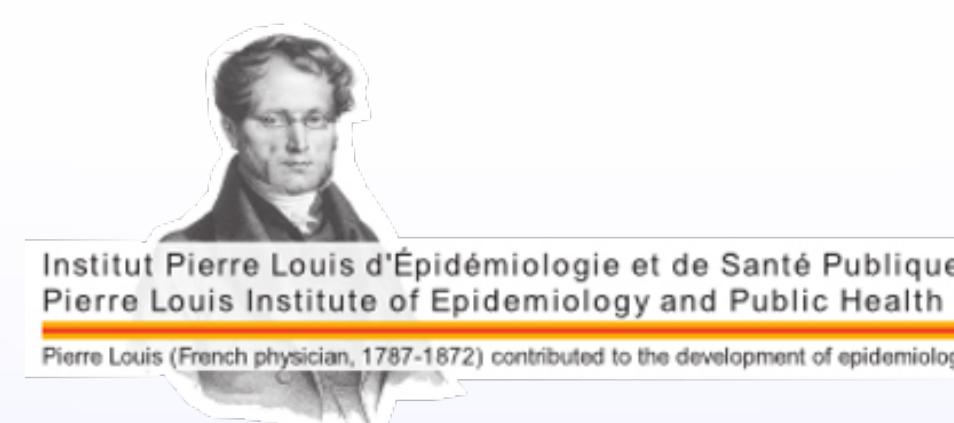


# Improving early epidemiological assessment of Zika virus epidemics using historical data

Julien Riou, Chiara Poletto, Pierre-Yves Boëlle

Sorbonne Universités, Inserm/UPMC, Institut Pierre Louis d'Epidémiologie et de Santé Publique (IPLESU UMRS 1136), Paris, France  
[ julien.riou@iplesp.upmc.fr ]



UPMC  
SORBONNE UNIVERSITÉS



## Introduction

In the early epidemic phase, model-based epidemiological assessment is often conducted to support decision-making by public health authorities. However, long-term forecast is generally unreliable at this point, as it relies on insufficient information recovered from the few data points available in real time.<sup>?</sup>

We propose to complement this information with **historical data on similar epidemics**, to be integrated in the analysis under the form of **informative prior distributions**.

## Materials and Methods

This approach was retrospectively applied to the **2015-2017 Zika virus epidemics in the French West Indies (FWI)**. We used a Bayesian dynamic discrete-time SIR model, where the observed incidence  $O_t$  depends on

$$O_t \sim \text{Neg-Bin} \left( R_0 \frac{S_t}{N} I_t^*, \phi \right) \quad (1)$$

where  $R_0$  is the **basic reproduction number**,  $S_t/N$  is the proportion of susceptibles (which depends on past incidence and the **reporting rate**  $\rho$ );  $I_t^*$  summarizes exposure (past incidence weighted by the discretized serial interval); and  $\phi$  is the dispersion parameter.

This model was implemented in *Stan*.<sup>?</sup> It was fitted to increasing amounts of incidence data from Guadeloupe, Martinique and Saint-Martin to obtain posterior distributions for  $R_0$  and  $\rho$  at different points of the epidemics. The remaining of the epidemic was then simulated stochastically, leading to a **predictive distribution of future incidence**.

Informative priors for  $R_0$  and  $\rho$  were obtained by combining the results of fitting a two-level **hierarchical** dynamic SIR model to historical data on earlier CHIKV outbreaks in the FWI (2013-2015) and CHIKV and ZIKV outbreaks French Polynesia (2013-2015), accounting for the similarities and differences between CHIKV and ZIKV transmission and reporting.<sup>?</sup>

Forecasts were thus conducted using each of three choices of prior distributions:

- **Non-informative priors** (baseline);
- **Regional priors** (island-averaged, from top-level parameters);
- **Local priors** (island-specific, from bottom-level parameters).

## References

<sup>1</sup> Hans Heesterbeek, Roy M Anderson, Viggo Andreasen, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339, 2015.

<sup>2</sup> Bob Carpenter, Andrew Gelman, Matt Hoffman, et al. Stan: a probabilistic programming language. *Journal of Statistical Software*, 2015.

<sup>3</sup> Julien Riou, Chiara Poletto, and Pierre-Yves Boëlle. A comparative analysis of chikungunya and zika transmission. *Epidemics*, 19:43–52, 2017.

## Results: eliciting priors from historical data

The **regional** informative priors, suitable for a ZIKV outbreak in a typical island of the French West Indies, were relatively wide, with 95% credible intervals between 0.5 and 2.5 for  $R_0$  and between 0 and 0.30 for  $\rho$ . The **local** informative priors carried specific information from past outbreaks in each island and had a narrower range, especially in Guadeloupe and Martinique.

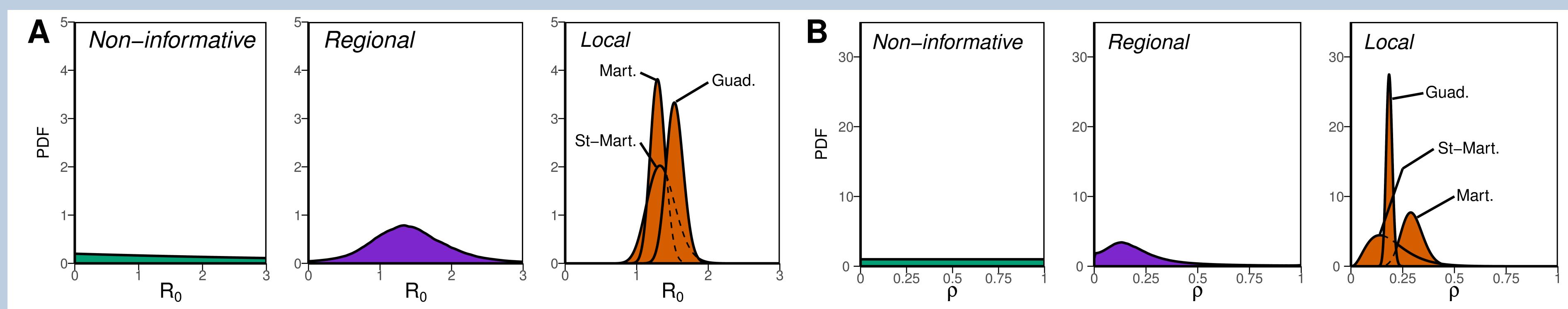


Figure 1: Three choices of prior distributions for the parameters  $R_0$  (panel A) and  $\rho$  (panel B).

## Results: epidemic forecast

An example of the predictive distribution of future incidence in Martinique is show in Fig. 2A. At this point in time, predictions with non-informative priors were largely off-target and showed considerable uncertainty. Using informative priors led to major improvements in both **accuracy** (distance with actual observations) and **sharpness** (width of the 95% prediction interval). These results were typical during the initial phases of the three epidemics. Once the peak of incidence was reached, the choice of the prior distributions was less critical to prediction quality.

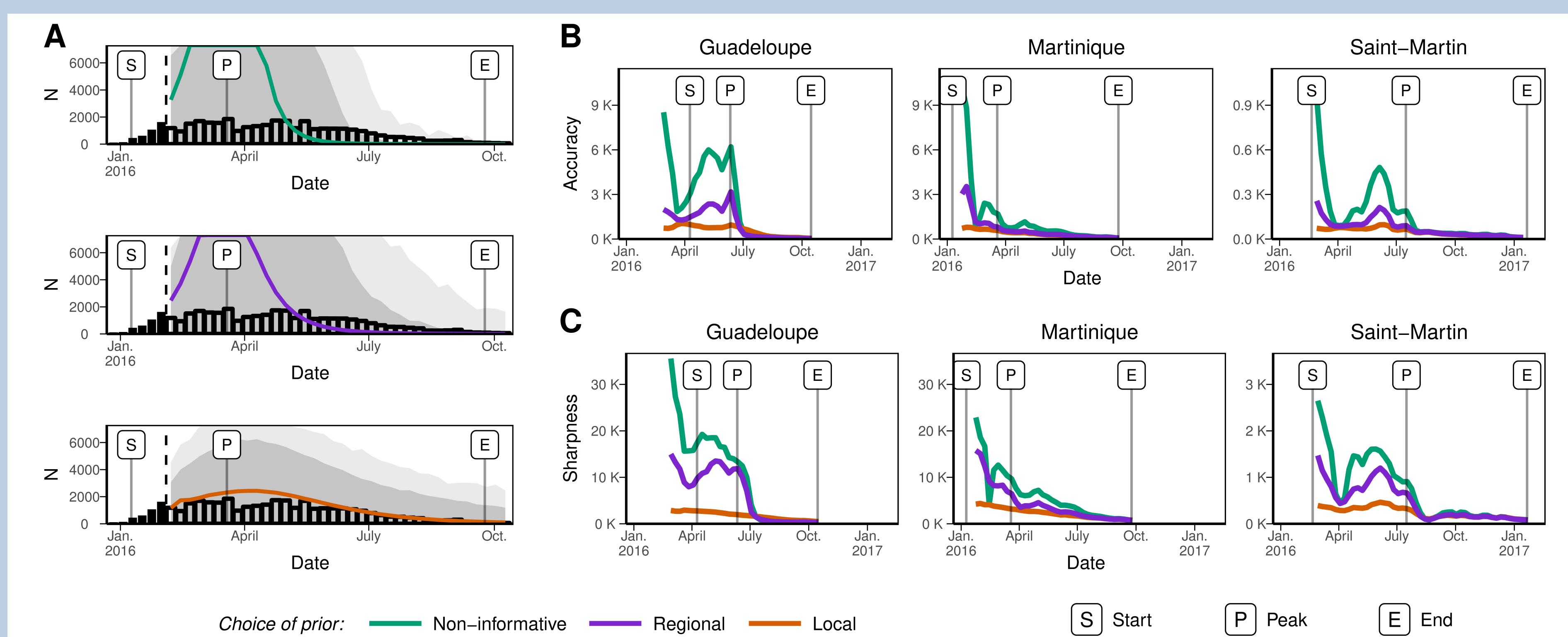


Figure 2: (A) Predictive distribution of future incidence in Martinique on January 31st, 2016. (B-C) Accuracy and sharpness of the predictive distribution of incidence according to the date of analysis.

Early forecasts of **epidemic size** and **maximal incidence** showed more accuracy and stability when using regional and particularly local priors. However, forecasts of the date of peak incidence and of the duration of the period of high epidemic activity were less affected by the choice of priors.

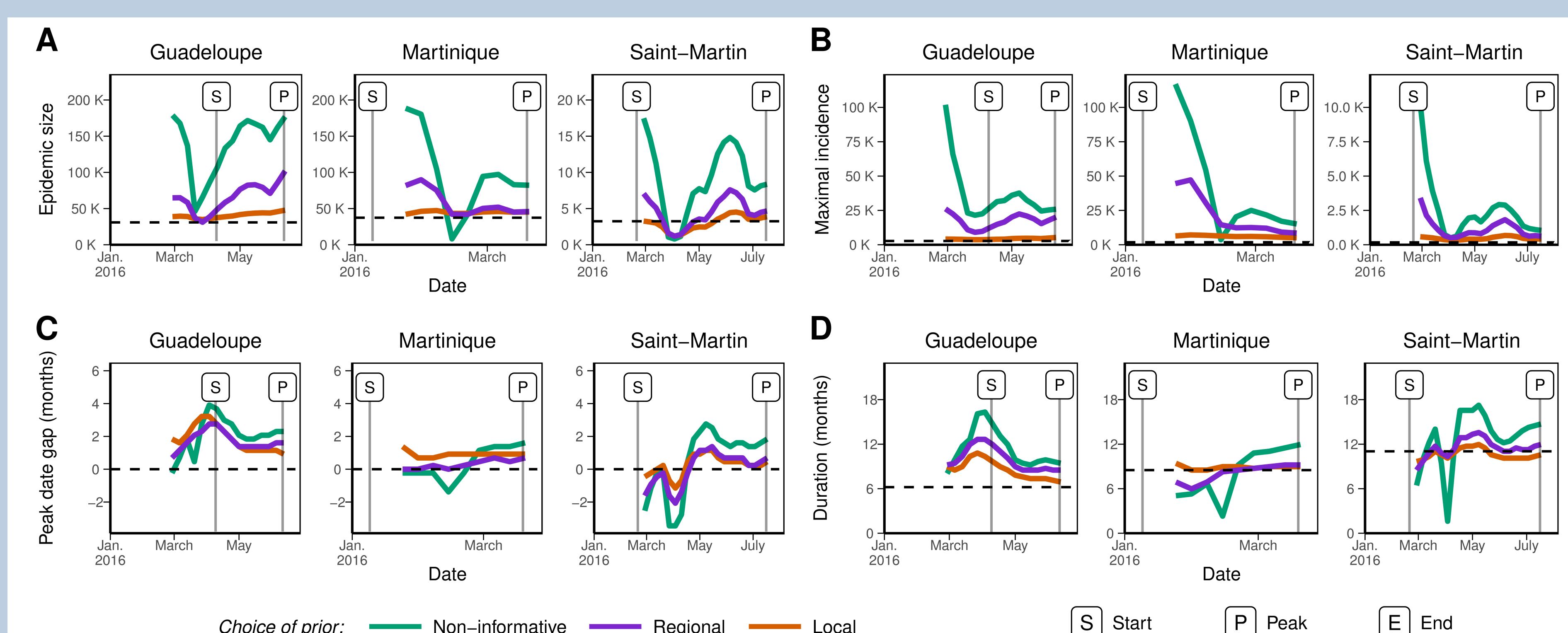


Figure 3: Early forecasts regarding four indicators of operational interest: (A) final epidemic size; (B) maximal incidence; (C) date of peak incidence (difference with the eventually observed date in months); and (D) duration of the period of high epidemic activity (in months). Dashed lines represent actual values.

## Conclusions

Our results suggest that during the initial stage of an epidemic, epidemiological forecasting may benefit from using informative priors on key parameters, increasing the accuracy and the stability of the predictions. If a theoretical framework allows it, these informative priors can be elicited from historical data on previous, similar epidemics using hierarchical models. This approach could provide public health authorities with reliable forecasts at an earlier stage than is usually possible.