RESET SEND

RESEARCH PROPOSAL 2015/2016

DUE DATE: WED., NOVEMBER 19, 2015

COMPLETE ALL 3 (THREE) PAGES.

Before starting a research project, a proposal must be submitted so that your project meets all requirements and guidelines of proper research. Your proposal will be submitted under the headings below. This proposal is interactive which means you will type directly on the form to provide the information needed. You will then print out the form. ***DO NOT*** **complete by hand. Print a copy for your records. TO SAVE, SAVE TO YOUR COMPUTER AS A BLANK COPY, THEN FILL IT OUT.** ***THIS FORM MUST BE COMPLETED PRIOR TO STARTING THE EXPERIMENTATION.***

Student's Name (Last, First, Middle): Kim, Nathaniel Jongmin

Period: ☐ 6 ☐ 7 ☑ 8 **Grade:** ☐ Freshman ☑ Sophomore ☐ Junior ☐ Senior

Proposed Project Title Predicting El Niño Indicators using Recursion, Data Trees and Evolutionary Algorithms

Category:

- ☐ Animal Sciences
- ☐ Behavioral & Social Sciences
- ☐ Biochemistry
- ☐ Cellular and Molecular Biology
- ☐ Chemistry
- ☑ Computer Science
- ☐ Earth Science
- ☐ Engineering
- ☐ Energy & Transportation
- ☐ Environmental Science
- ☐ Mathematics
- ☐ Medicine & Health Sciences
- ☐ Microbiology
- ☐ Physics and Astronomy
- ☐ Plant Sciences

Indicate if your project involves the following (check "YES" or "NO" for each). The page number refers to the page in the INTEL Rules and Guidelines Booklet. Refer to the booklet to get a description of each. The full booklet is at the website: http://www.societyforscience.org/Document.Doc?id=10

○ YES	◉ NO	Human Subjects	P.7
○ YES	◉ NO	Vertebrate Animals	P. 10
○ YES	◉ NO	Potentially Hazardous Biological Agents	P. 13
○ YES	◉ NO	Hazardous Chemicals	P. 17
○ YES	◉ NO	Hazardous Devices	P. 17
○ YES	◉ NO	Radiation	P. 17
○ YES	◉ NO	Prescription Drugs	P. 17
○ YES	◉ NO	Alcohol and tobacco	P. 17

Is this a continuation of a previous project? ☐ YES ☑ NO

Problem Statement or Scientific Question to be Answered:

Can predictive recursive models of El Niño indicators be developed with the use of data tress and an evolutionary algorithm?

Independent variable(s): List the factor(s) that you will control or manipulate. If more than one, put a star ***in front of*** the one(s) that will most likely be relevant to your research.

General structures of the algorithm and model; which factors (e.g. SOI, SST indices) are incorporated in the algorithm and model

Dependent variable(s): List the factor(s) that respond to the control or manipulation. These factors produce collectable data.

Specific model resulting from the algorithm; efficacy and accuracy of the model

Control (if applicable): Describe the standard of comparison you will use to assess experimental effects.

Other predictive models of El Niño currently in use; various evolutionary algorithm models incorporating different factors

PRESS BUTTON TO GO TO PAGE 2

Your Hypothesis/Engineering Goal: Formulate one or more statements that predict the possible outcome(s) of your research. These statement(s) should describe how changing the independent variable will affect the dependent variable.

El Niño indicators can be predicted to some degree using a data tree involving an evolutionary algorithm, and these predictions can be made more accurate by incorporating more factors.

Rationale and Justification Make a personal statement of why this problem is of interest (**the "SO WHAT?" factor**). If using vertebrates as test subjects, justify their use and explain the humane manner in which they will be treated. ***No vertebrates may be sacrificed for the sake of research.***

New prediction models of El Niño, which is very important to prepare for, are always valuable. Also, evolutionary algorithms have seen success in many other areas and I want to try to expand its potential use.

Number of Samples/Trials: 100+

Expected Results:

I expect that my model will predict various indicators and characteristics of El Niño to some degree.

Location of Your Research Work:

My house

Special Requirements: If performing your research at a site other than school or home, state how you expect to travel to your research site. Will a parent or guardian be driving and have they agreed to transport you as necessary? Are there any other special requirements that you will need or anticipate?

--

Mentor information: If you are working with a mentor, give the following information:

Mentor's name: --

Title: _____

Phone number: _____

E-mail: _____

Give me a brief outline about how you made contact with your mentor and how a working relationship developed that led to your working with your mentor.

PRESS BUTTON TO GO TO PAGE 3

Procedure: Use numbers to indicate the sequential steps of your procedure. Include the materials and equipment, specifying quantities, concentrations, amounts, time, etc. ***THIS SECTION SHOULD BE SO DETAILED THAT ANOTHER PERSON SHOULD BE ABLE TO DO THIS RESEARCH FOLLOWING YOUR DIRECTIONS. IT IS LIKE A LAB MANUAL THAT YOU USE IN CHEMISTRY, BIOLOGY, OR PHYSICS.*** Be as specific as possible, yet recognize that changes are likely to occur. Include the length of time the experiment or field research will take.

Let f(t) = SOI at a month t. Thus f(t-k) equals the SOI k months before the time t.

The models were represented using a data structure known as a tree. The root of the tree is an empty node, and all other nodes consisted of an operator and a value. The set of all operators was +, -, *, /, and ^. Values could either be constants, such as 0.5, 1, or 2.97, or previous SOI values such as f(t-60).

Each path directly from the root to a node represented a model. For example, if the nodes in such a path were "" (the root), "+f(t-1)", "*2", and "+f(t-3)", in that order, then the model was given by f(t) = f(t-1) * 2 + f (t-3). Note that the operator of the second leaf is disregarded.

The program, written in Python 3.4.3 for Windows 7, did the following:

1. The program first fetched monthly SOI values from 1991 to 2015 from the following link: https://www.longpaddock.qld.gov.au/seasonalclimateoutlook/southernoscillationindex/soidatafiles/MonthlySOI1887-1989Base.txt. To develop the models, only values from 1991 to 2010 were used as training data.

2. The program then generated a tree containing 100 random models with very low depth (no more than 2).

3. The program tested each model against the training data and calculated the accuracy of each one using an average of RSS (residual sum of squares), RMSE (root-mean-square deviation), and MSE (mean squared error). The accuracy of the most accurate model was recorded.

4. The program deleted the branches of the 50 worst models (i.e., those which have the lowest accuracy).

5. To replace these 50 models, the program duplicated the best 35 models and created 15 new models by attaching leaves to pre-existing nodes. The duplication was similar to reproduction, and the creation was similar to mutation.

6. Next, the program performed approximately 10 crossovers. A crossover consisted of swapping an entire subtree with another subtree, while still keeping these subtrees and the rest of the tree intact. This process preserved the number of leaves in the tree and thus the total number of models as well.

7. The program performed approximately 10 mutations. A mutation consisted of replacing random nodes in the tree having sufficiently high depth with new, randomly generated nodes. Again, this process also preserved the number of leaves in the tree and the total number of models. Note that these mutations differed from those described in step 4.

8. Steps 3 through 7 were repeated 99 more times in order to simulate 99 more generations (100 generations total).

9. Finally, the most accurate model and its accuracy was recorded. Furthermore, this model was compared against the data from 2011 to 2015, and its accuracy at predicting the SOI during that five-year stretch was also recorded.

The program took about one hour to run.

(Note: I will attach source code to my final project.)