Atmos. Meas. Tech. Discuss., 8, C5137–C5157, 2016
www.atmos-meas-tech-discuss.net/8/C5137/2016/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Self-Nowcast Model of extreme precipitation events for operational meteorology" *by* G. B. França et al.

G. B. França et al.

valdonel@lma.ufrj.br

Received and published: 24 January 2016

Dear Sir/Madam, Thank you so much for helping us to improve our manuscript. We recognize that our work is truly related to nowcasting for aviation application and then a new manuscript (attached as supplement) was produced as attached file. English native speaker generally revised the manuscript (see attached the editorial certificate as Fig. 1 of this form). Our comments/answers are below item-by-item Sincerely, Authors.

1 ATMOSPHERIC MEASUREMENT TECHNIQUES Manuscript amt-2015-247-discussions

Self-Nowcast Model of Extreme Precipitation Events for Operational Meteorology. Gen-

eral comments: This manuscript presents a neural network-based algorithm for the automated nowcasting of convective rain around Rio de Janeiro′s International Airport. Despite being an important subject that deserves more attention and investment by the meteorological community, there are several issues that must be addressed before considering this manuscript ready for publication. For that reason my recommendation is for a major revision. Below the authors will find more detailed comments.

Title: The term "Self-Nowcast" is not adequate to describe the automated model developed for nowcasting because it gives the (meaningless) idea that 'the model nowcasts itself'.

The term should be replaced by something like "An automated nowcasting model…". Secondly, this work does not really deal with extreme precipitation events (see my detailed comments in page 4 below), so the authors should refrain from using the expression "extreme precipitation events" in the title and throughout the text.

Authors′ response: accepted! Authors′ change: The new title is as follows: "An Automated Nowcasting Model of Significant Instability Events in the Flight Terminal Area of Rio de Janeiro – Brazil"

SECTION 1: Introduction:

Major issue 1: In this section the authors make an effort to describe an extreme meteorological event (EME) but their approach is biased towards extreme meteorological events associated with convective precipitation. This is partially understandable because heavy rain is the target of the study, but because not all EMEs are associated with extreme precipitation, the authors must rephrase several sentences in order to emphasize that their interest lies on the EMEs that cause heavy rainfall, and not on EMEs related to other variables (such as strong wind gusts). For example, in line 11 it is stated that "Teixeira and Satyamurti (2007) studied EME occurrences in southeastern Brazil…". This should be rephrased as "Teixeira and Satyamurti (2007) studied EMEs associated with heavy precipitation in southeastern Brazil…".

Authors′ response: We have accepted all comments/suggestions Authors′ change: The introduction was entirely rewritten at the end of this section.

Lines 5 to 7: "According to Marengo et al. (2004), an EME is defined as a rare meteorological phenomenon with very low statistical distribution in a particular place." Marengo et al. (2004) provides no definition for an extreme meteorological event and this sentence should be removed. Authors′ response: Accepted! Author s′ change: It was removed from the manuscript at the end of this section.

In addition, the authors cite several climatological studies addressing EMEs, while their study is on nowcasting. Therefore, there is a serious conceptual mismatch in their literature review. Authors′ change: The introduction was entirely rewritten at the end of this section.

Major issue 2: In lines 15 to 17, the authors indicate that "EME phases [. . .] fall into a nowcasting time scale, implying a short-term forecast." There are several problems with the statement above. First. The concept of EME is strongly associated with the climatology and statistical distribution of the meteorological variables for a given region. An EME is related with the idea of a rather rare event, and, therefore, only trough climatological studies can one determine what is an EME for a 2 given region and time of the year. In fact, when the authors first introduce the term "nowcasting" in the text, it comes right after a number of articles that investigate EMEs from a climatological standpoint. Now, how does nowcasting relate to those climatological/statistical studies? In operational nowcasting, do we need to determine if a given weather event is an EME (from a statistical perspective) in order to issue a warning? The authors submit a manuscript on the nowcasting of extreme precipitation events, but two-thirds of their Introduction offers a general overview on the concept of EME, citing a number of articles on the climate perspective of extreme precipitation. Is this review about EMEs really necessary for an article on nowcasting? Authors′ response: The EME concept was removed from manuscript as can be found in attached file. Authors′ change: See attached file (new manuscript).

Second: Only "EMEs" associated with convective weather (or winter weather) can be addressed through nowcasting in the operational setting. EMEs are not synonym for convective weather. In fact, the forecast of EMEs and their stages of evolution ïĆ¿ïĂăeven those associated with heavy precipitation ïĆ¿ïĂăcan be addressed through a myriad of approaches ranging from climate prediction (e.g., seasonal forecast; "Are we expecting more frequent and intense South American Convergence Zone this Summer?"), to medium-range forecasting (e.g., "How much rain should we expect from a given tropical/subtropical/extratropical cyclone?"), and finally down to the nowcasting range. Hence the statement "EME phases [. . .] fall into a nowcasting time scale, implying a short term forecast" is not accurate and should be eliminated, or entirely rephrased.

What are the nowcasting techniques available for convective storms capable of producing heavy rain? The authors only mention (briefly!) Mueller et al. (2003) and Mass (2012), while the nowcasting field has a vast literature, including the important topic of assimilation of radar data on NWP models (e.g., Sun and Wilson, 2003). Authors′ response: We agreed with all above comments and, thus, some references were removed and others pertinent ones included. Authors′ change: See new introduction below and attached file (new manuscript).

Lines 25 and 26: The authors state that "The present numerical prediction models do not satisfactorily model EMEs in location-specific and short-term scales." Is this an opinion from the authors, or is it a general result from past studies? If in the second case, where are the references? What are the objectives and shortcomings of high resolution NWP? Finally, how their approach to the nowcasting of heavy rain differs from previous studies? Authors′ response: Considering that, the new focus is to forecast -in the short-term for specific area - of significant instability and its impact for aircraft flow, the authors find unnecessary to respond the above inquiries.

In short, the authors must rewrite their entire Introduction section to better relate to the main topic of the study. Authors′s change: Considering some above suggests, the introduction was rewritten as follows: 1. Introduction Aviation is negatively or positively influenced by the atmospheric conditions at any place and time (Ahrens, 2008). In particular, the Terminal Area (TA) of an airport is the area where the aircraft are waiting for landing or take-off and, thus, is quite sensitive to weather conditions. The air traffic controllers and pilots require precise information about the weather conditions at the TA to make short-term decisions that fall into the time scale of nowcasting, which ranges from the interval of a few minutes up to 6 h. During the few last decades, various works associated with nowcastingâĂŤfor example, Wilson (1966), Wilk and Gray (1970) and othersâĂŤhave initially proposed nowcasting approaches based on extrapolations of radar data to generate nowcasting of thunderstorms. To follow up this idea, the convective tracking approaches were improved by including the cell evolution in time and intensity using radar data (Dixon and Wiener 1993). Wilson et al. (1998) presented a review of the nowcasting techniques developed during the 1960s and 1970s. The advancement of parallel computing and data availability allowed a numerical weather model to assimilate via rapid update cycle (and, more recently, via rapid refresh method) mesoscale data such as satellite and/or radar data to nowcast convective systems. Several authors have addressed the latter in the last two decades or soâĂŤe.g., Xue et al. (2003), Sun and Wilson (2003), Schroeder et at. (2006), Liu et al. (2008) and others. Mueller et al. (2003) proposed a sophisticated system to nowcast (up to 1 h) thunderstorm locations based on a combination of surface meteorological, radar, satellite data and numerical modelling, which considers the storm stages. Mass (2012) provided a comprehensive review of nowcasting including current developments and future challenges. Considering the aviation application, Isaac et al. (2006), Isaac et al. (2011) and Isaac et al. (2012) presented a sequence of works that resulted in a refined nowcasting system for aviation that uses data from numerical models, surface observations, radar, satellite and a microwave radiometer to generate nowcasts for principal airports in Canada up to approximately 6 h. In contrast, in Brazil, a meteorologist currently uses his experience to integrate different in situ meteorological observations and/or atmospheric model outputs using conceptual models

on how the atmosphere works to generate nowcasts at principal airports. In particular, the TA of Rio de Janeiro, the focus of this study, has five airports (see Figure 1) whose flights are significantly affected (by delays and trajectory changes), especially during the approximations for landing or take-off, by Significant Instability Events (SIE), which are normally associated with convective weather. Groisman et al. (2005) presented evidence that the incidence of convective weather has increased approximately 58% per year in south-eastern BrazilâĂŤwhere the Rio de Janeiro TA is locatedâĂŤsince the 1940s. Therefore, the objective here is to present an Automated Nowcast Model (ANM) to generate short-term and local-specific predictions of SIEs, based on neural network techniques, for the fight TA of Rio de Janeiro, Brazil.

SECTION 2: Meteorological data sets and study region:

There is no mention about the period of study in the text; only Table 1 provides this information. This information must be indicated in the text as well.

Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

TEMP and METAR are not time series per se, but meteorological codes. The authors should replace "…four time series" by "…four data sources" or "…four datasets". "TEMP represents the upper atmospheric profile for…" Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

No. TEMP is the meteorological code employed to report profiles of atmospheric variables. "The TEMP time series was obtained…" Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section. Should be replaced by "Time series of TEMP-coded data was obtained…" "…where SB and GL mean Brazil and Galeão, respectively." Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

3 Not a relevant information. Remove this sentence. "…SBGL is the only one of the

stations that collects atmospheric profiles in a daily basis. . ." Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

Poor writing. Replace by ". . .SBGL is the only station where hourly meteorological data are reported regularly. . ." Authors′ response: It was removed Authors′ change: See below new text of section 2 at the end of this section.

The expression "atmospheric profiles" makes no sense if one refers to METAR data which contain surface reports only. Replace "metropolitan region" by "metropolitan area", which is a more common usage. The network of 29 rain gauges is operated by/belongs to which institution? Providing the internet link is not enough; the authors must name the institution. Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section. ". . .distributed around the Rio de Janeiro metropolitan region." Poor writing. Replace by ". . .distributed over Rio de Janeiro metropolitan area." Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

Caption of Figure 1 is badly drafted and does not provide credits to the image provider. It should read: "Satellite image of Rio de Janeiro′s metropolitan area. Yellow triangles [Red squares] indicate location of the twenty-nine rain gauges from {add here the institution that runs that network} [five airport meteorological stations]. Satellite image from Google Earth." Authors′response: The new caption of Figure 1 is: Figure 1 - Satellite image of Rio de Janeiro′s metropolitan area. Yellow triangles [red squares] indicate location of the twenty-nine rain gauges from Alerta Rio's system that belongs the City Hall of Rio de Janeiro [five airport meteorological stations]. Source: Adapted from www.google.com.br/maps.

". . .lightning reports [. . .] characterize each occurrence by its location. . ."

Poor writing. Replace by ". . .lightning reports [. . .] indicating location. . ." Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section. "Table 1 summarizes. . ." Poor writing in this full sentence. Replace by: "Table

1 summarizes the data sources {or datasets} utilized to train and validate the nowcasting algorithm in this study." Authors′ response: done! Authors′ change: See below new text of section 2 at the end of this section.

2. Meteorological datasets and study area

This study used four datasets from 1 January 2007 to 31 December 2008, as follows:

ïĆğ TEMP is the meteorological code used to report profiles of atmospheric variables and is normally generated daily at 0000 UTC and 1200 UTC on all radiosonde stations, one of which, in this work, is located at Galeão′s Airport, whose international aviation code is SBGL, where SB and GL denote Brazil and Galeão, respectively (see Figure 1). The TEMP-coded dataset was obtained online from http://weather.uwyo.edu/upperair/sounding.html; ïĆğ METAR and SPECI are meteorological codes employed to report hourly surface meteorological conditions and significant change (decline or improvement) in the weather condition, at any time from the full hour. Figure 1 shows the locations of five surface meteorological stations (represented by red icons) in the Rio de Janeiro metropolitan area. The SPECI data were used only for the model validation. The stations (or airports) are Galeão (SBGL), Santa Cruz (SBSC), Santos Dumont (SBRJ), Jacarepaguá (SBJR), and Afonsos (SBAF). The data were obtained at the URL address mentioned above; ïĆğ rain rate (RR) is obtained from twenty-nine rain gauges (represented by yellow triangles in Figure 1) distributed over the Rio de Janeiro metropolitan area. The data were obtained from http://www.rio.rj.gov.br/alertario/ and collected by Alerta Rio's System, which belongs the City Hall of Rio de Janeiro; and ïĆğ lightning reports, regularly collected by the National Integrated Lightning Detection Network (RINDAT), characterize each occurrence by indicating location (latitude, longitude), intensity polarity (cloud to ground or ground to the cloud), and time (UTC with accuracy in milliseconds). ELETROBRAS FURNAS Company kindly made the data available.

Table 1 summarizes all information on the datasets used for ANM training, testing and validation in this study. Figure 1 shows the study region and the flight terminal area of Rio de Janeiro.

———————————————-

Caption of Table 1 is also very badly drafted. It should read: "Datasets and meteorological variables used in the distinct stages of development of the neural network-based automated nowcasting algorithm. Lightning data is used only during the validation stage of development." Information such as "This is not important, since it is used only for validation" is irrelevant and, thus, unnecessary in the caption. Authors′ response: done! Authors′ change: The new caption is: Table 1 - Datasets and meteorological variables used in the distinct stages of development of the neural network-based automated nowcasting model. It covers a period from 1 January of 2007 to 31 December 2009. Table 1: Replace "Time series" by "Dataset". What is "Data-time"? Do the authors mean "Date-time"?

Authors′ response: "Data-time" - It was remove from the manuscript

Third-line Third-column: wind is missing in the list of variables. Footnote 1: K-index formulation uses (T@700hPa – Td@700hPa), not (T@700hPa – Td@500hPa) as indicated in the footnote. The definition of lapse rate is mistaken.

Authors′ response: The mentioned footnote was removed. The updated section 3.2.1 describes the input variables. The updated Table 1 has only resulting twelve variables - used to train/test and validate the ANM - divided into eight primary and four derived variables as listed in columns three and four, respectively.

Authors′ change: See new text of section 3.2.1 and Table 1.

Observation: (Major issue 3) Henceforth, I will not provide further suggestions for improving poor writing in the manuscript because there are just too many passages to improve/correct. Not only are there grammatical errors, but also the authors abuse in the use of unnecessary comments, such as "The problem with this is…". Other ex-

pressions are translations of idioms typically used by native portuguese speakers, but that are very unusual (or sound very strange) in english. The authors ought to hand the manuscript to a native english speaker or to a professional translator to improve the text.

Authors′ response: - The sentence - "The problem with this"- was removed from the manuscript. - English native speaker generally revised the manuscript (see attached the editorial certificate)

4 SECTION 3: Method. {Very badly written} The title of this section should read: "Methodology and algorithm description". Authors′ response: It was replaced Authors′ change: See below new text of section 3 at the end of this section.

"Figure 2 represents a typical neural network." Figure 2 alone does not help describing a neural network at all. This sentence should be removed. Authors′ response: done! Authors′ change: See below new text of section 3 at the end of this section and new manuscript attached (PDF).

Variables and coefficients in equation (1) must be described in more detail. What are the weights? What do the indices M and D represent? What is ïĄş? Authors′ response: done! Authors′ change: See new text of section 3 at the end of this section.

3.1 Data processing.

Page 6, line 9: "…consisted of three simple tasks." It is not clear what the authors mean by "simple tasks". This must be rephrased or clarified. Authors′ response: It was modified. Authors′ change: See new text of section 3.1 at the end of this section.

Page 6, line 10: "…and their consistency observed…" It should read "…and their consistency checked…"; but how was data consistency checked?

Was there also a quality control procedure applied to the datasets? This is relevant information that deserves to be addressed in the text.

Authors′ response: It was clarified. Authors′ change: See new text of section 3.1 at the end of this section.

Page 6, line 11: "...meteorological recordings." Replace by "...meteorological records." Authors′ response: It was replaced Authors′ change: See new text of section 3.1 at the end of this section.

Major issue 4: Page 6, lines 11 to 13: "...the rain rate time series, based on RR h-1, was used to classify the meteorological recordings into four classes..."

The authors did not describe the methodology used to determine the thresholds for null, light, moderate and heavy rain episodes. The 9.9 mm h-1 threshold for heavy rain does not seem appropriate (note: Teixeira and Satyamurty (2007) present a brief but good review of criteria usually employed to define 'heavy rain').

Since the authors are interested in highlighting the nowcasting of extreme precipitation events, a detailed reasoning must be presented in support of the 9.9 mm h-1 threshold for heavy rain. Wilks (2006) describe distinct procedures to characterize extreme weather events based on objective statistical methods, but it appears to me that the authors have chosen a subjective approach that led to a far than adequate criterion to discriminate extreme precipitation. As a matter of fact, if the 9.9 mm h-1 threshold is to be used to discriminate 'heavy rain', then the authors must refrain from using the expression 'extreme precipitation events' when referring to these precipitation episodes. Authors′ response: Considering that the updated model outputs do not included RR, above comments have no more sense. Now, the outputs are defined as weather condition reported in METAR code and divided in two classes 0 and 1, as in Table 2. The classes 0 and class 1 mean no significant weather condition (or nonexistence of significant instability) and thunderstorm (i.e. weather condition of METAR code as T, TL, TRW-, TRW, TRW+), which assume presence of significant instable in the terminal area of Rio de Janeiro, respectively.

Authors′ change: See new text of section 3 at the end of this section.

3.1 Input and outout. Major issue 5: Page 6, lines 24 to 26: The authors state that "The SNM's purpose is to nowcast EMEs; therefore all input (or predictors) should indicate EME phases, i.e., initialization, growth and decay." First, the authors do not explain how the different phases or stages of the precipitation events were determined based on the dataset they have. Without that information it is impossible to complete an assessment of the methodology used. Second, are the set of predictors utilized to train the neural network to forecast the distinct stages of evolution of the precipitation episode? In other words, was it also a goal to predict the demise of the precipitation event? Third: the authors did not explain why they did not include, in the screening stage, candidate predictors coming from numerical model output or weather radar data (Rio de Janeiro state does have operational meteorological radars), and non-local candidate predictors. All these issues must be clarified in the methodology section. 5 Authors′ response: See new text of section 3 at the end of this section.

For which atmospheric layer(s) is computed the lapse rate? The authors do not indicate this information in the text, only in a 'hard-to-read' footnote of Table 1 (and there is a typo in that definition). Authors′ response: The footnote text was moved to the main text as can be found as new text of section 3.2.1 at the end of the section.

Page 7, line 6: What do the authors mean by "quite constructive"? Page 7, line 7: "After a simple correlation test..." First question: a "correlation test" of what sort? Second question: what exactly is the predictand variable? Accumulated rainfall? The time derivative of hourly rainfall? This stage in the development of the algorithm has to be better described. Authors′ response: Most of above questions are answered in the new text of section 3 (at the end of this section).

Page 7, lines 9 and 10: "These variables were initially judged the best data set to transmit atmospheric conditions during neural network training"

This sentence must be fully rephrased because it does not make sense. For example, what do the authors mean by "to transmit atmospheric conditions"? And there should be a table indicating all fifty seven atmospheric variables/parameters that 'survived' the first predictor screening.

Authors′ response: Considering above suggest, section 3 was rewritten (at the end of this section).

Page 7, lines 15 to 17: "The latter is responsible for converting the input (or predictors) in the event that all four RR classes occur." I could not understand this passage, especially where it reads "in the event that all four RR classes occur". The authors must clarify this methodological approach.

Authors′ response: Considering that, section 3 was rewritten (at the end of this section).

3.3 Neural network training. Page 7, lines 20 and 21: It is not clear what the authors mean by "It requires previous knowledge of the phenomenon in conjunction with the experience of the training team." I can only guess that they refer to the stage when atmospheric variables were assessed as predictors from a physical basis, but this needs to be clarified. Authors′ response: Considering that, section 3 was rewritten (at the end of this section).

Page 7, lines 21 and 22: "EMEs are characterized by thermodynamic atmospheric patterns represented by local meteorological recordings." Not only EMEs are characterized by 'atmospheric patterns'. Any meteorological event, including non-EMEs, can be related to an atmospheric pattern. In addition, the characterization of atmospheric patterns, either in the synoptic scale or in the mesoscale, require a (non local) two dimensional analysis of the meteorological variables. The authors do not explain how this is performed given that datasets they employ.

Page 8, lines 2 and 3: "The EME is defined as a nowcast corresponding to "yes=class three (RR > 9.9mm h-1)" or "no=class one, two, or three". I think there is a mistake here. I think the authors meant to say "no = class zero, one, or two". I am convinced that at the end of this subsection the reader will not feel well informed about how the

neural network algorithm was trained. Authors′ response: The new section 3 describes all steps (1 to 4) of the algorithm until it reaches the optimal results. Please see the new section 3 at the end of this section.

3.4 Validation and other procedure steps. I am not sure if the expression "other procedure steps" makes sense. Authors′ response: It was removed!

Major issue 6: In this subsection the authors describe the method with which the neural network algorithm was validated. Three distinct data sources were used to verify the neural network′s automated forecasts, but the methodology is flawed. The authors group a set of METAR observation codes (namely, R+, R+F, RW, RW+, T, TL,TRW-, TRW, TRW+; Table 2) into the same 'class 3' in which the hourly rainfall rate (as measured with rain gauges) is above 9.9 mm h-1.

However, what is the relation between any of those METAR observation codes with quantitative precipitation? Why should one consider the observation code R ('moderate rain' in the METAR code) to be representative of the same 'class 2' in which, according to the authors, the hourly rainfall rate is below 9.9 mm hr-1? Continuous moderate stratiform rain can produce a rainfall amount reaching more than 10 mm within 1 hour, and an aerodrome observer could still report it correctly as moderate rain, receiving the METAR code R. Even worse, while the observation code T alone means 'thunderstorm with no rain being reported', this observation is grouped in the same 'class 3' as rainfall rate above 9.9 mm hr-1 ("extreme precipitation event"). This is contradictory. Moreover, the authors also added 'lightning reported inside a 50 km radius centered at Galeão Airport during a one-hour period' as representative of a 'class 3' event. Again, what is the relation between lightning occurrence and quantitative precipitation? How about lightning flash rates? The authors do not mention anything about flash rates. This suggests that, regardless of the number of lightning flashes occurring within the hour in a 50 km radius, the authors grouped the lightning event in the same 'class 3' even if the nearby thunderstorm is producing little rain (and, thus, not representing an "extreme precipitation event"). As a matter of fact, the authors do acknowledge this potential inconsistency further ahead in the text, in subsection 4.1.3. In addition, the authors state that "weather conditions reported in a METAR represent an observation by the meteorologist in an instant of time (ten minutes before the hour); therefore, sometimes it does not correctly represent an entire one-hour period, which is the minimum time interval for an SNM forecast". This is not entirely correct. Meteorological observers in aerodromes do not have to wait until the top of the hour to report significant weather conditions (as thunderstorms, for example). A SPECI weather report, which follows the same METAR coding, is issued immediately whenever significant weather conditions occur at or around the airport. Therefore, the dataset used by the authors can be improved by including SPECI reports.

In summary, the methodology chosen by the authors to discriminate "extreme precipitation events" in the validation of the algorithm is not conceptually coherent with the idea of an extreme precipitation event, and has serious implications for the interpretation of the results.

Authors′ response: Thus, the section 3 was entirely rewritten, as below, considering the clarification of above questions and the removal of possible contradictions. ————————————————————- 3. Methodology and algorithm description

Meteorologists have limited windows of time in which to integrate all available data and generate a nowcast, as stated by Mueller et al. (2003). Therefore, the idea is to create an automated nowcast model in which a neural network algorithm is used for data fusion, similar to the work performed by Cornman et al. (1998) for detecting and extrapolating weather fronts. At present, one may find applications of neural network in numerous fields of science, such as modelling, time series investigations, and image pattern recognition, owing to their capability to learn from input data (Haykin, 2002). Normally, stages of neural networks are denoted by a global function (Equation 1), as described by Bishop (2006)âĂŤfor example: Equation (1) is here!

where xi and yk are the input and output, respectively; (1), (2) and Wji, Wkj represent

the input layer, hidden layer and the connection weights (that should determinate) between input and hidden layers and hidden and output layers, respectively; D and M are the number of inputs and number of neurons in the internal layer, respectively; and σ and h are linear and no linear transfer functions between the neural network layers, respectively. Thus, determination of the output via Equation 1 crucially depends on the values of the weights that are worked out, similarly as in a multiple linear regression using a set of inputs and outputs; however, instead, to minimize the distance as in nonlinear regression, the neural networks attempt to minimize the cost function. Given that the SIE forecast problem requires a categorical output, it was decided to use probabilistic neural networks, initially proposed by Specht (1990, 1991), which is based on radial-basis function (RBF), A RBF network consists of three layers: the input layer; the second layer (or hidden), apply a non-linear transformation, denoted as h that, here, is Gaussian function, of the input space to the hidden space. The third layer, the outgoing, is linear (σ), providing the network response. Further details about neural networks and their applications may be found in Pasini et al. (2001), Haykin (2002), Pasero and Moniaci (2004), Bremnes and Michaelide (2005), Bishop (2006), Haupt et al. (2009) and Hsieh (2009).

Figure 3 depicts a general flowchart for the proposed automated nowcasting model. It has four major steps: (1) data processing; (2) definitions of input and output variables; (3) training and testing; and (4) validation. These steps are described below.

3.1 Step 1âĂŤData processing:

All datasets were sorted chronologically, and their statistical consistency was observed, resulting in 63,320 h of meteorological records. Based on weather conditions reported by METAR, each meteorological record was classified into two classesâĂŤ"0" and "1", representing nonexistence of important weather conditions (low impact to flight flow) and the existence of significant atmospheric instability (or SIE, as previously defined) for flights in the TA of Rio de Janeiro, respectively. Table 1 shows all weather conditions reported in terms of METAR code and their classification per class.

3.2 Step 2âĂŤInput and output definition:

ANM data fusion is based on a neural network, which must be sequentially trained, tested and subsequently validated to forecast the presence or absence of SIEs. The latter corresponds to the learning process of a neural network. The input and output variables play an important role in ANM data fusion and should be previously defined.

3.2.1 Input variables

These variables are the predictors of ANM and indicate the atmospheric stages of SIEs in the study area that are used by the ANM during its learning process. A meteorological record is composed of primary and derived variables that are extracted from METAR, TEMP, and RR and calculated using primary variables. The purpose of ANM is to nowcast SIEs and other weather conditions; therefore, all inputs (or predictors) should thermodynamically represent the presence or absence of SIE, which are embedded in the meteorological records utilized to train/test and validate the ANM. The latter should be able to classify or forecast weather conditions of classes numbered as “0” and “1”, and its performance is evaluated by cross-validation with observations as presented later. The criterion to select input (primary and derived) variables is based on a conceptual model of how the atmosphere worksâĂŤparticularly during SIE occurrence, which have typical atmospheric patterns. Several input variables are usedâĂŤfor example, atmospheric instability indices, i.e., K-index (K) = (T850- T500)+Td850-(T700-Td500), where Tz and Tdz represent temperature and dew point, respectively, in Celsius degrees, and z is the given atmospheric pressure in hPa); Total Totals (TT) = T850+Td850- 2T500; Lapse Rate (LR), represented by LR = 1000(T500- T700)/(GPH500- GPH700), where GPH denotes the geopotential height; and others defined in columns three and four of Table 1. At the beginning, many inputs were generated. However, with regard to the neural network training, it is necessary to adopt a method to prune collinear inputs that bring no new information and, thus, could reduce the network performance. Pasini and Ameli (2003) have investigated heuristic pruning methods. Here, autocorrelation was selected and enforced to remove collinearity of

the input. Twelve variables then remained, divided into eight primary and four derived variables as listed in columns three and four of Table 1, respectively.

3.2.2 Output variables

The output is defined as weather conditions reported in METAR codes and divided into two classes, “0” and “1”, which represent the absence and presence of SIEs, respectively, as shown in Table 2. In other words, classes 0 and 1 indicate nonexistence of significant instability and existence of significant instability (i.e., weather condition of METAR code as T, TL, TRW-, TRW, TRW+) in the TA of Rio de Janeiro, respectively.

Following Pasini (2015) and aiming to avoid the overfitting problem during the learning process of the neural network, which is represented by step 3, the meteorological records were divided into three subsets: training, testing and validation. Figure 4 (a) shows the initial training and testing datasets representing 70% of the original records (or 44,324) with 30% (or 18,996) for validation, as shown in Figure 4 (b).

3.3 Step 3âĂŤNeural Network Training and Testing

The internal number of neurons (previously defined as M) of probabilistic neural networks is here determined based on cascade-correlation algorithm suggested by Fahman and Lebiere (1990). Figure 2 shows generally an example of a cascade forward network for five inputs and one output. The training and testing are performed in an iterative cycle composed of a looping of two phases, which are executed using a specific dataset (initially the one in Figure 4 (a), which could be artificially modified until the optimal dataset is reached, as described in step 4), and a constant number of inputs (defined as D is equal to twelve). The two phases are described as follows:

i) It starts with a minimal (only one neuron) internal layer of the neural network (represented generally by Equation 1) and automatically adds new hidden neurons one at a time, in each round, finally resulting in a multilayer structure with the input connection frozen (represented by squares in Figure 2); and ii) The follow-on neural network is applied to the test dataset, and the error is calculated. There are then two options: first, return to (i) if the test error has not increased from the previous round and the number of neurons in the internal layers is less than 150; or second, to go step 4, which means that the final (or that could be an optimum) neural network configuration (or ANM) has been obtained.

3.4 Step 4âĂŤValidation:

This step compares the SIE forecasts (output) of ANM with the true observations, which are assumed to have at least one of two conditions:

a) weather conditions (class 1 of Table 2) reported by METAR or SPECI (corresponding the validation dataset in Figure 4 (b)); and/or

b) lightning reported inside a 50-km radius centred at Galeão airport during a 1-h period. The lighting data are included in the validation because the weather conditions reported in METAR or SPECI represent an observation by the meteorologist at an instant of time; therefore, sometimes it does not correctly represent an entire one-hour period, which is the minimum time interval for an ANM forecast, and the lightning data will be continuously generated during the entire ANM forecast time and beyond the METAR observation, which depends on the meteorologist's observation skills. The lightning data allow the ANM forecast verification to be spread out to encompass the entire flight terminal area of Rio de Janeiro. Moreover, it is assumed in this work that the presence of lighting is related with SIE. Therefore, these two conditions will certainly permit a better ANM validation, which is accomplished via a two-dimensional contingency table. The calculation of five categorical statistics used to verify the frequency of correct and incorrect forecasted values is performed as follows: 1) proportion correct (PC), which shows the frequency of the ANM forecasts that were correct (a perfect score equals one); 2) BIAS, which represents the ratio between the frequency of ANM estimated events and the frequency of ANM observed events (a perfect score equals one); 3) probability of detection (POD), which represents the probability of the occasions when

the forecast event actually occurred (hits), and the scale varies from zero to one, where one indicates a perfect forecast; 4) false-alarm ratio (FAR), which indicates the fraction of ANM-predicted SIEs that did not occur (a perfect score equals zero); and 5) threat score (TS), which indicates how the ANM forecasts correspond to the observed SIEs (a perfect score equals one). In particular, the TS is relatively sensitive to the climatology of the studied event, tending to produce poorer scores for rare events, such as an SIE. Therefore, the model is considered to be optimal when it creates SIE nowcasting with scores as near perfect as possible for the five statistics described (Wilks, 2006).

Finally, if the validation results of the ANM do not indicate satisfactory performance, a normal procedure is to rearrange the representativeness of the target class one in the training data (i.e., modifying the training/testing dataset) and then go to step 3 and repeat step 4 in Figure 3. Otherwise, the optimal model is reached. The ANM training strategy and results are discussed in the next section.

——————————————— Authors′ final comment: Considering the referees comments and suggestions, the manuscript was revised as in attached file (PDF).

Please also note the supplement to this comment:
http://www.atmos-meas-tech-discuss.net/8/C5137/2016/amtd-8-C5137-2016-supplement.pdf

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 10635, 2015.

AMERICAN JOURNAL EXPERTS

EDITORIAL CERTIFICATE

This document certifies that the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts.

Manuscript title:
An Automated Nowcasting Model of Significant Instability Events in the Flight Terminal Area of Rio de Janeiro - Brazil

Authors:
Gutemberg Borges França*, Manoel Valdonel de Almeida, and Alessana C. Rosette

Date Issued:
January 20, 2016

Certificate Verification Key:
AD45-4DAA-C963-1628-2150

This certificate may be verified at www.aje.com/certificate. This document certifies that the manuscript listed above was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts. Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page. If you have any questions or concerns about this edited document, please contact American Journal Experts at support@aje.com.

American Journal Experts provides a range of editing, translation and manuscript services for researchers and publishers around the world. Our top-quality PhD editors are all native English speakers from America's top universities. Our editors come from nearly every research field and possess the highest qualifications to edit research manuscripts written by non-native English speakers. For more information about our company, services and partner discounts, please visit www.aje.com.

Fig. 1. certificate