

FastGLVM: Big Data Ordination Towards Intensive Care Event Count Cases

Rezzy Eko Caraka (✉ rezzyekocaraka@gmail.com)

Seoul National University Hospital

Rung Ching Chen

Chaoyang University of Technology

Youngjo Lee

Seoul National University

SU-WEN HUANG

Taichung Veterans General Hospital

SHYUE-YOW CHIOU

Taichung Veterans General Hospital

Bens Pardamean

Binus University

Research Article

Keywords: GLLVM, Poisson, Negative Binomial, Laplace, Variational Approximation

Posted Date: March 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-277089/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

FastGLLVM: Big Data Ordination Towards Intensive Care Event Count Cases

REZZY EKO CARAKA^{1,2,4,6*}, RUNG-CHING CHEN¹, YOUNGJO LEE^{2,3}, SU-WEN HUANG⁵, SHYUE-YOW CHIOU⁵, BENS PARDAMEAN^{6,7}

Abstract

Background: In the heart data mining and machine learning, dimension reduction is needed to remove the multicollinearity. Meanwhile, it has been proven to improve the interpretation of the parameter model. In addition, dimension reduction is also can increase the time of computing in high dimensional data.

Methods: In this paper, we perform high dimensional ordination towards event counts in intensive care hospital, following emergency department (ED 1), First Intensive Care Unit (ICU1), Second Intensive Care Unit (ICU2), Respiratory Care Intensive Care Unit (RICU), Surgical Intensive Care Unit (SICU), Subacute Respiratory Care Unit (RCC), Trauma and Neurosurgery Intensive Care Unit (TNCU), Neonatal Intensive Care Unit (NICU) which use the Generalized Linear Latent Variable Models (GLLVM's).

Results: During the analysis, we measure the performance and calculate the time computing of GLLVM with employing variational approximation and Laplace approximation, and compare the different distributions including Negative Binomial, Poisson, Gaussian, ZIP, and Tweedi, respectively.

Conclusions: In a nutshell, GLLVM's leads as best performance reaching the accuracy 98% comparing other methods. In line with this, we get the best model negative binomial and Variational approximation which provides the best accuracy by accuracy value of AIC, AICc, and BIC. In a nutshell, our best model is GLLVM-VA Negative Binomial with AIC 7144.07 and GLLVM-LA Negative Binomial with AIC 6955.922.

Keywords: *GLLVM; Poisson; Negative Binomial; Laplace; Variational Approximation*

Background

Big data is a collection of huge data and more complex, especially from new data sources [1]. The data set is large enough so that software for traditional data processors not good enough to manage it, but this massive amount of data can be used to overcome a variety of business problems that previously could not be solved and the need for decision making [2]. The most straightforward and obvious explanation is that Big Data is the collection and use of information from various sources to provides important information. Big Data is also a concept of the ability to collect, analyze, and

understand a large amount of data on a very broad range of activities.

Big Data is beneficial for the hospital service system itself. One of the classic problems which have existed that there are too many staff or too few staff so that the hospital will risk incurring additional costs than they should. Not only that, hospitals that lack staff also will jeopardize the quality and performance of services provided. If a few teams handle many patients, this will have a direct impact the services provided to these patients will be of poor quality and unsatisfactory. The primary key to implementing hospital orientation is the patient. Then, the patient satisfaction is the success of a hospital in managing health care services.

Customer satisfaction is an abstract thing, and the results are very varied.

However, perceptions depend on each person and tend to be different. The availability of medical personnel with high knowledge and skills is essential for patients in choosing a health service as a place that can help them recover from the disease. The core business of the hospital is to provide health services. A good hospital cannot only offer professional medical personnel but also provides the best facilities and an excellent patient-care system [3]. At the same time, monitoring patient clinical status is essential, particularly in intensive care units (ICUs) [4]. During that time, the teleporter plays the role of "facilitator" and "supporter." It is one of the medical team's valuable members and the connection window between the unit and the department. The transmission staff is responsible for assisting the patient's medical treatment or acting as a helper for the family to take care of the patient. It must have sufficient resilience to respond to the emergency that may occur, and the transmission process must strictly follow the actual transfer and relevant safety rules. The mastery, accuracy, and completeness of the delivery service time relate to the smooth connection of medical services, so it must have a certain degree of job sensitivity and excellent communication skills. Furthermore, with the increasing needs and desires of patients in obtaining the best services, it is necessary to do the right planning, especially in the intensive care center room.

The most crucial point is to place appropriate medical personnel in the intensive care center. If the placement of medical staff is proper, hospital services will be better, and patients will be treated faster. Then another thing is to provide training to improving the work of medical personnel. If the human resources are of high quality and in line with company expectations, then the company has high

competitiveness. Therefore, the products and services produced are also of high quality.

Intensive care units (ICUs) of university hospitals and advanced medical centers are indispensable for providing critical and intensive care for patients who have undergone major surgery or have received emergency care. Hospitals can obtain higher revenue from national insurance by a short admission in the ICU than by admission to other hospital departments. Intensive care units are the foremost part and are very important in the hospital. Intensive care units act as the main gate of entry for emergency patients and patients with mild conditions. Good or bad service in the intensive care unit will give an overall impression of hospital services. Analysis of the number of events in the ICU is also essential to do to analyze. The cost estimation and a profit and loss analysis are necessary for health care practice [5].

A significant part of this work is to decide whether ICU procedures of care can improve results for those identified as frail. Instances of procedures that may have a differential way in the individuals who are slight incorporate wholesome help; sedation rehearses the force of assembly/restoration. In other words, an analysis of the number of medical personnel needs is essential due to in the ICU room, a first aid kit is needed quickly and temporarily given to a person suffering from an injury or sudden illness. A fundamental objective of first aid is to provide care and health services that will benefit these people in preparation for further treatment.

The emergency is a condition that is related to a disease or other life-threatening conditions. In contrast, a crisis is a sudden and unforeseen condition, an immediate or urgent need [6]. Due to the emergency room's operational nature must be fast, precise, and not limited by the time [7]. At the same time, we need to be concerned about the ideal performance of the emergency room is highly dependent on human resources and proper work procedures. Moreover, supporting examination

facilities to support the diagnostic process, adequate drug support, and medical consumables clear patient in and outflow, ready operating room, and ambulance transport support that focuses on patient safety.

Big Data Analysis offers an excellent opportunity for improving the strategic unit management and to handling of concrete clinical cases [8] [9]. Moreover, different biomedical and medicinal services devices produce a primary measure of information. We must think about and evaluate what can be accomplished by utilizing this information [10]. The problem that cannot be separated from big data is the selection of methods for large-dimensional data, many attributes, and causing some algorithms not to get good performance. Therefore, the solution offered is to do feature selection or dimension reduction by using PCA [11], K-means [2], CCA [12], Factor analysis [13] [14], XGBoost [15] [16], Bayesian [17][18]. Nowadays, there is a challenge to measure statistical parameters in vast data sets, and most traditional statistical methods cannot handle high dimensional data and including large numbers of parameters [18] [19] [20]. Assuming, casually, we let p mean the unknown element and let n as a known element. Then it will reach until limited to small p and large n . This situation additionally normally mirrored the contemporary impediments of computing. In short, this research will get an ordinance of intensive care hospital rooms so that it can be used to calculate and predict daily and hourly how many patients are expected to be in the room. The remainder of the paper is organized as follows. Section 2 explains the methods. Section 3 presents the application of a

high dimension. Finally, conclusions and future research directions are indicated in Section 4.

METHODS

GENERALIZED LINEAR LATENT VARIABLE MODELS

The classical linear model was initially more widely used in the field of statistics or better known as straight-line equations [21]. The traditional linear models were commonly used in the field of statistics, especially for modeling data count [22]. The simplest classical linear model is defined in Eq (1).

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}\mathbf{x} + \boldsymbol{\varepsilon} \quad (1)$$

Where y is the dependent variable whose value depends on the independent variable x , β is an unknown parameter in the model. At the same time, ε is a random variable that differs from the actual value of y with its estimated value [23]. The random variable ε is assumed to follow the normal distribution $(\mathbf{0}, \boldsymbol{\sigma}^2)$. The development of linear models was very rapid after the discovery of the normal distribution, until the beginning of the 19th century. To begin with, [24] published a research in the field of agriculture using design experimental. Simple GLM is the development of a classic linear models (LM's) with many predictors or called Multiple Linear Regression [25]. The least-square method by Gauss remains the basis for estimating model parameters. However, the classical assumptions on LM's also carry over to GLM is $\boldsymbol{\varepsilon}$ follows the normal distribution $(\mathbf{0}, \boldsymbol{\sigma}^2)$. The predictor does not have to be continuous. Category predictors also underlie Fisher's research in experimental design. Under the auspices of the normal distribution assumption, linear models

can be written in general, or general terms define GLM's as in Equation (21).

$$Y_n G_n = X_n B H_n + Z_n \Theta Q_n + E_n \quad (2)$$

Where Y_n is a random matrix of size $n \times m$. G_n is the matrix X_n, H_n, Z_n and Q_n are $n \times p, q \times k, n \times u$ and $v \times k$ matrices, respectively. B is the matrix $p \times q$ of the unknown parameter. Θ is a random matrix $u \times v$ and E_n is the matrix $n \times k$ of the random error that is normally distributed $(0, \Sigma)$. The model in Eq 2 represents a GLM that various linear models. Such as linear regression (including simple or multiple) and Multivariate regression [26]. The parameter estimation technique is also developing. Regarding to the least square [27] the parameter estimation can be obtained by using the H-likelihood [28], Double H-likelihood [29], and Bayes estimation approach [30]. In Eq (1), we can apply to negative binomials $Y \sim NBin(k, \pi)$ [31] so that it can be written in Eq (3) as follows.

$$f(y) = \binom{y-1}{k-1} (\pi)^k (1-\pi)^{y-k} \quad (3)$$

Then, Eq(3) will be changed in the Exponential Family form. $f(y; \theta, \phi) = \exp \{k \log \pi +$

$$(y-k) \log(1-\pi) + \log \left(\frac{y-1}{k-1} \right)\}$$

Moreover, can be written in Eq(23):

$$f(y; \theta, \phi) = \exp \left\{ y \log(1-\pi) + k \log \left(\frac{\pi}{1-\pi} \right) + \log \left(\frac{y-1}{k-1} \right) \right\} \quad (4)$$

From Eq (4) we can determine the value of Expectations of Eq (5), Variance Eq (6) and Deviance Eq (7), respectively:

$$\begin{aligned} E(Y) &= b'(\theta) \\ b(\theta) &= -k \{ \log(1-e^\theta) - \log(e^\theta) \} \\ E(Y) &= -k \{ \log(1-e^\theta) - 1 \} \\ E(Y) &= -k \left\{ \frac{e^\theta}{(1-e^\theta)} - \frac{(1-e^\theta)}{(1-e^\theta)} \right\} \\ E(Y) &= -k \left\{ -\frac{1}{(1-e^\theta)} \right\} = \frac{k}{\pi} \end{aligned} \quad (5)$$

$$\begin{aligned} E(Y) \text{Var}(Y) &= b''(\theta) a(\phi) \\ \text{Var}(Y) &= k \left\{ -(1-e^\theta)^{-2} (-e^\theta) \right\} \\ \text{Var}(Y) &= k \left\{ \frac{e^\theta}{(1-e^\theta)^2} \right\} \\ \text{Var}(Y) &= \frac{k(1-\pi)}{\pi^2} \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{D(y; \hat{\mu})}{\phi} &= 2[l(\hat{y}, \phi; y) - l(\hat{\mu}, \phi; y)] \\ \frac{D(y; \hat{\mu})}{\phi} &= 2 \left\{ \left[k \log y \right. \right. \\ &\quad + (y-k) \log(1-y) \\ &\quad + \left. \log \left(\frac{y-1}{k-1} \right) \right] \\ &\quad - \left[k \log \hat{\mu} \right. \\ &\quad + (y-k) \log(1-\hat{\mu}) \\ &\quad + \left. \log \left(\frac{y-1}{k-1} \right) \right] \left. \right\} \\ &= 2 \{ k \log y - k \log \hat{\mu} + (y-k) \log(1-y) \\ &\quad - (y-k) \log(1-\hat{\mu}) \} \\ &= 2 \left\{ k \log \left(\frac{y}{\hat{\mu}} \right) + (y-k) \log \left(\frac{1-y}{1-\hat{\mu}} \right) \right\} \quad (7) \end{aligned}$$

GLLVM is the extended version of GLM's with a latent variable [32]. The marginal density of

the manifest variables can be rewritten as Eq (7) [33].

$$L(\omega, \phi) = \int \exp \left(p \sum_{i=1}^n \frac{1}{p} \left[\frac{y_{ij} \theta_{ij} - b(\theta_{ij})}{\phi_j} + c(y_{ij}, \phi_j) \right] h(\mathbf{u}_i) \right) d\mathbf{u}_i \quad (8)$$

Since we assume the latent variable \mathbf{u}_i to follow a standard normal distribution such that we have Eq(9) and Eq(10), respectively.

$$h(\mathbf{u}_i) = \frac{1}{(2\pi)^{\frac{q}{2}}} \exp\left(\frac{-\mathbf{u}_i' \mathbf{u}_i}{2}\right) \quad (9)$$

$$\log h(\mathbf{u}_i) = \frac{-\mathbf{u}_i' \mathbf{u}_i}{2} - \frac{q}{2} \log(2\pi) \quad (10)$$

Model Selection

The model selection criteria are statistical tools that identify an “optimal” statistical model from among a set of models. Meanwhile, the set is usually called a set of candidate models. A model is considered as an optimal model if it satisfies three essential features: generalizability, parsimony, and goodness-of-fit. The principle of generalizability is the capability of the fitted model to describe or predict new data. The purpose of statistical modeling should be to predict new data as opposed to precisely characterizing the actual model that generated the data. On the other hand, the candidate models are significant in analyzing the selection criteria.

The criteria can be used following Akaike Information Criterion (AIC), Akaike Information Criterion Correction (AICc), and Bayesian Information Criterion (BIC) [34]. Lastly, the selection of models should take the generalizability, parsimony, and goodness-of-fit into account. The motivation behind measurable demonstrating ought to be that of anticipating new information rather than unequivocally describing the genuine model that created the information. Where k is the

dimension of the parameter θ , and n is the sample size.

$$AIC = -2 \ln f(\mathbf{y}, \hat{\theta}) + 2k. \quad (11)$$

$$BIC = -2 \ln f(\mathbf{y}, \hat{\theta}) + k \ln(n) \quad (12)$$

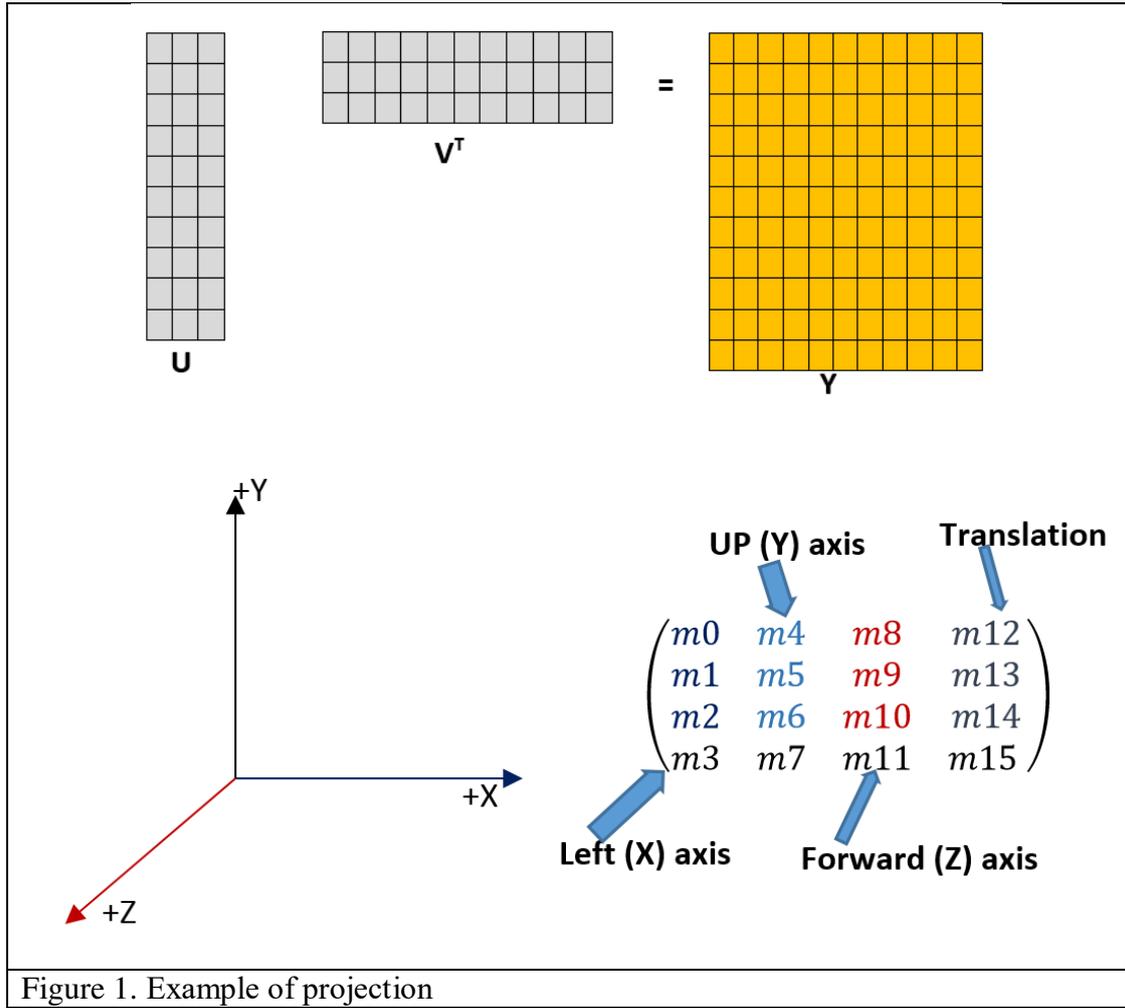
$$AICc = -2k \left[\frac{n}{n-k-1} \right]. \quad (13)$$

However, the researcher leans toward BIC to AIC, since BIC may possible prompts choosing a more closefisted fitted model than AIC. It demonstrates that BIC is steady, yet it is not asymptotically productive. In addition, AICc is useful in the small dataset

HIGH DIMENSION DATA

In this paper, we use the event count data that occur in the intensive care center to meet the needs of medical operations. The operations include pushing hospitalized patients for hemodialysis treatment, receiving emergency treatment drugs, transferring specimens, and collecting blood, and related services such as respirators, oxygen cylinders, and other equipment or items required for the treatment. The specifications are as follows: Emergency Department (ED 1), First Intensive Care Unit (ICU1), Second Intensive Care Unit (ICU2), Respiratory Care Intensive Care Unit (RICU), Surgical Intensive Care Unit (SICU), Subacute Respiratory Care Unit (RCC), Trauma and Neurosurgery Intensive Care Unit (TNCU), Neonatal Intensive Care Unit (NICU).

The data starts from June (33561 cases), July (31557 cases), August (35689 cases), September (34293 cases), and October (35310 cases). In total, the matrix dimension is (170410 x 7). Because in this paper used eight types of ICU rooms. To get the ICU ordination per room will be transposed to (7 x 170410). Then the dimension matrix is reduced again to retrieve the total daily occurrence data. Also, we get a matrix (153 x 7) and the projection can be represents in Figure 1



RESULTS AND DISCUSSION

As explained in the previous section, we are using the daily data of the number of cases of incentive care rooms. Then, the matrix dimension is quite large. So computation [35] time will be calculated on selected distributions such as negative binomials, Poisson, Gaussian, ZIP, and Tweedie. We successfully compared with two types of optimization including of variational approximation, and Laplace approximation. Also, we make a comparison with the number of latent variables. Table 1 explains that the best model is the smallest AIC, AICc, and BIC values for the negative distribution of GLLVM-VA and GLLVM-LA binomials. Figure 2a and 2b are explains that

information. In general, VA (1) gives promises to complete computing time compared to LA (2).

Based on this simulation, we get the insight the difference in the number of latent variables used does not affect the accuracy results. Besides, the recognizable proof of the estimation model is that there are sufficient for each latent variable. The decision of connection capacity ought to be founded on hypothetical contemplations and model fit. The scope of qualities it creates for the mean $\mu_i = g^{-1} v_i$ can be contemplate when picking link function. For example, the logit and probit interface capacities are the regular decision when the reaction variable is two-fold since they limit the likelihood μ_i within the interval $[0, 1]$ Another factor to consider identifies with the

understanding of the relapse parameters [35]. However, $\mathbf{v}_i = \mathbf{x}_i' \boldsymbol{\beta}$ utilize an identity link function relates to additive impacts of the covariates on the mean and a log link compares to multiplicative impacts. Another significant thing in GLLVM's the decision of the dissemination. The decision of dissemination depends on the kind of reaction variable, the

procedure that may produce the reaction and the state of experimental dispersion. For instance, for parallel reactions, the undeniable decision is the Bernoulli dissemination while for counts. In line with this, the Poisson dispersion is regularly picked for fitting the model.

TABLE 1. GLLVM Performance

	LV	Family	Selection Criteria			DF	log-likelihood:	Time Computing
			AIC	AICc	BIC			
GLLVM-VA	1	Negative Binomial	7144.07	7151.123	7207.709	21	-3551.035	00.13,23
	2	Negative Binomial	7309.096	7321.192	7390.918	27	-3627.548	00.05,85
	3	Negative Binomial	7472.096	7489.696	7569.07	32	-3704.048	00.06,11
	1	Poisson	7693.37	7699.733	7753.978	20	-3826.685	00.01,22
	2	Poisson	7693.37	7699.733	7753.978	20	-3826.685	00.03,60
	3	Poisson	7685.202	7695.439	7760.963	25	-3817.601	00.12,20
	1	Gaussian	7260.324	7267.378	7323.964	21	-3609.162	00.01,32
	2	Gaussian	7409.686	7421.782	7491.508	27	-3677.843	00.03,42
	3	Gaussian	7570.224	7587.824	7667.198	32	-3753.112	00.09,10
GLLVM-LA	1	Negative Binomial	6955.922	6962.976	7019.562	21	-3456.961	00.41,30
	2	Negative Binomial	-	-	-	27	3381132785	00.47,11
	1	Poisson	7736.851	7739.895	7779.277	14	-3854.426	00.05,32
	2	Poisson	7387.098	7393.462	7447.707	20	-3673.549	00.44,47
	3	Poisson	7227.31	7237.546	7303.07	25	-3588.655	01.17,90
	1	Gaussian	7107.34	7114.393	7170.979	21	-3532.67	00.01,72
	2	Gaussian	7103.672	7115.768	7185.494	27	-3524.836	00.02,22
	3	Gaussian	7110.665	7128.265	7207.639	32	-3523.333	00.01,88
	1	ZIP	7637.484	7644.538	7701.123	21	-3797.742	00.31,38
	2	ZIP	7326.899	7338.995	7408.721	27	-3636.449	00.59,18
1	Tweedie	7010.549	7022.645	7092.371	27	-3478.275	45.11,33	

Long story short, we are using different distributions such as Negative Binomial (1), Poisson (2), Gaussian (3), ZIP (4), and Tweedie (5). As shown in Figure 5, running a Tweedie distribution will take a very long time. The power parameters are vital to discuss. In tweedy probability density cannot be closed from, so it is slow to finish computing. To solve this

problem, quasi and pseudo-likelihood can be used for Tweedie.

For the Tweedie distribution can only be analyzed with use Laplace approximation GLLVM. Indeed Variational approximation is a bayesian inference to solve complex statistics. For a more precise explanation about Variational approximation, refers to Ormerod

[36]. On the other hand, bayesian along these lines[37], certainly relies upon the researcher's capacity to compute integrals concerning the posterior distribution. Be that as it may, this is a troublesome issue and separated from the conjugate models, the explicit type of the

thickness a posteriori is regularly accessible just to a factor

$$\pi(x|y_1, \dots, y_N) \propto p_0(x)p(y_1, \dots, y_N|x)$$

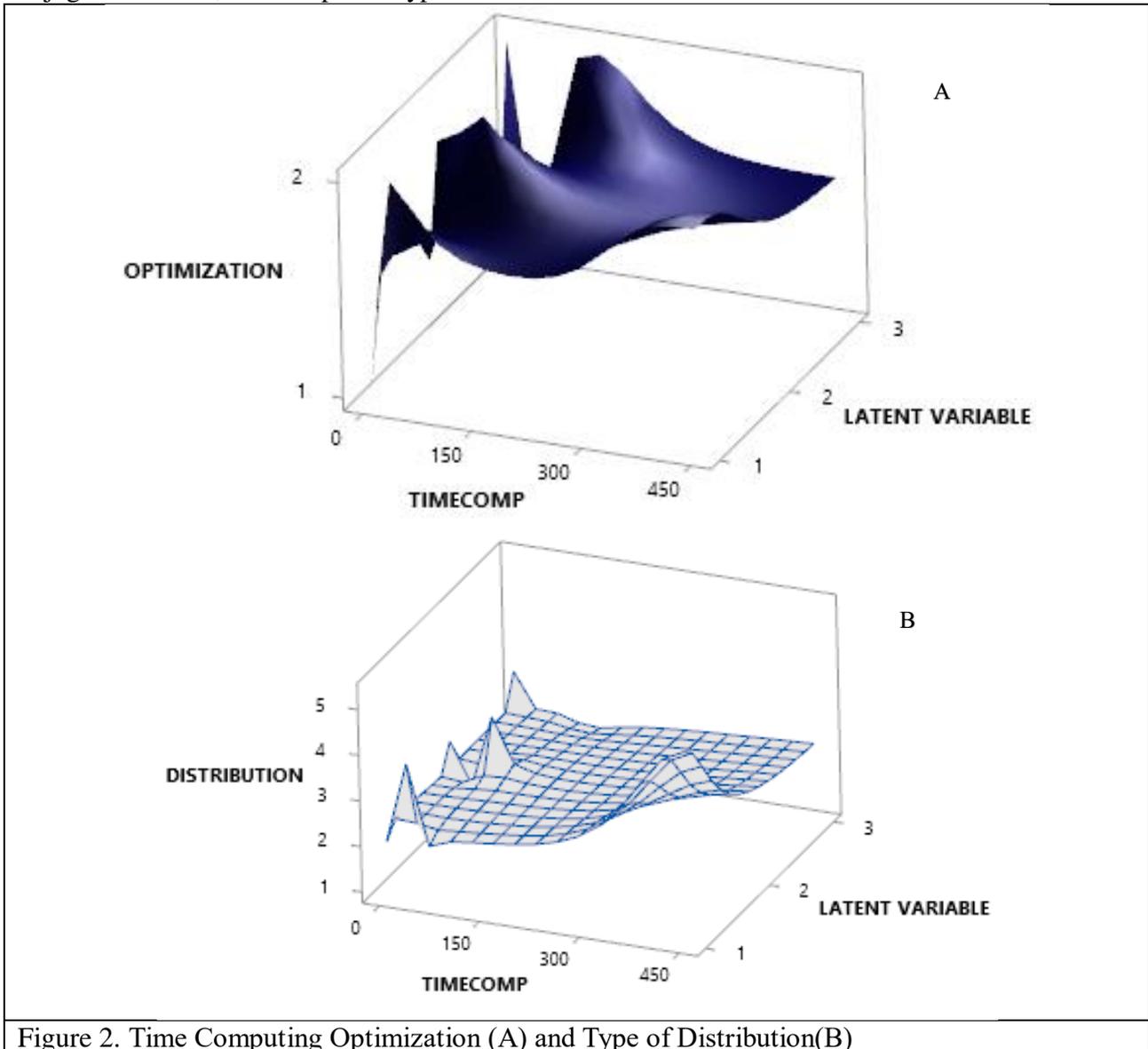


Figure 2. Time Computing Optimization (A) and Type of Distribution(B)

During the experimental, we compare GLVVM's to PCA, Factor Analysis Extraction Maximum Likelihood, K-Means 2 cluster, Canonical Correlation Analysis, and Global Multidimensional Scaling. However, by using K-means only use 2 groups in accordance with the number of groups that have been previously

determined. To determine the group members can be done by calculating the minimum distance of the object. The value obtained in the membership of data at the distance matrix is 0 or 1, the value 1 for data allocated to group A while the value 0 for data allocated to group B. In this simulation obtained distance centroid

(Cluster 1 to Cluster 2 = 24.6436). Table 2 provides of information (%) of each methods.

TABLE 2. Comparson Performance

Methods	Information (%)
PCA	78.5%
Factor Analysis Extraction Maximum Likelihood	75.2%
K-Means 2 cluster	51.070%
*Ours (GLLVM)	98%
Canonical Correlation Analysis	70.2%
Global Multidimensional Scaling	68.5%

Meanwhile, two significant methodologies have shown up in measurements, such as approaches dependent on the characterization of the posterior and approximation. For a differential condition whose arrangement is not easy work, at any rate, the Laplace approximation can tell the arrangement is the inverse Laplace likewise; the underlying conditions are folded into the strategy for the arrangement from the beginning. Nevertheless, with bayes, we do not have the entirety of the underlying derivatives, so we need to keep some of them around as free parameters. The Laplace for the most part, is not in nonlinear issues, because we do not receive a decent arithmetical condition in return [38] [39]. One exception is that the Laplace change of a

convolution is only an item, which is helpful [40]. The data matrix is usually a proximity matrix (a matrix that has a distance between objects) and including ordinal data types. This result is expected to be robust because the configuration results are obtained from its iteration. However, due to the reduction in dimensions, the process will also lose some information. The ordination is also useful for reducing the dimensions of data from several variables so that there are new variables that are no longer correlated and have as much information as possible from the original data. After getting the best model that is negative binomial on two different optimizations VA and LA, it is necessary to find linear predictors with residuals in both of these models.

Figure 3a and 3b are represent of scale location. The line starts off horizontal at the beginning of our predictor range, slopes up to around 2, and then slopes down around 3. In contrast with Laplace approximation, the line is flattened around 2.5 because the residuals for those predictor values are not more spread out. The development of the GLLVM ordination will continue by using Variational approximation with the assumption that it provides speed in computing with accuracy differences that are not significant as Laplace approximation. Figure 4 a and b are explain how linear these predictors are at residuals. Then the normal QQ-plot describes theoretical quantiles following the normal distribution and the points forming a roughly straight line.

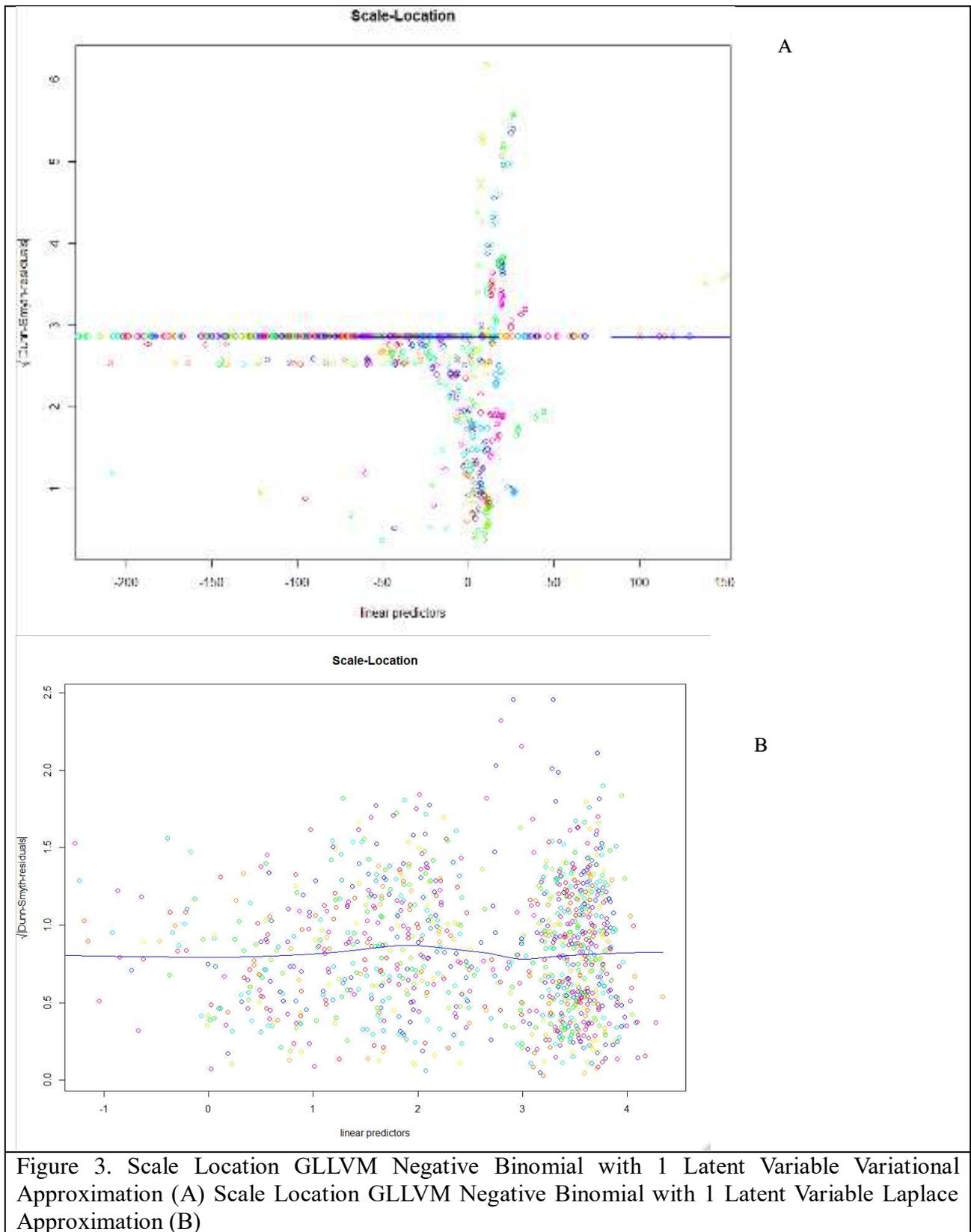
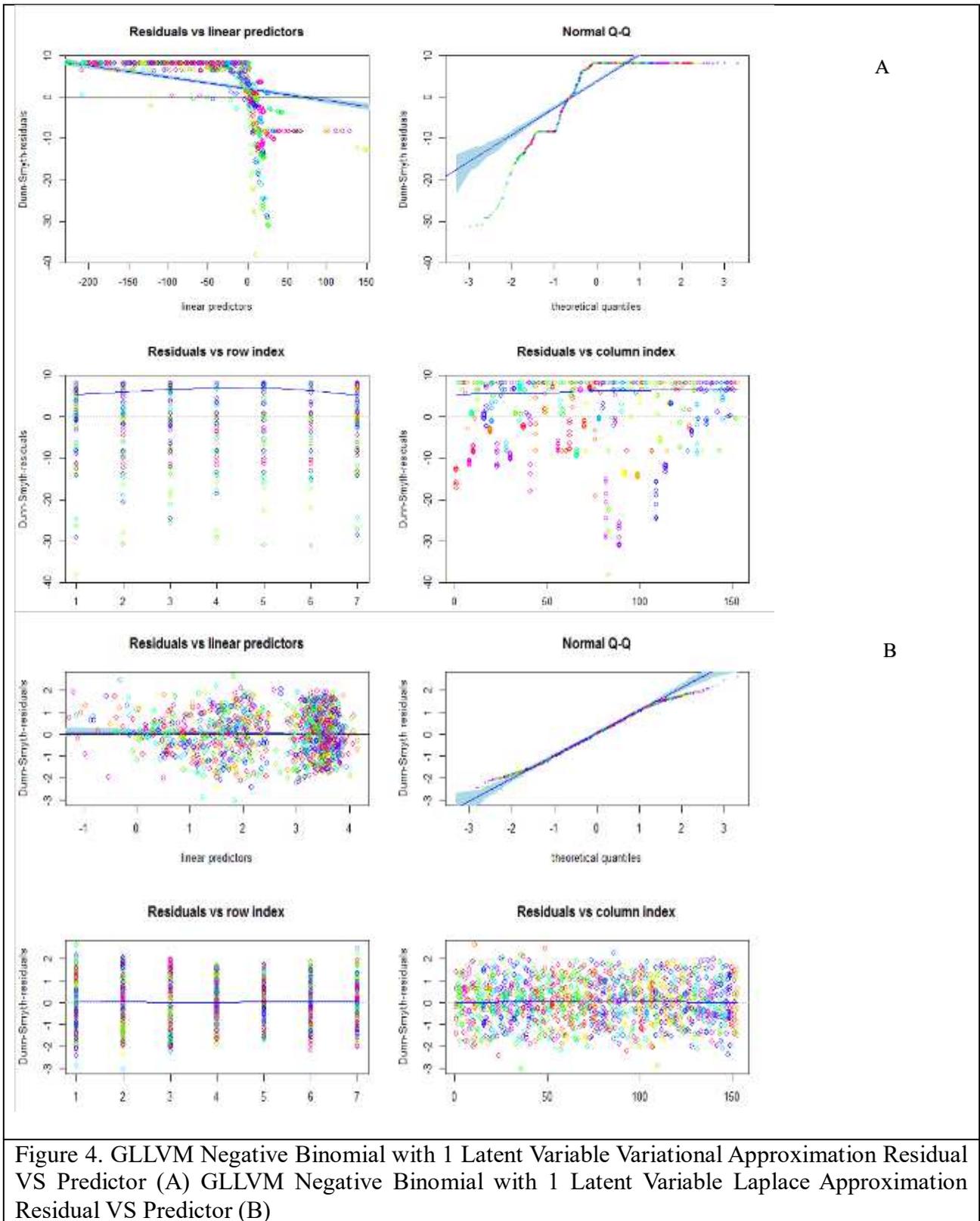


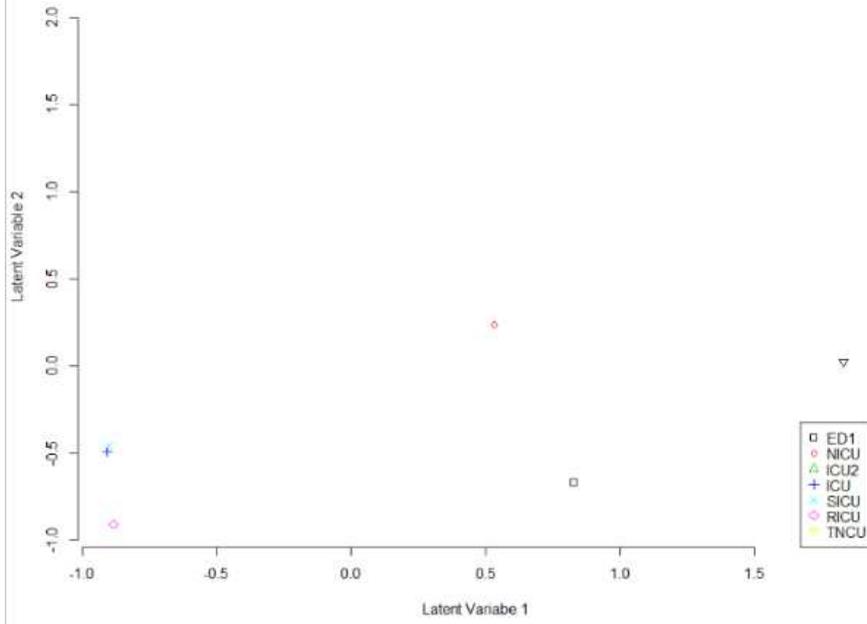
Figure 3. Scale Location GLLVM Negative Binomial with 1 Latent Variable Variational Approximation (A) Scale Location GLLVM Negative Binomial with 1 Latent Variable Laplace Approximation (B)



However, Figure 5 a explains the ordination in 7 different room types. It seems so clear that each room has a different ordination. In addition, Figure 5 b represents the number of manpower based on best model. The type of ICU room requires more manpower than other rooms. Nevertheless, visually ICU and RICU rooms have the same characteristics compared to the others. Overall different ordinances are ICU2 rooms and different ordinances in the RICU room. At the same time, the ordinances look similar in ED1 and TNCU rooms, respectively. Otherwise, Figure 6 explains the distribution of frequency of events data in the Intensive Care Units' room if there are several similarities between one day and another. The

highest number of cases occurred on Monday and Saturdays and Sundays decreased quite far. If the hospital wants to focus on maximum service, it might be better to consider the appropriate number of medical staff on a specific day. Each room has its tasks such as the First Intensive Care Unit (ICU1), Second Intensive Care Unit (ICU2). The content of the hospital's transfer staff is to transfer patients to outpatients, wards, inspections and other units. The transfer methods include leadership, bed and wheelchair push, and the receipt and transfer of medicines, blood, specimens, articles, instruments and stationery to other units.

A



B

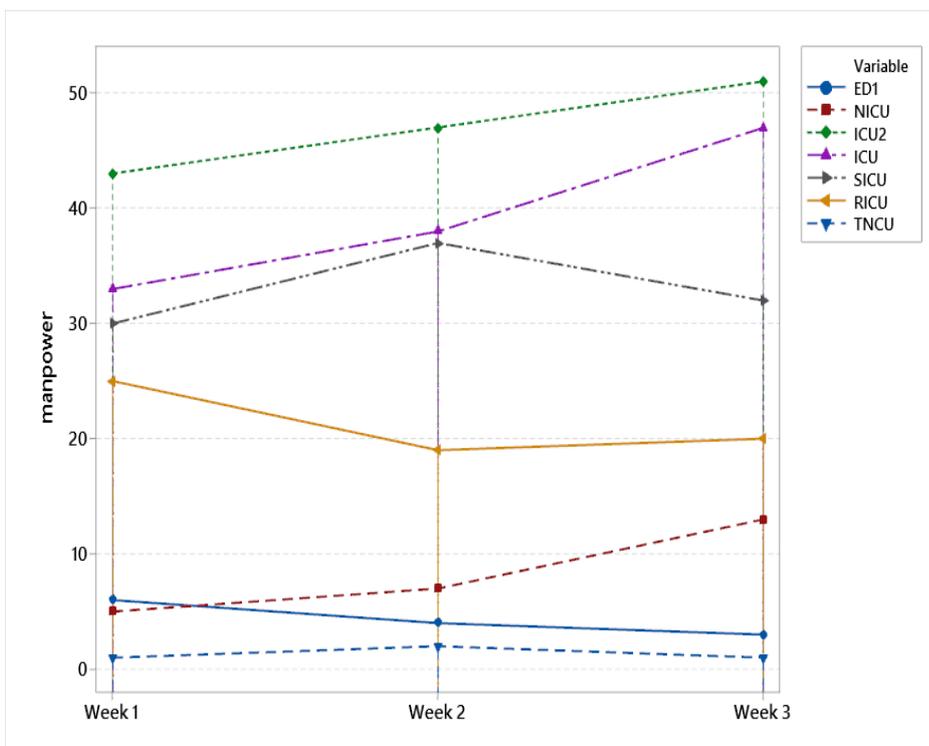


Figure 5. GLLVM Ordination (A) and Prediction manpower (B)

The outsourcing business of the hospital's labor service is also to maintain the business activity . The staff is responsible for it, and it includes the ward, medical department, or particular operation unit's internal labor service. It is fixedly dispatched to the demand unit. Non-medical care services, such as ward

replenishment, hand sanitizer, and redemption of infectious devices, medicine ladders, cleaning of dirty clothes, extra isolation clothes, etc., work items will follow the general ward. Moreover, emergency characteristics of intensive or special wards and departmental treatment units may be different.

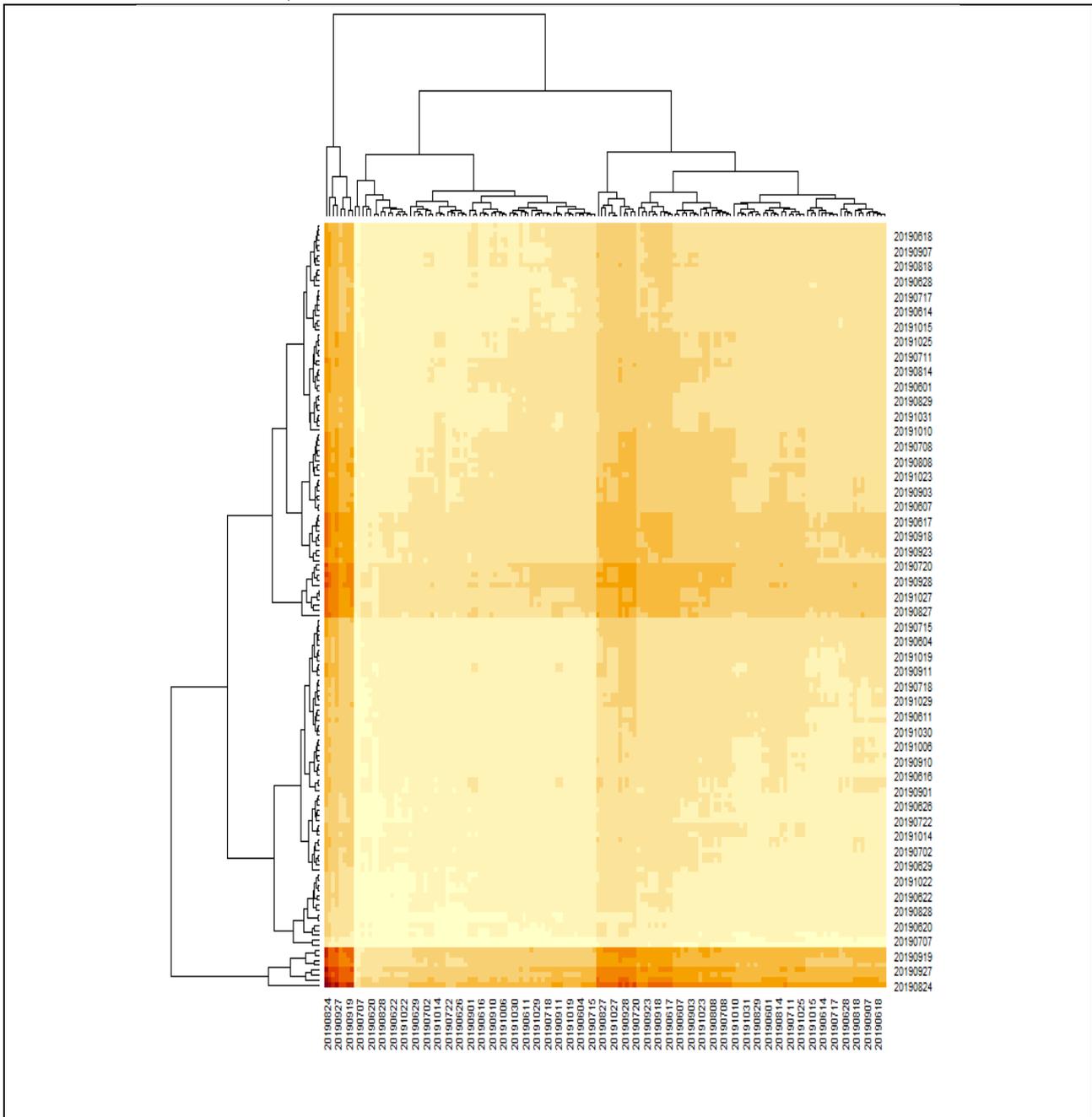
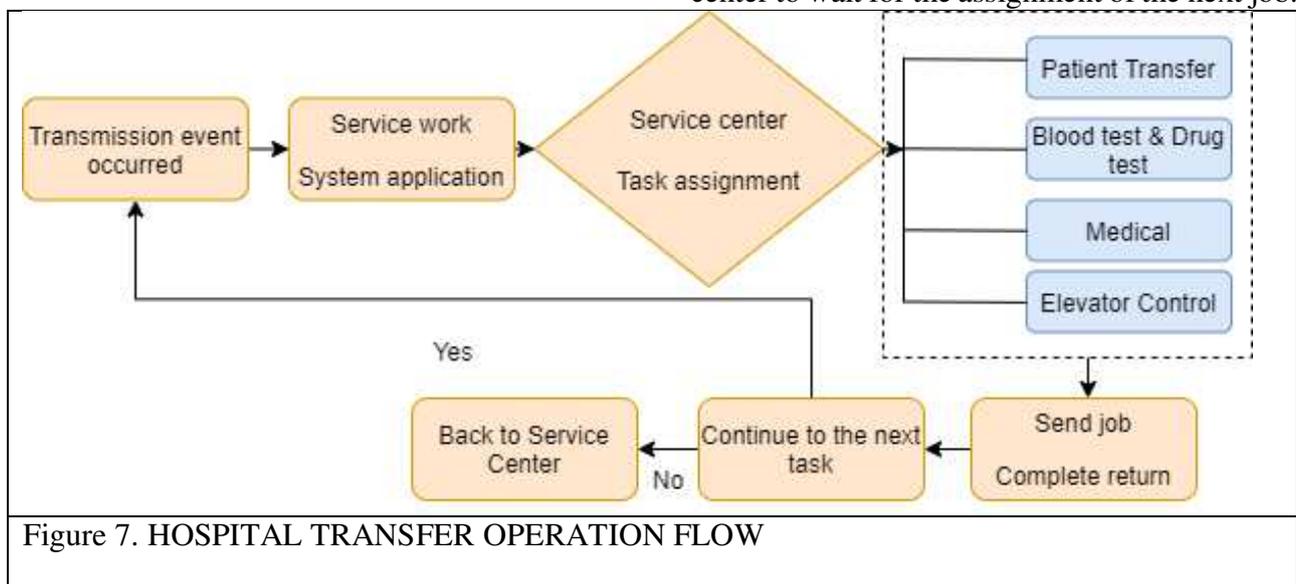


Figure 6. HEATMAP EVENT COUNTS

Still, the work they perform is non-medical affairs, and responsible for such work belongs to internal staff. This mode's transmission requirements are mainly related to the relevant operational processes required for the treatment of inpatients. The examinations are X-rays, ultrasound, electrocardiograms, computed tomography (CT) examinations or anesthesia visits before the operation of the patient; or pushing inpatients for blood Dialysis treatment, receiving emergency treatment medicines, transferring specimens, and related operations

such as respirators, oxygen cylinders, and other Equipment or items required for treatments. Thereupon, Figure 7 represents the transfer process is the application of the "Hospital Transfer Operating System" by the ward nursing station, and the dispatching method is based on the delivery center. The cases are general, urgent, or scheduled categories. The application event is transmitted to the service center to print the document. The service center dispatches personnel to perform the transmission operation build upon the priority of the event transmission or the application sequence. When the transmission staff completes the task, they return to the service center to wait for the assignment of the next job.



An ICU is an Intensive Care Unit, and CCU, for the most part, represents the Cardiac Care Unit. An emergency unit a basic consideration unit that concedes therapeutic and careful patients who are fundamentally sick or harmed, while a Cardiac Care Unit concedes patients with heart issues, generally medicinal cardiovascular issues. Respiratory intermediate care unit (RICU) ought to be practically incorporated with the intensive care hospital room, the general ICU, and the restorative or different wards. These units ought to be described by higher self-sufficiency than the

checking units because of the more elevated level of care [41]. Subsequently, patients with intense on the incessant respiratory disappointment of any level of seriousness ought to be admitted to these units except for the individuals who are as of now intubated. Moreover, fundamentally sick patients with weaning issues could be admitted to the RICU. On the other hand, The Surgical Intensive Care Unit provides care for patients who have undergone a myriad of critical surgical procedures. SICU will cover of Pediatric Vascular, Gastrointestinal Liver, Renal, Renal-

Pancreas Transplantation Orthopaedics, Plastics, Otolaryngology, Urology Thoracic, Surgical Oncology, Oral Maxillo-Facial Obstetrics, and Gynaecological Surgery. Management of patient trauma is essential, and this treatment is carried out at the trauma care centre plus (TNCU).

Traumatic patients need airway evaluation and management, respiratory support, handling of bleeding cases, rapid, swift. Patients who come to the emergency unit must go through triage, which is the process of evaluating the patient's condition to determine the emergency level. Patients will be treated according to the category of triage, videlicet, triage one, patients with conditions that are life-threatening or loss of limb function and require immediate action or intervention with a waiting time of 0 minutes.

Then, triage two is a patient with a non-life-threatening condition, but has a potential threat to limb function and requires prompt medical intervention or action with a waiting time of 0-5 minutes. Triage three are patients with acute conditions, but not urgent (mostly stable), there is no potential to experience worsening, and do not require immediate medical intervention or intervention with a waiting time of 5 to 15 minutes. NICU stands for neonatal intensive care unit is an intensive care room in the hospital that is explicitly provided for newborns who experience health problems [42].

Generally, babies are placed into the NICU room in the first 24 hours after birth. The length of stay in the NICU room varies, depending on the condition of each baby. The more serious the health problem is experienced, the longer he will be in the NICU room. There are many reasons why babies need to be cared for in the NICU room but aim to get the child under intensive supervision and care. The NICU room is a sterile area that cannot be entered by just anyone. Each hospital has a different blueprint regarding the number and hours of parents visiting the NICU room. However, all hospitals must provide soap or

hand sanitizers to ensure that visiting guests are sterile. In general, NICU room conditions are tranquil because the babies in it are susceptible to sound and light. The babies in the NICU room are usually in the incubator to keep their body temperature stable. In terms of hospital delivery business, it is roughly divided into first, patient escort: during the patient's medical treatment process, the patient is pushed for examination, surgery, kidney dialysis or related treatment. Additionally, Non-patient transmission: similar transmission of specimens, drugs, blood, documents, medical records or medical supplies.

This research transmission business is aimed at the business of front-open patient escort and non-patient transmission. According to the different work attributes of each ward, medical department or operating unit, the required human resources are divided into four categories, and various types of human resources are ordered according to their complexity or danger.

Acknowledgement

This work in part of Ministry of Science and Technology, Taiwan, under Grant MOST 109-2622-E-324 -004. This research supported by National Research Foundation of Korea grants [NRF-2019R1A2C1002408].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author Contributions:

Conceptualization: REC,RCC,YL

Methodology: REC,YL

Project Administration:

RCC,YL,SWH,SYC,BP

Software: REC, YL

Validation: REC, YL

Visualization: REC

Writing – original draft, review and editing:

REC,RCC,YL.

Writing – original draft, review and editing:

REC,RCC,YL.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Author's information

¹ *Department of Information Management, College of Informatics, Chaoyang University of Technology, Taiwan (ROC), 41349.*

² *Lab Hierarchical Likelihood, College of Natural Science, Seoul National University, 56-1 Mountain, Sillim-dong, Gwanak-gu, Republic of Korea.*

³ *Department of Statistics, College of Natural Science, Seoul National University, 56-1*

Mountain, Sillim-dong, Gwanak-gu, Republic of Korea.

⁴ *College of Medicine, Seoul National University, 103 Daehak-ro, Ihwa-dong, Jongno-gu, Seoul, Republic of Korea.*

⁵ *Taichung Veterans General Hospital, Taiwan (ROC)*

⁶ *Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia*

⁷ *Computer Science Department, Bina Nusantara University, Jakarta 11480, Indonesia*

Corresponding author: rezzy94@snu.ac.kr

REFERENCES

- [1] Q. Li *et al.*, “A Framework for Big Data Governance to Advance RHINs: A Case Study of China,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2910838.
- [2] X. D. Wang, R. C. Chen, F. Yan, Z. Q. Zeng, and C. Q. Hong, “Fast Adaptive K-Means Subspace Clustering for High-Dimensional Data,” *IEEE Access*, vol. 7, pp. 42639–42651, 2019, doi: 10.1109/ACCESS.2019.2907043.
- [3] J. L. Raheja, Dhiraj, D. Gopinath, and A. Chaudhary, “GUI system for elders/patients in intensive care,” 2014, doi: 10.1109/ITMC.2014.6918605.
- [4] G. Hever, L. Cohen, M. F. O’Connor, I. Matot, B. Lerner, and Y. Bitan, “Machine learning applied to multi-sensor information to reduce false alarm rate in the ICU,” *J. Clin. Monit. Comput.*, vol. 34, no. 2, pp. 339–352, 2020, doi: 10.1007/s10877-019-00307-x.
- [5] P. Cao, S. ichi Toyabe, T. Abe, and K. Akazawa, “Profit and loss analysis for an intensive care unit (ICU) in Japan: a tool for strategic management,” *BMC Health Serv. Res.*, vol. 6, p. 1, 2006, doi: 10.1186/1472-6963-6-1.
- [6] A. Agresti, *An Introduction to Categorical Data Analysis*. 2007.
- [7] M. Capuzzo, R. P. Moreno, and R. Alvisi, “Admission and discharge of critically ill patients,”

- Curr. Opin. Crit. Care*, vol. 16, no. 5, pp. 499–504, 2010, doi: 10.1097/MCC.0b013e32833cb874.
- [8] I. Do Ha and Y. Lee, “Estimating Frailty Models via Poisson Hierarchical Generalized Linear Models,” *Journal of Computational and Graphical Statistics*, 2003, doi: 10.1198/1061860032256.
- [9] I. Do Ha, M. Noh, and Y. Lee, “FrailtyHL: A package for fitting frailty models with h-likelihood,” *R J.*, vol. 4, no. 2, pp. 28–37, 2012, doi: 10.32614/rj-2012-010.
- [10] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, “Big data in healthcare: management, analysis and future prospects,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0217-0.
- [11] J. Gower, S. Lubbe, and N. le Roux, “Principal Component Analysis Biplots,” in *Understanding Biplots*, 2011.
- [12] C. J. F. ter Braak and P. F. M. Verdonschot, “Canonical correspondence analysis and related multivariate methods in aquatic ecology,” *Aquat. Sci.*, 1995, doi: 10.1007/BF00877430.
- [13] M. Noh, Y. Lee, J. H. L. Oud, and T. Toharudin, “Hierarchical likelihood approach to non-Gaussian factor analysis,” *J. Stat. Comput. Simul.*, vol. 89, no. 3, pp. 1555–1573, 2019.
- [14] S. Jin, M. Noh, and Y. Lee, “H-Likelihood Approach to Factor Analysis for Ordinal Data,” *Struct. Equ. Model.*, vol. 25, no. 4, pp. 530–540, 2018, doi: 10.1080/10705511.2017.1403287.
- [15] R. Mitchell, A. Adinets, T. Rao, and E. Frank, “XGBoost: Scalable GPU Accelerated Learning,” 2018. [Online]. Available: <http://arxiv.org/abs/1806.11248>.
- [16] R. C. Chen *et al.*, “An End to End of Scalable Tree Boosting System,” *Sylwan*, vol. 165, no. 1, pp. 1–11, 2020.
- [17] R. E. Caraka, N. T. Nugroho, S.-K. Tai, R. C. Chen, T. Toharudin, and B. Pardamean, “Feature Importance of The Aortic Anatomy on Endovascular Aneurysm Repair (EVAR) using Boruta and Bayesian MCMC,” *Commun. Math. Biol. Neurosci.*, vol. 2020, no. 1–23, 2020, doi: <https://doi.org/10.28919/cmbn/4584>.
- [18] I. M. Johnstone and D. M. Titterington, “Statistical challenges of high-dimensional data,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2009, doi: 10.1098/rsta.2009.0159.
- [19] A. Choiruddin, J. F. Coeurjolly, and F. Letué, “Convex and non-convex regularization methods for spatial point processes intensity estimation,” *Electron. J. Stat.*, vol. 12, no. 1, pp. 1210–1255, 2018, doi: 10.1214/18-EJS1408.
- [20] A. Choiruddin, F. Cuevas-Pacheco, J. F. Coeurjolly, and R. Waagepetersen, “Regularized estimation for highly multivariate log Gaussian Cox processes,” *Stat. Comput.*, pp. 1–14, 2019, doi: 10.1007/s11222-019-09911-y.
- [21] S. Kwon, S. Oh, and Y. Lee, “The use of random-effect models for high-dimensional variable selection problems,” *Comput. Stat. Data Anal.*, vol. 103, no. 1, pp. 401–412, 2016.
- [22] D. I. Warton, “Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data,” *Environmetrics*, vol. 16, no. 3, pp. 275–289, 2005, doi: 10.1002/env.702.
- [23] J. del Castillo and Y. Lee, “GLM-methods for volatility models,” *Stat. Modelling*, vol. 8, no. 3, pp. 263–283, 2008, doi: 10.1177/1471082X0800800303.
- [24] D. A. Lury and R. A. Fisher, “Statistical Methods for Research Workers.,” *Stat.*, 1972, doi: 10.2307/2986695.
- [25] W. R. Dillon and M. Goldstein, “Multivariate Analysis. Methods and Applications,” *Geological Magazine*, 1985, doi: 10.1017/S0016756800031678.
- [26] P. Shi and E. A. Valdez, “Multivariate negative binomial models for insurance claim counts,” *Insur. Math. Econ.*, vol. 55, no. 1, pp. 18–29, 2014, doi: 10.1016/j.insmatheco.2013.11.011.
- [27] D. M. Bates and D. G. Watts, “Review of Linear Regression,” *Nonlinear Regres. Anal. Its Appl.*, pp. 1–31, 1988, doi: 10.1002/9780470316757.ch1.
- [28] I. D. Ha, Y. Lee, and J. K. Song, “Hierarchical-Likelihood Approach for Mixed Linear Models with Censored Data,” *Lifetime Data Anal.*, vol. 8, no. 2, pp. 163–176, 2002, doi: 10.1023/A:1014839723865.
- [29] Y. Lee and J. A. Nelder, “Double hierarchical generalized linear models,” *J. R. Stat.*

- Soc. Ser. C Appl. Stat.*, 2006, doi: 10.1111/j.1467-9876.2006.00538.x.
- [30] A. Buchholz, A. Buchholz, and S. T. Université, “High dimensional Bayesian computation,” ENSAE-CREST Paris, 2018.
- [31] J. M. Hilbe, *Negative binomial regression*. 2007.
- [32] J. Niku, W. Brooks, R. Herliansyah, F. K. C. Hui, S. Taskinen, and D. I. Warton, “Efficient estimation of generalized linear latent variable models,” *PLoS One*, vol. 14, no. 5, pp. 1–20, 2019, doi: 10.1371/journal.pone.0216129.
- [33] R. E. Caraka *et al.*, “Ecological Show Cave and Wild Cave: Negative Binomial Gllvm’s Arthropod Community Modelling,” *Procedia Comput. Sci.*, vol. 135, pp. 377–384, 2018, doi: 10.1016/j.procs.2018.08.188.
- [34] D. Lee *et al.*, “Optimal likelihood-ratio multiple testing with application to Alzheimer’s disease and questionable dementia Data analysis, statistics and modelling,” *BMC Med. Res. Methodol.*, vol. 15, no. 1, pp. 1–11, 2015, doi: 10.1186/1471-2288-15-9.
- [35] R. Herliansyah and I. Fitria, “Latent variable models for multi-species counts modeling in ecology,” *Biodiversitas*, 2018, doi: 10.13057/biodiv/d190538.
- [36] J. T. Ormerod and M. P. Wand, “Explaining variational approximations,” *Am. Stat.*, vol. 64, no. 2, pp. 140–153, 2010, doi: 10.1198/tast.2010.09058.
- [37] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, “The variational approximation for Bayesian inference,” *IEEE Signal Process. Mag.*, 2009, doi: 10.1109/msp.2008.929620.
- [38] Z. Shun and P. McCullagh, “Laplace Approximation of High Dimensional Integrals,” *J. R. Stat. Soc. Ser. B*, 1995, doi: 10.1111/j.2517-6161.1995.tb02060.x.
- [39] A. Adibi and M. SHARIFI, “Semiconductor Device Simulation by a New Method of Solving Poisson, Laplace and Schrodinger Equations,” *Int. J. Eng.*, vol. 13, no. 1, pp. 89–94, 2000.
- [40] Z. Mohammadpoory and J. Haddadnia, “Speech Enhancement Using Laplacian Mixture Model under Signal Presence Uncertainty,” *Int. J. Eng.*, vol. 27, no. 9 (C), pp. 1367–1376, 2014, doi: 10.5829/idosi.ije.2014.27.09c.06.
- [41] A. Corrado *et al.*, “Respiratory intermediate care units: A European survey,” *Eur. Respir. J.*, vol. 20, no. 5, pp. 1343–1350, 2002, doi: 10.1183/09031936.02.00058202.
- [42] E. Losiouk *et al.*, “A telemonitoring service supporting preterm newborns care in a neonatal intensive care unit,” 2017, doi: 10.1109/RTSI.2017.8065890.

Figures

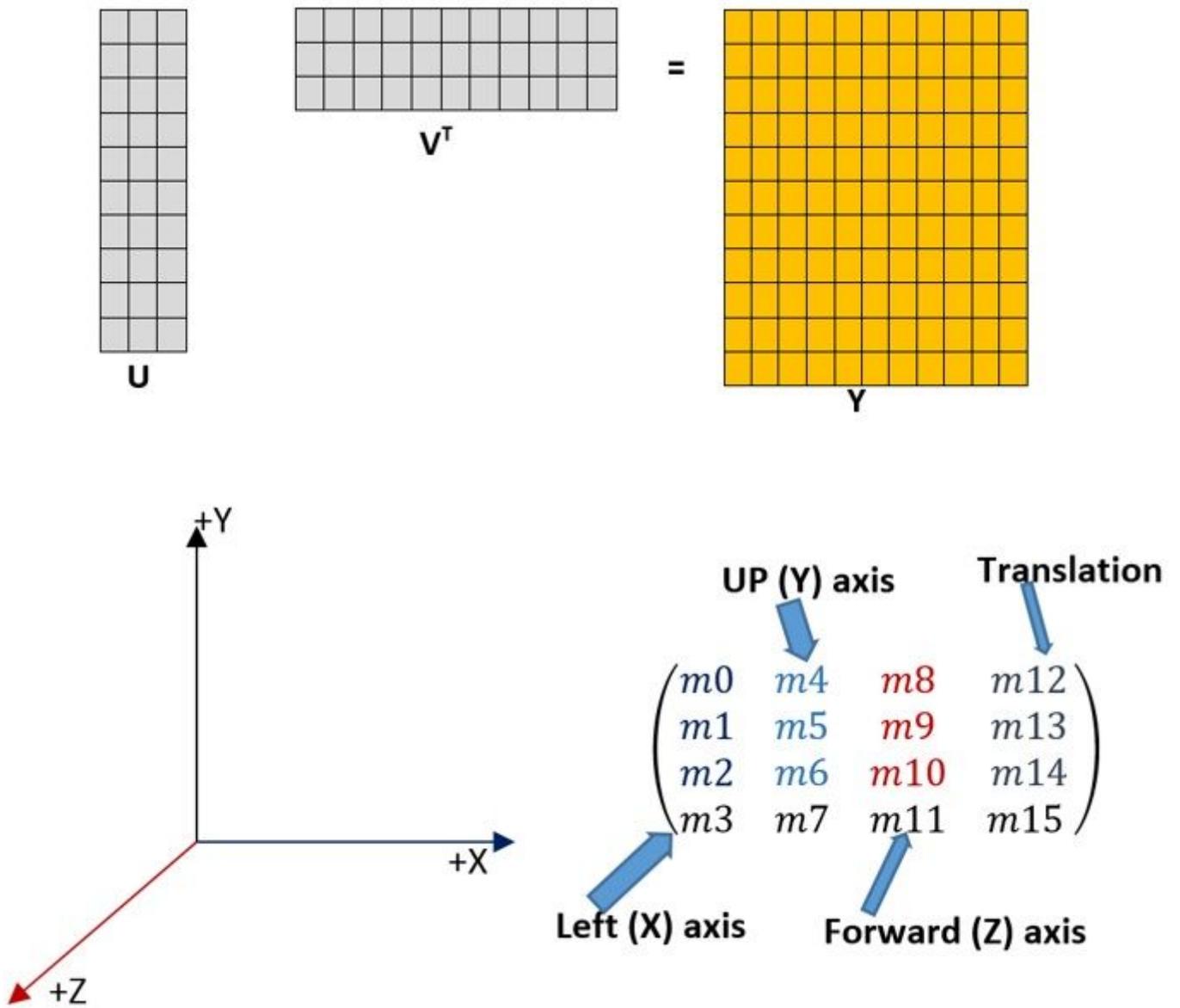


Figure 1

Example of projection

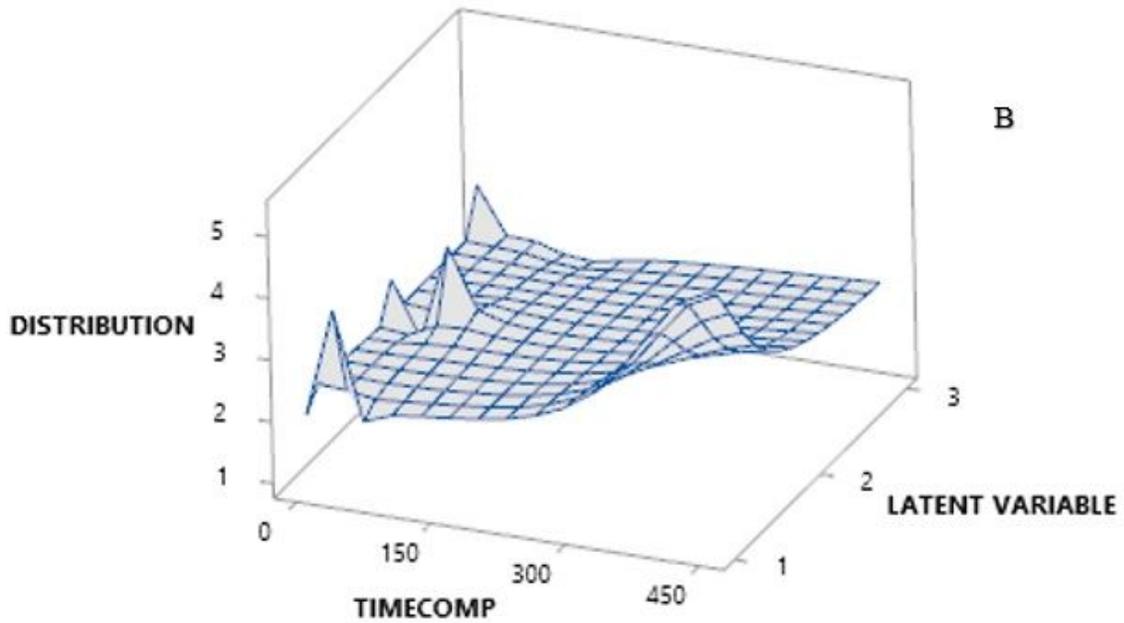
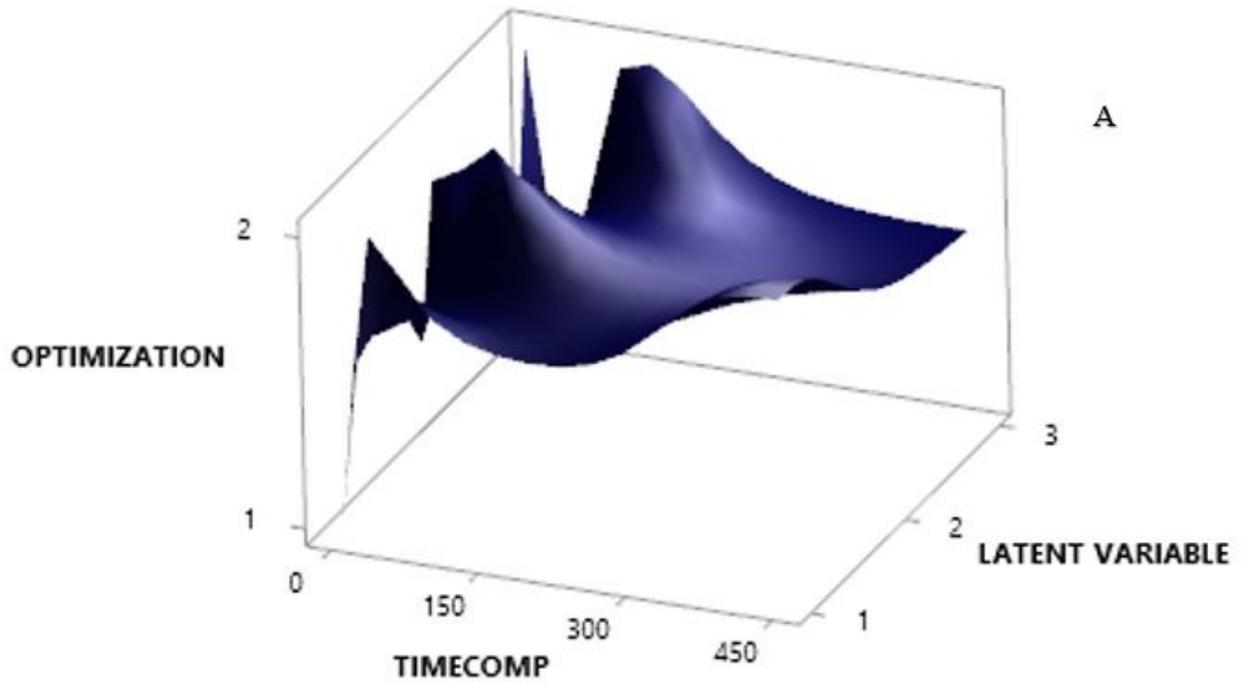


Figure 2

Time Computing Optimization (A) and Type of Distribution(B)

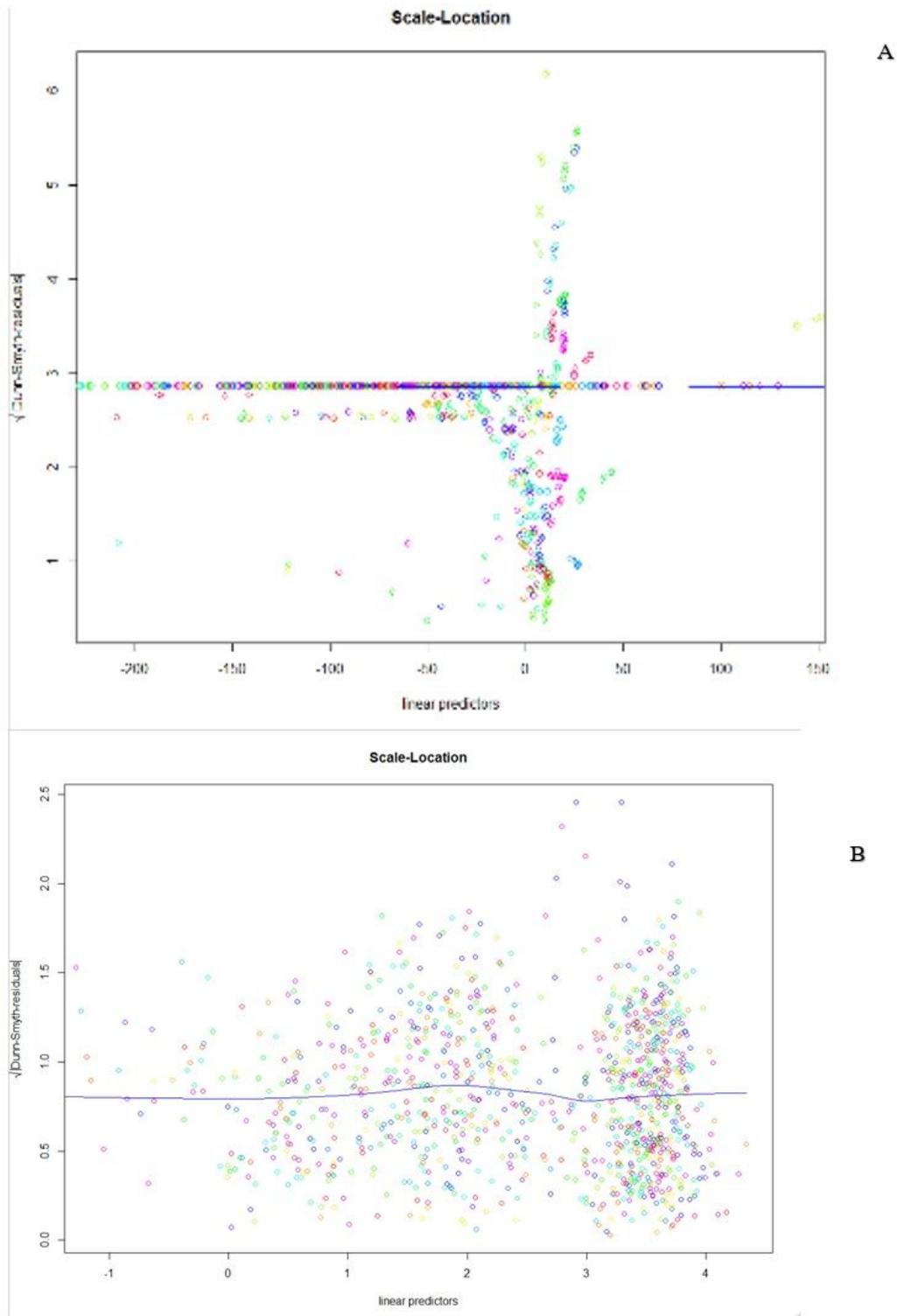
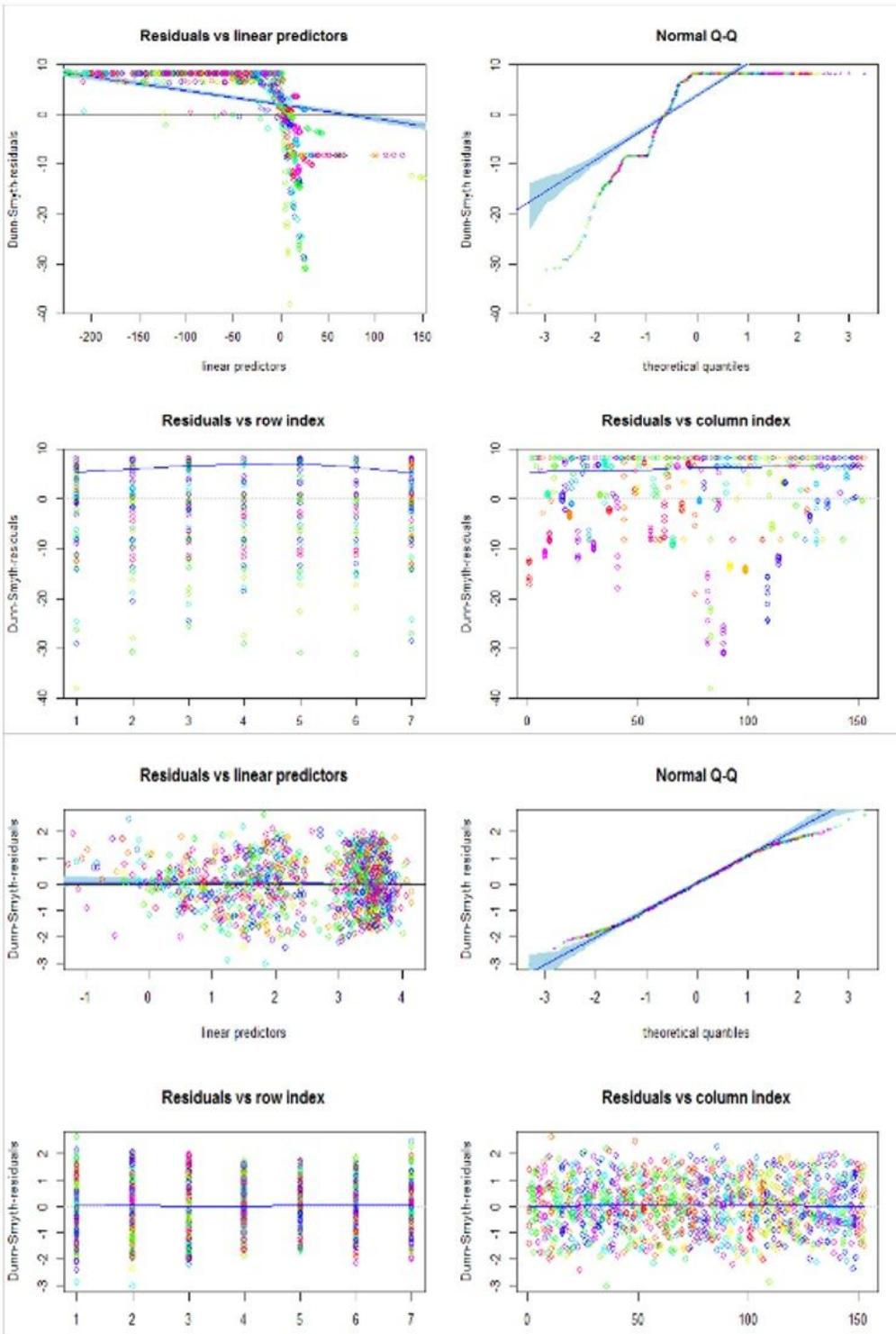


Figure 3

Scale Location GLLVM Negative Binomial with 1 Latent Variable Variational Approximation (A) Scale Location GLLVM Negative Binomial with 1 Latent Variable Laplace Approximation (B)



A

B

Figure 4

GLLVM Negative Binomial with 1 Latent Variable Variational Approximation Residual VS Predictor (A)
 GLLVM Negative Binomial with 1 Latent Variable Laplace Approximation Residual VS Predictor (B)

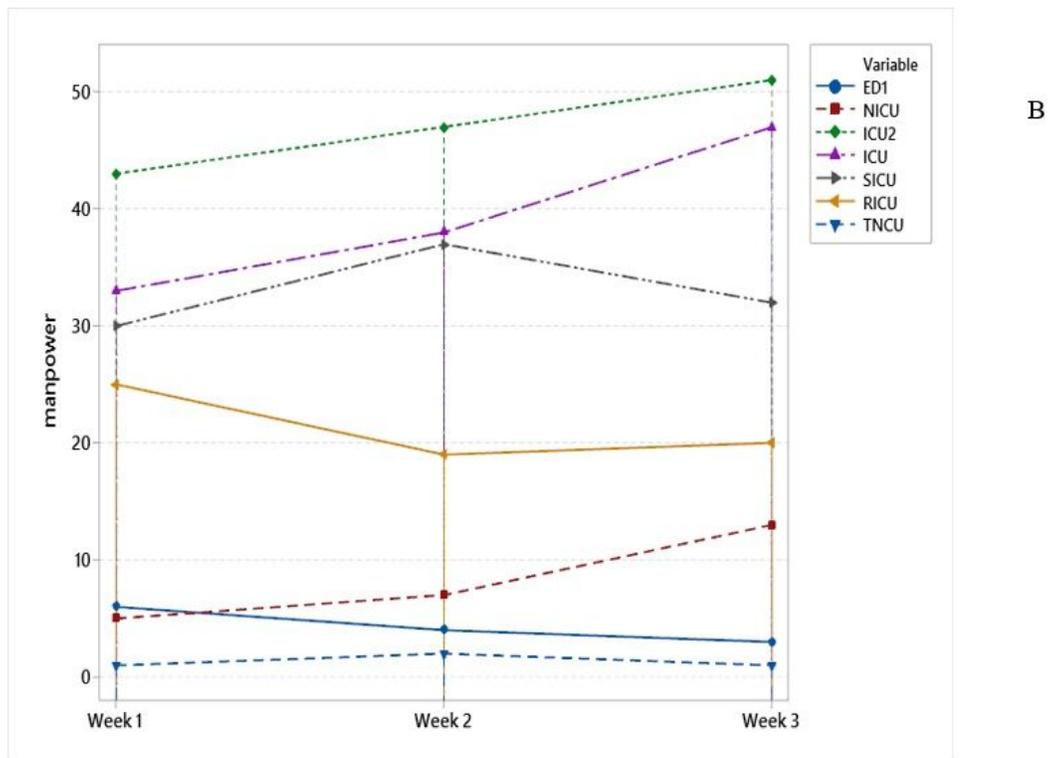
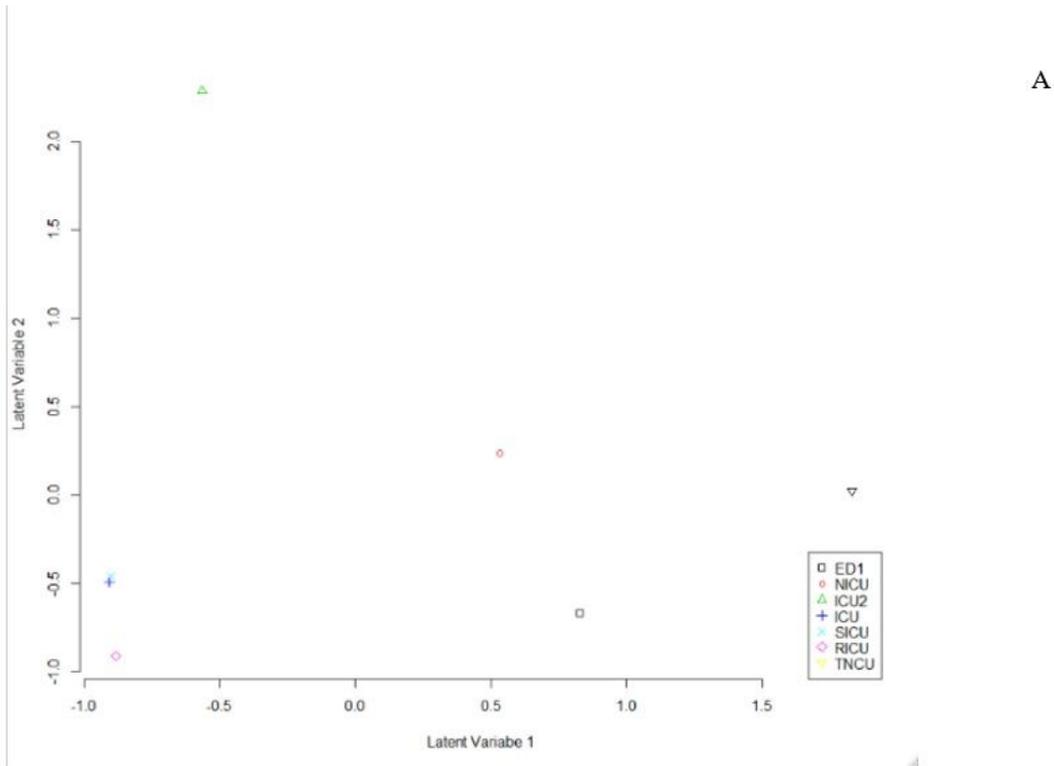


Figure 5

GLLVM Ordination (A) and Prediction manpower (B)

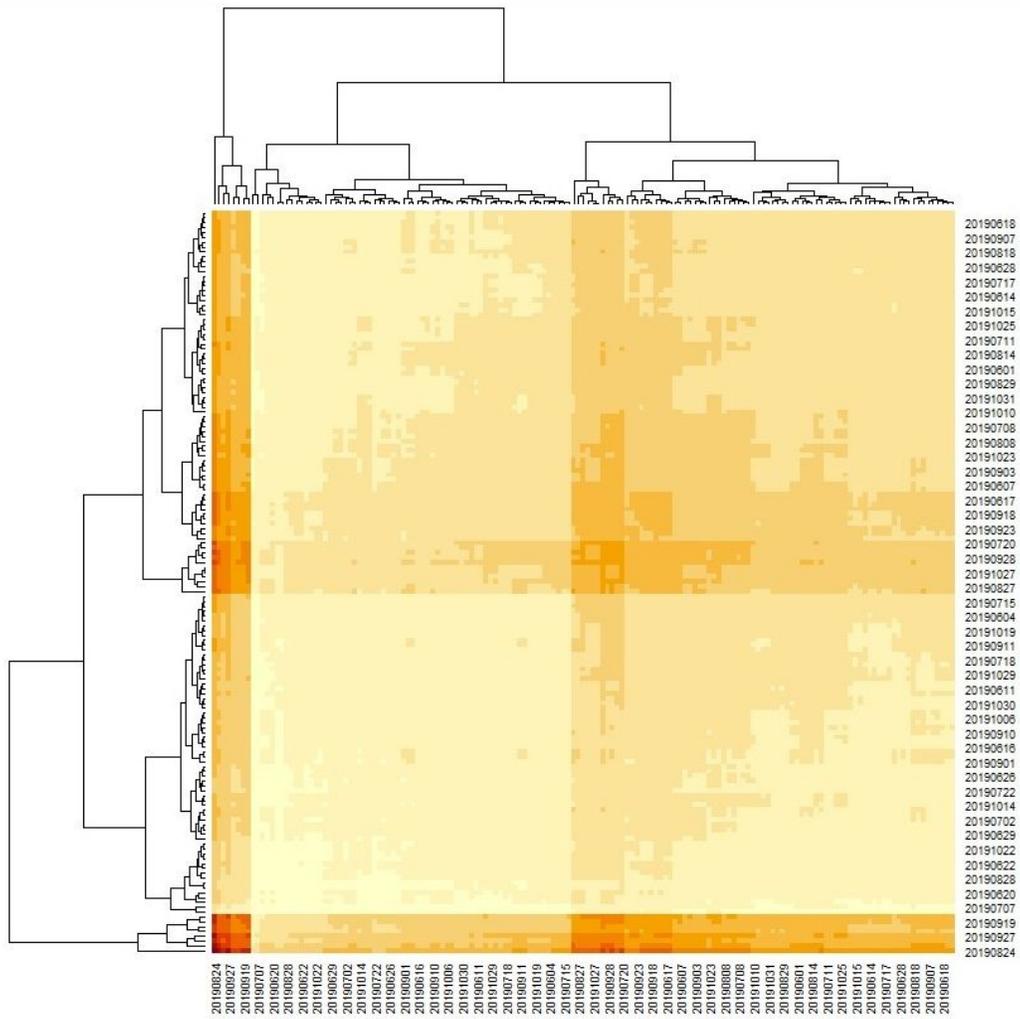


Figure 6

HEATMAP EVENT COUNTS

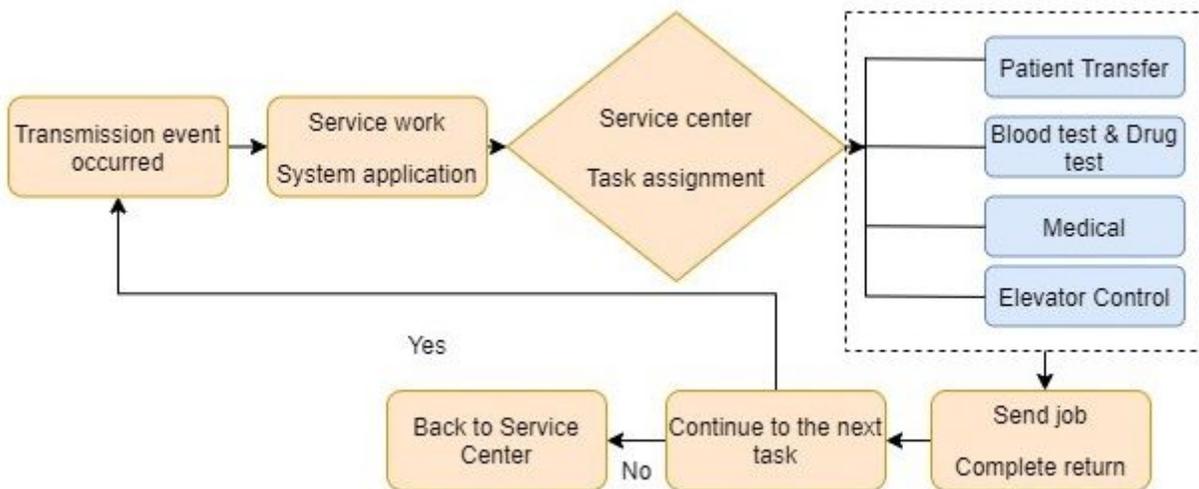


Figure 7

HOSPITAL TRANSFER OPERATION FLOW