Reliability: Theory and Applications,
Special Issue № 6 (80), Volume 19, December 2024

PART-2
REDUCING THE RISKS OF CLIMATE-RELATED NATURAL DISASTERS

RT&A, Special Issue № 6 (81), Part-2,
Volume 19, December 2024

A HYBRID FORECASTING MODEL BUILDING STRATEGY

Vugar Aliyev[1], Merab Pkhovelishvili[2], Natela Archvadze[3]

•

[1]AMIR Technical Services Company, Baku, AZERBAIJAN
[2]Muskhelishvili Institute of Computational Mathematics, Georgian Technical University,
[3]Ivane Javakhishvili Tbilisi State University, GEORGIA
prof.vugar.aliyev@gmail.com
m.pkhovelishvili@gtu.ge
natela.archvadze@tsu.ge

Abstract

The article discusses the classification of existing models for predicting given events according to the statistics of all their forecasts and an algorithm for using the intersection or combination of sets of forecasts projections with other models to find the best pairs. The following terms were introduced: "necessary", "sufficient", "unnecessary" and "insufficient" forecasting models. This is discussed…this is discussed in the earthquake forecasting examples, and insufficient similarly be used to predict other events.
We explained what "necessary" and "sufficient" models are. For "unnecessary" models, an algorithm is given for how to choose a hybrid model - the intersection of two or more models that together give a forecast with a higher probability. We also discussed "sufficient" models and an algorithm for selecting such "sufficient" models, the combination of which completely covers all past events, that is, the combination of such "sufficient" models becomes "necessary".
Discusses how one can obtain a "sufficient" or "nearly sufficient" forecasting model by combining "necessary" models, and by combining "sufficient" models to obtain a "necessary" or "almost necessary" models. Also unnecessary models is discussed. In our earlier work, these models were not taken into account; such models were removed from the model database. Similarly, when considering "sufficient" models, if the forecast of the occurrence of an event is specified redundantly, such a model can be excluded from the set of "sufficient" models.
From "sufficient" models we obtain the "necessary" model, which will be both "sufficient" and "necessary" at the same time. In addition, we combine enough models to get the "necessary" model.
All of this uses forecast statistics to strategically select a hybrid model.

Keywords: classification of forecasting models, intersection and combination of forecasting models, necessary, sufficient, unnecessary, and insufficient forecasting models

I. Introduction

When modeling any processes, when it is necessary to test hypotheses about asynchronous processes, forecasting models are often used, which makes it possible to analyze their effectiveness [1,2]. We aim to create a new hybrid model based on existing models that will improve forecast accuracy. This increases the relevance of the topic under discussion.

The most popular existing forecasting model is Bayesian, mainly used to prove asynchronous hypotheses [3,4]. We aim to compare all existing forecasting models with the model we presented and thus obtain a new model (building a forecasting model with parallel data) that further improves forecasting accuracy.

In general, the prediction of asynchronous processes is often not justified because these models are mostly "necessary" and “unsufficient”. “Sufficient” models for predicting defined asynchronous processes either do not exist or are very rare.

Forecasting models can be divided into two groups: “necessary“ models and “sufficient“ models [5,6]. “Necessary” forecasting models are those models whose set of forecasts always includes, those from events that have already occurred. Such models often make incorrect predictions, but they predict every event that occurs. “Sufficient” forecasting models are those whose forecasts are always correct, but they cannot predict all events that occur.

If enough models predict that an event will occur, then that event will happen, but there may be events that they do not predict. In practice, there may be too few such “sufficient” models (for example, in earthquake forecasting) or too many (for example, in economics).

Forecasting models are characterized by a probability of success, which is not unique to statistical models [7,8]. The probability of validity of a forecasting model is the ratio of the number of events With the number of forecasts predicted by a given model, expressed as a percentage, that is, the relative frequency of occurrence (occurrence of an event), expressed in %.

The question arises of when to select the best model from among the “necessary” models, as well as when to identify “sufficient” models. The algorithm in [9] will initially analyze all existing models and identify a pair of models, a triple, a quadruple, etc. models to obtain the appropriate number of “necessary” models, the joint use of which (the intersection of models) gives the best result. Also, from “sufficient” models, a union of “sufficient” models is obtained, which is closer to predicting all occurring events.

Obviously, after each event, a situation may arise when we already have new “necessary” models, or from the old “necessary” models it turns out that some of them are no longer “needed”, that is, they are not predictable. all events I which have already occurred, the models are then reviewed again, removing this unnecessary model from their set and starting to search for new pairs.

As for sufficient models, after each event, new “sufficient” models may appear, whose predictions give better results when combined with the predictions of other models.

Let us consider algorithms for constructing hybrid models using the example of earthquake forecasting. As characteristics of each earthquake, we took the earthquake magnitude, date of occurrence, time, and name of the epicenter.

II. Developing necessary forecasting models

Note: $\mathrm{Mod}_{1,}\mathrm{Mod}_{2,}$... and so on earthquake forecasting models, which provide some predictions based on their predecessors (such as when an earthquake will occur, in what location, and with what magnitude). From these models, only the “necessary” models should be selected; this condition in the case of the earthquake problem means the following: if, for example, during T time an earthquake occurred, n only those models that predicted all these earthquakes should be considered. Let's assume the following models: $ModN_{1,}ModN_{2,}$ …, $ModN_n$ (n=5 in the following example), what each model is and what earthquake precursors it predicts are not important when considering the algorithm.

For each model, it is necessary to count the number of forecasts, and the number of justified and unjustified forecasts, and calculate the probability of justification for each model. It is clear that the sum of justified and unjustified forecasts gives the total number of forecasts. As for the probability of justification, it is calculated for each model and determines how many times an earthquake was predicted and how many times an actual earthquake occurred (see Table 1).

Table 1: *Calculation of justification probabilities for one "necessary" model*

Model	Number of forecasts	A successful number of predictions (m)	Unsuccessful number of predictions	Probability of success (%)(P)
$ModN_1$	92	2	90	2.17
$ModN_2$	80	2	78	2.50
$ModN_3$	81	2	79	2.46
$ModN_4$	97	2	95	2.06
$ModN_5$	82	2	80	2.43

In Table 1, m the number of events that occurred, and P_i indicates A_i the number of times all events predicted by the model will occur. From this table, we calculated probabilities of forecasting success for 5 models (%) .

When predicting earthquakes, the author of each model claims that his model is the best, and explains this by the fact that his model predicted all the earthquakes that occurred. Neither of them gives the number of incorrect predictions and therefore does not calculate the probability of their being correct, which are quite small numbers.

The probability of a model's prediction being justified may be quite small but for this model, there is another model that, together with the probability of success, gives us a better result. Let us show the correctness of this with our example.

At the next stage of the algorithm, we need to consider pairs of models. For each model, it is necessary to count the number of realized forecasts, and the number of justified and unjustified forecasts, and also calculate the probability of justification for each pair. Let us introduce the following notation: – $P_{i,j}$is the probability of the forecast cross-section being justified $ModN_i$ and $ModN_j$.

The following table shows these values (see Table 2).

Table 2: *Probabilities of justifying forecast intersections*

$P_i \cap P_j$	P_1	P_2	P_3	P_4	P_5
P_1		9	10	22	27
P_2	9		5	74	40
P_3	10	5		65	37
P_4	22	74	65		63
P_5	25	40	37	63	

If you specify the probability $P_{i,j}$ joint justification two models $ModN_i$ and $ModN_j$, then we obtain the following values (see Table 3):

Table 3: *Probabilities of validity of model predictions for couples*

$P_{1,2}$	$P_{1,3}$	$P_{1,4}$	$P_{1,5}$	$P_{2,3}$	$P_{2,4}$	$P_{2,5}$	$P_{3,4}$	$P_{3,5}$	$P_{4,5}$
9	10	15	25	5	74	40	65	37	63

Let us analyze the constructed table using the corresponding diagram (see **Figure 1**), from which it can be seen that the best result $P_{2,4}$ gives a combination of two models $ModN_2$ and $ModN_4$. The likelihood of their joint acquittal has already increased to 74%. Although individually these models have much lower acquittal rates of 2.50% and 2.06% than the others. In the example discussed, two or more pairs of models may show the same result. At this point, the expert must decide which one should be used.

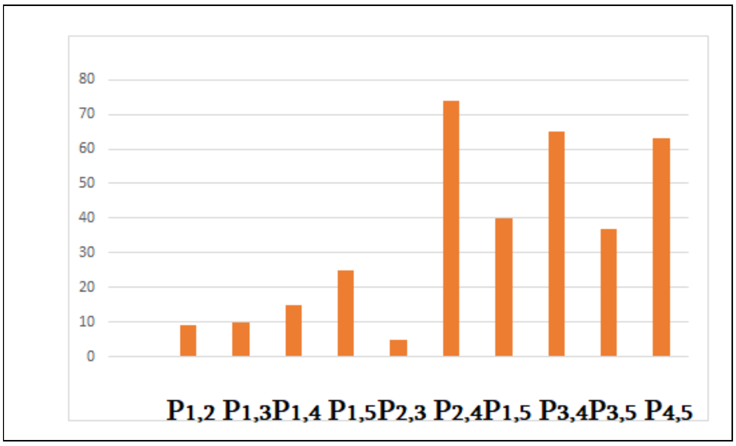

Figure 1: *Probability diagram for justifying pairs of models*

Let us analyze the constructed table using the corresponding diagram (see Figure 1), from which it can be seen that the best result $P_{2,4}$ gives a intersection of of two models $ModN_2$ and $ModN_4$. The likelihood of their joint acquittal has already increased to 74%. Although individually these models have much lower acquittal rates of 2.50% and 2.06% than the others. In the example discussed, two or more pairs of models may show the same result. At this point, the expert must decide which one should be used.

In addition to the "necessary" models, the example of earthquake prediction also included "sufficient" models, for example, there were three models: $ModS_1$, $ModS_2$ and $ModS_3$(see **Table 4**):

Table 4: *Calculation of success probabilities for individual "sufficient" models*

Model	Number of forecasts	A successful number of predictions	Unsuccessful number of predictions	Probability of success (%)
$ModS_1$	7	10	7	100
$ModS_2$	5	10	5	100
$ModS_3$	4	10	4	100

Each "sufficient" model predicted the occurrence of an event, but could not fully predict the occurrence of all events. $ModS_1$ The model predicted the occurrence of the event only in 7 cases out of 10, the second model $ModS_2$ _ðა in 5 cases and the third model $ModS_3$ – ðაin 4 cases. Even if we combine the predictions of these three "sufficient" models, their sum will be 7+5+4>10, but this does not mean that the combination of these models predicted all ten events.

Probabilities need to be calculated successfully for each pair of "sufficient" models, as in the case of "necessary" models (see **Table 5**):

Table 5: *Probabilities success of a combination of forecasts*

$P_i \cup P_j$	P_1	P_2	P_3
P_1		8	9
P_2	8		10
P_3	9	10	

This applies not only to the prediction of earthquakes but also to the prediction of any other event, both static (most often this is the task of forecasting natural disasters) and dynamic prediction, for example, economic fields [1].

It should be said that existing forecasting systems do not divide models into "necessary" models and "sufficient" models. They work with all models, which, if it does not complicate the task, then distorts the accuracy of the forecast.

III. Processing "unnecessary" forecasting models

We discussed a new approach to the forecasting process and demonstrated that forecasting is improved by using "necessary" models of forecast pairs, triplets, quadruples, etc., with a joint review [10,11]. Algorithms have also been formulated on how to determine the "necessary" models, and how to obtain from "necessary" models to „almost sufficient" models. Even for "sufficient" models it is determined how to obtain an „almost necessary" model from them.

Theoretically, it seems that one of the pairs of models may be a "unnecessary model", and the other also a "unnecessary" model, but it is also possible to find a pair (three) that will be close to the "sufficient" model. Let's consider algorithms for processing "unnecessary model " forecasts.

Let's say we are considering a case where an event (for example, an earthquake) occurred several times (for example, 5) and there are three models for predicting this event: *Model I* (I), *Model II* (II).) and *Model III* (III). It is necessary that the model predicts The model must predict at least one…at least one of the events and the set of forecasts has an intersection with the set of forecasts of other forecasting models. *Model I* predicted 3 events, *Model II* predicted 4 events, and *Model III* predicted 1 event (see **Figure 2**). The circles in the figure indicate the events that occurred, and the numbers indicate the names of the models that predicted the occurrence of this event:

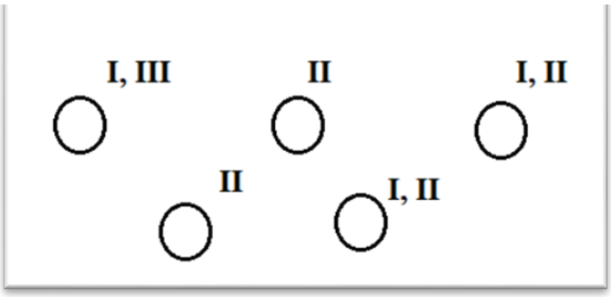

Figure 2: *Forecasting events by forecasting events*

Let us consider the set of events predicted and not predicted by each model, in the case of three models (see Figure 3). Events that these models could not predict are indicated by dots, events that were predicted by these models are indicated by circles.

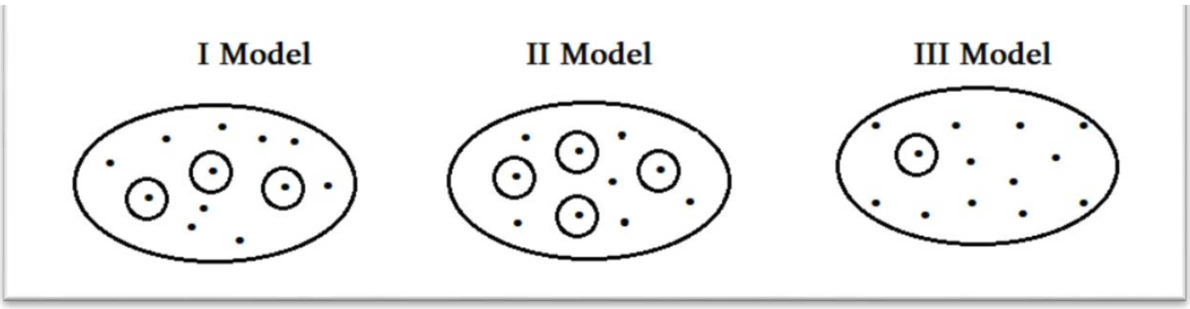

Figure 3: *"Unnecessary" models.*

Let us denote P_i by P_j and N not obligatory models forecasts and consider their intersection P_{ij} (see Figure 4):

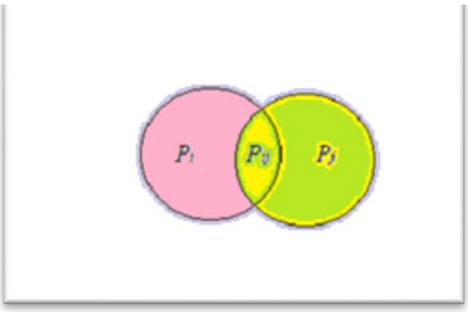

Figure 4: *Intersection of predictions of two models*

If we consider the model's predictions as a possible intersection of sets of events, then three cases can arise here (see Figure 5): when the intersection is greater than the events that have already occurred, when the intersection is smaller, and when they coincide. The event set in the image is marked in blue, and the events in the intersection set are marked in yellow.

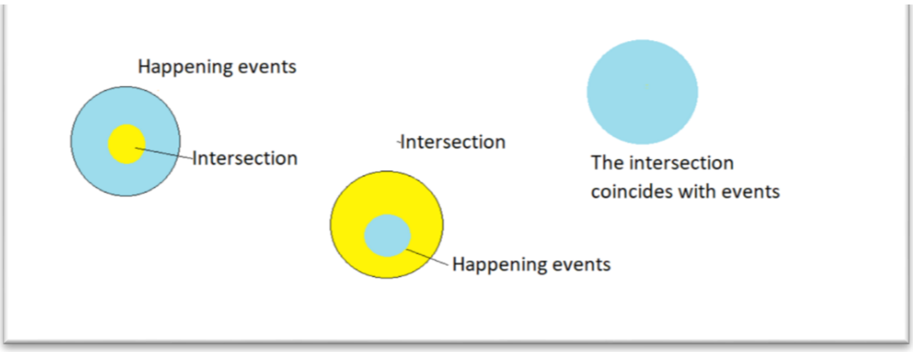

Figure 5: *Occurring events and forecasts of unnecessary models*

If the intersection of models coincides with the set of already existing events, this is the best case, since the set of models involved in the intersection becomes a "sufficient" model for prediction.

Let's look at an example. For each model, count the number of forecasts made, the number of events, and the number of justified and unjustified forecasts, and calculate the probability of justification for each model (see Table 6).

Table 6: *Calculation of justification probabilities for individual "unnecessary" models*

Model	Number of forecasts	A successful number of predictions (m)	Number of events that occurred	Unsuccessful number of predictions	Probability of success (%)(P)
$ModNN_1$	92	3	5	2	3.2
$ModNN_2$	80	4	5	1	5.0
$ModNN_3$	81	1	5	4	1.2

As the next step of the algorithm, we need to consider pairs of models. For each model, it is necessary to count the number of realized forecasts, the number of justified and unjustified forecasts, and also calculate the probability of justification for each pair. Let us introduce the following notation – $P_{i,j}$ $P_{i,j}$ this is the probability of the forecast intersection being justified $ModN_i$ and $ModN_j$. The following table shows these values (see **Table 7**):

Table 7: *Probabilities success of a combination of forecasts in the case of " unnecessary" models .*

$P_i \cap P_j$	P_1	P_2	P_3
P_1		5	6
P_2	5		3
P_3	6	3	

From Table 7 we can conclude that the intersection of two unnecessary models leads to an "almost sufficient" model.

This way we can use "unnecessary" models if we select "unnecessary" models from models of forecasting with a certain filter. It is possible to use "unnecessary" models in some positions, because the intersection of unnecessary models can give us "almost sufficient" models.

IV. Discussion

We explained what "necessary" and "sufficient" models are. For "unnecessary " models, an algorithm is given for how to choose a hybrid model - the intersection of two or more models that together give a forecast with a higher probability. We also discussed "sufficient" models and an algorithm for selecting such "sufficient" models, the combination of which completely covers all past events, that is, the combination of such "sufficient" models becomes "necessary".

Discusses how one can obtain a "sufficient" or "nearly sufficient" forecasting model by combining "necessary" models, and by combining "sufficient" models to obtain a "necessary" or "almost necessary" model.

In our earlier work, these models were not taken into account; such models were removed from the model database. Similarly, when considering "sufficient" models, if the forecast of the occurrence of an event is specified redundantly, such a model can be excluded from the set of "sufficient" models.

From "sufficient" models we obtain the "necessary" model, which will be both "sufficient" and "necessary" at the same time. In addition, we combine enough models to get the "necessary" model.

All of this uses forecast statistics to strategically select a hybrid model.

References

[1] Yanchenko, A.K., Tierney G, Lawson J. "Multivariate dynamic modeling for Bayesian forecasting of business revenue." - Models in Business, 2023.

[2] Chelidze T., G.Melikadze, T.Kiria, T.Jimsheladze, G.Kobzev. Statistical and Non-linear Dynamics Methods of Earthquake Forecast: Application in the Caucasus. Frontiers in Earth Science, 2020, https://doi.org/10.3389/feart.2020.00194.

[3] Bansal, N. K., Hamedani, G. G. & Maadooliat, M. (2015). Testing Multiple Hypotheses with Skewed Alternatives. Biometrics, 72(2): 494-502.

[4] Huynh C. (2014) Estimation of Type III Error and Power for Directional TwoTailed Tests Using PROC POWER. Ph.D., Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada.

[5] Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. Reducing risks through improvement of preduction models. Theory and Applications. 2022, 17, pp. 197–202.

[6] Prangishvili, A., Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. Predicting Events by Analyzing the Results of the Work of Predictive Models. 2022, 1562 CCIS, pp. 64–78.

[7] Phkhovelishvili, M., Archvadze, N., Gasitashvili, Z. Peculiarities of the Hybrid Model Built Using Parallel Data. Proceedings - 2022 7th International Conference on Mathematics and Computers in Sciences and Industry, MCSI, 2022, pp. 66–70.

[8] Aliyev, V., Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. Algorithm for building a hybrid mode of the existing risk model. Reliability: Theory and Applications. 2022, 17, pp. 106–110.

[9] Pkhovelishvili, M., Archvadze, N., Shetsiruli, L. Optimization of Expert Solutions for Events Forecasting.

[10] Proceedings - 2023 International Conference on Control, Artificial Intelligence, Robotics and Optimization, ICCAIRO 2023, 2023, pp. 82–85.

[11] Aliyev, V., Pkhovelishvili, M., Archvadze, N., Gasitashvili, Z., Tsiramua, Z. Hybrid Prediction Method Using Experts and Models. 2023 5th International Conference on Problems of Cybernetics and Informatics, PCI 2023, 2023.