About the Dengue Forecasting Tool

Gwen Rino

5/25/2018

OVERVIEW

No forecasting method can perfectly predict the future, but data science offers tools that can make estimates of what is most likely to happen by analyzing patterns from past events. These forecasts, uncertain as they are, can offer valuable information to help organizations make critical decisions about the best use of resources.

On behalf of the Departamento de Salud de Puerto Rico, we are able to offer a tool for forecasting the number of weekly cases of dengue disease with enough accuracy to substantially improve efficiency in staffing the San Juan public health clinic.

USING THE FORECASTING TOOL

For instructions on installing and using the tool, please see ~CSX415.1-project/deploy/README.md.

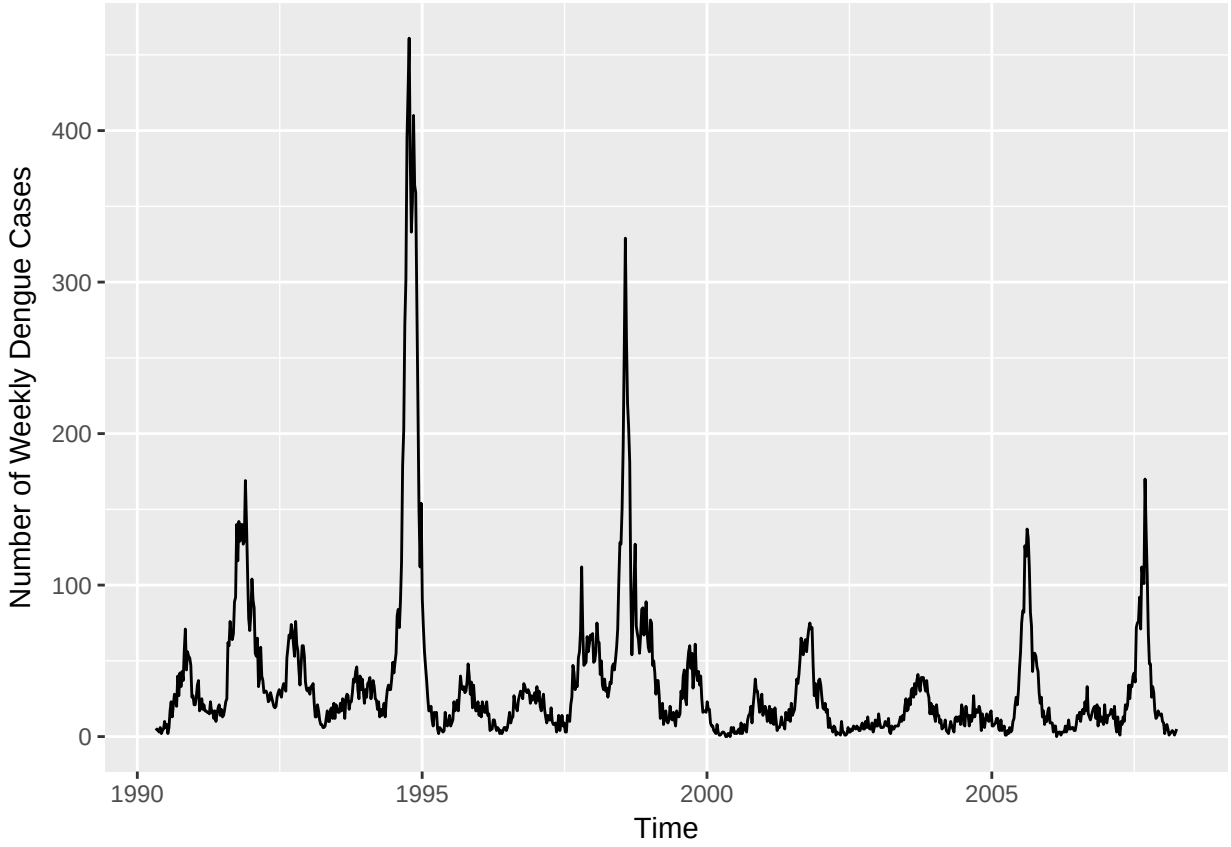
HOW THE FORECASTING TOOL WORKS

The forecasting tool is based on a predictive model that makes use of two types of information about the past, both of which are quite intuitive, to forecast the future.

The first type of information in the model is about weather, which is makes sense given the fact that dengue is a mosquito-borne virus. Indeed, the Departamento de Salud's current staffing model, which adjusts staffing levels according to the season of the year, is based on the well-understood seasonal relationship between weather patterns, mosquito populations, and disease. Our mathematical model refines and focuses this insight.

The second type of information is recent number of cases of dengue disease. This is also intuitive; you would not expect to see a precipitous increase or decrease in number of cases from one week to the next most of the time; today's incidence of disease is related to yesterday's and to tomorrow's. Our mathematical model takes advantage of this strong predictive pattern (a phenomenon called autocorrelation).

A plot of the number of weekly dengue cases over time suggests both of these relationships. There is a clear annual cycle with disease levels rising in the summer and fall, and falling in the winter and spring, which suggests the correlation between weather and disease. Furthermore, the number of cases doesn't oscillate and generally doesn't jump wildly from one observation to the next; it rises for a time and then falls for a time, suggesting that the recent past is strongly correlated with the near future.



The forecasting tool forecasts the number of cases of dengue disease for a time horizon of the user's choice, and returns both a forecast of the number of cases for each future week and also the number of staff that would be needed to meet the required 1:9 staffing level for the forecasted number of cases.

It should be emphasized that the forecasting tool gives much more dependable results for near-future predictions than long-term ones; the uncertainty rises dramatically for longer time horizons. The average error (positive or negative) in predicted number of dengue cases looking ahead one week is just 1.06; looking ahead six weeks the error rises to 7.98; looking ahead six months the error is 18.03. (For more information about the accuracy of the model at different time horizons, please see the document 4.DynRegEval.pdf, which can be found at dengue/reports/ModelEvaluations/4.DynRegEval.pdf)

THE IMPACT OF THE FORECASTING TOOL

In order to understand the potential impact of the forecasting tool, it is helpful to compare what has actually happened in the past with what would have happened had the forecasting tool been in use at that time.

(As a reminder, the current staffing model for the San Juan clinic calls for 1 dengue specialist in the spring, increases that number to 5 during the summer and fall when disease is more prevalent, and reduces the number to 2 during the winter. The clinic fills in with per diem nurses as needed in order to meet the minimum required ratio, 1:9, of caregivers to patients. Overstaffing costs $500 per week per extra health care worker, and understaffing costs $600 per week for each per diem health care worker.)

Example 1: Forecasts for March 18 - April 22, 2008

Here we examine how the tool would have performed forecasting the past six weeks, a period in early spring with low levels of disease.

The table below shows, for each week, the actual number of dengue cases, the number of staff that were required to meet the 1:9 ratio, the number of staff hired for the week based on the current staffing model, the difference between the number required and the number hired, and the cost of that difference.

The error in the current staffing model cost the Departamento de Salud $500 in the past six weeks.

WeekOf	ActualCases	RequiredStaff	CurrentStaff	CurrentError	CostCurrentError
2008-03-18	3	1	2	1	500
2008-03-25	4	1	1	0	0
2008-04-01	3	1	1	0	0
2008-04-08	1	1	1	0	0
2008-04-15	3	1	1	0	0
2008-04-22	5	1	1	0	0

The table below shows, for each week, the actual number of dengue cases, the number of staff that were required to meet the 1:9 ratio, the number of staff that would have been recommended by the forecasting tool, the difference between the number required and the number recommended, and the cost of that difference.

Had it been in use during this six-week time period, the forecasting tool would have provided the Departamento de Salud a savings of $500 over the current staffing model.

WeekOf	ActualCases	RequiredStaff	RecommendedStaff	RecError	CostRecError
2008-03-18	3	1	1	0	0
2008-03-25	4	1	1	0	0
2008-04-01	3	1	1	0	0
2008-04-08	1	1	1	0	0
2008-04-15	3	1	1	0	0
2008-04-22	5	1	1	0	0

Example 2: Forecasts for August 27 - October 1, 2007

Here we examine how the tool would have performed forecasting a six-week period in early fall 2007, a period with rapidly rising levels of disease.

As you can see in the table below, the current staffing model greatly underestimated the number of health care workers that were required by the actual number of cases of dengue disease during this time period. The cost to the Departamento de Salud in making up the staffing shortage with per diem health care workers was very high, $27,000 over six weeks.

WeekOf	ActualCases	RequiredStaff	CurrentStaff	CurrentError	CostCurrentError
2007-08-27	92	11	5	-6	3600
2007-09-03	71	8	5	-3	1800
2007-09-10	112	13	5	-8	4800
2007-09-17	106	12	5	-7	4200
2007-09-24	101	12	5	-7	4200
2007-10-01	170	19	5	-14	8400

The table below shows that the forecasting tool also (mostly) underestimates the number of health care workers that were required by the actual number of cases of dengue disease during this time period, but not nearly as significantly as the current staffing model did. The total cost to the Departamento de Salud of the difference between the staffing levels suggested by the forecasting tool and the actual required number of

staff was $13,700 over six weeks.

Had it been in use during this time period, the forecasting tool would have provided the Departamento de Salud a savings of $13,300 over the current staffing model.

WeekOf	ActualCases	RequiredStaff	RecommendedStaff	RecError	CostRecError
2007-08-27	92	11	9	-2	1200
2007-09-03	71	8	9	1	500
2007-09-10	112	13	9	-4	2400
2007-09-17	106	12	9	-3	1800
2007-09-24	101	12	9	-3	1800
2007-10-01	170	19	9	-10	6000

Example 3: Forecasts for November 5 - December 10, 2007

The quite severe dengue season of 2007 peaked in early October and then fell steadily through the late fall and winter to typical low levels by the spring of 2008.

Here we examine how the tool would have performed forecasting a six-week period in late fall 2007, when disease levels were falling.

The table below shows the current staffing model's performance during this six-week period, with a total cost from over- and underestimating staffing needs of $5,600.

WeekOf	ActualCases	RequiredStaff	CurrentStaff	CurrentError	CostCurrentError
2007-11-05	48	6	5	-1	600
2007-11-12	26	3	5	2	1000
2007-11-19	33	4	5	1	500
2007-11-26	29	4	5	1	500
2007-12-03	17	2	5	3	1500
2007-12-10	12	2	5	3	1500

The table below shows the forecasting tool's performance during the same six-week period, with a total cost from over- and underestimating staffing needs of $3,600.

Had it been in use during this time period, the forecasting tool would have provided the Departamento de Salud a savings of $2,000 over the current staffing model.

WeekOf	ActualCases	RequiredStaff	RecommendedStaff	RecError	CostRecError
2007-11-05	48	6	5	-1	600
2007-11-12	26	3	5	2	1000
2007-11-19	33	4	4	0	0
2007-11-26	29	4	4	0	0
2007-12-03	17	2	4	2	1000
2007-12-10	12	2	4	2	1000

Summary of Examples

In each of the three examples, the forecasting tool performed better than the current staffing model, saving signficant resources ($500 - $13,300 over six weeks) that the Departamento de Salud could have directed to

other public health projects, including education, vector management, and research.

As discussed above, no forecasting tool can be perfect, the results are less and less precise the further out you forecast, and there are no guarantees that this tool's forecasts will be accurate enough to offer cost savings over the current staffing model during any particular time period.

However, the forecasting tool is able to offer predictions of weekly cases of dengue for relatively short time horizons (1-6 weeks) that are, on average, good enough to substantially improve the staffing efficiency of the San Juan clinic.

NEXT STEPS

The current version of the forecasting tool is a beta model. It does not include a mechanism for updating the data regarding weekly weather and number of cases; therefore, it is not useful in the future, but "frozen in time." The next step in the development of the tool is to build a process for updating the weekly data that feeds the model.