Lesson Title	Creating Remote Sensing Models from Data
Length of Lesson	1 day
Created By	Matthew A. Lee, Henry Stauffenberg, and William Funderburk
Subject	Physics
Grade Level	11-12
State Standards	Physics: 1
DOK Level	DOK 4
DOK Application	Critique, analyze, design
National Standards	9-12: B (Physical Science)
Graduate Research Element	Choosing where to sample ground truth data given limited time constraints and using that limited ground truth data to develop the best general model.

Student Learning Goal:

Physics: 1. Apply inquiry-based and problem-solving processes and skills to scientific investigations.

a. Use current technologies such as CD-ROM, DVD, Internet, and on-line data search to explore current research related to a specific topic.
b. Clarify research questions and design laboratory investigations.
c. Demonstrate the use of scientific inquiry and methods to formulate, conduct, and evaluate laboratory investigations (e.g., hypotheses, experimental design, observations, data analyses, interpretations, theory development).
d. Organize data to construct graphs (e.g., plotting points, labeling x-and y-axis, creating appropriate titles and legends, for circle, bar, and line graphs) draw conclusions and make inferences.
e. Evaluate procedures, data, and conclusions to critique the scientific validity of research.
f. Formulate and revise scientific explanations and models using logic and evidence (data analysis).
g. Collect, analyze, and draw conclusions from data to create a formal presentation using available technology (e.g., computers, calculators, SmartBoard, CBL's, etc.).

National Science Education Standards of Content 9 – 12

B (Interactions of Energy and Matter): Each kind of atom or molecule can gain or lose energy only in particular descrete amounts and thus can absorb and emit light only at wavelengths corresponding to these amounts. These wavelengths can be used to identify the substance.

Materials Needed (supplies, hand-outs, resources):

Students and the teacher will need to use Excel (data provided, see teacher's notes) and some program to view images. In class, the students can be put into groups and each group will need a computer.

Lesson Performance Task/Assessment:
Students will learn how to make the best use of limited time to collect ground truth data that is most representative of the general phenomenon. Students will use Excel to create a linear regression model based on the ground truth data they collect, and test their model. They will learn how to present and defend their model.

Lesson Relevance to Performance Task and Students:
The students will learn how to design experiments, and how they should sample data to create the best model.

Anticipatory Set/Capture Interest:
You (the students) are researchers at TurfTech, an imaginary company that uses remote sensing and other technology to improve the quality of landscaped turf. They have clients such as golf courses, sports fields, parks, and the government. In the course of your research, you discover that NASA has developed a sensor that can sense the health of plants and deployed it in a satellite. NASA is mainly interested in studying the health of the Earth's rainforests, but they have some imagery distributed in various places around the world that they provide for free, and they will capture images of any place in the world for a fee of $15.00 per square mile. You mentioned this to your boss (the teacher), and he is excited to hear about this, and asks if you could develop a system that uses this satellite to determine where grass needs to be watered. TurfTech currently determines this by having people walk through the grass because vehicles would damage the grass too much. This typically requires 2.5 man hours per square mile, so the company stands to profit a lot since it pays $8.00 per hour for this work. Fortunately, your boss is able to convince his boss, the CEO, to give enough funding to purchase a single image from NASA and a maximum of 20 ground truth samples. He asks you collect ground truth in the form of soil moisture samples, and analyze the data to see if your method will work. Also, the CEO wants you to prepare a presentation of your findings and recommendations, and present it to him when you finish your analysis. This is your moment to shine. If you do well, your CEO and boss will be very pleased and there may be a promotion for you, but if you do poorly, your CEO and boss will not be very happy with you for wasting the companies money.

Guided Practice:
The class should have a discussion about what questions the CEO might ask.

- Will this even work?
- If it does work, why does it work?
    - These two questions should be answered by considering that plants generally grow better when there is enough soil moisture in the dirt they

are growing in, and poorly when there is not. So the health of the plants is likely correlated to the soil moisture.

- Will this make money for the company?
- Will the investors like the idea?
- What should he tell the customers to justify using this technique?
    - This can only be justified if the expected utility of the new technique is greater than the utility of doing it the old way. The expected utility will depends on several factors that this lesson does not cover, but they are:
        - The money saved by not paying people to walk through the grass minus the cost of the satellite.
        - The cost of the water saved when the system correctly detects the soil doesn't need watering minus the cost of the water when the system detects dry soil when it is moist.
        - The value of the quality improvement or degradation.

- Is the health of the grass related to the soil moisture?
    - The students will try to determine this, but it is good to start think about.

- What kinds of things can fool the sensor?
- Can we expect the health of different plants to correlate to soil moisture the same way grass health does? How much soil moisture does a cactus need to be healthy, and how much soil moisture does grass need to be healthy?
    - This is an open ended question for the students to ponder. There are perhaps an infinite number of things that can go wrong, but only a few are very likely. One common thing that can fool a satellite is atmospheric conditions. The description never said whether the satellite can see through clouds or pollution. Another thing to think about is that you can have wet or dry soil and not have any plants. Finally, the second question brings up the point that different plants are healthy with different soil moisture concentrations.

- Where should we collect our ground truth so that we get the most accurate model of the relationship between grass health and soil moisture?
    - Since we are trying to develop a linear model, we want to try to get a wide range of soil moisture. If all our data points are close together, it may be possible to fit a line through our data that does not match the general trend. Also, since we are only interested in correlating grass health to soil moisture, we should confine our samples to grassy areas. Also, the size of the pixel from the satellite might be large, so we want to make sure that any pixel that includes our sample can only contain grass. This means we should avoid boundaries of the grass.

- How can we tell how accurate our model is?
  - To do this, the students will likely need more statistics than they can learn in this lesson and they will probably not have enough data. However, we can measure the error in our model. Often, error is measured by squared error, which is $error^2 = \sum_{i=1}^{n}(y_{measured} - y_{predicted})^2$. A second technique comes from assuming that soil moisture (y) is related to plant health (x) by the model $y = mx + b + \varepsilon$, where m and b are the slope and y-intercept of a line and ε is a random variable with zero mean and some unknown variance. We can estimate the variance of ε, which is $var(y_{measured} - y_{predicted})$.

Independent Practice:
The students will use the Excel file with the data to collect 20 "ground truth" points by copying and pasting data from the NDVI image and the soil moisture table in the Excel file. The file has a spread sheet for them to place their data on, and when they do, a scatter plot will be filled in, and a trend line will be computed. The students can then use the trend line function to compute the predicted values by inputting the function into cells. Then they can use this result to compute the variance of ε and $error^2$. The students will then be asked to explain they think their soil moisture predictions are good or bad, and explain why they chose their ground truth points. The students need to be careful to copy the same pixel/ground truth point location for their data points. The coordinates on the image are the same as the coordinates for the ground truth points.

Remediation and/or Enrichment:
Remediation: individual IEP; partner help throughout the lesson; the teacher can observe the students and intervene during the independent practice.

The students can make PowerPoint presentations of their work.

Check(s) for Understanding:
The teacher should use informal checks such as questions and discussions to check for understanding.

The most important thing to check for is that they can explain variance, error, and how data points should be chosen.

Closure:
The students should present what they discover about the data. You can discuss with them reason why they might have gotten different results for variance and error.

Possible Alternate Subject Integrations:
Statistics

Teacher Notes:
The data I used in designing this lesson comes from an aircraft mounted sensor called Hyperspectral Digital Imagery Collection Experiment (HYDICE). It is an image of the Washington, D.C. Mall. Some of the kids might notice this. You can see the Capital building at the bottom of the image, and the Lincoln Memorial at the top of the image. The Washington Memorial is in the middle of the image. The buildings show up as black areas because they are not plants, but you can see the shadow of the Washington Memorial in the NDVI image. I have included JPG files to help understand the Excel file. There is a real color image and a NDVI image. I recommend that the students be instructed to use the grassy area around the Washington Memorial to collect data, but you are welcome to use any area you want. I have noticed that there is a green area that appears to be grass in the real color image above the Capital building, but it is black in the NDVI, which means it is not grass. (I hope I didn't just out some national secret.) I strongly recommend that you not use this area. Also, you might want to tell the students that the NDVI image is real, but the soil moisture data is fake.