How to Run Model Validation with the EPIDEMIA system

For epidemia-demo v2.1 using epidemiar v2.1.0

Dawn Nekorchuk, Michael Wimberly, and EPIDEMIA Team Members
Department of Geography and Environmental Sustainability, University of Oklahoma
dawn.nekorchuk@ou.edu; mcwimberly@ou.edu

Updated December 05, 2019

Running Model Validation and Assessment

Validation and assessment have been built into the epidemiar package in the function `run_validation()` for on-demand evaluation for any historical period.

Evaluation can be made for one through n-week ahead predictions, and include comparisons with two naive models: persistence of last known value, and average cases from that week of the year.

Building validation into the early warning system provides more opportunities to learn about the model via the validation results. Locations where the models perform well and where they do not can be identified with geographical grouping-level results.

With on-demand implementation and time-range flexibility, one can also investigate how accuracy changes over time, which is of particular interest in places like Ethiopia with changing patterns and declining trends due to anti-malarial programs.

Specific Arguments

The `run_validation()` function takes 4 arguments, plus all the `run_epidemia()` arguments.

- `date_start`: The week to begin validation, can be built with `epidemiar::make_date_yw()` and isoyear and isoweek numbers (or epiweeks, with appropriate modifications).
- `total_timesteps`: The number of weeks from `week_start` to run the validation.
- `timesteps_ahead`: To validate 1 through *n*-week ahead predictions (the number of weeks into the future the predictions are made).
- `reporting_lag`: Default of 0 weeks, but can be adjusted for different assumptions about the length of the lag in data reporting. Enter the number of timesteps to simulate reporting lag. For instance, if you have weekly data, and a `reporting_lag` of 1 week, and are working with a timesteps_ahead of 1 week, then that is the functional equivalent to reporting lag of 0, and timesteps_ahead of 2 weeks. I.e. You are forecasting next week, but you don't know this week's data yet, you only know last week's numbers.
- `skill_test`: TRUE/FALSE on whether to also run the naive models for comparison: Persistence - last known value carried forward n-appropriate weeks, and Average Week - the average cases from that week of the week (per geographic grouping). This will create skill scores showing the relative improvement of the forecast model over the naive model (negative: naive performs better, 0: no improvement, positive values up to 1: relative improvement).

Other Arguments & Adjustments

The `run_validation()` function will call `run_epidemia()`, so it will also take all the arguments for that function. The user does not need to modify any of these arguments (e.g. event detection settings, `forecast_future`), as `run_validation()` will automatically handle all of thse adjustments.

It is envisioned that users can take their usual script for running EPIDEMIA forecasts, and simply sub in the validation function with those five validation settings for doing model assessments.

Validation Output

Statistics

Validation statistics included Mean Squared Error (`MSE`), Mean Absolute Error (`MAE`), and R^2 (`R2`, variance explained). Where 'obs' is atual observed value and 'pred' is the predicted forecast values:

- $MAE = mean(|obs - pred|)$
- $RMSE = sqrt(mean((obs - pred)^2))$
- $R^2 = 1 - (SSE/TSS)$
  - $SSE = sum((obs - pred)^2)$
  - $TSS = sum((obs - mean(obs))^2)$

Skill scores are calculated per statistic. The forecast accuracy statistic value (score$_{\mathrm{fc}}$) is compared against the naive model statistic (per naive model, score$_{\mathrm{naive}}$) in regards to a perfect (no error) value (score$_{\mathrm{perfect}}$):

- $Skill = (score_{fc} - score_{naive})/(score_{perfect} - score_{naive})$

The skill metric has an upper bound of 1. Skill of 0 is no improvement of the forecast model over that naive model. Skill between 0 and 1 shows the relative improvement of the forecast model over that naive model. Lower bound of the skill statistic depends on statistic.

Results will be returned summarized at the model level and also at the geographic grouping level.

Format

Results are returned in a list.

1. `skill_scores`: The skill score results of the forecast model compared against the naive models, if `skill_test = TRUE` was selected

- `skill_overall`: The skill scores at the overall model level
- `skill_grouping`: The skill score results per geographic grouping

2. `validations`: The validation accuracy statistics per model (name of the base model and the naive models if run with skill test comparison). Each model entry will have three items:

- `validation_overall`: Overall model accuracy statistics per timestep_ahead (week in the future)
- `validation_grouping`: Accuracy statistics per geographic grouping per timestep_ahead
- `validation_timeseries`: In beta-testing. Early version of a rolling validation results over time

3. `metadata`: Metadata on the parameters used to run validation and the date it was run.

Demo scripts

Run script

1. In the `epidemia-demo` folder, click on the `epidemia_demo.Rproj` to open the project in RStudio.
2. In RStudio, click on the `run_validation_amhara.R` script in the `validation` folder to open it.
3. You can edit the evaluation start week, total number of weeks, number of weeks ahead, reporting lag, and skill test parameters in Section 1.
4. Sections 2 and 3 load the libraries and the data. These are the same as if you were about to run a report.

5. Section 4 & 5 will run the validations per species. The two species (P. falciparum + mixed, and P. vivax) have been separated into their own section. Because this runs the forecasts for *each week* in the evaluation period, this will take a while to run. Each section will also create and save the formatted pdf report as well (using the Sweave formatting script `validation/epidemia_validation.Rnw`). The files will automatically save with names indicating the case field, year-week start, and number of weeks evaluated.