

Hierarchical linear modelling of risk assessment of petroleum installations

Syed T. Tirmizi^{1*}, Syed R.U.H. Tirmizi²

¹Petroleum Engineering Department, NED University of Engineering and Technology, Karachi, Pakistan

²PNEC – National University of Sciences and Technology, Islamabad, Pakistan

Email: talhatirmizi@outlook.com

ABSTRACT

The present-day configuration of industrialized civilization well recognizes the importance of petroleum for its sustenance and its role in the country's economy. Constantly working in highly vulnerable environment, oil and gas companies have to endure high level of risks encompassing environmental, operational, health and safety, as well as organizational and human hazards. Risk assessment has significantly grown to prominence in petroleum industry with an intention to mitigate these hazards. This research provides deep insight about the application of hierarchical linear modelling, a widely accepted tool in the field of risk assessment. Consequently, the proposed multilevel risk analysis will have a comprehensive approach than any other risk estimation methods. The present research infers to managing the confronting risks by studying the arbitrary relationships between the hazards in multilevel hierarchical modelling. To ensure the safety of the whole system, a statistical model is developed that would have a probabilistic approach towards the minimization of the potential hazards in the industry, quantifying the risk associated with the activity process.

Keywords: Hierarchical Linear Modelling, Oil and Gas Industry, Risk Assessment.

1. INTRODUCTION

The potential hazards associated with the oil and gas industry is unavoidable and the production of highly combustible materials cause a great security concern for the workers in close proximity to it. The oil companies may endure significant financial losses and environmental damage if fire and accidents interrupt the oil production. The focus for oil and gas companies is on how to manage all kinds of risk in order to rein in costs. And that task is more daunting than ever for the upstream sector of the oil and gas industry. [1] Risk assessment has emerged as a result of worldwide interest in different aspects of hazards. Risk is often defined as possible loss from the potentially harmful phenomena in the prescribed area. [2] Risk Assessment delivers the structural basis for the oil and gas industry in mitigating the hazards while ensuring that risk has been reduced to minimal level economically. The present research incorporates the significance of multilevel hierarchical modelling into the risk analysis of oil and gas industry, providing a statistical model to help in the understanding of the risk. Hierarchical linear modelling (HLM) has emerged as a prominent statistical analysis that has attracted the attention of many researchers and it forms the basis of risk assessment in this research. HLM requires datasets arranged in groups and form a sense of grouping hierarchy with different levels. In multilevel modelling, 2-level model is usually preferred, and the present research is

based on 2-level model as demonstrated in Table 2. During HLM analyses, a normally distributed parameter is identified as the single dependent variable which is of great interest during the study. The data is arranged in a chronological order on the basis of their dependency. At level-1, there are independent variables measured at a specific time interval. The level-2 parameters serve as the between-group predictor variables. The regression equations at each level is provided in methodology section. The advancement in computing technologies has greatly influenced the HLM statistical techniques in a sense that larger datasets could be evaluated with improved accuracies. The evaluation of cross-level interaction is of high interest in HLM in which slope is randomly varied and to obtain level-1 covariate we may include level-2 predictor in the slope formula. In brief, HLM is a very flexible statistical technique which find its applications in various fields and the present research intends to contribute towards its flexibility and help in the understanding of risk analysis in OG industry statistically. In the present study, HLM is imperatively utilized to develop a risk model for exhibiting the hazards associated with the petroleum installations. It is envisaged that this research would contribute towards providing a substantial model for risk management by quantifying the risks inherent in the operations of petroleum installations and delivering better understanding of it.

2. METHODOLOGY

Data was obtained from project titled “Accident statistics for offshore floating units on the UK Continental Shelf (UKCS)” [3]. This project was carried out with the sole purpose of aggregation of data for the later use in risk assessment projects. The data obtained, revealed the accident frequencies for the accidents and incidents which had occurred on the fixed units engaged in oil exploration and production on the UKCS in the period 1990-1999. In the present research, only four petroleum installations were considered as adequate information was available about them which was sufficient for the evaluation process. The classification of units is mentioned below:

Table 1. Classification of units

Category	Type of unit- description
Drilling	Operations on drilling platform
Production	Manned production platforms which also includes drilling, production and accommodation facilities.
Compression	Gas Compression Platforms.
Wellhead	Serving only as well support.

The classification of events has been done as per WOAD concept which shows that one accident may be linked to a series of consecutive events e.g. explosion, fire and oil spill can occur as a result of blowout. This means that one single event may give rise to several occurrences. Thus, a sense of hierarchy has been established and the multilevel modelling is evaluated to further derive results from the data. The total number of occurrences will thus be higher than the total no. of accidents. In table 2, this concept is presented as number of occurrences per unit year. In the following table, each variable is assigned its particular variable id for the sake of convenience when dealing with the equations. The factors at each hierarchical level that affects the Accidents is mentioned in the table below:

Table 2. Multilevel dataset

Hierarchical Level	Variable ID	Variables
Level-2	α_1	Anchor Failure
	α_2	Blowout
	α_3	Capsize
	α_4	Collision
	α_5	Contact
	α_6	Crane
	α_7	Explosion
	α_8	Falling Object
	α_9	Fire
	α_{10}	Foundering
	α_{11}	Grouping
	α_{12}	Helicopter
	α_{13}	Leakage
	α_{14}	List

α_{15}	Machinery
α_{16}	Off position
α_{17}	Spill/Leakage
α_{18}	Structural
α_{19}	Towing/towline
α_{20}	Well problem
α_{21}	Other

Level-1	ACCIDENT*	Accidents
	NO (used as short form for number of observed installations)	Number of observed Installations where accidents had taken place each year.

Notes: * indicates the outcome variable which is mostly a level-1 variable.

The level-1 data is illustrated in figure 1 by means of the stacked area graph which demonstrates the number of accidents for each installation per year, with production installation having high level of accidents. The complexity of level 2 data with respect to its numerous parameters have made it difficult to illustrate the data in graphical manner but figure 1 provides apparent assumption to visualize the level-2 data as it depends directly on the number of accidents occurred in level-1 data.

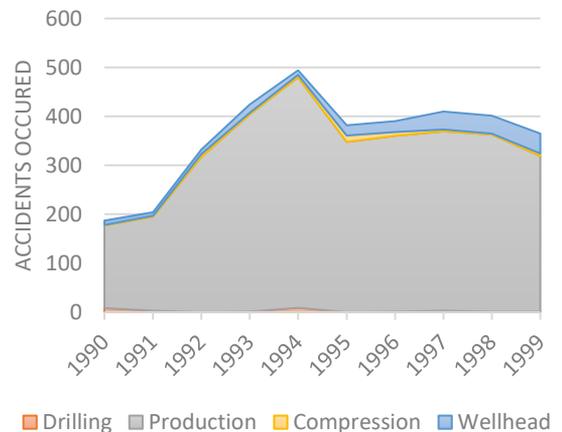


Figure 1. Level-1 data of no. of accidents per year

Hierarchical linear models use nested regression equations to investigate associations between variables at different levels. This accounts for how observations can be related in groups within a hierarchy. [4] In multilevel modelling, four models were evaluated from the data which includes unconditional model, random intercept model, mean outcome model and slope & intercept model. The results obtained from these models were compared with each other. In the initial phase of research, the data was categorized in four installations as obtained, namely, drilling, production, compression and wellhead. The multilevel modelling was then performed on each of the four installations individually. Variance component model, also called random intercept model, can be fitted as the simplest multilevel model. This model serves as an extension of linear model where the effects of accident frequencies has on the accidents occurred per year are treated as random variables. We can write it as

$$y_{ij} = \beta_0 + X_{ij}^1 \beta^1 + X_j^2 \beta^2 + \mu_j + e_{ij} \quad (1)$$

where,

$\beta_0 =$ global intercept,

$\beta^1 =$ set of level 2 effects

$\beta^2 =$ set of level 1 effects

$\mu_j =$ set of level 2 residuals

$e_{ij} =$ set of level 1 residuals

Eq(1) is modified in which the fixed part of the model is referred to the combined effect of first three terms while μ_j and e_{ij} are referred to as random part of the model. This modified version of the Eq(1) is written as:

$$y_{ij} = X_{ij}\beta + \mu_j + e_{ij}, \quad (2)$$

where,

$$\mu_j \sim N(0, \sigma_\mu^2)$$

$$e_{ij} \sim N(0, \sigma_e^2).$$

Multilevel residuals were defined from the normal distributions assumed and σ_μ^2 and σ_e^2 are the variances of the level 1 and level 2 residuals, respectively. The percentage variance in a data is defined by intra-class correlation or ICC.i.e.

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad (3)$$

It is interested to find the value of ICC as it explains the percentage variance in the data by higher level residuals.

Multilevel models follow complicated estimation procedures unlike linear models which can be estimated by using simple matrix formulae. Several algorithms exist for its estimation; we will use the one described by HLM (Centre for Hierarchical Linear Modelling). When dealing with multilevel data, ordinary regression approaches can take three forms. First, individual data grouped at level 1 is analyzed without regard for group structure. This method ignores group level 2 variability and disrupts OLS regression's main assumptions. Second, separate un-pooled analyses can be conducted within each group. This approach can be useful for examining between-group variability, but it requires relatively large samples for each group and it is cumbersome when the number of groups becomes large. Third, aggregate analysis can also be conducted at the group level alone, but this approach ignores within-group variability and requires a relatively large number of groups for analysis. [5]

Our first model, unconditional model, which is only an intercept estimate with no predictor variables. In this model, the intercept is estimated by regressing the variance in Accidents onto a unit vector, which provides the mean variance. This is done by partitioning the within and between-group variance in Accidents, which allows for the estimation of ICC.

Mixed Model:

$$ACCIDENT_{ij} = \gamma_{00} + u_{0j} + e_{mj}$$

Hierarchical Model:

$$ACCIDENT_{mj} = \psi_{0j} + e_{mj}$$

$$\psi_{0j} = \gamma_{00} + u_{0j}$$

Once, the above model is satisfied, HLM can proceed with the following three models where there exists significant variance in intercept and slopes across groups.

The second model is referred to as random intercept model which includes the number of observed installations predictor. This model is significant to answer the question whether the years with more number of observed installations also have high number of accidents. In terms of regression equations, it can be stated as:

Mixed Model:

$$ACCIDENT_{mj} = \gamma_{00} + \gamma_{10} * NO_{mj} + u_{0j} + e_{mj}$$

Hierarchical Model:

$$ACCIDENT_{mj} = \psi_{0j} + \psi_{1j} * (NO_{mj}) + e_{mj}$$

$$\psi_{0j} = \gamma_{00} + u_{0j}$$

$$\psi_{1j} = \gamma_{10}$$

Regression with mean outcome model is the third model which includes the effects of level 2 predictors. This model is motivated to answer the questions related to the correlation between intercepts and slopes. The equations for this model can be stated as under:

Mixed Model:

$$ACCIDENT_{mj} = \gamma_{00} + \sum_{i=1}^n \gamma_{0i} * \alpha_j + u_{0j} + e_{mj}$$

Hierarchical Model:

$$ACCIDENT_{mj} = \psi_{0j} + e_{mj}$$

$$\psi_{0j} = \gamma_{00} + \sum_{i=1}^n \gamma_{0i} * \alpha_j + u_{0j}$$

$$\psi_{1j} = \gamma_{10}$$

The fourth model is referred to as intercept and slope model. This model is used to achieve an explanatory model for the variability in the data. Generally, this model is used to test the interactions between the level-1 and level-2 predictor variables and this model also helps to verify the two previously discussed models because this model tests the main effects of predictor variables on both the levels instead of evaluating them separately. In terms of regression equations, it is written as:

Mixed Model:

$$ACCIDENT_{mj} = \gamma_{00} + \sum_{i=1}^n \gamma_{0i} * \alpha_j + \gamma_{(n+1)0} * NO_{mj} + u_{0j} + e_{mj}$$

Hierarchical Model:

$$ACCIDENT_{mj} = \psi_{0j} + \psi_{1j} * (NO_{mj}) + e_{mj}$$

$$\psi_{0j} = \gamma_{00} + \sum_{i=1}^n \gamma_{0i} * \alpha_j + u_{0j}$$

$$\psi_{1j} = \gamma_{10}$$

In brief, for each multilevel model considered, we have incorporated prior distribution to create equivalent models to

determine unknown parameters and inferred on the resulting posterior distributions, that is explained in the following section.

3. DISCUSSION

Initially, a one-way analysis of variance (ANOVA) is evaluated to validate the variability in the outcome variable which is significant than being zero, referred to as Unconditional Model. This provides the basis for the evaluation of further models as to whether there is difference in outcome variable at group level. In unconditional model, the result which holds the most significance is the chi-square test (χ^2) found under the heading of Final estimation of Variance Components, in the HLM output report. If this result is significant, it indicates there is a considerable variance in the outcome variable in level-2 groupings. This also provides justification for carrying out further model analyses. The result for the unconditional model indicates that for the drilling and production installations, chi-square test result is $\chi^2(9) = 11.13288, p < .001$ while in the case of wellhead and compression installations, the chi-square test result is $\chi^2(9) = 9.00000, p < .001$. The fore-mentioned results of chi-square test provides the justification for evaluating further HLM analyses since the values obtained show significance. Additionally, Intra-Class Correlation (ICC) can be determined which is the percentage variance in the data. Here, ICC was calculated from the formula discussed in the methodology section. The results for ICC can be determined by using the $\sigma^2(\text{level-1})$ and τ (level-2) and ICC value was found out to be 0.2 which means that 20% of the variance in the Accidents is at the group level and 80% variance exists at individual level.

Now, proceeding further to the Random Intercept model, we intent to test the relationship between the outcome variable and level-1 predictor. i.e. no. of observed installations in a year. For this model, level-1 predictor variable was entered as group centered variable to get more accurate intercept estimates and study the predictor variable independently, at group level. The outcome variable remained untouched in the model as it was in the unconditional model. Considering the HLM output report for random intercept model, it is significant to note the non-standardized final estimation effects which should be similar to the standardized values with robust errors otherwise, it may be concluded that there has been violation of assumptions e.g. normality, in which case, final estimation standardized values should be used. The results of the present analysis support the relationship between Accidents and no. of observed installations. To calculate the measure of effect size, variance (τ^2) is explained by level-1 predictor variable in outcome variable using the following equation.

$$\tau^2 = \frac{\sigma_{null}^2 - \sigma_{random}^2}{\sigma_{null}^2}$$

From the results obtained, it is concluded that while considering the drilling installation type, no. of observed installations in a year explains a variance of 2.5% in Accidents. In table 3 below, this relationship has been explained for the remaining installation types under consideration.

Table 3. Results for random intercept model

Installation Type	τ^2	Variance %
Drilling	.025	2.5 %
Production	.411	41.1 %
Compression	.1396	13.96 %
Wellhead	.872	87.2 %

Subsequently, means as outcomes model is developed for testing the relationship between outcome variable and level-2 predictor variables. If level-1 and level-2 predictor variables are being considered separately then centering is not necessary as it won't change the regression but will only influence the intercept value. In the present study, at level-2, the mean score of zero is not significantly important since the score ranges from 1 to nearly 5 in most of the cases. The analysis of current model reveals the following results:

Table 4. Results for means as outcome model

Installation Type	Variance components	Standard deviation (b)	τ^2	Between measures variance (%)
Drilling	2.85640	1.69009	-.427	-42.7 %
Production	383.277	19.57747	.757	75.7 %
Compression	5.02513	2.24168	.599	59.9 %
Wellhead	4.17325	2.04285	.896	89.6%

Notes: In the table 4 above, τ^2 is obtained by null-model testing.

Table 5. Results for slope and intercept model

Drilling		
Fixed effect	Coefficient (β_s)	t-ratio (t)
Contact	3.75	0.348
Falling	1	0.301
Fire	3.83333	0.475
Production		
Fixed effect	Coefficient (β_s)	t-ratio (t)
Crane	3.499139	1.008
Spill	1.588279	1.929
Wellhead		
Fixed effect	Coefficient (β_s)	t-ratio (t)
Contact	14.22285	2.02
Crane	4.309573	0.64
Spill	1.276177	1.384
Well problem	9.616222	1.97
Compression		
Fixed effect	Coefficient (β_s)	t-ratio (t)
Crane	0.243	0.817
Fire	3.01	0.03

The results displayed in table 4 infer that there is still variation in the accidents that could be further explained. Therefore, it is highly obliged to study a more complex model called slope and intercept model which includes other

independent variables, in order to explain the variability more comprehensively.

Finally, the slope and intercept model is run to test the interaction between the predictor variables of level1 and level2. The inter-level effects of the variables were tested inclusively. In brief, the slope and intercept model could be used to test the two models discussed earlier, instead of evaluating them separately, but there will be variation in the obtained results, since the maximum likelihood estimation methods will vary. The slope and intercept model revealed the significance of different variables in terms of their fixed effects. This result is displayed in table 5 below:

The table 5 demonstrates the results for only those fixed effects which showed significant coefficient values. In the drilling installation, the main effects of Contact on Accidents is significant. i.e. ($\beta_s = 3.75, t = 0.348$). That is, the grand mean of Accidents increases or decreases by 3.75 units with 1 unit increase or decrease in Contact. This concept applies in a similar manner to the rest of the other fixed effects. The main effects of different significant variable can be compared across the four installations with the help of figure 2.

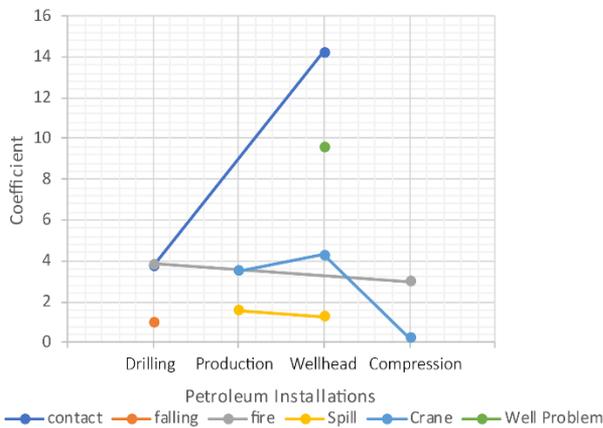


Figure 2. Comparison of main effects of different predictor variables by slope and intercept model

In figure 2, the comparison of different variables could be done by visual inspection. From the figure, the main effects of crane variable is seen prevalent in three installations, namely, production, wellhead and compression, whereas in wellhead installation, contact variable has the most significant value of coefficient ($\beta_s = 14.22285, t = 2.02$). It can be concluded that all the activities involved with crane operations are more exposed to accidents in three of the installations, but the activities associated with contact variable are anticipated to have more dire consequences due to their higher coefficient values.

4. RESULTS

Since the analyses are complete, it is now possible to summarize the results. Hierarchical linear modelling (HLM) is used to statistically analyze a data structure where different hazardous parameters (level-2) are nested within the level-1 predictor variables comprising of the number of accidents occurred and the observed installations in each subsequent year. The relationship between the multilevel predictor variables were of specific interest throughout the study. Model

testing progressed in 4 phases: unconditional model, random intercepts model, mean outcome model and intercept & slope model.

The ICC of the unconditional model resulted .20. Therefore, 20% of the variance in Accidents is between-group level and 80% of variance between level-1 variables. Since, at each level, the variance had existed, it is required to add each predictor variable separately. The random intercept model was tested using the no. of observed installations each year, as predictor variable. It could be inferred that the accidents were seen prevalent in the years when the number of observed installations were also higher. It revealed that results obtained while considering the afore-mentioned level-1 predictor variable was insufficient to explain the variance in the outcome variable and henceforth, provide justification for the further endeavors. In means as outcomes model, the predictor variables at level-2 were added which explained that those variables which were significant, greatly influenced the value of outcome variable. The slope and intercept model was run in the last to test the simultaneous interaction of predictor variables with the outcome variable. The cross-level interaction demonstrated significant results revealing the main effects of the predictor variables across the installations. It is clearly understood from the figure 2, that the crane variable is seen prevalent throughout three installations type and contact variable had the most significant value in wellhead installations. Through this conception, we can further derive the vulnerability of other significant predictor variables that have influential effects on the hazardous accidents in different installations. The approach applied throughout the research provided a different perception of the hazards involved in the OG industry which significantly indicated the areas of vulnerabilities which could not have been obtained from a conventional perspective.

5. CONCLUSION

The present research has demonstrated that the present hierarchical modelling methodology applied for risk appraisal provide pluralistic approach towards optimized decisions resulting in the increased preparedness of infrastructural components in OG industry from unforeseen hazards. It is envisaged that the research has also provided a framework which will form an important module in prioritizing components of substantial hazards on different petroleum installations.

REFERENCES

- [1] Oracle White Paper - How to Reduce Costs and Manage Risk in the Upstream Oil & Gas Industry with Enterprise Project and Portfolio Management Solutions (2011). Oracle Corporation, California, USA.
- [2] Tirmizi S.T., Tirmizi S.R.U.H. (2017). GIS based risk assessment of oil and gas infrastructure in Sindh, Pakistan, *Environmental and Earth Sciences Research Journal*, Vol. 4, No. 3, pp. 55-59. DOI: [10.18280/eesrj.040301](https://doi.org/10.18280/eesrj.040301)
- [3] Funnemark E., Young E. (2003). Accident statistics for fixed offshore units on the UK Continental Shelf 1980-2001, Health and Safety Executive, Hovik, Norway.

- [4] McMahon S.M., Diez J.M. (2007). Scales of association: hierarchical linear models and the measurement of ecological systems, Vol. 10, No. 6, pp. 437-452. DOI: [10.1111/j.1461-0248.2007.01036.x](https://doi.org/10.1111/j.1461-0248.2007.01036.x)
- [5] Johnson B., Multilevel analysis in the study of crime and justice, Handbook of Quantitative Criminology, Piquero A. R., Weisburd D. L., Springer, New York, pp. 615-648. DOI: [10.1007/978-0-387-77650-7_30](https://doi.org/10.1007/978-0-387-77650-7_30)
- [6] Osborne J.W. (2000). Advantages of Hierarchical Linear Modelling, *Practical Assessment, Research & Evaluation*, Vol. 7, No. 1.