

LA-UR-15-24598

Approved for public release; distribution is unlimited.

Title: 2015 IC Report: w14_biosurveillance

Author(s): Fairchild, Geoffrey
Del Valle, Sara Yermimah
Priedhorsky, Reid
Mniszewski, Susan M.

Intended for: Report

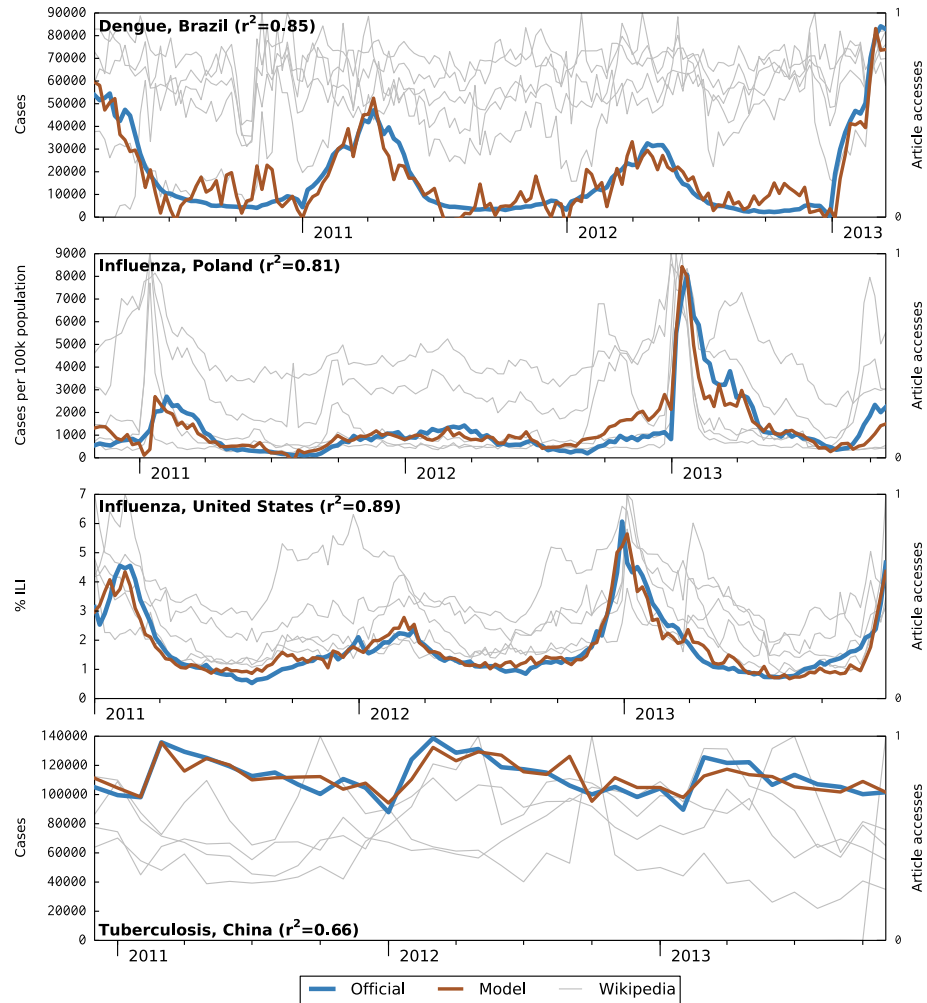
Issued: 2015-06-18

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

2015 IC Report: w14_biosurveillance

- Our primary use of the institutional computing clusters has been for Wikipedia-based disease surveillance and forecasting work.
- We've expanded *breadth*: the number of diseases and locations that can be surveilled using internet data.
- We've expanded *forecasting*: surprisingly little work towards forecasting disease spread with quantified uncertainty.
- We've expanded *big data processing techniques*: the Wikipedia access logs are 5TB of compressed text; we're working on efficient ways to process and store time series.
- Our work received significant media coverage and was highlighted as LANL's best science story of 2014.



These graphs show official epidemiological data and nowcast model estimate (left Y axis) with traffic to the five most-correlated Wikipedia articles (right Y axis) over the 3 year study periods. The Wikipedia time series are individually self-normalized.