During this lecture we will talk about a few concepts. First we will cover model evaluation and validation. This topic is one that is in constant debate within the scientific community and we will read a paper from Oreskes in which she attempts to tackle this debate. We will then focus our attention on model uncertainty.
I would like to begin by reviewing the purpose of a model prediction is to reduce the range of likely answers to “what if” questions. For example, the purpose of modeling the effects of silvicultural treatment on fire behavior is not to demonstrate how bad or good our plan is, but to try and see what is most likely to happen if we implement a particular treatment. With this definition in mind, it is easy to see how this fits into the prediction process we discussed last week.
From this point we can explore how the validation and evaluation debate has come to be. Let’s begin by looking at two definitions of valid. The first states that being valid is without obvious flaws or defects, the second suggests that valid implies being supported by truth.

The differences in these two meanings are of great importance. The first meaning is what the scientific community typically is referring to when they suggest that a model is valid. That is to say, it appears to have no major flaws, it has been tested against independent data, and it’s shown to confine to some standard of statistics. A key point under this definition is that just because a validated model produced accurate results once it may not do so again under slightly different conditions.

The second definition is typically how the general public thinks about the term valid. That is to say when we use the term valid we are implying (to the public) that a model is supported by truth.

If we use the first definition of valid it is easy to see that a model prediction would have little relevancy in assisting a decision. In fact, several management or policy agencies have defined “validation” as to imply it will accurately represent the system being modeled. From this standpoint, these agencies are treating the prediction as a piece of information and not as part of a bigger process. When we use the term “valid” to mean there is a truth to the model, we are implying it is absolutely correct (there should never be any deviation from the suggested outcome), so we need to think about what we say and how we communicate.
This idea of validation has come into debate among the scientific community in recent years. Dr. Karl Popper argued that we can never prove a scientific model or theory right - we can only prove it wrong. The implications of using the term “validation” in models goes against Popper’s theory; “evaluation” is a more scientific way of assessing a model. Using this theory we can only show that under a given set of conditions our model met a standard of accuracy we have defined. This does not prove the theory; it only suggests that it works within our acceptable limits for this case.

In addition, there is also the issue that a model cannot be validated because there are known flaws in every model. Even in a lab setting we will make known assumptions about a real system and include these into the model. For example, we may assume a frictionless surface.

With these two points of view it is easy to see why the term validation is debated. Instead of this term many have proposed to use the term evaluation. Evaluation implies an assessment in which both positive and negative outcomes are possible, where as validation implies only a positive outcome.

In fact several authors have argued that we should not ever validate a model, and the money spent on this could be better used in another way. Although this is an interesting point, a process of model evaluation is essential, particularly in areas where the predictions will be used to make a decision. It is in our best interest as decision makers to know as much about a prediction as we can.
Let’s start to think about how a model should be evaluated. Many have suggested that a model should always be evaluated in terms of its purpose, context and domain. When considering these aspects of a model there are two parts we should evaluate. First, look at the model composition, which is the manner in which the model formulates hypothesis. In other words, does the model show reasonable behaviors and process? This is the internal review of the model; we want to know if this model fits with our understanding of the real world.

Second, we must consider model performance, which is the acceptability and usefulness of the model predictions for an intended task. This is a little harder to do since many models are developed for a scientific purpose and not for a management purpose in mind, and even if there are studies evaluating this it is likely that not every purpose for every set of criteria has been tested.
In most cases there are three components that are conducted in a model evaluation. These are scientific peer review, quantitative analysis of model predictions compared to field observations, and sensitivity analysis.

We will not talk any further about peer review, and for now you should be aware that sensitivity analysis shows the degree to which the model prediction is affected by changes in selected input parameters. The final part of the evaluation is the comparison between predictions and field data. We will cover this idea in more detail in another lecture so let's move on.
Uncertainty simply means that, given our current knowledge, there are multiple possible future states of nature. Typically uncertainty can be broken down into two types as I mentioned in the past lectures:

- Aleatory uncertainty – reflects the nature of random processes
- Epistemic uncertainty – is incomplete knowledge of processes that influence events

Total uncertainty - the sum of epistemic and aleatory uncertainty

Aleatory uncertainty reflects the nature of random events, and is studied using statistics.

There is also epistemic uncertainty, which is incomplete knowledge of processes that influence events. This incomplete knowledge results from the complexity of the world in which we are dealing with. The total uncertainty will be a combination of both the aleatory uncertainty and the epistemic uncertainty.
These three cases show a hypothetical plot where the prediction is put on the x axis and the actual event is placed on the y axis. For example, the event we could be predicting might be flame length during a crown fire.

The top left example has a correlation coefficient of .20
The bottom picture has a correlation coefficient of .50
The top right picture has a correlation coefficient of .80

A correlation coefficient measures the strength of a linear relation between x and y. The higher the correlation, the better x predicts y. This is just one example of how we can measure uncertainty.

In the case of the top right picture the results are better correlated with the actual event; thus the predictions are better and more desirable. However, in terms of the prediction process, any one of these models might have lead to a good decision and would therefore be okay to use in a given case.
The following table lists several common methods used to explain aleatory uncertainty. You have probably encountered several of these statistical methods; in particular, the r squared value and 95 percent confidence intervals are often used in scientific papers. It is outside the scope of this class to review the assumptions and methods used in each of these tests but you may find information about these in any statistics text book or on the internet.
Due to the nature of our work (land managers), I want to point out that the very use of a scientific prediction allows for the following situation. Model predictions, the methods that were used to generate the prediction, the data input into those methods, the application of the prediction, and the decisions we generate from the prediction will potentially be under constant review and scrutiny by the public, the scientific community and the judicial community.
Of particular importance to the use of models as a tool in our planning is the Data Quality Act. This act requires that nearly all federal agencies prepare guidance to maximize the quality, objectivity, utility, and integrity of information the agency publicizes or disseminates.

The guidelines in the Data Quality Act affect many processes, such as documents created under the following:
- National Forest Management Act
- National Environmental Policy Act
- Endangered Species Act

These guidelines are far-reaching in their effects on how we do business in natural resources. The data quality act will effect any analysis, we conduct under the following: National Forest Management Act, the National Environmental Policy Act and the Endangered Species Act.
The USDA guidelines which cover the Data Quality Act state that the agency will:

- Use sound analytical methods
- Use reasonably reliable and timely data
- Ensure transparency of analysis and documentation of data sources, including uncertainty and limitations
- Where appropriate, employ external peer review
- Document models and other estimation or forecasting techniques to describe the data sources used and methodologies and assumptions employed

The USDA guidelines which cover the data quality act state that all agency information will: use sound analytical methods, use reasonably reliable and timely data, ensure the transparency and accurate documentation of models, data sources, uncertainty, limitations, and methodologies, and, where appropriate, employ external peer review as cited in the Office of Management and Budget recommendations. Notice that these guidelines are very similar to the keys to improving the decision making process we have discussed already and to the keys to model evaluation.

In your reading it states that there have been 11 data quality act requests reported by the USDA Forest Service. I have checked the web site and there are still only 11 although this has not been updated in almost 2 years. So for now we will assume that none of these have questioned the use of a particular model.
I would like to review the two court cases covered in your readings as I think these are very important in understanding and putting context to the recommendations we have been discussing so far in class. We will also use these court cases as a basis for our discussions this week.

The first case I would like to talk about is the Sierra Club v. USFS in 1993. As far as I can find in the literature and with an internet search this is the only court case where an agency has been questioned about its dependencies upon a model in the decision process.

The result of this case provides some validity to the recommendations we have discussed so far in this class. Specifically that our assumptions, limitations and uncertainties need to be openly discussed, and that we should not base our decision on the model itself.
In the state of Ohio v. the US EPA (1986), the EPA was sued for failing to demonstrate the accuracy of a computer model “without adequately validating, monitoring or testing its reliability.” The court ultimately ruled in favor of the state, declaring that, in order for a model to be useful, it must accurately predict the behavior of the system being modeled.

Also interesting to this class is that a poor decision was made as a result of using the model, and human health was put at risk. Had the EPA followed the guidelines we have laid out, would the model have been called into question? I would argue that, had the model assumptions, limitations and uncertainties been communicated and used within the prediction process, a different decision might have been made by the EPA, and this case never would have gone to court.

Regardless, it leaves us with the question as to how much site-specific testing and monitoring must occur to meet legal and community standards.
<table>
<thead>
<tr>
<th>Review and Recommendations for the Use of Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Model assumptions, limitations, and uncertainties should be stated</td>
</tr>
<tr>
<td>• The decision-making variables should not only depend upon the model and should be disclosed</td>
</tr>
<tr>
<td>• The degree of scientific acceptability should be reported</td>
</tr>
<tr>
<td>• A model should go through peer review</td>
</tr>
<tr>
<td>• Empirical testing of the model should be described</td>
</tr>
</tbody>
</table>
As we conclude our introduction to evaluation and uncertainty, I would like to conclude with the following thoughts.

The purpose of modeling is to reduce the range of likely outcomes to a set of “what if” questions.

We should try to use the term “evaluation” and not “validation” when we review models.

Models should be open (transparent) and we should report the model assumptions, limitations, and uncertainties in terms of the intended use and within the larger prediction process.