New attitude to aviation forecast verification and incurred challenges

David Sládek

Annotation

Příspěvek je shrnutím a prezentací výsledků disertační práce obhájené v srpnu 2021. Hlavním tématem je metoda verifikace leteckých kódovaných předpovědí. Článek shrnuje nejpodstatnější závěry s odkazy na již publikované výstupy. Práce zahrnovala analýzu již publikovaných přístupů použitých u nás I v zahraničí, rozbor platných předpisů a odhalení některých obtížně interpretovatelných míst. Práce a dříve publikované výsledky též odhalují některé části předpisu, které si navzájem protiřečí. Prostřednictvím odkazů na publikované příspěvky práce komentuje analýzu spolehlivosti vydávaných předpovědí TREND. V závěru představuje vlastní navrženou softwarovou aplikaci v modulu Meteochart Visual Weather, která má za úkol analyzovat pozorování a předpovědi na leteckých stanicích a okamžitě indikovat předpověď nevyhovující přesnosti.

This paper is a summary and presentation of the results of a dissertation thesis presented in August 2021. The main topic is a method for verification of aviation coded forecasts. The paper summarizes the most important conclusions with references to already published results. The work included an analysis of already published approaches used at home and abroad, analysis of current regulations and detection of some difficult to interpret points. The work and previously published results also reveal some parts of the regulations that contradict each other. Through references to published papers, the thesis comments on the analysis of the reliability of the TREND forecasts issued. Finally, it presents a self-proposed software application in the Meteochart Visual Weather module to analyze observations and forecasts at aviation stations and immediately indicate a forecast of unsatisfactory accuracy.

Keywords

ICAO Annex 3, L3 Meteorologie, METAR, TAF, TREND, verification, IBL Visual Weather, Meteo Chart

1 Introduction

Checking and verifying the accuracy and quality of the forecast is an integral part of the meteorologist's job as a forecast maker. The conceptual work of the Hydrometeorological Service should also be directed towards improving the quality of forecasts. Not only is continuous monitoring of forecasts required by current regulations, but the results of quality testing can serve as feedback to forecasters. This applies to quality characterised by the accuracy, formal correctness or unambiguity of the forecast being checked.

For the user, the determination of the quality of the forecast can be an indicator of the professionalism of the service provided, for the supervisors of the forecasters a measure of the quality of the performance of the subordinates. For the forecasters, a sophisticated quality control method can highlight problematic weather situations in a retrospective evaluation. When evaluated in real time, the evaluation can completely prevent the validity of poorly accurate forecasts.

This paper offers a perspective on some of the problems with the interpretation of mandatory regulations, an evaluation of some of the methodologies used, a proposal for its own methodology for evaluating aircoded forecasts, and a discussion of future research opportunities in the area.

2 Directive and reviewed methods

In the Czech Republic and in general in states committed to ICAO standards, the issuance of forecasts and the required accuracy of forecasts and observations are mainly governed by the ICAO Annex 3 standard and its national legal implementation. In the Czech Republic this implementation is the Meteorology L3 regulation [1]. The second common material for the provision of meteorological services is also the agreement between the meteorological service provider and the operator. This is embodied, for example, in the form of an amendment or part of the airport regulations, which is based on the needs of the primary users at each airport.

There are other regulations for the production and issue of weather reports and forecasts that are followed in practice. These are, for example, the WMO No. 306 Manual on Codes [2], or the L8400 [3], to which the Annex 3 and L3 regulations refer in some parts. However, the content of the documents does not bring significant changes to the topic of forecast verification or quality control compared to the Annex 3 rules. Rather, they provide clarification of contentious situations or technical code specifications when compiling forecasts or observation reports.

Several papers have been published in the field, some [4] are only usable in theoretical terms and most methodologies do not adhere to ICAO criteria. This can be considered problematic given the use in flight operations.

1.1 Directive review

One example of discrepancy in the directive is the change in visibility along with the use of the mist phenomenon (BR). For this combination, three cases can occur as shown in the figure (Figure 1).

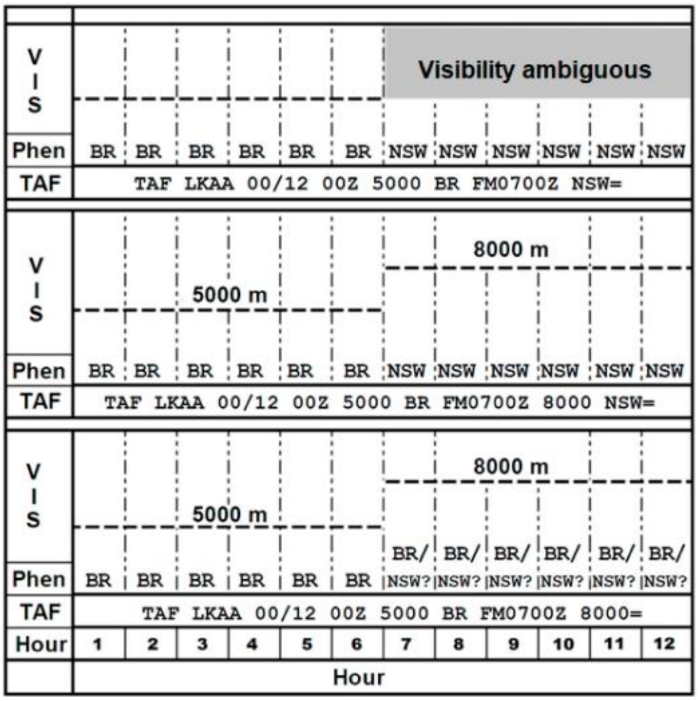

Figure 1: Ambiguous application of the change in visibility and classification of mist in conformity with Annex 3 [5]

In the first example, without a visibility value included, the values are ambiguous and the visibility cannot be determined. Such an error is of course unacceptable. In the second case, although the visibility value is present, the occurrence of the phenomenon is terminated with the abbreviation NSW (No Significant Weather), but the conditions for the inclusion of change groups are not met. The third case indicates a change in visibility, but the phenomenon is not terminated with the appropriate abbreviation. For the last two cases the error is not so serious, but it does provide potential for ambiguity.

An obvious example of a discrepancy between the accuracy requirements and the recommended thresholds for issuing a correction or change group may be the case of visibility. Regulation L3, Appendix 5, Section 1.3.2 recommends thresholds for the assignment of a change group or correction to a forecast for horizontal visibility as follows: 150, 350, 600, 800, 1500, 3000 and 5000 m, based on the ICAO Annex 3 [7]. In contrast, the required accuracy up to a visibility value of 800 m is ± 200 m, and above 800 m an error of up to 30 % is allowed, as mentioned in the directive [7]. The exact example of TAF is in previously published articles [5],[8].

1.2 Other methods review

Of the methods evaluated, the following three are the most suitable for use:

(a) Austrocontrol
(b) Sharpe et. al.
(c) IBL Slovakia

Austrocontrol

Austrocontrol is the Austrian governmental organisation responsible for air traffic control. It currently provides the option of purchasing software or licensed access to a web-based application to verify TAF forecasts. The main principles of the methodology used in the software were published by G. Mahringer in 2008 [6].

The advantages of the method are its very detailed statistical apparatus and its successful practical application. However, the actual detail of the results can be presented in a rather complex manner. The file available online as a demo output for those interested in the analysis shows that for one of the analysed elements (wind direction, wind speed, visibility, phenomena, cloud cover), the user receives a set of eleven tables of results, partial calculations and calculations in the form of a table for each hour. Thus, for five elements and one 24-hour TAF, there may be more than 150 tables.

The method is commercially used, but it is based on what appears to be a subjective basis. In the published work [6], the author does not fully explain the reasons for the choice of the set of meteorological phenomena entering the verification process. However, in terms of translatability into an algorithm, it is clear that strictly following the regulations would be difficult to say the least.

Sharpe et al.

The main difference between RMetS studies and, for example, Mahringer's method is the combination of probabilistic and deterministic approaches. The bivariate analysis for change groups adds a third dimension, which is the probability dimension. The basic idea of both Harris' and Sharpe's study is that not all values of the interval covered by the predictions are of equal weight. Quite logically, therefore, the methods divide the predicted values that are valid at the same time and the probability with which they occur.

In particular, Sharpe develops the method by defining deterministic groups, especially TEMPO, where he defines exactly how many times with what probability the conditions will occur. The paper is concerned only with evaluating visibility and other elements are not addressed. Thus, it is not even clear whether the proposed approach would be suitable for other elements of the forecast. The paper also does not offer a solution to the situation where short-term values occur multiple times in one interval (between observation dates).

It is also evident from the paper [4] that Sharpe compares his methodology with Mahringer's. He refers to the case of his own interpretation and shows the biggest weakness of Mahringer's approach. This, according to him, lies in the treatment of change groups and the use of only boundary conditions. However, on close examination of both methods, one cannot help but feel that Sharpe is using Mahringer's method contrary to the author's intention. Because of the misinterpretation, Sharpe arrived at the wrong intention and his method does not achieve significantly different results.

IBL Slovakia

IBL, whose software is used by both the Czech Army and the Czech Hydrometeorological Institute, offers in its Aero Weather product continuous monitoring of forecasts issued at selected airports and checks them against issued METAR and SPECI reports. It alerts the user if the criteria for the recommended issuance of a correction are exceeded or if dangerous meteorological phenomena occur. Alerts are issued via the application's main dashboard and the criteria can be modified when requested directly by the user.
Another option of the Aero Weather program is monthly evaluation and success rates for specified parameters. It also allows the accuracy to be evaluated for individual personnel issuing TAF forecasts [9].
The manual provided by IBL itself shows that the software offers full user orientation to the methodology and criteria. The Verification method can be selected from *Ranges*, i.e. intervals defined by regulations or by the user, or *Tolerance*, where the default values are set according to Annex 3, but again user-definable.
The software tools offered by the company also offer the possibility to be alerted when the values for issuing alerts are predicted in the TAF. While the interconnectedness of these two products is overlooked, it is true that it is useful to also alert on potential alert service activity in the case of forecast generation.

Conclusion on methods

Based on a detailed research and comparison of already known approaches to the evaluation of coded aviation forecasts, it is clear that in many cases the authors have simplified some sub-problems or subjectivized the evaluation method. In order to establish an objective procedure, one rather administrative and several other technical requirements for the designed method need to be identified:

1) Comply with the requirements of ICAO Annex 3 or L3 standards and ICAO 9783 Manual on Quality [2].
2) Establish criteria for the evaluation of individual elements where mandatory provisions are ambiguous or insufficient.
3) Establish a procedure for the evaluation of concurrent change groups.
4) Use appropriate software and statistical tools to evaluate TAF and TREND predictions.
5) Propose a way to interpret the results and reduce the complexity of the evaluation outputs.

By formulating these requirements, we are now able to create a fully functioning application.

2 Development of new TAF verification method and tools

During the development of the application it was found through basic statistical indicators that the TREND forecast is neglected by meteorologists in the Czech Republic. Therefore, it was decided that the TAF forecast would be the primary controlled forecast.

In addition to the need to apply ICAO and WMO standards, it is necessary to define application requirements that should benefit both the user and the forecaster. Based on the previous chapters and considerations, the requirements for the final and practicable product have been defined. The main attributes should be:

1) minimum time devoted to the verification process,
2) immediate feedback,
3) elimination of inaccurate predictions at the time of validation,
4) warning of potentially dangerous inaccuracies in the forecast,
5) the possibility of continuous monitoring and control,
6) implementation in systems already in use (IBL VW, IBS...).

As part of the research, two working Python web applications were also created to highlight the problems of backwards evaluation.

The disadvantage of web applications is a certain amount of time. The meteorologist has to turn on the application, perform the evaluation, back up the results to the database and perform the evaluation for a longer period of time at a set interval, e.g. once a month. This comes with another drawback of this approach, namely the time lag of the evaluation. It is justifiable that the forecaster should be confronted with the verification immediately and the origin of the error analysed. This is because, at the time of the forecast's validity, he or she has the best overview of the situation and can examine in detail the influences leading to the inaccuracies. Hindsight has advantages especially in terms of statistical outputs, where a review would reveal, for example, the tendency of meteorologists to be too 'optimistic' or 'hedging'. This means that the meteorologist sets too wide intervals in the forecast in order to increase the probability of a successfully predicted value. Immediate warning of an incorrect forecast value adds the major benefit that the precise use of such a product will completely prevent the occurrence of valid inaccurate forecasts, because such forecasts will be corrected immediately. This has not always been the case, as shown in practice and published papers, e.g. [5],[8],[10].

2.1 Criteria

The only two possible criteria were those mentioned in the ICAO Annex 3 Code, which are the Accuracy Requirements in Attachment B or the intervals for the inclusion of change groups and TAF AMD.

The main disadvantage of fixed intervals is that even a very accurate value may not be evaluated as successful if it is on the opposite side of the threshold. For this case, tolerances calculated from the observed value are preferably more appropriate.

Element to Be Forecasted	Operationally Desirable Accuracy of Forecasts	Minimum Percentage of Cases within Range
Wind direction	±20°	80%
Wind speed	±5 kt	80%
Visibility	±200 m up to 800 m ±30% between 0.8 and 10 km	80%
Visibility	Occurrence or nonoccurrence	80%
Cloud amount	One category below 450 m (1500 ft) Occurrence or nonoccurrence of BKN or OVC between 450 m (1500 ft) and 3000 m (10,000 ft)	70%
Cloud height	±100 ft up to 1000 ft ±30% between 1000 ft and 10,000 ft	70%
Air temperature	±1 °C	70%

Figure 2: Operationally Desirable Accuracy of Forecasts according to Annex 3 [5]

Despite the apparent difficulties in interpreting the criteria from the regulation, the values are the only ones applicable under strict conformity with ICAO regulations. They are not dependent on the terms of local agreements between the service provider and users, such as SPECI or TAF AMD criteria.

2.2 User interface

The application of the specified criteria is done through algorithms consisting of built-in Kernel functions in the Meteochart module of the IBL Visual Weather software. This is more of a display module and the user has no significant input into it unless they want to change the functions and cell values. If the values of more complex multiple nested functions are returned in individual cells, modifying them without proper knowledge is not recommended. The built-in functions are not designed to work with cycles, and therefore some of the functions are very extensive and completely unintuitive for the uninformed user.

In the application, the user will find primarily an evaluation of:
1) wind direction
2) wind speed,
3) prevailing visibility,
4) phenomena,
5) amount and height of the cloud base.
6) As additional series, a check of colour codes and flight categories for instrument and visual (IMC/VMC) flights and a display of wind gusts in METAR and predicted in TAF have been included.

The following figure (Figure shows the evaluation of the TAF forecast for the station Náměšť' (ICAO: LKNA). The colour interpretation is intuitive and the red colour immediately catches the attention of the employee.

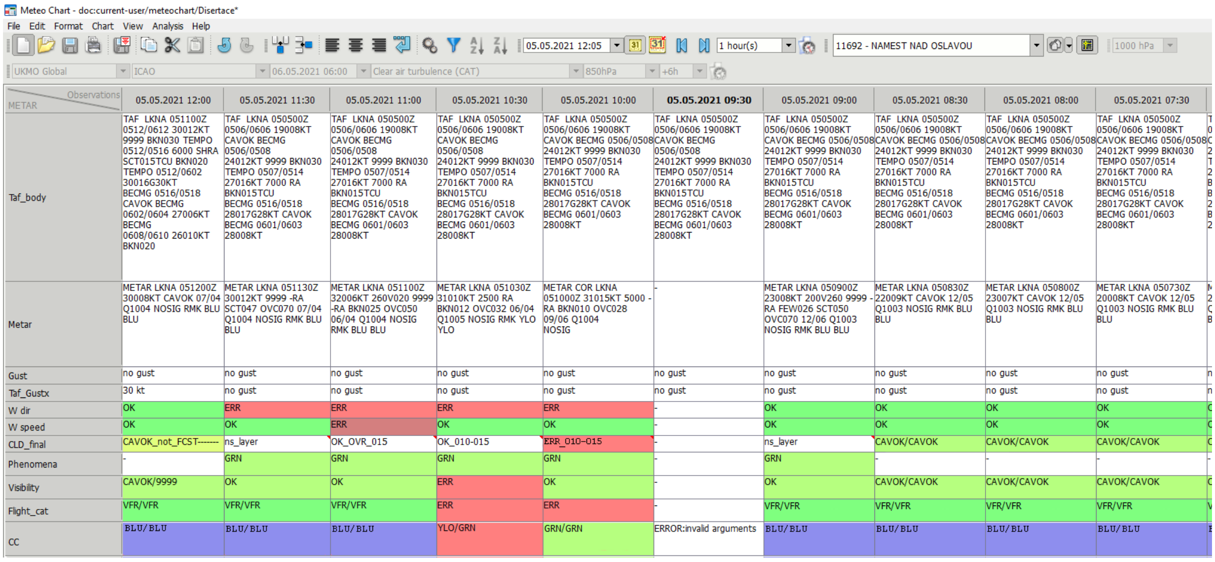

Figure 3: User interface of the verification application for Namest airbase (ICAO: LKNA) of 5.5.2021

On the same day, the forecast success rate was slightly lower at Tallinn Airport in Estonia. The meteorologist did not respond to the inadequate wind direction forecast, nor to the wind speed on several dates (9:00, 10:00, 11:30 UTC). On the last two dates, the cloud cover forecast also failed.

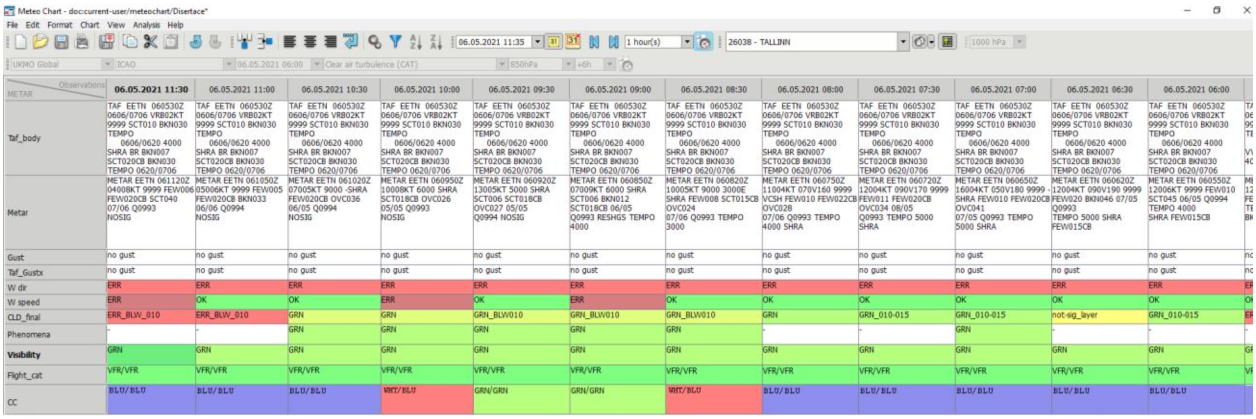

Figure 4: User interface of the verification application for Tallinn airport (ICAO: EETN) of 5.5.2021

The assessment for Tallinn Airport of 5/05/2021 was not very favourable. The wind direction is mostly assessed with an error value because a variable direction "VRB" is predicted. Similarly, the tolerance is not met by the wind speed in the last few dates. However, from experience and practice, it must be added in defence of the EETN service that cloud cover in FEW coverage and wind direction at such low speeds are often not assessed. However, this makes it possible to illustrate how a subjective evaluation can be in direct and very frequent contradiction with the applicable regulations.

3 Discussion

The main, practically usable, output is an application in the Meteochart module of the IBL Visual Weather software. It will be tested in practice directly by meteorologists and according to their feedback modified into the final form. The output in the form of a table applies the proposed algorithms, respects the established criteria and displays the returned values together with colour symbols. The intent is also to create a message to warn of an inaccurate forecast or to warn of exceeding the criteria for an AMD type correction.

An unresolved issue that has arisen during the course of the solution is the ability to compare verification results between stations and meteorologists. Even a completely objective methodology will not ensure complete objectivity in the comparison because different weather situations manifest themselves differently and the course of their elements varies. Therefore, further research could additionally objectify the assessment of the complexity of the conditions predicted by the meteorologist. The verification method could also serve as a basis for research on verifying the quality of automatically generated forecasts.

Another question that has arisen in the course of this paper focuses on TREND forecasts. Considering the fact that in the Czech Republic they are published basically in the form of "persistent forecasts", where meteorologists pay minimal attention to the issuing of forecasts, the question is whether some simple algorithm using the simplest statistical methods could not automatically generate TREND forecasts and at the same time be more accurate than persistent forecasting.

3 References

[1] *L3 Meteorologie*. Change no. 5. Prague: Civil avation authority, Czech republic, 2020. Č. J.: 584/2008-220-SP/4.

[2] *Manual on the Quality Management System for the Provision of Meteorological Service for International Air Navigation: Doc 9873*. In: . International Civil Aviation Organization, 2010, číslo 2.

[3] *Zkratky a kódy: L8400*. Change no. 33. Civil avation authority, Czech republic, 2019. Č.J. 710/2007-220-P/2.

[4] SHARPE, M., BYSOUTH a TRUEMAN. Towards an improved analysis of Terminal Aerodrome Forecasts. *Meteorological Applications* [online]. 2016, 23(4), 698-704. DOI: 10.1002/met.1593. ISSN 13504827. Available from: http://doi.wiley.com/10.1002/met.1593

[5] NOVOTNÝ, J., DEJMAL, K., RÉPAL, V., GERA, M.,SLADEK, D. Assessment of TAF, METAR, and SPECI Reports Based on ICAO ANNEX 3 Regulation. *Atmosphere* [online]. 22 January 2021. Vol. 12, no. 2, p. 138. DOI 10.3390/atmos12020138. Available from: http://dx.doi.org/10.3390/atmos12020138

[6] MAHRINGER, G. Terminal aerodrome forecast verification in Austro Control using time windows and ranges of forecast conditions. *Meteorological Applications* [online]. 2008, 15(1), 113-123. DOI: 10.1002/met.62. ISSN 13504827. Available from: http://doi.wiley.com/10.1002/met.62

[7] International Civil Aviation Organization (ICAO). 2010. *Meteorological Services for International Air Navigation. Annex 3 to the Convention on International Civil Aviation*, 15th edn. ICAO: Montreal.

[8] SLÁDEK, D. Attitudes Comparison of TAF Forecast Quality Assessment. In: *2019 International Conference on Military Technologies (ICMT)* [online]. IEEE, 2019, 2019, s. 1-7. ISBN 978-1-7281-4593-8. Available from: doi:10.1109/MILTECHS.2019.8870081

[9] Aviation forecasting made easy. *IBL* [online]. 2019 [cit. 2020-08-25]. Available from: https://www.iblsoft.com/products/visualweather/aviation-forecasting-made-easy/

[10] SLÁDEK, David. Weather phenomena and cloudiness accuracy assessment in TAF forecasts. In: Kolar P. 2021 *8th International Conference on Military Technologies,* ICMT 2021 - Proceedings. Brno, Czech Republic: Institute of Electrical and Electronics Engineers Inc., 2021, p. 1-6. ISBN 978-1-6654-3724-0.