



HAL
open science

Causality Analysis Between Climatic Factors And Dengue Fever Using The Granger Causality

Nguyen Hoang Pham, Jean Daniel Zucker, Marc Choisy, Ho Tuong Vinh

► **To cite this version:**

Nguyen Hoang Pham, Jean Daniel Zucker, Marc Choisy, Ho Tuong Vinh. Causality Analysis Between Climatic Factors And Dengue Fever Using The Granger Causality. 12th IEEE RIVF Conference on Computing and Communication Technologies (RIVF 2016), Nov 2016, Ha Noi, Vietnam. 10.1109/RIVF.2016.7800268 . hal-01548254

HAL Id: hal-01548254

<https://hal.sorbonne-universite.fr/hal-01548254v1>

Submitted on 27 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Causality Analysis Between Climatic Factors And Dengue Fever Using The Granger Causality

Pham Nguyen Hoang^{*†}, Jean Daniel Zucker^{†‡§}, Marc Choisy[¶] and Ho Tuong Vinh^{*†}

^{*}International Francophone Institute, Vietnam National University, Hanoi, Vietnam

[†]Research Institute for Development, UMI 209, UMMISCO, Bondy, France

[‡]Institute of Cardiometabolism and Nutrition, ICAN, Assistance Public Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Paris, France

[§]Sorbonne University, UPMC University Paris 6, UMR_S 1166, ICAN, NutriOmics Team, Paris, France

[¶]MIVEGEC - (Laboratoire Maladies Infectieuses et Vecteurs, Ecologie, Génétique, Evolution et Contrôle), UMR CNRS 5290/IRD 224, Université de Montpellier, Montpellier, France

^{||}OUCRU - (Oxford University Clinical Research Unit), Hanoi, Vietnam

Email: pnhuang@cit.ctu.edu.vn, jdzucker@gmail.com, mchoisy@gmail.com, vinhht@vnu.edu.vn

Abstract—Dengue is a mosquito-borne tropical disease caused by one of the four serotypes of dengue virus. There are about 2.5 billion people living in endemic areas. The potential influence of environmental and climatic factors on the transmission of mosquito-borne diseases are expected in many situations but the exact real influences of climatic variables remain uncertain in most of the cases. In the present study, we applied the Granger causality to analyze the relation between climatic variables and the monthly incidences of dengue syndromes in Vietnam for 64 provinces from 1994 to 2010. The experimental results shows a close relationship between climatic factors and dengue incidence in each province and their neighbourhoods.

Keywords—dengue fever, Granger causality, climatic drivers, Vietnam

I. INTRODUCTION

Dengue fever is the first human arbovirose and is widely prevalent in tropical countries. According to the World Health Organisation (WHO, 2007), about 40% of global population in 112 countries in the world is exposed to the virus. There is an estimate of 390 million new infections every year [1]. A number of studies have aimed at characterizing the influences of climate, human demography and behaviour on the epidemiology of infectious diseases, using different methodologies. Zhang et al. [2], Cumming et al. [3], Barreto et al. [4] have used mathematical models to investigate the effect of climate change, demographic transition and the urban structure on the transmission of dengue fever. An overall result is that the quantitative relationships between climate and vector-borne diseases are inconsistent between different studies. Some evidence has a strong relationship between dengue incidence and El Niño in the Pacific islands [5], [6] and in Thailand [7], whereas other found that increasing population density and inadequate water source are the main drivers of dengue epidemiology [8], [9]. In analysing the monthly data of dengue hemorrhagic fever (DHF) in Thailand from 1983 to 1997, Cumming et al. [10] showed that rural areas with low population density can also experienced serious epidemics.

Southeast Asia is a region of particular intense dengue circulation [11], and this is the case also for Vietnam. Vietnam has this additional particularity of having an important

diversity of climates on a relatively small area (330,000 km²). This climatic diversity is generated by a wide latitudinal range orthogonal to a large range of elevations from sea level on the East coast to more than 3,000 m on the West border. Schmidt et al. [12] performed a cohort study and a spatial analysis in Vietnam and showed that the risk of dengue fever was higher in rural areas with poor water supply than in urban areas with high population densities. Do and collaborators [13], [14] studied dengue fever in the city of Hanoi between 2004 and 2009 and evidenced a strong seasonality of dengue epidemiology with high incidences between June and November, which is the rainy season that also corresponds to high temperatures. Le et al. [15] has analysed the influence of tourism on dengue incidence in Cat Ba island (northern Vietnam) from September to November 2013. They showed that dengue transmission is unlikely to be sustained on Cat Ba Island and that the 2013 epidemic likely originated through the introduction of viruses from the mainland, most likely Hanoi. Cazelles and collaborators [16] analyzed the dynamics of dengue incidence in Binh Thuan province in southern Vietnam. They used wavelet decomposition to detect and quantify the periodicity of dengue between January 1994 and June 2009 and to describe the profile of synchrony in time and space. They also used this method to explore the relationship between dengue incidence and the El Niño Southern Oscillation (ENSO), and found a strong non-stationary association between ENSO indices and climatic variables for the period of 2-3 years.

All the studies performed in Vietnam and summarized above are carried out at a small spatial scale and thus do not encompass the great spatial variability of climates that is found in Vietnam. In this paper, we present the first analysis on the relationship between climate and dengue syndromes conducted at the scale of the whole country of Vietnam. Furthermore, we explore, for the first time to our knowledge, the potential of the Granger causality methodology to identify causal relationships between climatic variables and infectious diseases epidemiology. The analysis is based on monthly time series of incidence of dengue syndromes in 64 provinces and 7 climatic variables from 67 climatic stations.

II. DATA AND METHODS

A. Data

The number of DHF cases used in this research are the monthly report of 64 provinces of Vietnam [11], [17] from January 1998 to September 2010. Monthly climatic data in 67 climatic stations are also collected from the same time period and contain 7 different variables: average, maximal and minimal temperatures, absolute and relative humidities, rainfall and hours of sunshine. Global Administrative Areas (GADM, www.gadm.org) data of Vietnam were used for geography visualisation. Population sizes per province from the census of 2009 were retrieved from the General Statistics Office of Vietnam (GSO, www.gso.gov.vn).

B. Method

Granger has defined a concept of causality based on the concept of vector autoregression (VAR) models that have been widely used in recent years. This method consists in a statistical test to determine the relationships between time series by measuring the ability to predict the future values of a time series using prior values of another time series. The key idea of Granger Causality Analysis (GCA) is that “ Y_t is causing X_t if we are better able to predict X_t using all available information than if the information apart from Y_t had been used” [18], [19]. If this condition is satisfied, it means that Y_t contains exclusive information that is not present in all the other time series and that this information is useful to predict the time series X_t of interest.

Granger causality relationship is based on two principles :

- The cause happens prior its effect;
- The cause produces unique changes in the effect. In other words, the causal time series contains unique information about the effect time series that is not available otherwise.

Let $I(t)$ be the set of information containing all the relevant information in the universe available up to time t and $I_{-X}(t)$ is the set of all information in the universe excluding X up to time t . Giving two time series X and Y , Granger proposed to test the following hypothesis for identification of a causal effect of X on Y

$$\mathbb{P}[Y(t+1) \in A|I(t)] \neq \mathbb{P}[Y(t+1) \in A|I_{-X}(t)]$$

where \mathbb{P} denotes a probability and A is an arbitrary non-empty set. If we can prove the above hypothesis or if we can reject the null hypothesis

$$\mathbb{P}[Y(t+1) \in A|I(t)] = \mathbb{P}[Y(t+1) \in A|I_{-X}(t)]$$

we say that X Granger-causes Y .

More specifically, consider two stationary time series X_1 and X_2 with a linear autoregressive model :

$$\begin{aligned} X_1(t) &= \sum_{j=1}^p A_{11j} X_1(t-j) + \sum_{j=1}^p A_{12j} X_2(t-j) + E_1(t) \\ X_2(t) &= \sum_{j=1}^p A_{21j} X_1(t-j) + \sum_{j=1}^p A_{22j} X_2(t-j) + E_2(t) \end{aligned} \quad (1)$$

where p is the maximum number of lagged observation included in the model, A is the matrix containing the coefficients of the model (e.g. the contributions of each lagged observations) of the predicted values of $X_1(t)$ and $X_2(t)$. E_1 and E_2 are predictions errors for each times series. If the variance of E_1 (or E_2) is reduced by the inclusion of the X_2 (or X_1), then it is said that X_2 (or X_1) Granger-causes X_1 (or X_2). In other words, X_2 Granger-cause X_1 if the coefficient A_{12} are significantly different from zero. This can be tested by performing an F-test of the null hypothesis that $A_{12} = 0$, given the assumption of covariance stationarity on $X_1(t)$ and $X_2(t)$. The magnitude of a G-causality interaction can be estimated by the logarithm of the corresponding F-statistic [20].

The Granger causality framework can be extend to solve the multivariable problem: given p time series, $X_1, X_2, \dots, X_{p-1}, X_p$, we are interested in identifying which time series Granger cause X_i . There are two ways to solve this problem :

- Examining the pairwise relationships in each pair of time series;
- Performing a vector autoregression as :

$$X_i(t) = \sum_{j=1}^p A_{i,j}^T X_j^{t,lagged} + E \quad (2)$$

where $X_j^{t,lagged} = [X_j(t-L), \dots, X_j(t-1)]$ is the lagged time series and $A_{i,j}$ is the coefficient vector. Then we test whether $A_{i,j}$ are significantly different from zero. A time series X_i is called a Granger cause of another time series X_j if at least one of the elements $A_L(i, j)$ for $L = 1, \dots, p$ is significantly larger than zero.

We applied this methodology to climatic and dengue incidence data in Vietnam from January 1998 to September 2010. Monthly numbers of new dengue cases in each province were divided by the province population size. Both dengue incidence and climatic time series were then linearly interpolated for missing values and normalised. We applied the Granger causality to analyse the relationship between DHF incidence and climatic variables in each of the 64 provinces of Vietnam. The original Granger causality test was performed using the `vars` R package [21]. Each province was matched to the closest climatic station to its population centroid. In a first step we performed a Granger causality test to find out the influence of climatic factors on dengue incidence for each province. In a second step, we added to the analysis dengue incidences in neighbouring provinces as potential Granger-causes of dengue epidemiology in the focal province. Neighbouring provinces where identified as the ones being directly adjacent to the focal province.

III. RESULTS

Figure 1 shows the infection rates of dengue for each month, averaged over the 13 years of the study period and figure 2 shows the same for the 7 climatic variables of the 63 climatic stations of Vietnam. We can see that the means of infection rate strongly increases between June and November in many provinces in the southern and south-central areas.

During this period of time, the three temperature variables (maximal, minimal and average) also dramatically raises and reach their nadir in June and July (the middle of rainy season), before decreasing from August to December (the end of rainy season and the beginning of dry season). During the rainy season (from May to October), rainfall and relative humidity rapidly increase while the number of hours of sunshine strongly decrease.

The apparent change of climatic factors and the increase of dengue incidence between June and November may have some relationship. To investigate this formally, we applied the Granger causality test to test their causal relationships. First we analysed the relationship between each climatic factor and dengue incidence for each province in Vietnam. For each province, we tested the null hypothesis H_0 that the climatic factor does not Granger-cause the dengue epidemiology. Figure 3 presents the output of the test at the $\alpha = 0.05$ threshold for each climatic variable and each province. The results in this figure 3 suggest that:

- temperature strongly impacts dengue incidence, mostly in the central and south areas of Vietnam;
- rainfall and relative humidity weakly impact dengue incidence in some provinces;
- hours of sunshine strongly impact on the influence of the central area.

Accounting for the dengue incidence in the neighbouring provinces, figure 4 shows overall the same results as figure 3, the only difference being that the number of provinces for which the test is significant is slightly lower. Figure 5 and figure 6 shows the p-values of the result of two tests.

IV. DISCUSSION

Our analysis presents an overview of the influence of several climatic factors on DHF incidence in 64 provinces of Vietnam from January 1998 to September 2010. We show that temperatures strongly impact dengue incidence in most of the provinces and that rainfall and relative humidity impact dengue incidence too but in some scattered provinces. The number of hours of sunshine strongly impacts the incidence in the central area. Furthermore, accounting for dengue incidence in the neighbouring countries did not substantially affect the results. The fact that temperature and hours of sunshine strongly impact dengue incidence suggests that these climatic factors have a strong influence on the dynamics of the mosquitoes population. This is not surprising since effect of temperature on the mosquito population dynamics has been shown both in field and experimental studies [22]–[26]. The effect of sunshine may be more surprising. However, it may not be a direct effect but simply a confounding one due to the fact that hours of sunshine and rainfall are (as expected) negatively correlated. Finally, the fact that including incidence in neighbouring provinces to the analysis does not substantially affect the results suggests that dengue local dynamics are more affected by local drivers (such as environmental and climatic ones) than by a spatial dynamics of the diseases between provinces.

This study greatly exemplifies the potential of the use of Granger causality in epidemiology in order to unravel epidemiological drivers and to weigh the effects of local versus

distal such drivers. Future work includes an extensive use on Grange causality on a much broader Asian dataset so as to further analysis the relationships between Climatic Factors and Dengue Fever at a broader scale and for different Asiatic countries.

REFERENCES

- [1] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh *et al.*, “The global distribution and burden of dengue,” *Nature*, vol. 496, no. 7446, pp. 504–507, 2013.
- [2] Y. Zhang, P. Bi, and J. E. Hiller, “Climate change and the transmission of vector-borne diseases: a review,” *Asia-Pacific Journal of Public Health*, vol. 20, no. 1, pp. 64–76, 2008.
- [3] D. A. Cummings, S. Iamsirithaworn, J. T. Lessler, A. McDermott, R. Prasanthong, A. Nisalak, R. G. Jarman, D. S. Burke, and R. V. Gibbons, “The impact of the demographic transition on dengue in thailand: insights from a statistical analysis and mathematical modeling,” *PLoS Med*, vol. 6, no. 9, p. e1000139, 2009.
- [4] F. R. Barreto, M. G. Teixeira, M. C. Costa, M. S. Carvalho, and M. L. Barreto, “Spread pattern of the first dengue epidemic in the city of salvador, brazil,” *BMC Public Health*, vol. 8, no. 1, p. 51, 2008.
- [5] S. Hales, P. Weinstein, and A. Woodward, “Dengue fever epidemics in the south pacific: driven by el niño southern oscillation?” *The Lancet*, vol. 348, no. 9042, pp. 1664–1665, 1996.
- [6] S. Hales, P. Weinstein, Y. Soares, and A. Woodward, “El nino and the dynamics of vectorborne disease transmission,” *Environmental Health Perspectives*, vol. 107, no. 2, p. 99, 1999.
- [7] B. Cazelles, M. Chavez, A. J. McMichael, and S. Hales, “Nonstationary influence of el nino on the synchronous dengue epidemics in thailand,” *PLoS Med*, vol. 2, no. 4, p. e106, 2005.
- [8] M. G. Teixeira, “Dengue fever: a call for local, national, and international action,” *Emerg Infect Dis*, vol. 12, pp. 887–93, 2006.
- [9] D. J. Gubler, “Cities spawn epidemic dengue viruses,” *Nature medicine*, vol. 10, no. 2, pp. 129–130, 2004.
- [10] D. A. Cummings, R. A. Irizarry, N. E. Huang, T. P. Endy, A. Nisalak, K. Ungchusak, and D. S. Burke, “Travelling waves in the occurrence of dengue haemorrhagic fever in thailand,” *Nature*, vol. 427, no. 6972, pp. 344–347, 2004.
- [11] W. G. van Panhuis, M. Choisy, X. Xiong, N. S. Chok, P. Akarasewi, S. Iamsirithaworn, S. K. Lam, C. K. Chong, F. C. Lam, B. Phommasak *et al.*, “Region-wide synchrony and traveling waves of dengue across eight countries in southeast asia,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 42, pp. 13 069–13 074, 2015.
- [12] W.-P. Schmidt, M. Suzuki, V. D. Thiem, R. G. White, A. Tsuzuki, L.-M. Yoshida, H. Yanai, U. Haque, D. D. Anh, K. Ariyoshi *et al.*, “Population density, water supply, and the risk of dengue fever in vietnam: cohort study and spatial analysis,” *PLoS Med*, vol. 8, no. 8, p. e1001082, 2011.
- [13] W. Hu, P. Q. Thai, L. N. Hoat, P. Wright, P. Martens *et al.*, “Hot spot detection and spatio-temporal dispersion of dengue fever in hanoi, vietnam,” *Global health action*, vol. 6, 2013.
- [14] T. T. T. Do, P. Martens, N. H. Luu, P. Wright, and M. Choisy, “Climatic-driven seasonality of emerging dengue fever in hanoi, vietnam,” *BMC public health*, vol. 14, no. 1, p. 1078, 2014.
- [15] T. Le Viet, M. Choisy, J. E. Bryant, D. V. Trong, T. P. Quang, P. Horby, H. N. Tran, H. T. T. Kieu, T. N. Vu, K. N. Van *et al.*, “A dengue outbreak on a floating village at cat ba island in vietnam,” *BMC public health*, vol. 15, no. 1, p. 1, 2015.
- [16] K. T. Thai, B. Cazelles, N. Van Nguyen, L. T. Vo, M. F. Boni, J. Farrar, C. P. Simmons, H. R. van Doorn, and P. J. de Vries, “Dengue dynamics in binh thuan province, southern vietnam: periodicity, synchronicity and climate variability,” *PLoS Negl Trop Dis*, vol. 4, no. 7, p. e747, 2010.
- [17] W. G. M. Choisy, X. Xiong, N. S. Chok, P. Akarasewi, S. Iamsirithaworn, S. K. Lam, C. K. Chong, F. C. Lam, B. Phommasak *et al.*, “Region-wide synchrony of dengue incidence in southeast asia during a historic el nio event,” *Nature*, 2014.
- [18] C. W. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

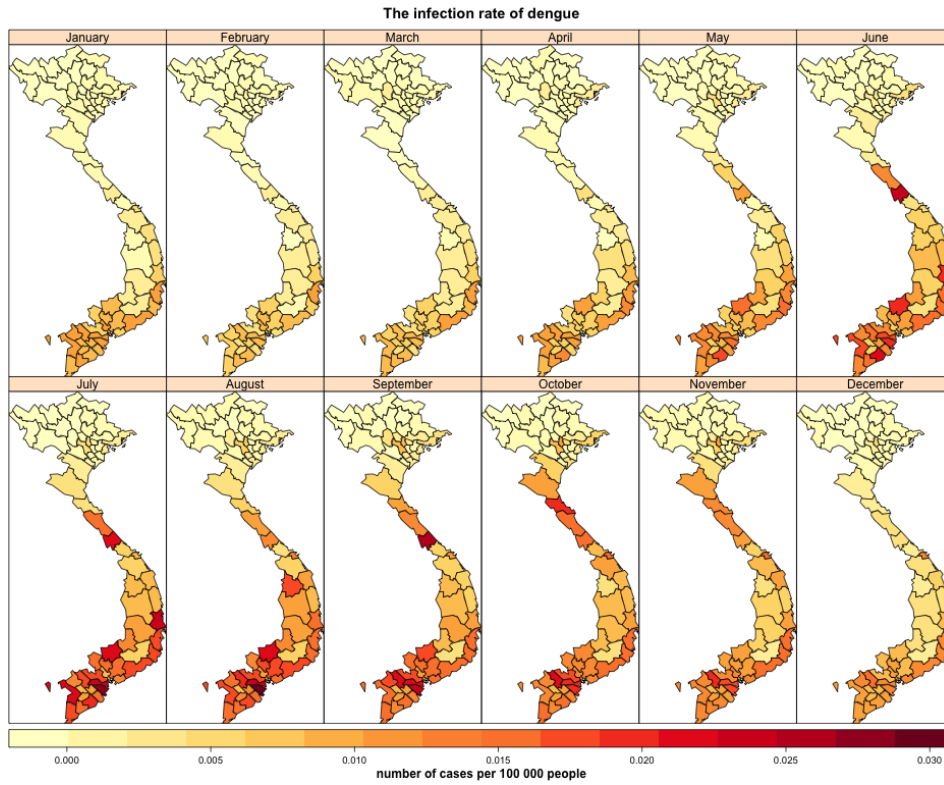


Fig. 1: Dengue incidence by month and province in Vietnam averaged over the 13 years from January 1998 to September 2010.

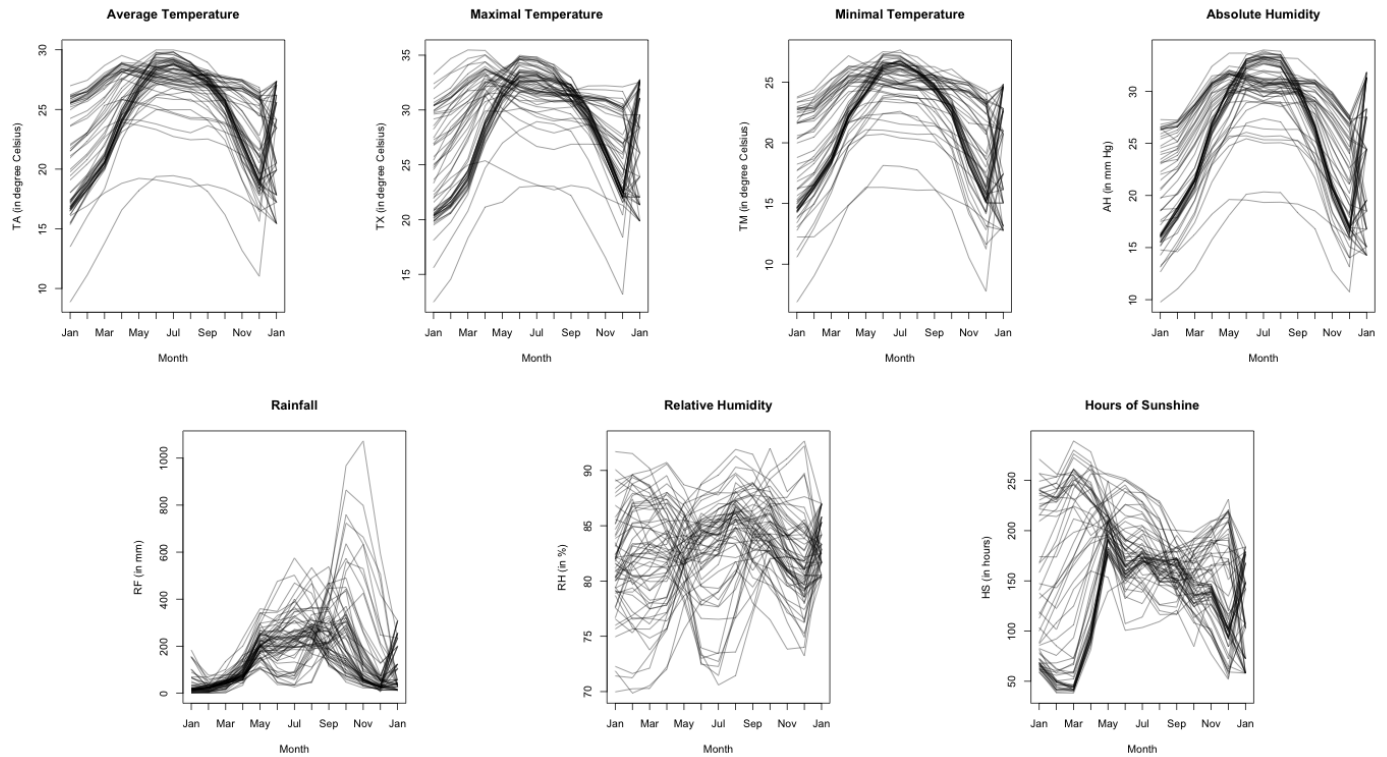


Fig. 2: Variations of the 7 climatic variables by month and climatic stations averaged over the 13 years from January 1998 to September 2010.

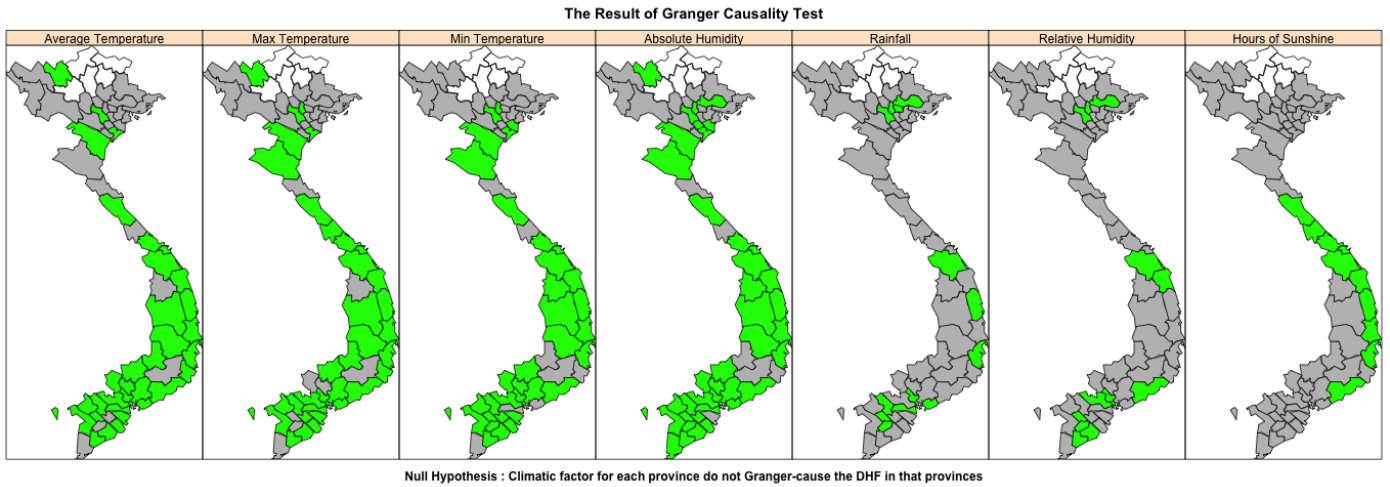


Fig. 3: Results of the Granger causality test for each climatic variable and each province. Green indicates the provinces for which the test is significant at the $\alpha = 0.05$ level, grey indicates the provinces for which the test is not significant and white indicates provinces for which the test could not be performed because of lack of data.

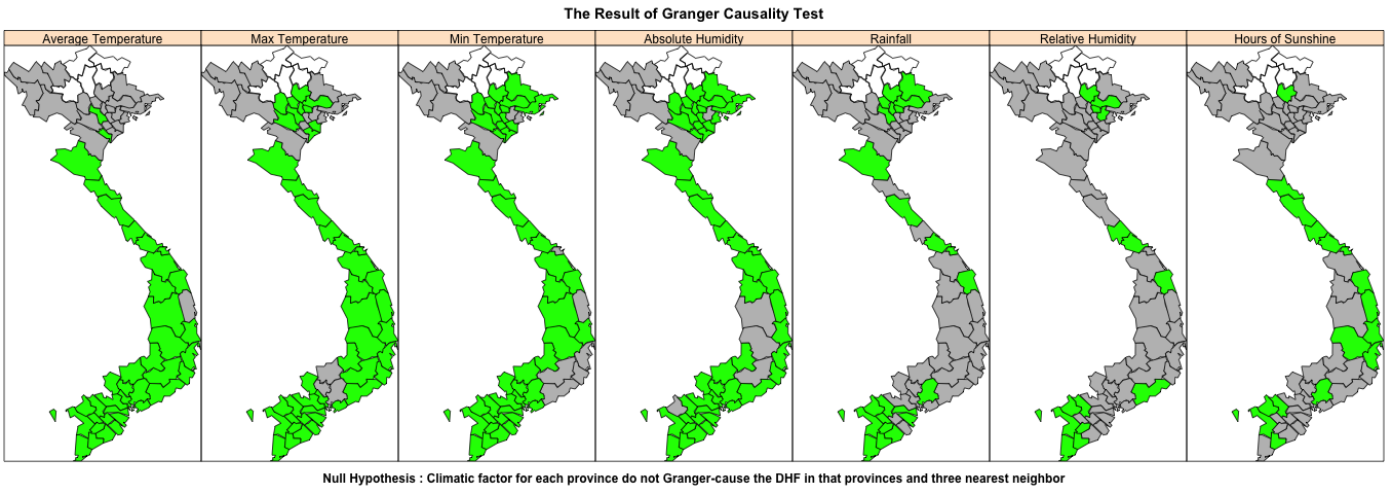


Fig. 4: Same as figure 3, adding dengue incidence in the neighbouring provinces as a potential Granger-cause of the dengue epidemiology in the focal province. The colour code is the same as for figure 3.

- [19] —, “Testing for causality: a personal viewpoint,” *Journal of Economic Dynamics and control*, vol. 2, pp. 329–352, 1980.
- [20] J. Geweke, “Measurement of linear dependence and feedback between multiple time series,” *Journal of the American statistical association*, vol. 77, no. 378, pp. 304–313, 1982.
- [21] B. Pfaff, M. Stigler, and M. B. Pfaff, “Package ?vars?” 2013.
- [22] H. Yang, M. Macoris, K. Galvani, M. Andrighetti, and D. Wanderley, “Assessing the effects of temperature on the population of aedes aegypti, the vector of dengue,” *Epidemiology and infection*, vol. 137, no. 08, pp. 1188–1202, 2009.
- [23] L. Lambrechts, K. P. Paaijmans, T. Fansiri, L. B. Carrington, L. D. Kramer, M. B. Thomas, and T. W. Scott, “Impact of daily temperature fluctuations on dengue virus transmission by aedes aegypti,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7460–7465, 2011.
- [24] T. W. Scott, P. H. Amerasinghe, A. C. Morrison, L. H. Lorenz, G. G. Clark, D. Strickman, P. Kittayapong, and J. D. Edman, “Longitudinal studies of aedes aegypti (diptera: Culicidae) in thailand and puerto rico: blood feeding frequency,” *Journal of medical entomology*, vol. 37, no. 1, pp. 89–101, 2000.
- [25] D. M. Watts, D. S. Burke, B. A. Harrison, R. E. Whitmire, and A. Nisalak, “Effect of temperature on the vector efficiency of aedes aegypti for dengue 2 virus,” DTIC Document, Tech. Rep., 1986.
- [26] E. Massad, F. A. B. Coutinho, L. F. Lopez, and D. R. Da Silva, “Modeling the impact of global warming on vector-borne infections,” *Physics of Life Reviews*, vol. 8, no. 2, pp. 169–199, 2011.

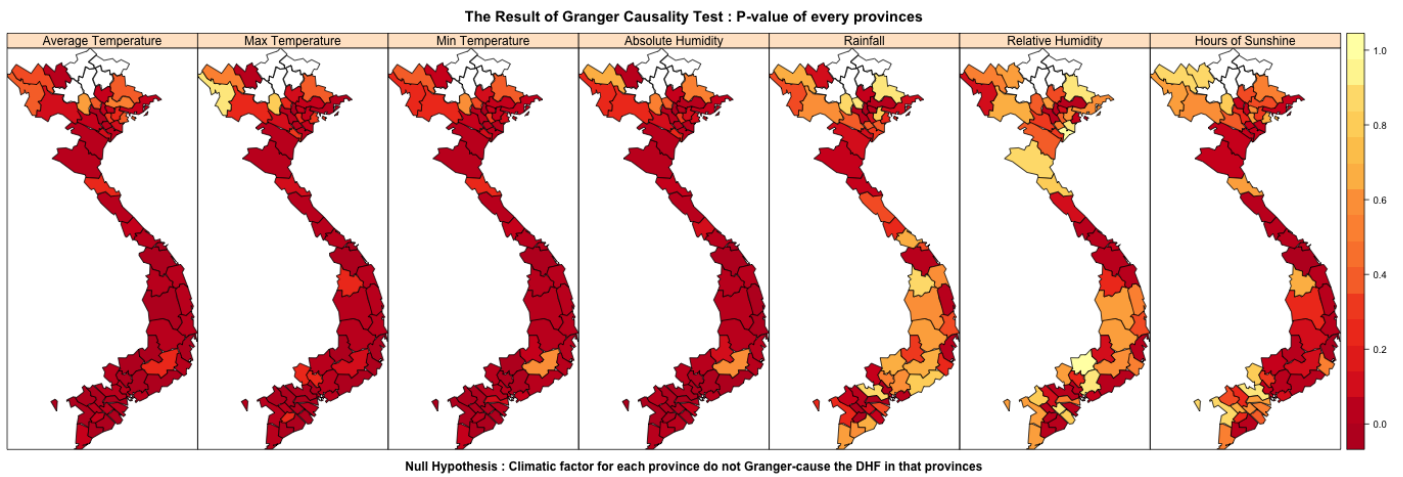


Fig. 5: Same as figure 3, showing the actual values of the p-values.

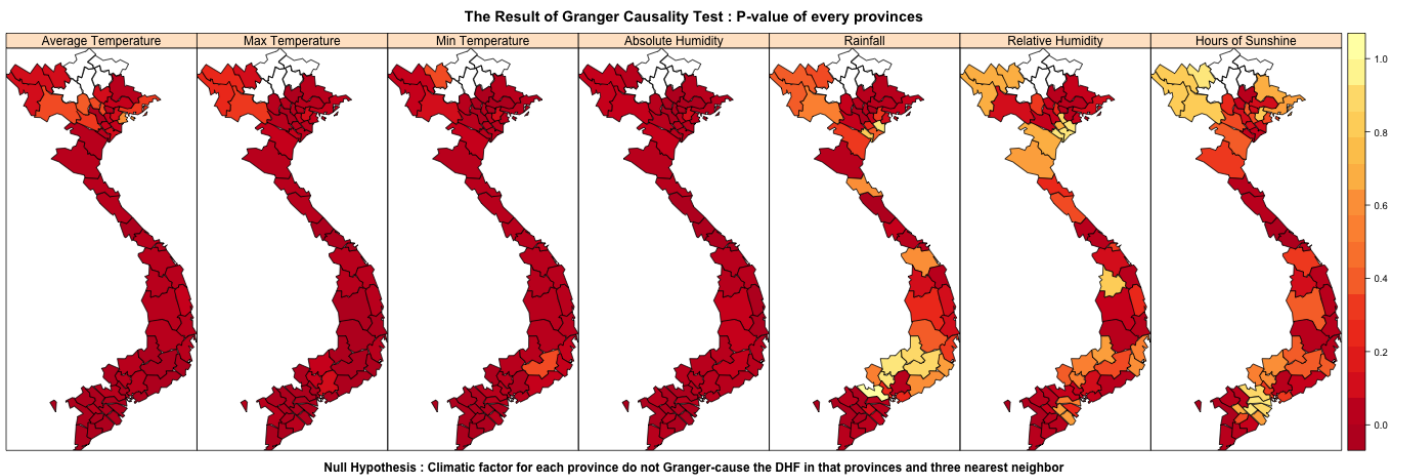


Fig. 6: Same as figure 3, showing the actual values of the p-values.