T-Cube Web Interface
in Support of Real-Time Bio-surveillance Program

Artur Dubrawski [1], Maheshkumar Sabhnani [1], Michael Knight [1], Michael Baysek [1], Daniel Neill [1], Saswati Ray [1], Anna Michalska [2], and Nuwan Waidyanatha [3]

[1] *Auton Lab, Carnegie Mellon University, USA;*
[2] *Institute of Computer Science, University of Warsaw, Poland;*
[3] *LIRNEasia, Colombo, Sri Lanka*

Abstract

T-Cube Web Interface is a generic tool to visualize and manipulate large scale multivariate time series datasets. The interface allows the user to execute complex queries quickly and to run various types of statistical tests on the loaded data. We show its utility in an important application scenario: **real-time bio-surveillance system designed to support rapid detection and mitigation of bio-medical threats in developing countries**.
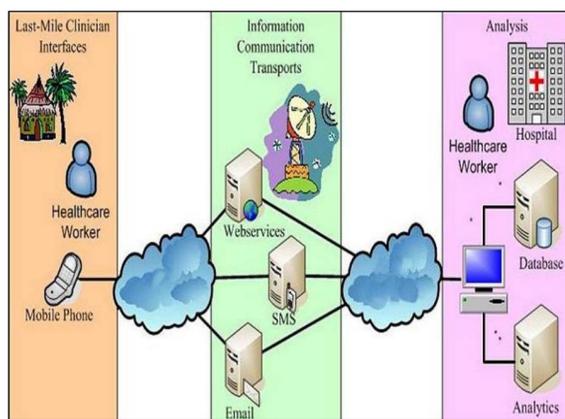
RTBP: Real-Time Bio-Surveillance Program (under support from IDRC)

Purpose: Mobilize healthcare workers in the rural settings with mobile phones to record and submit health-related information for the purpose of consolidating national health data for detecting disease outbreaks before they reach epidemic states.

Strengthen existing disease surveillance & detection communication systems, reduce latencies in detecting and communicating disease information, and set a stand interoperable protocol for sharing disease information with national and international health-related organization in the region.

T-Cube Web Interface and Computationally Efficient Algorithms for Bio-event Detection

Allow the users to execute complex queries quickly and to perform various types of statistical tests against the public health data. The underlying data representation technology, the **T-Cube**, is an in-memory data structure designed to **improve the response time to ad-hoc time series queries against large datasets**.

T-Cube Web Interface goes beyond univariate analysis and performs temporal and spatio-temporal analyses with visual results on the map to assist in finding spatial outbreaks using **Bayesian Spatial Scan** and **Temporal Scan** methods.

Fast Spatio-Temporal Scanning

OTC drug sales

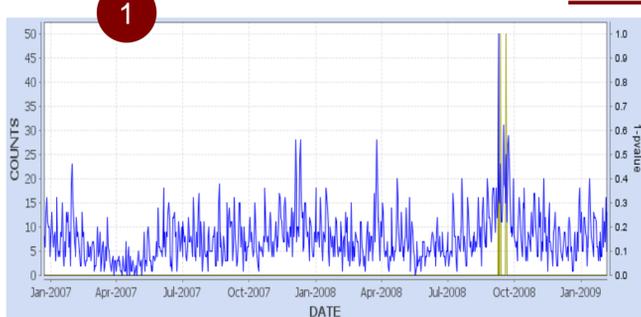Time series of counts for each analyzed region
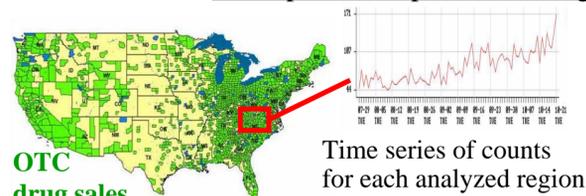
Multivariate Bayesian Spatial Scan analyzes multiple streams of data to maximize detection power while enabling disambiguation among possible causes of outbreaks.

Exhaustive Temporal Scan
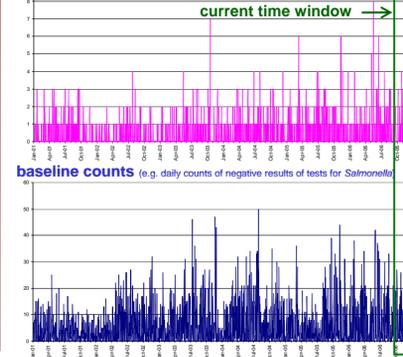
Temporal Scan alerts of unusual increases in counts of events of interest (e.g. patients reporting recently with Dengue fever) which cannot be explained by the changes in baselines (e.g. total number of patients reporting). It performs a large number of such tests very quickly.

Example: Leptospirosis Event in Colombo, Sri Lanka

Data: Reportable disease counts collected by the Sri Lanka Ministry of Healthcare and Nutrition

blue: daily time series of disease occurrences
olive: alerts generated by temporal scan

Within seconds of loading the data, the analysts found the Leptospirosis event

Above: Time series of daily counts of all reportable disease occurrences in Colombo. The alerts generated by temporal scan indicate days when the increases in these counts exceeded the nationwide trends.
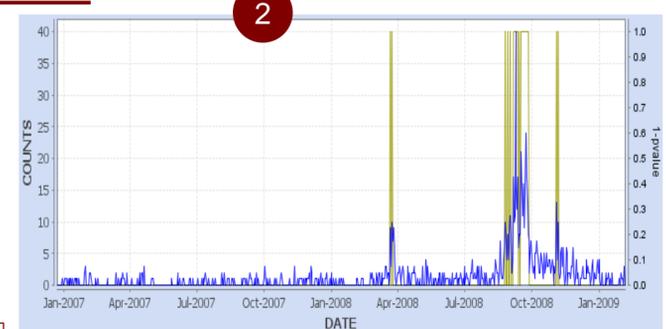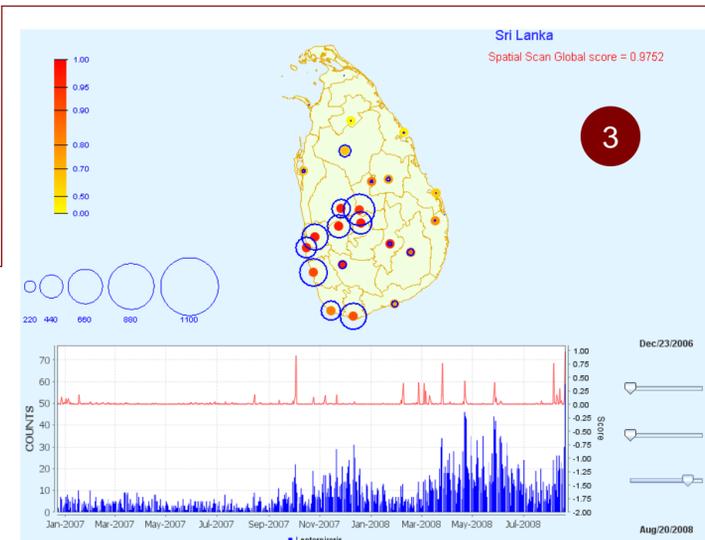
Right: Apparently, the Leptospirosis event was not restricted to Colombo but as indicated by Spatial Scan analysis, it has shortly spread to 7 other cities. High probability of the locations being affected by it is shown with red circles on the map.

Above: Running temporal scan against public health data from Colombo identifies Leptospirosis as the main contributing factor to the observed unusual trend. Note that even though the disease outbreak lasted for more than 45 days, the first alert was generated within 2-3 days of its onset.

Left: Spatial Scan tracks probabilities of Leptospirosis outbreak anywhere in the nation for all days in the past data (red line). On Aug 20, 2008, it was over 97%.