

Prediction of MPB outbreaks using Markov process logistic regression

Past and future project milestones

Mario Pineda

MPK Analytics, Inc.

Suite 300, 8507 - 112 Street

Edmonton, AB T6G 2L7

Canada

Version: 2012-10-17 13:40:00

Abstract

This vignette provides an abbreviated log of milestones that have been reached and that still need to be reached.

About this document

This document was created using Sweave, a literate programming tool for R. To generate this documentation (i.e. weaving) may require a UNIX platform to allow for system calls. Weaving may take a long time. To speed up repeated weaving intermediate data and figures may be cached. For system requirements see end of document. This document (incl. the source) is available at <https://github.com/mariopineda>.

Copyright (c) 2012 Mario Pineda

MPK Analytics, Inc.
Suite 300, 8507 - 112 Street
Edmonton, AB T6G 2L7
Canada

Ph: 866.406.0187
Fax: 780.800.6929
Email: info@mpkanalytics.com

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Admin

[**DONE**] Set up of password protected backup (sirius) for the project and data.

[**DONE**] Look into running the model under the Enterprise R version using the lab HPC infrastructure. Explored options together with Mark Duller but model ran slower and this approach was abandoned in favour of optimizing the model to run on my laptop.

[**IN PROGRESS**] Clean up and organize project files and data backed-up located on sirius. This is the version of the project that will be handed over to Devin. Mario will not continue working on this version of the project.

[**IN PROGRESS**] Develop a series of R packages and continue development of project.

Coding

[**DONE**] Develop Empirical Bayes model in C++.

[**DONE**] Computational optimization of predictive model.

Modelling

[**IN PROGRESS**] Perform model selection. We would like to use BIC. Mark had an argument for why BIC would be preferred over AIC in this case. Request a copy of AP's write-up on the virtues of BIC.

[**TODO**] Run model using the CSI variables (i.e. by replacing the temperature variables).

[**TODO**] Run model using temperature predictions (from Subhash Empirical Bayes).

[**TODO**] Run model for 100% of BC.

[**TODO**] Figure out what results we are interested in, i.e. what message should the manuscript have? There seems to be two alternatives, (1) along the lines of the old manuscript (i.e. predictive model) or (2) focusing on the role of winter temperature on the spatio-temporal outbreak dynamics. The second option may be of more general interest.

[**TODO**] Calibration of the empirical Bayes temperature prediction model.

Data

[**DONE**] Obtained raw red top data from Fangliang He with an email confirmation its Open Data status. Cleaned up, prepared and analyzed red top data. (See vignette)

[**DONE**] Clean up, preparation and analysis of weather station data that was obtained from Tomas. (See vignette)

[**DONE**] Interpolations of temperature variables (annual degree days and minimum winter temperature). (See vignette)

[**DONE**] Generated summer Climate Suitability Indices (CSI) data with Barry using the Powell and Logan recipe.

[**DONE**] Obtained new DEM data and BC outline data from GIS Analyst. Prepared and analyzed the data. (See vignette)

[**DONE**] Obtained pine cover data from Barry, GIS Analyst performed preprocessing and Mario did the final clean up and analysis.

[**TODO**] Check on CSI data integrity. BioSim flaked out over parts of BC in the initial runs, need to check that the corrupted data has been removed and replaced by the correct version.

[**TODO**] Fix data inconsistencies at the edges of the DEM. This will have no effect on the results but is required for the model to run for all of BC.

[**ON HOLD**] Follow up on CSI data conversion with GIS Analyst.

[**TODO**] It appears the `nn1` and `nn2` variables include the focal location in the summation. Figure out how to deal with this.

[**TODO**] For the manuscript we need to update red top data and temperature data for post-2003 years and feed model with it.

Session information

- R version 2.15.1 (2012-06-22), i386-apple-darwin9.8.0
- Locale: C/en_US.UTF-8/C/C/C/C
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Loaded via a namespace (and not attached): tools 2.15.1

The R code in this chapter took 0 seconds to execute, the last time the source file was modified was on 2012-10-17 13:40:22 and the last time the PDF file was modified was on 2012-10-17 13:40:00.



MPK Analytics — Transforming Data Into Insight
Suite 300, 8507 - 112 Street, Edmonton, AB T6G 2L7, Canada
Ph: 866.406.0187, Fax: 780.800.6929
Email: info@mpkanalytics.com