



## What is the probability that new epidemic phenomena could occur in future? A probabilistic big data analysis to prevent the emergence and spread of future epidemics

*Field: Mathematical Models/Big Data*

*Study by: F. De Cecco – Senior Consultant, G. Orlando – Senior Consultant*

### Abstract

After the dramatic effects caused by almost unstoppable spread of Sars-Cov-2, one of the most recurring topic addressed by studies and governments all over the world has become avoiding it a second time. The aim of this article is to address this issue from a different perspective, looking at the root causes that led to the current situation, by understanding which is the probability that new epidemic phenomena could occur in the future.

Starting from the results of a prediction of zoonotic spillover coming from mammals, which provides an overview of the geographical places where the pathogens transmission between species are the most likely, the main mobility flows will be analyzed in order to determine the possible spread of a future pandemic.

In particular, the aim of the analysis is to preview possible spread scenarios, based on human mobility networks and the global spread of influenza-like viruses.

### Introduction

Outbreaks of highly contagious animal infections such as swine fever, avian influenza and the novel coronavirus occur with serious public health implications and socioeconomic consequences. Revised understanding of the factors facilitating the introduction and subsequent spread of these viruses is crucial for future effective control.

The majority of human infectious diseases, which caused recent epidemics, are zoonosis, originating in wild mammals. Therefore the comprehension of the patterns of viral diversity in wildlife and the determination of the successful cross-species transmission, or spillover, are the goals for pandemic surveillance. Olival et al [1] proposed a study in which a model is developed with the aim to predict where in the world and in which host species is most likely to have zoonotic transmission. The author's outcome shows the geographical hotspot of "future spillover" in South and Central America, Africa, and Southeast Asia (with variations depending on the species analysed and the total mammal richness). On this conclusion, the goal of the current paper is to suggest an analysis of new pathogens infection spatial spread in order to produce geographic risk maps.

The spread of the pathogens causing a pandemic in people is particularly driven by human mobility flows and for this reason human mobility patterns continue to be analyzed in order to determine the spatiotemporal dynamics of an epidemic. A model, called Global Epidemic and Mobility (GLEaM), that integrates demographic analysis and population mobility data in a stochastic approach has been developed to simulate the spread of epidemics at worldwide scale [2]. The work of Balcan et al., affirms that, given a set of initial conditions for a local outbreak of a new influenza-like virus, the timelines of the arrival of the epidemic in each country and the consequent infected peaks are mainly affected by the human mobility network that connect the different world region.

At worldwide scale, international movements of passengers by air travel appears to be the most relevant factors for dissemination of the pathogen [3] and govern the global pandemic behavior. With focus on European countries it is possible to see how Merler et al. [4] using a simulation model of the spread of influenza pandemic determine that the importation of the first case from abroad is more likely to occur in the population high mobility areas such as the Western part of Europe.



These computational approaches can be combined within a simplified theoretical framework where the areas with major viral richness are compared with the patterns of human mobility to evaluate the country where the probability of importing the first case is higher and to propose possible epidemics scenarios.

### Data Evaluation

Olival et al. [1] considers the areas with zoonotic virus richness in a number of mammals wild species with the aim to evaluate the risk map for spillover probability. One of the results reported in their study is shown in the Figure 2: and highlights the areas where the total viral richness is observed. In particular the Central and South America, Central Africa and South-Est of Asia are the regions identified by the model as the areas where a new animal infection is more likely to occur having the higher viral richness.

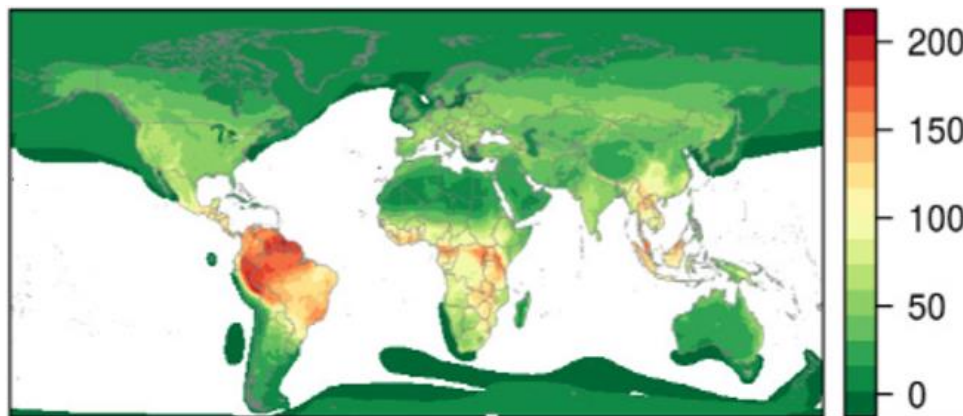


Figure 1: Mammal richness for species [1]

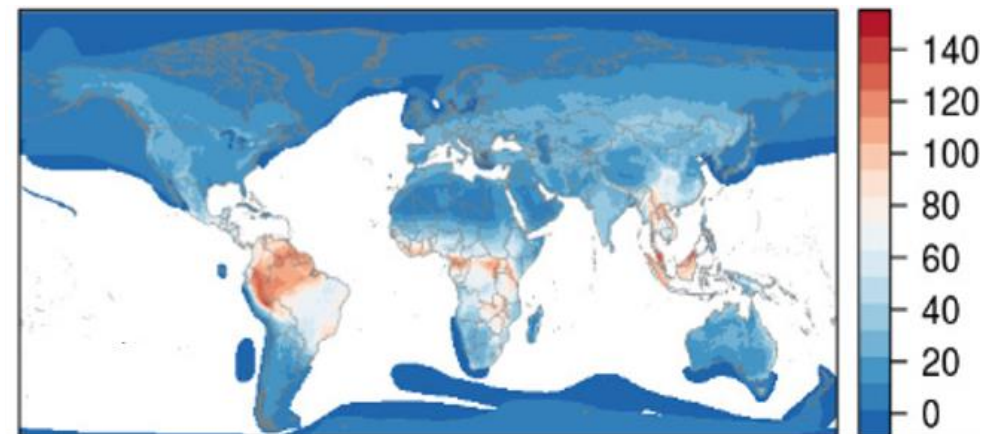


Figure 2: Areas with poor viral surveillance in mammal species [1]

Based on these images, it is clear that the above mentioned areas could be considered as the next “birthplace” of a new epidemic. For this reason, our aim is to cross these results with the transportation data available, in order to make a prediction of the possible evolution of a new outbreak. The results of this analysis could be used as a map to avoid an uncontrolled spread, as already happened with SARS-CoV-2.

Airline transportation network shrinks the geographical space by reducing travel time between the world’s most populated areas and defines the main channels along which emergent diseases will spread [5]. These mobility factors are considered in the GLEaM [2] model, a computational big data



analysis model used to predict the evolution of a hypothetical epidemic behavior and to analyze potential outbreak distribution patterns.

Liu et al. [6] considering different possible models to predict the spread of SARS-CoV-2, shows that GLEaM model “is capable of reliably forecasting COVID-19 activity even when limited historical disease activity observations are available”.

For this reason we apply the above mentioned model through a simulator. Starting from the results of Olival et al. [1], we made a simple consideration, evaluating the most likely starting location for a new hypothetical epidemic. Since the spread pattern of an outbreak departing from South Eastern Asia was already evaluated after the dramatic events caused by SARS-CoV-2, and since Central Africa have a considerably lower number of connection with the rest of the world, we have decided to consider the possible scenario of an epidemic starting from South America. The following frames captured from the simulation shows the evolution of a possible outbreak.

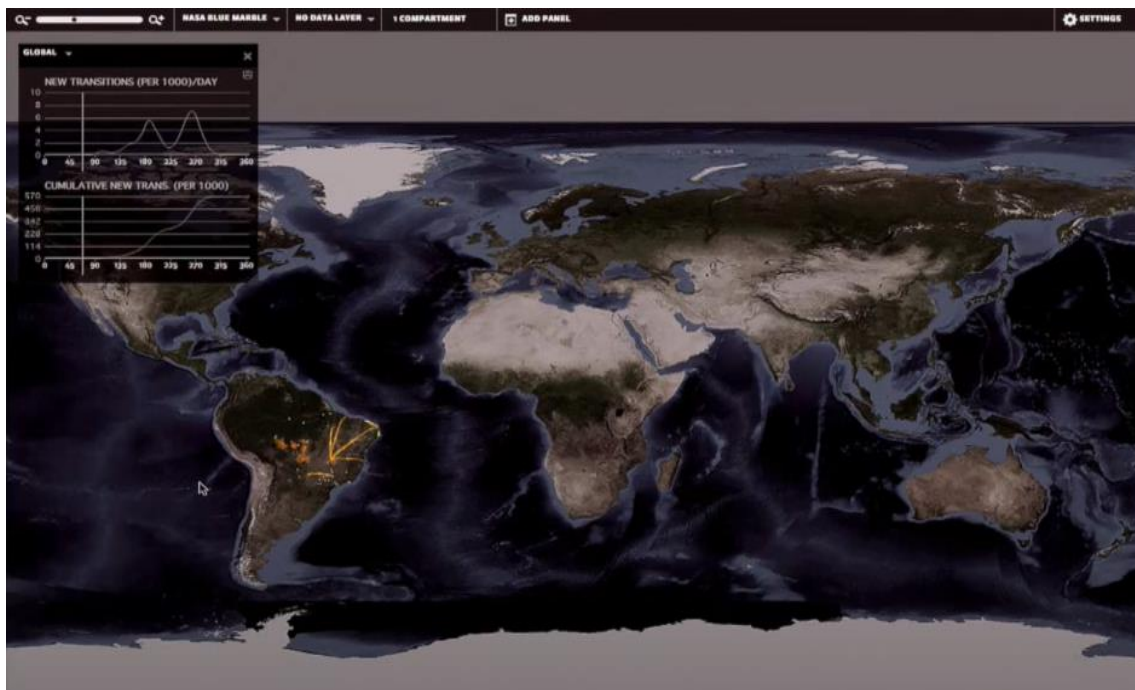


Figure 3: GLEAM Simulator: Epidemic transmission pattern, first frame captured

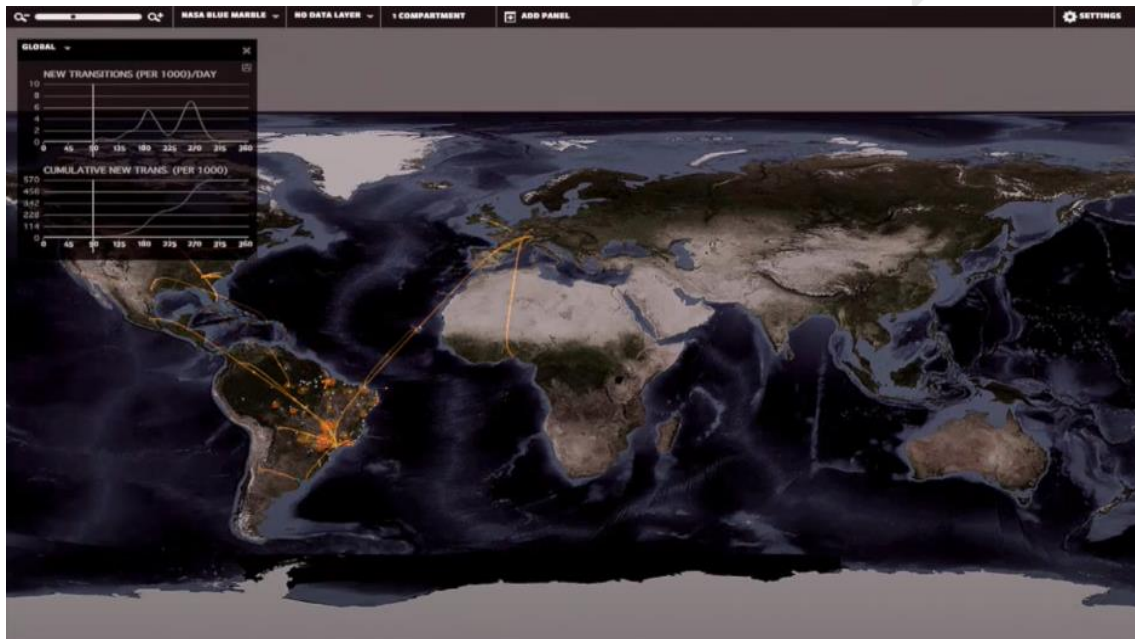


Figure 4: GLEAM Simulator: Epidemic transmission pattern, second frame captured

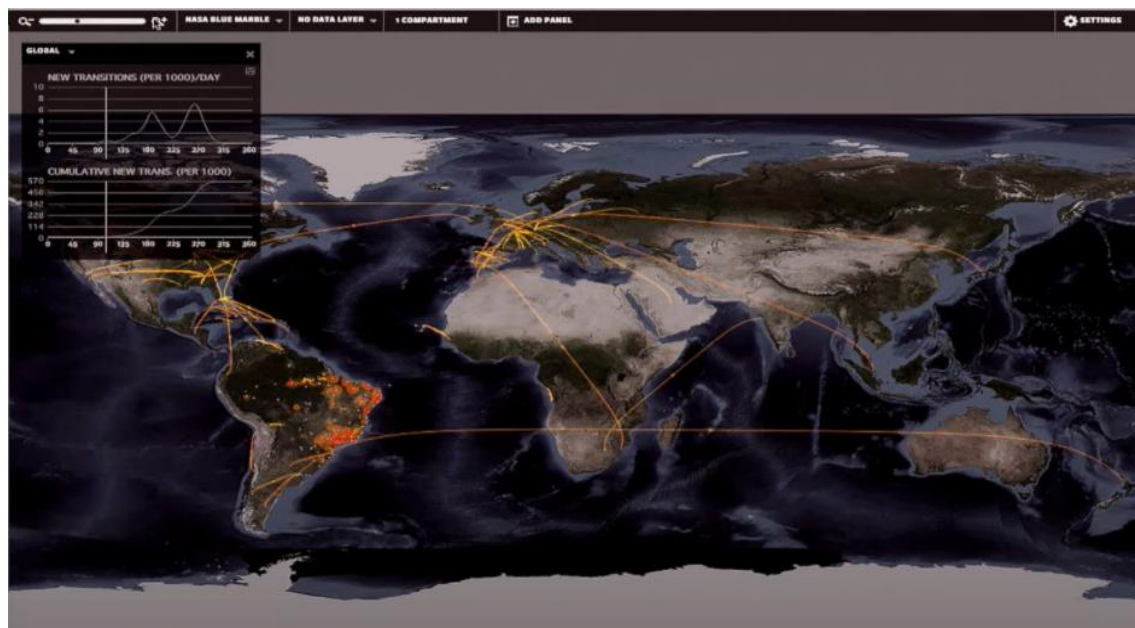


Figure 5: GLEAM Simulator: Epidemic transmission pattern, third frame captured

As it is possible to see from the simulation, the epidemic dynamics should follow the main mobility patterns and spread in North America first and then in Europe area. Flight connection between South America and Asia regions can be considered as negligible and this consideration reflects in the simulation analysis in a delayed spread of the epidemics from one area to the other one.

According to the previously mentioned factors is possible to assume that the global distribution of a contagious infection will impact earlier the areas destination of the majority of the flight routes.



Figure 6: Flight and connections representation

Moreover, to enforce the applicability of mobility model to epidemics spread is possible to see that the same patterns can be detected in the behavior of the novel coronavirus outbreak, where the imported case from China have been observed firstly in Europe and North America, then in Africa and South America that have a minimal percentage of flight routes from China, so the virus spread has been delayed.

## Conclusions

The results of this analysis inevitably lead us to consider two types of risk:

- A risk that we could define as **direct**, which is related to the spillover likelihood, that affects mostly the areas with a growing urbanization
- A second type of risk, on the other hand, is the one we can define as **indirect**, which takes into account the connections and mobility captured by the GLEaM model

On this basis, it is necessary to put in place different kind of mitigation actions that will reduce the above mentioned risks. In order to reduce the direct one, it is possible to think prospectively, by proposing monitoring programs with the aim to control the different channels that lead to the birth of a new infectious disease. Going more in detail, a suggestion for example is to intensify the controls on food distribution and animal farms in growing urbanization areas.

As far as it concern the so called indirect risk, it should be possible to apply mathematical models such as the GLEaM to identify the likely path that an epidemic phenomena will follow and avoid the uncontrolled spread of an infection by blocking the strategic mobility and communication channels.

The approach proposed by the present paper aims to be a basic tool to reduce the dramatic effects of a future pandemic event and to prevent what unfortunately already happened for SARS-CoV-2.



REFERENCES:

1. Olival K. J. , Hosseini P. R., Zambrana-Torrel C., Ross N., Bogich T. L., Daszak P., Host and viral traits predict zoonotic spillover from mammals, *Nature* Vol. 546 no 7660, pp 646-650, 29 (2017)
2. Balcan D., Gonçalves B., Hu H., Ramasco J. J., Colizza V., Vespignani A., Modeling the spatial spread of infectious disease: the Global Epidemic and Mobility computational model, *Journal of Computational Science*. Vol 1, no 3, pp 132-145 (2010)
3. Apolloni A., Poletto C., Colizza V., Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic, *BMC Infectious Disease* 13:176 (2013)
4. Merler S. & Ajelli M., The role of population heterogeneity and human mobility in the spread of pandemic influenza, *Proceedings of the Royal Society B* 277, 557-565 (2010)
5. Colizza V., Barrat A., Barthelemy M., Vespignani A., “The Modeling of Global Epidemics: Stochastic Dynamics and Predictability”, *Bulletin of Mathematical Biology*, 68: 1893–1921 (2006)
6. Liu D., Clemente L., Poirier C., Ding X., Chinazzi M., Davis J. T., Vespignani A., Santillana M. “A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models” arXiv preprint arXiv:2004.04019, 2020

GLOSSARY:

Terms and acronyms	Description
Coronavirus	Any of a group of RNA viruses that cause a variety of diseases in humans and other animals
GLEAM	Global Epidemic and Mobility
Sars-Cov-2	Severe acute respiratory syndrome coronavirus 2
Spillover	Capacity of a virus to spread from an animal species to another one
Stochastic	Having a random probability distribution or pattern that may be analysed statistically but may not be predicted precisely
Viral Richness	Number of pathogens presents in an animal species in a determinate geographical area