International Journal of English Language Teaching
Vol.11, No.4, pp.,7-19, 2023
Print ISSN: 2055-0820(Print)
Online ISSN: 2055-0839(Online
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

An Innovate Comparative Research Method to determine the convergence (/or divergence) of the Infective Virus like COVID-19 & Common Flu

Lam Kai Shun (Carson)
British National Oversea
B.Sc (HKU).,
M.Com.Engine., M.I.T.E.,
University of Hong Kong, Hong Kong
Fellow of Scholar Academic Scientific Society, India
Email: **sonsonshek@gmail.com**

doi: https://doi.org/10.37745/ijelt.13/vol11n4719 Published July 26 2023
Citation: Shun L.K. (2023) An Innovate Comparative Research Method to determine the convergence (/or divergence) of the Infective Virus like COVID-19 & Common Flu, ***International Journal of English Language Teaching***, Vol.11, No.4, pp.,7-19

ABSTRACT: *Infectious disease has a long history with our human beings. These diseases have been well known from the common flu in Spring to the most recent one such as the COVID-19 etc. People always want to find a cure such as vaccine to them. However, as the diseases' viruses have thousands or millions of mutation in their DNA or RNA, it is therefore difficult to predict the next popular species that may be spread among people in the coming year. Mathematically, the combinatoric plus the optimization techniques may be the most feasible way to solve the problem. In the present thesis, this author suggests an innovate comparative research method that includes research group (HKLam regression) and the control group (other regressions) to the infected virus such as COVID-19 or common flu etc. The aforementioned method is used to determine these virus(es)'s convergence (/or) divergence mutations such that we may further genetic edit them to obtain several feasible vaccine candidates for further investigation in human beings etc. In other words, we are trying to artificially developed a tailor made "cow-pox" for these infected viruses and hence may be save millions of our lives.*

KEYWORDS: innovate comparative research method, convergence, divergence infective virus, covid-19, common flu

INTRODUCTION

Humans have suffered from the influenza virus (i.e., the flu) for hundreds of years. The situation has historically become even worse when sudden outbreaks occur, such as the 1889 Russian flu (1 million deaths), the early 20th Century Spanish flu (20–100 million), the 1950s Asian flu (1–1.5 million), the 1960s Hong Kong flu (0.75–1 million), and the 2009 flu pandemic (swine flu), which accounted for several hundred thousand deaths [1]. Although the number of people infected has decreased since the introduction of the influenza vaccine, there is still a chance that future vaccines might be ineffective. Hence, a more accurate prediction mechanism is required in order to prepare for new strains of the influenza virus. Based on Hong Kong's present public healthcare system, this will decrease the social risk & nervous if there is another outbreak. All influenza viruses have two types of proteins: H and N, which range from (1–16) and (1–9) and can eventually mutate through mathematical combinatorial processes etc. The various combinations of virus mutation are extremely large, and therefore it is

extremely difficult to predict future flu types. As such, finding the right vaccine is crucial. In the present research proposal, the applicant will try to develop a novel method for finding the convergent species of infected viruses, such as influenza and SARS-CoV-2 through my self-developed HKLam theory together with some other regression methods. Then, we may perform scientific observation/theoretical data-based comparisons for regression convergence calibration (this will be done in order to search for the minimum of the expected gradient descent model [31] such that one may finally obtain a relatively stable mutated infectious virus, etc. [34], [35], [36]) The aim is to apply genetic editing techniques such as CRISPR-Cas9 (or precisely gene prime editor) to remove harmful DNA genes for the convergent species virus. One may then develop an ultimate vaccine that is similar to the case of cowpox (or it is now the genetically modified human-made type cowpox for influenza and SARS-CoV-2) and may hence finally eliminate smallpox, which has been around since the last century.

LITERATURE REVIEW

The Present Vaccine Technology – (Inactivated/Killed Virus and Other

According to the Centers for Disease Control and Prevention (CDC) [2], there are three common types of vaccine technology. They are known as the egg-based flu vaccine, cell-culture based flu vaccine, and the recombinant one. This applicant will describe the basic ideas behind these vaccine technologies. To begin, for the egg-based vaccine, the manufacturer must first select the candidate vaccine viruses (CVVs) and then grow them in chicken eggs, with reference to Food and Drug Administration (FDA) regulatory requirements. These CVVs are then injected into fertilized chicken eggs and incubated for several days such that the viruses are allowed to replicate. The fluid that contains the virus is then harvested from the eggs. Hence, in the case of inactivated influenza vaccines, the vaccine virus is then killed together with the antigen, then there is also an purification. The manufacturing process continues with quality testing, packing, and distribution, etc.

Second, for cell culture-based flu vaccines, the CDC or its laboratory partners usually relies on the influenza viruses that have been previously grown in cultured mammalian cells to make CVVs, which are then sent to the manufacturer [2]. Instead of inoculating CVVs into chicken eggs, the manufacturer must place them in cultured mammalian cells, and hence let CVVs make copies for a few days. These processes are similar to what is done for the chicken eggs. The advantage is that a cell culture-based vaccine may be faster than what can be achieved with chicken eggs.

Finally, recombinant flu vaccines are created synthetically [2]. First, scientists will obtain the virus's gene that is responsible for the genetic instructions to make hemagglutinin or the HA. In practice, HA is an antigen that helps our immune system create antibodies working against the targeted virus. The immediate result is the baculovirus that combines with the HA gene and then the infected invertebrates. The recombinant baculovirus will then deliver the genetic instructions into an FDA qualified host cell line to make the flu's HA antigen. The antigen collection process is similar to the above two aforementioned methods. The advantage of the recombinant vaccine is its production process is the fastest, as it can bypass the growth of CVVs in either eggs or cell culture-based vaccines.

Proposed Vaccine Technology – (Human-Made Genetic Modified Cowpox)

Any matrix (including mathematics graphs, or to be precise, the phylogenetic tree graph in terms of matrix form and its representation [3]) with immediately linear transformation (solution) vector **x** and the resulted vector, **b** can be approximated by the regression [4] and various linked causality philosophy variables [5]. The vice versa is also true. I have named this type of regression as HKLam Regression.

For the major contribution of the HKLam regression, every mutated influenza and SARS-CoV-2 variants can be approximated by their corresponding HKLam Regression and will generate different virus mutation regression models [4]. Next, we may compare these variant models by using different error parameters [6]. Next, we may select the most feasible mutated virus model as the next popular infection species. By continuing the above process a few more times until the mutated convergent [7], we may finally develop the corresponding optimized vaccine — the backward mutation technique [8] to practically repair the harmful parts of the virus. Hence, the repaired harmless virus will act as an ultimate vaccine for any possible future virus infection breakthrough. It is therefore possible to generalize a type of human made 'cowpox' like vaccine for destroying infected viruses, just like the elimination of small pox in the last century. In practice, the aim of the proposed study is to apply the HKLam regression model for both influenza and SARS-CoV-2 etc of the phylogenetic tree of the infected virus, such that one may construct a new type of regression model (rather than the traditional phylogenetic regression) to approximate their mutations. Once we have modeled different variants, we may finally get the most optimized one, and through backward mutation, a genetic-editing technique could be used to rebuild the virus and get the desired ultimate vaccines. Our research questions are the following:

i) How may one develop the various models of the virus variants' phylogenetic tree and predict the next feasible variants?
ii) What is the best optimized next model for a particular virus, and how may it be found by comparing model error parameters?
iii) In which way may we get the mutated convergent species of the variant?
iv) How may we rebuild the full phylogenetic mutation tree of the virus?
v) How may we develop the ultimate vaccine for the targeted virus?

RESEARCH GAP IN VACCINE TECHNOLOGY

The research gap of the proposed PhD project is that most of the recent vaccine technology is based on inactivated viruses in order to stimulate the human immunology system to gain antiviral power. However, present inactivated virus vaccine techniques cannot handle continuing mutated viruses, and will soon lose its efficiency. The key to overcoming endured mutation is to find the (lower bounded of the death rate) convergent species of the virus, such as the COVID-19 Omicron variant (as its variant species are now **gradually approaching to a mutation convergence of becoming a less-toxic & lower deadly rate virus** [10] & [11]).

The proposed novel vaccine technology aims to find this type of convergence virus species for cases such as the common flu and SARS-CoV-2 through the HKLam Theory and various regression methods. Next, we may perform scientific observation/theoretical data-based comparisons for regression convergence (and/or business gradient descent machine learning [31]) calibration for

infectious viruses [34], [35], [36]. As such, one may mathematically and statistically foresee the wanted convergent species of the virus. By applying the latest genetic prime editing technology [9] (possibly supplemented by CRISPR-Cas9 technology) to remove harmful DNAs out of the **final convergent (boundary bounded) virus** in a particular mutation (or may be in multiple ways) and convert it into a vaccine, the outcome result will be an ultimate one. That means, **human beings will have the chance to develop a new man made genetic modified "cowpox" or an artificial one to fight against infectious viruses** like the influenza virus and SARS-CoV-2 (until the next significant change), although they will continuously mutate according to rule — Survival of the fittest in the human body and other environments.

In real life practice, the final convergence of a de-toxic virus-made vaccine may at least last for a longer period until the virus's next significant mutation (e.g., the change from the Wu Han high death rate original Covid-19 to the Omicron variant with a relatively mild in low death boundary rate but high spread rate [10] & [11]). In fact, the **key trend** is **an increasingly higher transmission rate with the offspring of the Omicron variant, a lesser death rate, and a higher chance of immune escape**. Thus, **multiple types of virus DNA's functional modification** (or the related genetic editing etc.) may be needed to apply as the steps aforementioned simultaneously.

As an extension, one may genetically modify a virus such that it may infect and thus kill the SARS-CoV-2 virus together with the antigens that let the human body's immune system find the SARS virus, which is similar to what happens with cancer [13]. Without a doubt, **a better method** to improve my suggested scheme is to **use the defective gene of the genetic modified vaccine virus for vaccination. One example is the use of the relatively stable parts [32] in the RNA of the feasible candidates for the COVID-19 vaccine virus – HCoV-229 [33] after a regression convergence computation whenever necessary. Corresponding genetic-editing may then be performed for a safer and more stable proportion of RNA as vaccination.**
This is because the modified virus may still have the chance to mutate into novel strains that may cause harmful effects in humans [29], [30].

Major Implementation Method

A. Control Group – Using both Phylogenetic Regression and Evolutionary Regression to Find the Solution x of the Matrix Equation Ax = b.

The main aim of the phylogenetic regression is to first establish the phylogenetic correlation matrix from the phylogenetic trees of the virus variants. We may then compute the corresponding regression equation (theoretical values) according to the (*) with a known method regarding how to compute the statistical parameters β and ε of the equation $y = X\beta + \varepsilon$ that have been shown in the literature review. At the same time, we may apply evolutionary allometric regression with other known observational parameter values β' and $r(X_1,\ldots,X_m)$. Hence, by performing the scientific data-based research and comparing both theoretical and observational values in the phylogenetic and allometric regressions (such as the observational changes in the characters of the spike protein of mutated SARS-CoV-2), one may then easily find out the values of the solution **x** to the matrix equation A**x** = **b**, [34],[35],[36]. By further adjusting the value of **b**, and continuing/repeating the above process together with the matrix mathematics operations, we may finally get the convergent species of the investigated variants. In a vice versa way, with known values of next **b** and **x**, one may get the updated

next matrix A recursively etc. [16]. In short, we are continuing to perform the regression convergence (such that one may guess and find the minimum of the expected gradient descent model [31]) iteration/calibration to the final and relatively more stable mutated infectious virus [14] as a feasible vaccine candidate. Hence, in such a way, both recursive and forward-backward philosophy is naturally formed. Furthermore, by comparing the model error parameters of different variants, one may then select the best species of the ultimate convergent variant species between various kinds of virus variants. It is worth noting that this procedure may take up to half of a year.

Research Group – HKLam Regression plus (Genetic) Trend (and Drift) Regression

The major goal of the HKLam Regression method is to build a regression model equation for investigating influenza or the SARS-CoV-2 virus. We first make use of the phylogenetic correlation matrix of the investigated virus species. Then we work with the trend (and drift) regression where the mutated genes are considered as the experimental observation (I propose that this type of of regression should be named as a genetic trend that is inspired from [14]). The mutated genes will be considered as the y-variable or the y-axis while the time variable may be viewed as the x-variable or the x-axis. When a time variable is used, it may then turn into a bio-statistics survival analysis [22]. We may then plot the corresponding genetic trend variation graph with several y-intercept. The best optimal intercept value should be with the least y-valued intercept one (say **c**). The value **c** is then used together with the phylogenetic correlation matrix as the initialization data for the starting of the computation for the linear transformation (solution) vector **x**, and thus we may compute another new vector **b'**. Then, by using the HKLam regression method, **b'** may then be expressed as the simplified Gram-Schmidt process in [2]. Next, we may combine the above method together with [16] and continue the recursive process using matrix mathematics as described in [16]. Then, we may finally get the convergent species of the investigated virus. With comparison to the control group, one may then adjust and check whether the convergent variant's species is the best ultimate final one. This procedure may take up to half of a year.

Using Artificial Intelligence to Simulate the Backward Mutation Through a Genetic Algorithm

In order to get back to the present popular species of the variants, we may apply artificial intelligence (AI) to simulate the backward mutation [18] through the genetic algorithm [19]. Practically, the genetic algorithm consists of a cycle status process as evaluation, selection, crossover, and mutation [19]. With the help of AI applied to the genetic algorithm for the final convergent species [20], one may select and determine the best path backward to the present popular species (or even train the AI to search for the origin of the virus if there is enough data). Indeed, both of the popular virus species variants that mutated to the final convergent ultimate species, together with the convergent ultimate species backward to the popular virus variant species, constitute a kind of biological forward & backward mutation philosophy. This procedure of programming may take at least nine months or one year.

A Brief Schedule of the Project

1. Elementary data collection and Literature review – half a year;
2. Control group research – half a year to nine months;
3. Research group investigate – half a year to nine months;
4. Simulation of the virus genetic algorithm – nine months to a year;
5. A final mix-up – three months to half a year.

The applicant will first spend half a year to collect elementary data and write the corresponding literature review from both of the web and library searching. I will then use half a year to nine months

for the control group research together with the research group investigation. I will spend a year for the simulation of the virus genetic algorithm through computer programming software such as R and JASP. Finally, I will use half a year for a final mix-up such as the final writing of the present PhD thesis and other related research reports. A detailed schedule has been shown through the associated Gantt-Chart.

CONCLUSION

Finally, in the real case, it is important to note that one can mathematically model the virus infection processes that lead to liver cancer using Wnt signaling (e.g., corresponding to patterns in metabolism in colon cancer) [21]. The result being that one might be able to find a suitable virus, such as an oncolytic virus, for the treatment of colon cancer [22]. Moreover, it might be possible to 'reconstruct' the selected virus using CRISPR (according to the mathematical model computed previously) and train it to attack those cancer cells whenever humans cannot find a suitable virus that is suitably used as a vaccine for something like liver cancer. Hence, we might be able to develop the necessary drugs (that contain the suitable virus) to attack the respective liver cancer virus and balance the micro-organisms that co-existed in the surrounding cancer infected area. Theoretically, the tumor size will gradually be diminished, and we may even completely cure liver cancer caused by the hepatitis C virus. Similarly, we can apply the same mathematical and statistical modeling method for other diseases such as HIV and AIDS, and develop corresponding drugs using CRISPR re-construction for suitable viruses. An international study done by a South African professor found that when the COVID-19 virus meets AIDS in the human body, COVID-19 can mutate and further evolve into a more harmful virus. On the other hand, COVID-19 can be mutated and evolve in Netherlands minks to become less harmful, just like the case of small pox. This is an interesting result and may be used in the control of future virus mutations. In fact, the HKLam Theory can thus be applied as stated in the above section if the butterfly effects or the Lorenz attractor exists in COVID-19 mutations in order to find the saddle or equilibrium point [23]. Thus, it can then act as some form of control to harmful virus evolution. Finally, the SARS-CoV-2 virus mutates within HIV patients because the enzymes that copy RNA are prone to making errors [24]. This event implies the need to develop corresponding anti-enzymes of SARS-CoV-2 drugs to prevent continuous mutation in SARS-CoV-2 among HIV patients. Most of the present antiviral drugs in the USA for COVID-19 were made based on such a theory.

In a nutshell, in this project, I suggest that we may make a precise guess (no matter from the regression or the gradient descent with business machine learning [31] etc) about the infected virus's ultimate convergent species from the comparison between the control group – phylogenetic and the observed mutation allometric regressions with the research group – HKLam regression and the genetic trend (& drift) regressions (genetic mutation with the minimum y-intercept as the initial data for HKLam one). Hence, we have an in-depth idea and understanding about the mutation of the infected virus such as influenza and SARS-CoV-2 [25]. Then from the genetic editing techniques [26] applied to the present popular virus species, one may convert it into the virus without those harmless DNAs and be removed all dangerous mutated genes [27], one may develop and get the final wanted vaccine just like the cowpox (this time the artificial man-made one) that had already eliminated the smallpox [28] in the last century etc.

Indeed, from the comparison of observational and theoretical data in the section of implementation method, we may further develop the following general algorithm for my research:

1. Applying another regression once more (just like the case of many body problem for gravitational objects – may be expressed in terms of matrix approximated by my HKLam theory or a recursive regression in the present situation) – by using either principle of least square or M-estimator method etc for the regression data between the observational and theoretical regression for both of the control and research groups; (i.e. iterates for a feasible number of calibration – say 25 times etc.)
2. Using the Taylor series approximation (or the "poly-fit-n" function in the Matlab as the Gradient Descent model/function) to fit for the above two sets of data's difference;
3. Compute (may use Matlab code) the optimum (maximum/minimum) of the above Gradient Descent model (by using the steepest descent and/or the second variation method [37], to be precise, the differential corrections method [38]) that obtained from the above Taylor series;
4. Obtain the relatively stable and best optimized species of infectious virus; (N.B. Iterate convergent does NOT necessary imply the maximum/minimum (or optimum) of the data differences. It only shows the convergent – either by increasing/decreasing with a delta for an upper/lower bound.)
5. Use Prime editing method etc to modify the above convergent virus for detoxification as one of a feasible candidate of artificial-made "cowpox" vaccine.

This applicant wants to remark that by mirror reflecting the minimum point of the (converged) virus species found, theoretically, one may discover the maximum point of the infected virus well before the spreading or the source of the present widely affected influenza (or the common flu). Indeed, by continuing the above "gradient minimum-reflecting maximum" searching process with one by one levels recursively for every mutated generation, one may finally achieve the ultimate ancient origin DNA or the ancestor of the infected virus (as the earliest stated in the written historic documents) such as our human common flu etc.

To sum up, the main focus of this project is to employ the techniques of double regression through the virus data (theoretical and observational together with research and control group) convergence method (principle of least square or m-estimator etc) of machine learning (gradient ascent/descent) etc. In some cases, we may even perform high dimensional linearization from the related Jacobian matrix (indeed they constitute a forward-backward philosophy). The result may be the competing species model that leads to the dynamic phase plane etc. Hence, the most feasible final outcome or product may be an artificial "human tailor-made" cow-pox for our future infected virus such as the common flu or SARS-CoV-2 when we apply the prime editing technology for detoxification of such convergence infected virus etc. The significance of the present project is that we may introduce machine learning method for the developing of an artificial "human tailor-made" vaccine for the infected virus instead of using some other tools to guess the next year's randomized genetic mutation among the targeted virus etc. Indeed, for the case of SARS-CoV-2 virus, we may still apply the linearization and machine learning etc to get the competing species model and hence obtain the general pattern or the tendency of the divergent infected virus within several computational iterations (the meta-competing species model) for further research (in forecasting of virus mutation etc) after this project. For the divergent case of mutated virus, we may employ the financial time series (techniques) and compute the possible tendency line or even the approximation functions in some other divergent cases etc for the improved vaccine production. Indeed, the phylogenetic analysis of the virus protein structure may help us find the local ancestor of the present convergent species etc. But luckily, the divergence

mutation of the virus SARS-CoV-2 forms only the minority of our World's present diseases, so my project's main focus is still on those convergent virus data only, the divergence part is just for the future optional research etc.

Limitation of the Present Research

It is no doubt that I have developed a novel type of artificial vaccine which may revolutised the present vaccine technology. But we may need to go a step for the evaluation of the proposed human-made cowpox. Hence, there may be a need to have a comparison between old and my proposed novel type of vaccine technology. Details of other imaginary vaccines comparison have been shown in my paper named confusion matrix as an example [6]. One may require to find the model parameters like F1 score for the comparison from the confusion matrix. However, the full comparison between the old vaccine and the artificial one together with the forward & backward philosophy of confusion matrix and model parameters are out of the scopes for the present PhD research project as it is just a pioneered study. Thus, I may left these vaccine details to our future study after the present PhD project if there may be. Moreover, it should be noted that coevolution may happen when the virus has already adapted in the usual Darwinian sense by evading immunity through antigentic mutation. This applicant does not include such issue in the present pioneered research. Practically, coevolution and vaccination is a recursive process where the virus may evolute according to the variation of vaccine formulation. In such a case, we may attain a certain level of control to the mutation of the virus. It is because there may be correlation between the virus mutation and the vaccination or we may even control the locus of virus mutation (i.e. towards what direction the virus may mutate in the coming cycle) under suitable vaccination formulation in some testing animals etc. Actually, there may be a three world philosophy or the inter-relationship between the virus mutation, vaccine and the immunity. But this is another research issue that needs to be left for other biological professionals and I will only focus in the mathematics of convergence and divergence for the mutation of flu and COVID-19. In additional, one may need to feedback those real experimental data into the predicted for the calibration or the adjustment of the computational virus control system. Finally, one may also need to apply the comparison of the time series in the case for the divergence of the virus etc. All of the aforementioned are the necessary parts that requires our particular and specific attentions. Furthermore, for the prime genetic editing, one may apply the combinatoric optimization in finding the best fitted environmental factors – say "A*" for the controlled virus mutation with a particular genes – say "A". In a vice versa way, if the virus shows the particular genes – say "A", then this may imply the existence of such environmental factors "A*" etc. In such a way, we may have the chance of guessing the historical origin of the infected virus like influenza etc. Certainly, the combinatoric optimization for a controlled genetic virus mutation may violate the ethics in human religious or we – human beings is now doing the job of our God – in creating life. However, the above religious ethics may be reduced to a love issue – may our God determine our beloved another part before our birth or we may have the freedom to select who may be our beloved another part. Then this may be a question in human kind of long arguing philosophy – determinism or free will problems indeed. Anyway, what this author may believe is that there should NOT be a conflicts between science and our God's creation. We may ultimately find some ways to solve such kind of genetic-religious ethics problems etc.

Appendix – Mathematical Prelminary

1. What is Genetic Trend (& Drift) Regression? Indeed, the conceptual idea inspires from the trend (& drift) regression in [15]. One may consider the genetic one is inherited from the trends one. As a simple case study, people may usually consider the genetic mutation trends of the most recent SARS-

CoV-2 virus (i.e. Each of variants are with the first trends that increase of high infection rate and the second trend of a higher degree of immune escape). In other words, one may plot the regression graph between the phylogenetic trend Vs the infection rate or even the number of immune escape reported or temperature / humidity etc. From an experimental perspective, one may consider such of the above genetic trend idea with a small district area of population to determine the trend of virus variant's genetic change and then find the corresponding regression equation models etc. Or in a vice versa way, one may plot the regression graph from the different variants' mutation of the same virus and thus get the trends in genetic change. Hence, the aforementioned description precisely depicts how the genetic trend (& drift) regression may actually be implemented in the cases like influenza and SARS-CoV-2 infected viruses etc.

Phylogenetic Regression (Theoretical Values)

Without lost of generality, it is very difficult for us to analyze those comparative data of phylogenetically related species through the standard statistical procedures. This is because those mean values for a set of species or even their residuals from a statistical model are with the properties that they are both non-independent and identically distributed [7]. Although it is not feasible to analyze phylogenetic comparative data, there are several other methods developed to account for the phylogenetic non-independent species [8]. At the same time, there are some other alternatives such as the generalized least square (GLS) theory where the associated basic mathematical theories have been shown in the [9]. Thus, we may apply the this GLS framework for the phylogenetic history. The reason is meta-analysis can be modeled under the GLS family of statistics [10]. The advantage of doing so is we gain the flexibility to use different statistical models to effect size data. By using matrix notation, one may pool K number of effect sizes into a weighted-mean (pooled) effect size with this GLS regression equation:

$$\beta = \mu_{+} = (\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{Y} \text{ ---------- (*)}$$

The 95% confidence intervals (CI) for μ_{+} are calculated as follows:

$$95\%CI[\mu_{+} +/- 1.96[X^{T}WX)^{-1}]^{1/2}$$

Then the information on shared ancestry between ith and jth effect size is found in the phylogenetic correlation matrix **P**.

The elements of **P** are thus defined by the shared internode
branch-length distance between species on a phylogenetic tree like the figure 1:

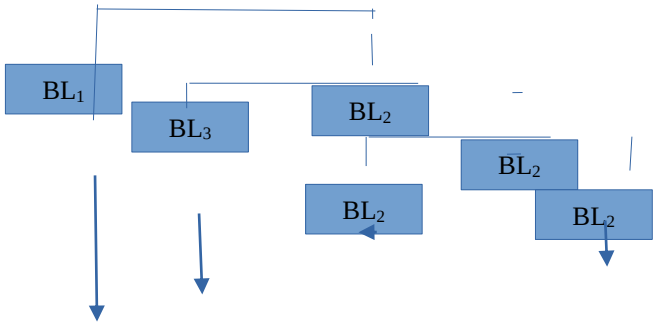

Figure 1: The phylogenetic correlation matrix P and how it is defined by the phylogenetic relationships of four species [1].

$$C = \frac{1}{BL_{max}} \begin{pmatrix} BL_1 & 0 & 0 & 0 \\ 0 & BL_3 + BL_2 & BL_2 & BL_2 \\ 0 & BL_2 & BL_5 + BL_4 + BL_2 & BL_4 + BL_2 \\ 0 & BL_2 & BL_4 + BL_2 & BL_6 + BL_4 + BL_2 \end{pmatrix}$$

If we further incorporate the parameter λ which can be estimated using maximum likehood, then we may get a more flexible residual error matrix [11] like figure 2:

$$C_\lambda = \frac{1}{BL_{max}} \begin{pmatrix} BL_1 & 0 & 0 & 0 \\ 0 & BL_3 + BL_2 & \lambda BL_2 & \lambda BL_2 \\ 0 & \lambda BL_2 & BL_5 + BL_4 + BL_2 & \lambda(BL_4 + BL_2) \\ 0 & \lambda BL_2 & \lambda(BL_4 + BL_2) & (BL_6 + BL_4 + BL_2) \end{pmatrix}$$

In fact, for the observation y (correspond to row of **X**), we may get the following phylogenetic linear regression equation that is related to the above 4 by 4 phylogenetic correlation matrix like the following:

$$y = \mathbf{X}\beta + \varepsilon$$

with β as shown in the above equation (*) and ε (or C_λ) = y – **X**β

3. Evolutionary Allometric Regression

Regressions of biological variables across species are rarely perfect [12]. Usually, there are residual deviations from the phylogenetic regression (or theoretical) value and the evolutionary regression value. Indeed, evolutionary regression (or experimental value) across species is one of the major statistical procedures used for the studying of the evolutionary relationship between biological variables and the adaptation to environmental variables [13]. Consider a biological trait Y that differs among a set of species. Natural selection will operate on Y, but the exact optimum value will depend on numerous ecological factors that all vary among species in different patterns. Assuming we knew the exact effects of all these factors in a species, one could make an exact prediction of Y for this species, as $Y = f(X_1, \ldots, X_m)$, for some function f, where the X_i are the exact states of the relevant factors in this species, but in reality, we may not have complete information about the states. If X_1 is a such focal variable, the evolutionary regression takes the form

$$Y = \beta_0 + \beta_1 X_1 + r(X_1,\ldots,X_m) \qquad \text{where } r(X_1,\ldots,X_m) = f(X_1,\ldots,X_m) - (\beta_0 + \beta_1 X_1)$$

are biological residuals of the model.

HKLam Theory & Regression Method [2]

In practice, the HKLam Theory & Regression Method was first self-developed by this applicant since 2004, from a statistical teacher training course and had been modified, updated from year 2015 – 2022 etc. Actually, what is the main context of my theory? The following may best describe what the theory may compose of and how it should be applied as a mathematical regression method:

My HKLam Theory : Any matrix (include mathematics graph in terms of matrix form representation) with immediately linear transformation (solution) vector x and the resulted vector can be approximated by the regression and various linked causality philosophy variables. The vice versa is also true indeed. — HKLam Theory plus its mathematical proof is practically a type of statistical model theory.
A brief proof for HKLam Theory:

$\underline{\mathbf{b}} = b''_1\underline{\mathbf{y}}_1 + b''_2\underline{\mathbf{y}}_2 + \ldots + b''_k\underline{\mathbf{y}}_k + c_1 + c_2 + \ldots + c_k$.

Or in general, for any resulted vector which is obtained from the space V through the linear mapping "T" to the space like W, we may have:

$\underline{\mathbf{w}} = \beta_0 + \beta_1\underline{\mathbf{x}}_1 + \ldots + \beta_n\underline{\mathbf{x}}_n + \underline{\boldsymbol{\varepsilon}}$ for some basis $\{\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_n\}$ in W and a vector $\underline{\boldsymbol{\varepsilon}}$ in W with some scalars β_i where i = 1,2, … ,n.

Indeed, one of the most famous case is the selection of dot product as the b'_i s in the orthogonal base with positive definite scalar product and more general case. With reference to Gram-Schmidt orthogonalization process [1], p.103, p.123, for any given arbitrary basis $\{\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \ldots, \underline{\mathbf{y}}_k\}$ of V, one may always select $\frac{\langle b'_{(i-1)}\overline{y}^T_{(i-1)}, b'_i\overline{y}_i\rangle}{\langle b'_i\overline{y}^T_i, b'_i\overline{y}_i\rangle}$ as our b''_i s such that

$$\underline{\mathbf{b}} = \underline{\mathbf{y}}_1 + \frac{\langle b'_1\overline{y}^T_1, b'_2\overline{y}_2\rangle}{\langle b'_2\overline{y}^T_2, b'_2\overline{y}_2\rangle}\underline{\mathbf{y}}_2 + \frac{\langle b'_2\overline{y}^T_2, b'_3\overline{y}_3\rangle}{\langle b'^T_3\overline{y}_3, b'_3\overline{y}_3\rangle}\underline{\mathbf{y}}_3 + \ldots + \frac{\langle b'_{(k-1)}\overline{y}^T_{(k-1)}, b'_k\overline{y}_k\rangle}{\langle b'^T_k\overline{y}_k, b'_k\overline{y}_k\rangle}\underline{\mathbf{y}}_k + \text{Error terms ------ (*)}$$

where $\langle b'_{(i-1)}\underline{\mathbf{y}}^T_{(i-1)}, b'_i\underline{\mathbf{y}}_i\rangle$ denotes the dot product between vectors $b'_{(i-1)}\underline{\mathbf{y}}_{(i-1)}$ and $b'_i\underline{\mathbf{y}}_i$ for I = 1,2,…,k and error terms will be defined as:

$$\mathbf{c}_i = b'_i\,\underline{\mathbf{y}}_i - \frac{\langle b'_{(i-1)}\overline{y}^T_{(i-1)}, b'_i\overline{y}_i\rangle}{\langle b'_i\overline{y}^T_i, b'_i\overline{y}_i\rangle}\underline{\mathbf{y}}_i$$

$$\underline{\mathbf{b}} = \underline{\mathbf{y}}_1 + \underline{\mathbf{y}}_2 + \underline{\mathbf{y}}_3 + \ldots + \underline{\mathbf{y}}_k + \text{Error terms ------ (*)}$$

where $\langle b'_{(i-1)}\underline{\mathbf{y}}^T_{(i-1)}, b'_i\underline{\mathbf{y}}_i\rangle$ denotes the dot product between vectors $b'_{(i-1)}\underline{\mathbf{y}}_{(i-1)}$ and $b'_i\underline{\mathbf{y}}_i$ for I = 1,2,…,k and error terms will be defined as:

$$\mathbf{c}_i = b'_i\,\underline{\mathbf{y}}_i - \underline{\mathbf{y}}_i$$

The above result is defined as my dot product (of the orthogonal projection) regression. Certainly, one may develop some other kind of regression by using the ideas of geometric mean and ordinary least square like the prescribed dot product one etc. This author remarks that one may go a further step by using the recursive Gram-Schmidt orthogonalization process to obtain the orthogonal basis and perform another type of projection etc.

REFERENCES

[1] Krammer F, Smith GJ, Fouchier RA, Peiris M, Kedzierska K, Doherty PC, et al. "Influenza". Nat Rev Dis Prim 2018;4:3.

[2] Centers for Disease Control and Prevention, U.S.A., How Influenza (Flu) Vaccine Are Made, **https://www.cdc.gov/flu/prevent/how-fluvaccine-made.htm**

[3] Rosenberg, M. S., Lajeunesse, M. J., & Jennions, M. D. (2013). Phylogenetic Nonindependence and Meta-analysis. In *Handbook of Meta-analysis in Ecology and Evolution* (p. 284–). Princeton University Press.
https://doi.org/10.23943/princeton/9780691137285.003.0017

[4] Carson, Lam Kai Shun (2022) A Mathematical Proof to the HKLam Theory by Linear/Tensor Algebra and Analysis, *International Journal of Mathematics and Statistics Studies*, Vol.10, No.5, pp.1-14

[5] Lam Kai Shun (2021) EVALUATION OF THE WEATHER-INFLUENZA PATTERN WITH A REGRESSION MODEL APPROXIMATION TO CAUSALITY**,** International Journal of Weather, Climate Change and Conservation Research, Vol.7, No.1, pp.26-44, 2021

[6] Frøysa, H. G., Fallahi, S., & Blaser, N. (2018). Evaluating model reduction under parameter uncertainty. *BMC Systems Biology*, *12*(1), 79–79.

https://doi.org/10.1186/s12918-018-0602-x
[7] Focosi, D. (2021). *SARS-CoV-2 spike protein convergent evolution : impact of virus variants on efficacy of COVID-19 therapeutics and vaccines*. Springer. https://doi.org/10.1007/978-3-030-87324-0
[8] Valentine, R. C. (Raymond C., & Valentine, D. L. (2015). *Human longevity : omega-3 fatty acids, bioenergetics, molecular biology, and evolution*. CRC Press, Taylor & Francis Group.
[9] Crunkhorn, S. (2021). Prime editing rescues mice from genetic disorders. *Nature Reviews. Drug Discovery*, *20*(10), 740–740. https://doi.org/10.1038/d41573-021-00149-w
[10] Finicial Times, 2022, Vaccines and Omicron mean Covid now less deadly than flu in England, https://www.ft.com/content/e26c93a0-90e7-4dec-a796-3e25e94bc59b
[11] Australian Bureau of Statistics, 2022, COVID-19 Mortality by wave, Analysis of COVID-19 Mortality by wave, with a focus on deaths occurring during the Omicron wave https://www.abs.gov.au/articles/covid-19-mortality-wave
[12] The Economics Times, Lab-made 'mild virus' that mimics coronavirus can help in developing antiviral drugs, 2020, https://economictimes.indiatimes.com/magazines/panache/lab-made-mild-virus-that-mimics-coronavirus-can-help-in-developing-antiviral-drugs/articleshow/77104646.cms?from=mdr
[13] National Cancer Institute, News & Event, 2018, Oncolytic Virus Therapy: Using Tumor-Targeting Viruses to Treat Cancer, https://www.cancer.gov/news-events/cancer-currents-blog/2018/oncolytic-viruses-to-treat-cancer
[14] Ghilani, C.D. and Wolf, P.R. (2006) Adjustment Computations: Spatial Data Analysis. 4th Edition, John Wiley & Sons, Hoboken.
[15] Castiglione, S., Serio, C., Mondanaro, A., Di Febbraro, M., Profico, A., Girardi, G., & Raia, P. (2019). Simultaneous detection of macroevolutionary patterns in phenotypic means and rate of change with and within phylogenetic trees including extinct species. *PloS One*, *14*(1), e0210101–e0210101. https://doi.org/10.1371/journal.pone.0210101
[16] Hazra, A., & Gogtay, N. (2017). Biostatistics Series Module 9: Survival Analysis. *Indian Journal of Dermatology*, *62*(3), 251–257. https://doi.org/10.4103/ijd.IJD_201_17
[17] Weisstein, Eric W. "Matrix Equation." From *MathWorld*--A Wolfram Web Resource. https://mathworld.wolfram.com/MatrixEquation.html
[18] Kitayimbwa, J. M., Mugisha, J. Y. T., & Saenz, R. A. (2013). The role of backward mutations on the within-host dynamics of HIV-1. *Journal of Mathematical Biology*, *67*(5), 1111–1139. https://doi.org/10.1007/s00285-012-0581-2
[19] Tao, J., Zhang, R., & Zhu, Y. (2020). *DNA computing based genetic algorithm : applications in industrial process modeling and control*. Springer.
[20] Mutingi, M., & Mbohwa, C. (2016). *Grouping Genetic Algorithms: Advances and Applications* (Vol. 666). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-44394-2
[21] Jung, Youn-Sang, and Park, Jae-Il. "Wnt Signaling in Cancer: Therapeutic Targeting of Wnt Signaling beyond β-catenin and the Destruction Complex." *Experimental & Molecular Medicine* 52.2 (2020): 183-91. Web.
[22] WEIN, Lawrence M, WU, Joseph T, and KIRN, David H. "Validation and Analysis of a Mathematical Model of a Replication-competent Oncolytic Virus for Cancer Treatment: Implications for Virus Design and Delivery."Cancer Research (Chicago, Ill.) 63.6(2003): 1317-324.Web.

[23] Shun CL, A Saddle point finding method for Lorenz Attractor through business machine learning algorithm, Journal of Physical Mathematics, Vol11:4,2020,https://www. hilarispublisher.com/open-access/a-saddlepoint-finding-method-for-lorenz-attractor-throughbusiness-machine-learning-algorithm.pdf
[24] Ewen Callaway, The coronavirus is mutating does it matter?, Sept 2020, Nature publishing. https://www.nature.com/articles/d41586-020-02544-6
[25] Maurer-Stroh, S., Li, Y., Bastien, N., Gunalan, V., Lee, R. T. C., Eisenhaber, F., & Booth, T. F. (2014). Potential human adaptation mutation of influenza A(H5N1) virus, Canada. *Emerging Infectious Diseases*, *20*(9), 1580–1582. https://doi.org/10.3201/eid2009.140240
[26] Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation*, *124*(10), 4154–4161. https://doi.org/10.1172/JCI72992
[27] Rossetti, M., Stella, L., Morlà Folch, J., Bobone, S., Boloix, A., Baranda, L., Moscone, D., Roldán, M., Veciana, J., Segura, M. F., Köber, M., Ventosa, N., & Porchetta, A. (2021). Engineering DNA Grafted Quatsomes as Stable Nucleic Acid Responsive Fluorescent Nanovesicles. *Advanced Functional Materials*, *31*(46), 2103511–n/a. https://doi.org/10.1002/adfm.202103511
[28] Fenner, F. J. (1983). Florey lecture, 1983. Biological control, as exemplified by smallpox eradication and myxomatosis. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *218*(1212), 259–285. https://doi.org/10.1098/rspb.1983.0039
[29] Carson Lam Kai Shun, 2021, A Causal (Regression) Analysis for Cancer Formation under a Corresponding Virus Attack, Chapter 24, Asian Journal of Mathematical Sciences
[30] Zhao, Hanjun et al. "Dual-Functional Peptide with Defective Interfering Genes Effectively Protects Mice Against Avian and Seasonal Influenza." *Nature communications* 9.1 (2018): 2358–14. Web. (or Team of University of Hong Kong, Department of Microbiology, 2018 https://www.hku.hk/press/press-releases/detail/18016.html)
[31] Carson Lam Kai Shun, A Saddle Point Finding Method for Lorenz Attractor through Business Machine Learning Algorithm, Hi-laris Publication, Vol. 11, Issue 4, June, 2021.
[32] https://pansci.asia/archives/194509
[33] https://www.twreporter.org/a/sars-cov-2-variants
[34] https://www.real-statistics.com/descriptive-statistics/m-estimators/
[35] John Fox & Sanford Weisberg (2013), Robust Regression, http://users.stat.umn.edu/~sandy/courses/8053/handouts/robust.pdf
[36] Dost Muhammad Khan, Anum Yaqoob, Seema Zubair, Muhammad Azam Khan, Zubair Ahmad, Osama Abdulaziz Alamri, "Applications of Robust Regression Techniques: An Econometric Approach", *Mathematical Problems in Engineering*, vol. 2021, Article ID 6525079, 9 pages, 2021. https://doi.org/10.1155/2021/6525079 https://downloads.hindawi.com/journals/mpe/2021/6525079.pdf
[37] Wander Almodovar Golfetto, Sandro da Silva Fernandes, A Review of Gradient Algoritms for Numerical Computation of Optimal Trajectories, J. Aerosp.Technol. Manag., Sao Jose dos Campos, Vol.4, No.2, pp. 131-143, Apr-Jun., 2012.
[38] J.D. Mireles James, Celestial Mechanics Problem Set 2: The State Transition Matrix and Method of Differential Corrections, December, 11, 2006.