th>NRI 2012 Real Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cropland</td>
<td>36728.178</td>
<td>36717.194</td>
<td>36593</td>
</tr>
<tr>
<td>Pastureland</td>
<td>3216.847</td>
<td>3239.688</td>
<td>3380</td>
</tr>
<tr>
<td>Forestland</td>
<td>6192.482</td>
<td>6167.369</td>
<td>6178</td>
</tr>
</tbody>
</table>

The RSSE for Model 2 is 212.4, and 187.7 for Model 1.

- Model 1 predicts 2012 more accurately than Model 2 for 38/46 states
  - Alaska and Hawaii are not considered.
  - Numerical maximization failed for Pennsylvania and New York so Model 2 was not fit to these states.
RSSE for Models 1 and 2
Model 2 relies on the CDL, which turns out to be fairly good at detecting cropland, but has a fairly high error rate for other categories.

- Some of this is clearly misclassification by the CDL.
- Some of this is also due to categorical mismatch between the NRI and CDL. For example the NRI defines parks in the middle of cities as Urban land, the CDL is likely to see grassland or trees.

**Table: NRI, CDL correspondence for Iowa 2011**

<table>
<thead>
<tr>
<th>NRI</th>
<th>CDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cropland</td>
</tr>
<tr>
<td>Cropland</td>
<td>1806</td>
</tr>
<tr>
<td>Pastureland</td>
<td>140</td>
</tr>
<tr>
<td>Forestland</td>
<td>8</td>
</tr>
<tr>
<td>Otherland</td>
<td>29</td>
</tr>
</tbody>
</table>
Recall that the sums of estimates across all four categories will not equal the state total. Ratio-adjusting the weights provides an easy fix for this.

- However if rotation points and pseudopoints behave differently than real core points, this may result in a biased estimate.

Possible fixes include further ratio adjustments, or a more sophisticated method of controlling to state totals.
Table: Model 1 Estimate under scaled weights for Illinois.

<table>
<thead>
<tr>
<th>LCU</th>
<th>Model 1</th>
<th>NRI 2012 Scaled Weights</th>
<th>NRI 2012 Real Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cropland</td>
<td>223556.000</td>
<td>222799.831</td>
<td>211780</td>
</tr>
<tr>
<td>Pastureland</td>
<td>19725.137</td>
<td>20579.439</td>
<td>18920</td>
</tr>
<tr>
<td>Forestland</td>
<td>37550.589</td>
<td>37615.319</td>
<td>33933</td>
</tr>
</tbody>
</table>
Variance estimation for Model 1 is straightforward. Note that although it uses a Markov model, the actual estimators are in essence design estimators. This means that it is reasonable to use standard survey variance estimation techniques for Model 1.

- In particular, the NRI produces $B = 29$ Jackknife replicate weights for each point. This provides a simple method for calculating a design variance for Model 1’s estimates of state totals.

- A second possibility is to compute the Jackknife variance estimates in a more traditional fashion, by deleting each observation in turn, recomputing the transition probability estimates and from them the estimated totals, then taking the variance of these replicate totals.
Variance estimation for Model 2 is somewhat more complex. One possibility is to use the delta method and asymptotic normality of the MLE to derive the asymptotic variance-covariance matrix of the estimates from that of the $\beta$ and $\gamma$ regression parameters. However this requires both some nontrivial calculus, and the validity of asymptotic normality for the estimates themselves may be questionable.

- A more attractive option for Model 2 is the parametric bootstrap. Simulating from this model is fairly easy, conditional on $X_0$. However the model is slow to fit, and given its generally poor performance compared to Model 1, it may simply not be worth the time.
Conclusions

▶ Model 1 offers a very easy and accurate way to predict the NRI one year in advance.
▶ Model 2 allows for the use of alternative sources of information. However the CDL has too many errors for it to be useful in most states. Use of county level effects may allow for better performance.
▶ Although Model 1 performs well,