264 TESTING OF A NEW GRAVITY WAVE BASED CLEAR-AIR TURBULENCE DIAGNOSTIC

Jacek M. Kopeć*
Institute of Geophysics, Faculty of Physics, University of Warsaw
Interdisciplinary Centre of Computational and Mathematical Modelling, Univerisity of Warsaw

Krzystof E. Haman
Institute of Geophysics, Faculty of Physics, University of Warsaw

Konrad Bajer
Institute of Geophysics, Faculty of Physics, University of Warsaw
Interdisciplinary Centre of Computational and Mathematical Modelling, Univerisity of Warsaw

1 Introduction

The coincidence between gravity waves (GW) and clear-air turbulence (CAT) has been evident for a long time (e.g. Bekofske and Liu, 1972). In recent years due to better data availability and improving understanding of GW dynamics possibilities of predicting CAT induced by GW are being researched (e.g. Kopec et al., 2011; Sharman et al., 2012; Knox et al., 2008). In this work we investigate application of an approach based on the work of Haman (1962). We assume that at some low altitude (e.g. top of the boundary layer) shallow convective motions or other processes generate small wave-like vertical displacement of the air. The idea behind our method is that in certain conditions the amplitude of this initially small disturbance will grow with altitude, until linear approximation ceases to be valid, and then break up generating turbulence. The equation governing vertical dependence of the vertical displacement of an air parcel due to a monochromatic GW characterized by certain wavenumber k and phase speed ϵ is of the form (Haman, 1962):

$$s''+\left(\frac{u'}{u-\epsilon}-\frac{\mathrm{g}}{\mathrm{c}^2}-\beta\right)s'+\left(\frac{\mathrm{g}\beta}{(u-\epsilon)^2}-\mathrm{k}^2\right)s=0 \quad (1)$$

Here z denotes altitude, ' denotes d/dz, $s(z)$ is the wave amplitude (vertical profile of vertical displacement from the mean flow), k= $|\boldsymbol{k}|$ is the length of the wave vector $\boldsymbol{k}$, $u = \mathrm{k}^{-1}\boldsymbol{u}\cdot\boldsymbol{k}$ is the wind speed component in the direction $\boldsymbol{k}$, ω is the angular frequency, $\epsilon = \frac{\omega}{\mathrm{k}}$ is the phase velocity of the wave characterized by $\boldsymbol{k}$ and ω, g is gravity acceleration, $\Gamma = (g/c_p)$ is the adiabatic lapse rate, $\beta = \frac{\Gamma+T'}{T}$ is the stability of the air, T is temperature, and c is the speed of sound. Of course real life disturbances are far from being a monochromatic waves and are rather linear combinations of infinite number of such waves. Yet the equation 1 shows that atmospheric profiles can be understood as a filter in case of gravity waves. Some parts of the spectrum are damped whilst other are amplified. To gain statistical knowledge about this filtering in certain background conditions we solve many Cauchy problems for equation 1. The parameters k and ϵ are chosen in such manner that ensures uniform and isotropic probing of a spectrum potentially associated with shallow convection clouds. That is wavelegths ranging from 500m up to 2000m and angular frequency ranging from -0.1rad/s to 0.1rad/s which correspond to typical spatial and temporal scales of shallow convection.

The last part of the necessary theory is the breaking criterion. Following Kopec et al. (2011) $\mathrm{K} = \left|\frac{u-\epsilon}{u}\right|((2\pi)^{-1}\mathrm{k}|s| + |s'|)$ is the ratio of the nonlinear terms magnitude to the magnitude of linear terms in the momentum equation used to obtain equation 1. Note that this quantity is only a rough estimate where only leading order terms were taken into account. We assume that linearization is no longer valid when K≥ 1. Moreover we will assume that the wave breaks into turbulence at the level this occurs and thus is not present above. It is removed from the spectrum. We will also assume that when a critical level is reached the same happens. It needs to be mentioned that this assumption is a simplification significantly different than the approach presented in Knox et al. (2008); McCann (2001). Yet, for simplicity we will hold this to be valid.

* *Corresponding author address:* Jacek Kopeć, Instytut Geofizyki, Wydział Fizyki, Uniwersytet Warszawski, ul. Pasteura 7, 02-093 Warszawa, Polska, email: jkopec@icm.edu.pl

In summary, our analysis assumes that:

1. At the top of the boundary layer a uniform and isotropic spectrum of disturbances is present.
2. Evolution of vertical displacement with height is given by equation 1.
3. The wave breaks into turbulence at the level where either it exceeds the linearity criterion K$< \ 1$ or it encounters a critical level, whichever is lower. Such wave is not present above - no secondary waves or crest turbulence are considered.

Accepting above assumptions a set of 1890 Cauchy problems is then solved for each atmospheric profile we are interested in (for a numerical weather model grid we would need to solve such set for each grid point to obtain a turbulence forecast). The result is a set of 1890 wave breaking heights. To convert this information into a single scalar, which would be related to CAT occurrence in Kopec et al. (2011) we simply calculated a density of broken waves per meter of altitude. In this paper we will use similar measure $N_{\Delta}(z)$ being the number of breaking waves in a layer of thickness 2Δ centered around an altitude z.

2 The data used for verification

The observation dataset used to assess the prediction efficiency of the method described above consisted of a set of 4011 AMDAR messages covering Europe in January, February and March 2010. Here AMDAR stands for Aircraft Meteorological Data Relay being a standard of automated messages generated by commercial aircraft during flight. Among measured background meteorolgical parameters AMDARs consist information about turbulence in the form of IT (graded on board accelerometer reading) and Derived Equivalent Vertical Gust (DEVG) (ARL, 1985) (for more information see WMO (2003)). The AMDARs were chosen so that each contains valid IT and DEVG records and was generated during cruise flight phase above 8500m amsl. The last two criteria serve to eliminate erroneous accelerometer reading during manouvers and to avoid cloud turbulence measurements (since no relative humidity reading was available). The geographical distribution of the data is illustrated in Fig. 1. Table 1 shows severity distribution of the observations. The IT and the DEVG records agree up to 30 cases of light turbulence (DEVG) or no turbulence (IT) therefore in the rest of this paper we will use IT. The background atmospheric profiles were provided by COAMPS model run operationally at Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw. The horizontal resolution of the forecast is 39km with 30 vertical levels.

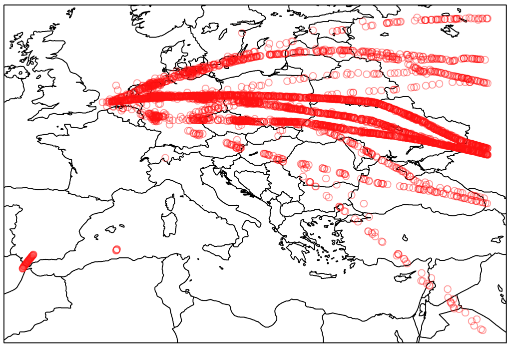

Figure 1: Geographical distribution of the observation set used for validation

3 Prediction test

The first question one would ask is which interval Δ is appropriate. Although the intuitive approach would be choosing a relatively small Δ since if the data were precise and the method at hand would be accurate in predicting turbulence the smallest reasonable interval would be the optimal solution. However our background data are gridded with vertical grid spacing changing with altitude and localization (terrain following vertical coordinates). To answer what is the preferred interval, for each of the 4011 sets of breaking heights a collection of N_{Δ} indices was calculated for 180 Δ intervals equally distributed in the range $(45m, 8995m)$ and centered at the altitudes reported in respective AMDARs. As an indicator of performance of such an index, the area under the ROC curve (AUC) was used. AUC values range from 0 to 1 with 0.5 meaning the index performance is as good as any random measure (ie. it bears no relevant information), 1 meaning the criterion is perfect and values lesser than 0.5 mean that the index is indicating inverted values (that is it bears relevant information but positive

Turbulence intensity	Number of observations (IT)	Number of observations (DEVG)
NOTURB	3730	3760
LIG	195	165
MOD	17	17
SEV	69	69

Table 1: Number of observations in each turbulence severity class (NO TURBulence, LIGht turbulence, MODerate turbulence and SEVere turbulence) according to IT and DEVG records

responses correspond to lower values). Since ROC curves are constructed for binary criteria we have tested the predictors for detection of moderate or greater (MOG) turbulence. The

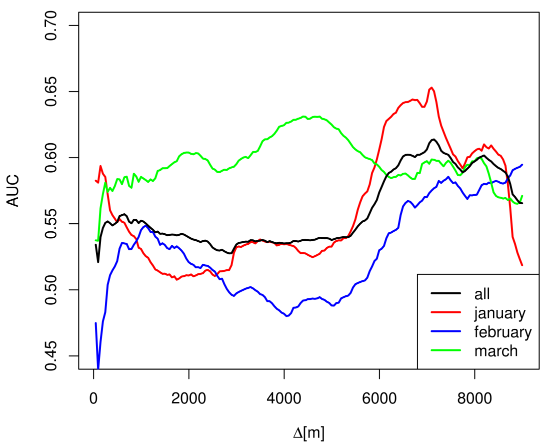

Figure 2: AUC for set of 180 values of Δ uniformly distributed in the interval $(45m, 8995m)$ for all data and separately for each of the three months

results of the monthly validation are illustrated in Fig. 2. One feature that is quite surprising is strong temporal variability in displayed prediction skill - for each month there is at least one N_Δ with AUC $\geq$ 0.6 but they are characterized by different interval values. The general skill distribution is very different for each of the three studied months. Even greater variability can be seen in Fig. 3. The other feature which is also unexpected and is visible in the three month evaluation is that indices corresponding to large Δ (approximately 6000m to 8000m) display significantly better skill than the ones corresponding to the lower values. This behaviour is visible in January and February validations. In March N_Δ characterized by highest AUC mostly are those with interval around 4900m. In general for all observations the index with the best skill

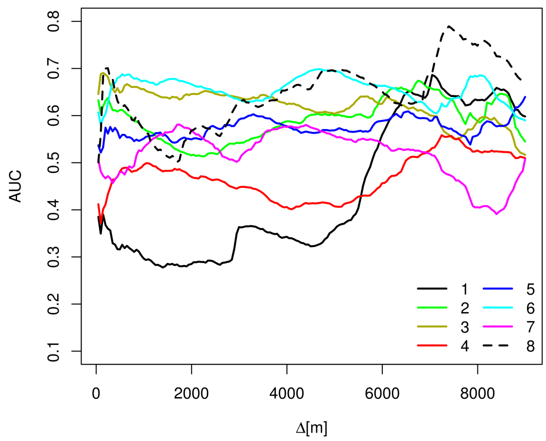

Figure 3: AUC for set of 180 values of Δ uniformly distributed in the interval $(45m, 8995m)$ for 11 day periods ranging from 01.01.2010 to 31.03.2010. The numbers in the legend denote the order of the periods.

(AUC=0.614) is the one with Δ=7145m.

These results show that this index must be used with care. The AUC associated with each N_Δ tends to approximately 0.55 (or 0.60 for larger Δ) for longer validation periods. However for shorter periods (especially see in 11-day periods verification in Fig. 3) the results vary strongly but for 5 out of 8 11-day verification periods AUC $\geq$ 0.673 and in cases where AUC was never larger than 0.600 minimum AUC values were smaller than 0.400. This leads to three conclusions. First, that in most cases there exists such interval length Δ that N_Δ bears information significantly related to the observed turbulence. Yet, it is not established whether the skill patterns as functions of Δ can be forseen, thus rendering N_Δ hard to use operationally. Second, that since the method originated from the idea that shallow convection excites the GW the N_Δ index may be meaningless when shallow convection is absent (which we do not know *a priori*). Third, that as a standalone index N_Δ is hard to use but it could be incorporated into predictors using a set of turbulence indices (e.g. Sharman et al., 2006).

4 Random forest training and performance analysis

Results shown in Fig. 2 and 3 suggest that N_Δ for different values of Δ could bear different information. A question arises whether it is so and could many N_Δ indices be combined into one index which is more reliable. As a possible solution for this problem we have used a random forest algorithm (Breiman, 2001) in the version ported to the R language (Liaw and Wiener, 2002). As the training period we have chosen January. There were in total 1416 observations in that month out of which 1332 were non-turbulent, 51 were light turbulence and 33 MOG turbulence (including only 3 cases of moderate turbulence encounters).

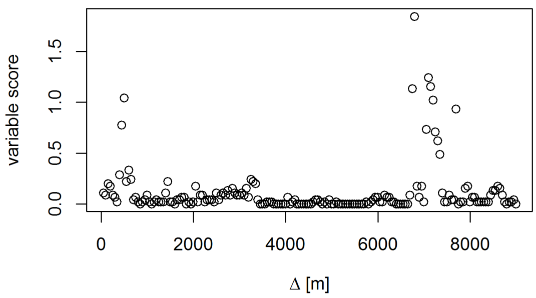

Figure 4: Boruta score for set of 180 values of Δ uniformly distributed in the interval $(45m, 8995m)$

The outcome we would like to obtain is a classification of a turbulence encounters into intensity classes. The best choice would be the standard four: no turbulence, light turbulence, moderate turbulence and severe turbulence. Yet the primary characteristic of January dataset is extreme lack of balance (this is also true for other months as shown in Tab. 1). Due to such distribution we have merged MOD and SEV classes due to low number of observations in the MOD class. Therefore our resulting classificator will have only three distinct CAT intensities: NOTURB, LIG and MOG (moderate or greater). Another consequence of strong imbalance in the dataset is a necessity to handle this issue in the use of random forest since this method is sensitive to unbalanced input. Our approach to the issue of unbalanced classes was in this case is composing a test set of balanced MOG, LIG and NOTURB classes. This is done by sampling the classes. Of course such procedure implies always choosing very small subset of NOTURB class that is why one can expect most of the singular results being biased by this choice. First of all we have run a model reduction test

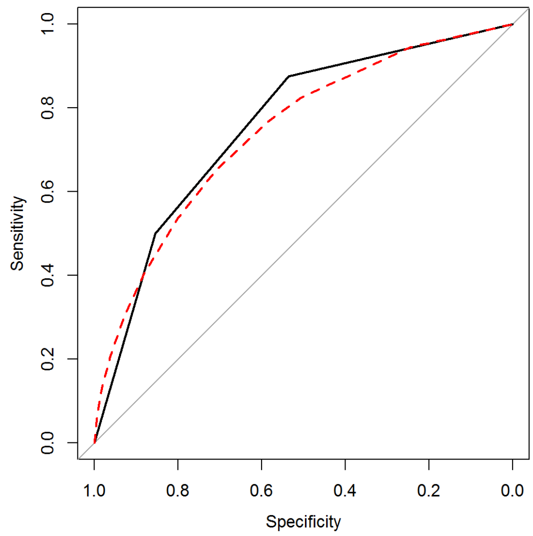

Figure 5: ROC curves for GTG (red dashed) and random forest predictor (black solid). Thin grey line is no-skill ROC.

in order to find a set of most useful of 180 potentially much correlated variables. To extract the most important variables (implicitly values of Δ) we have used Boruta procedure which is based on random forest algorithm (Kursa and Rudnicki, 2010). The Boruta algorithm tests a set of variables and then produces a classification of variable importance (three values are given: important, tentative and unimportant). Due to potential bias by NOTURB representatives we have repeated the procedure 45 times. Each time a numeric value has been assigned to algorithm result (0 - unimportant, 1 - tentative, 2 - important). Overall variable score is a mean of all 45 runs. The resulting variable importance classification is presented in Fig. 4. According to our expectations only small subset of 180 tested variables was sufficient for prediction. However, the most imporant variables according to this

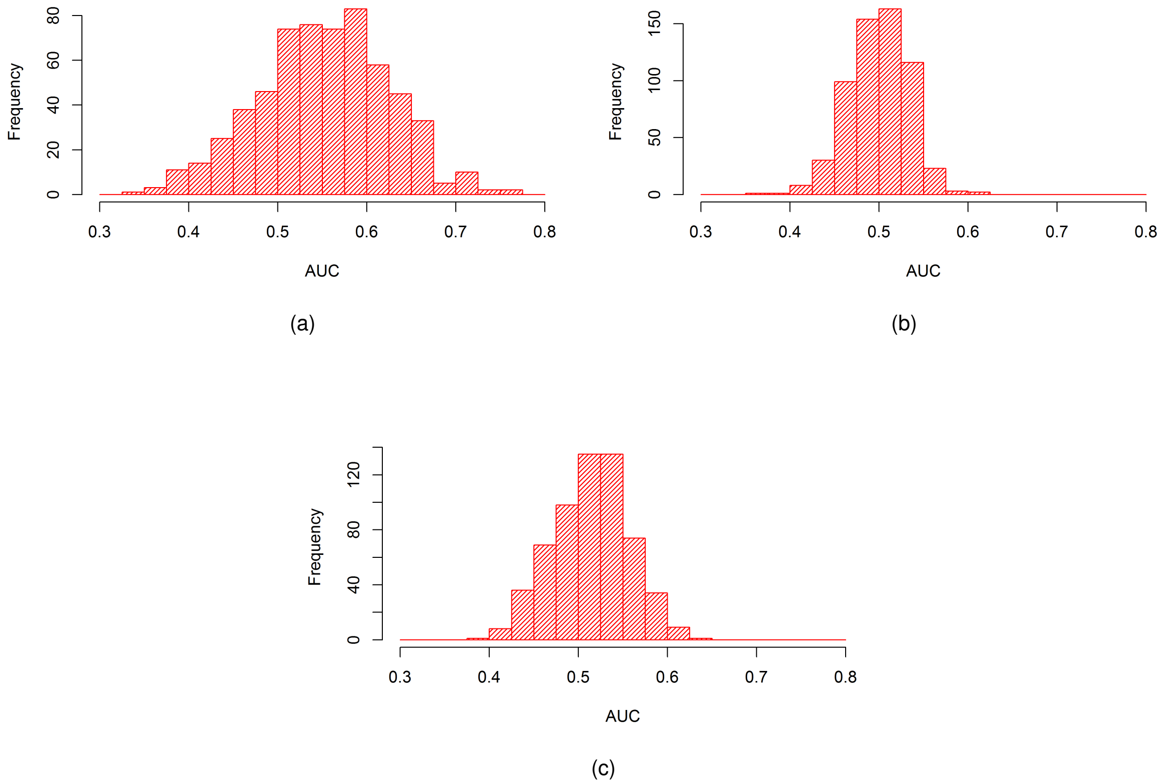

Figure 6: AUC distribution for tested 600 random forest predictors. 6a shows the test using January data, 6b using February data and 6c using March data.

method are not the ones with the highest AUC but rather the ones with somewhat less than maximal prediction skill. In this way a subset of 6 variables which have attained score greater than 1 has been selected. Those variables are corresponding to Δ values of 495m, 6745m, 6795m, 7095m and 7145m.

Based on those variables a set of 600 random forests were trained, each on an individual balanced training set. Three performance tests were run on the resulting predictors. First a control test on January data (minus training set) to see how the predictor captures potentially similar data. And two prediction tests using February and March as test sets. Results of each of those tests are presented in Fig. 6. For January AUC is, on average, slightly better than for individual predictors with mean AUC equal to 0.552, however there exist predictors with AUC$\geq$0.75 which is a significant improvement over individual indices. Fig. 5 shows a comparison of ROC curves* between the best random forest predictor and NOAA GTG1 which show that already the method described here has significant potential. Due to variable skill of individual N_Δ the skill of combined indices is low so that the distributions for February and March are centered around 0.500 and 0.515 respectively. Yet, for both months there exist indices trained in January for which AUC excedes 0.6. A curious feature of random forest predictor is that the best predictors for March and February are those which had AUC significantly less than 0.5 (that is they bear relevant information but tend to invert predicted turbulence class) in January and the best predictors for January are found among the ones with AUC much less than 0.5 for February and March. This phenomenon requires further investigation but seems to be a promising feature of the random forest predictors. We

*The data used for this comparison were obtained from website http://rtvs.noaa.gov/turb/stats/.

should be able to select *a priori* the random forest predictors showing potential for future forecasting just by taking the ones which have extreme AUC values in the initial tests.

5 Summary and further development

Based on work the of Haman (1962) we have introduced a new gravity wave based turbulence diagnostics N_Δ. Tests using filtered 3 month AMDAR data covering Europe have shown that their performance is not stable. That is, for most of time periods there exist an interval thickness Δ for which N_Δ has area under ROC curve for moderate or greater turbulence greater or equal than 0.67 yet this Δ changes in a way that can not be inferred from the described tests. A possible reason for this behaviour are a background weather conditions we did not know to sufficient extent. Because of this N_Δ could be used as one of many ingredients of an ensemble index but not as standalone turbulence diagnostic due to reasonably good yet unstable performance. A test of a random forest based ensemble index consisting of 6 selected N_Δ was performed. It showed that although the resulting predictor inherited much of the chaotic behaviour, this time we were able to formulate a condition to asses which among a collection of ensembles would show future potential for forecasting based only on a tests conducted using a fixed training set. It seems that understanding the behaviour of a collection of N_Δ indices is critical for this problem. We need to understand their variability and what new information those diagnostics introduce. This will require more tests using a bigger dataset. Another way to improve is tuning the random forest ensembles which show prospect of becoming a standalone index.

6 Acknowledgements

The research leading to these results has received funding from the Seventh Research Framework Programme of the European Union (FP7/2007-2013) under grant agreement n° 233801. We would also like to thank Christine Le Bot from Meteo France for providing AMDARs for testing. Finally we would like to thank Leszek Iżycki and Bogumił Jakubiak from Interdisciplinary Centre of Computational and Mathematical Modelling, University of Warsaw for help in accessing COAMPS data archive.

References

ARL, 1985: *The Australian implementation of AMDAR/ACARS and the use of derived equivalent gust velocity as a turbulence indicator*. Aeronautical Research Laboratories in Melbourne, structures Report – 418.

Bekofske, K. and V. C. Liu, 1972: Internal gravity wave-atmospheric wind interaction: A cause of clear air turbulence. *Science*, **178**, 1089–1092.

Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5–32.

Haman, K., 1962: On some possible causes of clear air turbulence. *Acta Geophysica Polonica*, **10**, 335–357.

Knox, J., D. McCann, and P. Williams, 2008: Application of the Lighthill-Ford theory of spontaneous imbalance to clear-air turbulence forecasting. *J. Atmos. Sci.*, **65**, 3292–3304.

Kopec, J., K. Haman, and K. Bajer, 2011: Clear-air turbulence induced by short gravity waves. *13th European Turbulence Conference*, proceedings to be published online in JPCS.

Kursa, M. B. and W. R. Rudnicki, 2010: Feature selection with the Boruta package. *Journal of Statistical Software*, **36**, 1–13.
URL `http://www.jstatsoft.org/v36/i11/`

Liaw, A. and M. Wiener, 2002: Classification and regression by randomforest. *R News*, **2**, 18–22.
URL `http://CRAN.R-project.org/doc/Rnews/`

McCann, D., 2001: Gravity waves, unbalanced flow, and clear air turbulence. *Natl. Wea. Dig.*, **25**, 3–14.

Sharman, R., C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268–287.

Sharman, R., S. Trier, T. Lane, and J. Doyle, 2012: Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. *Geophys. Res. Lett.*.

WMO, 2003: *Aircraft Meteorological Data Relay (AMDAR) Reference Manual*. World Meteorlogical Organization, WMO-No. 958.