The Reasonable Effectiveness of Data in ATM

Massimiliano Zanin
Innaxis Foundation & Research Institute,
Madrid, Spain
mz@innaxis.org

***Abstract*—Data Science has recently emerged as a fundamental paradigm for understanding the characteristics and dynamics of many real systems, enabling innovative approaches in such different scientific fields as astronomy or sociology. In spite of this success, far less attention has been devoted to the analysis of historical data in air transport and ATM, mainly due to the difficulties inherent the study of private and heterogeneous data sets. In this contribution, a simple case study is proposed, aimed at demonstrating the advantages of using such approach, especially in terms of novel knowledge that can be extracted even from basic analyses. Two different techniques are here compared, both aimed at assessing traffic densities: a classical one based on sectors, and a novel one based on multiple concentric circles. Results indicate that the latter yields more knowledge about the system, specifically about the appearance of safety-related events. More generally, this test case confirms the importance of going beyond human-based analysis, in order to *listening to the data*.**

***Index Terms*—Data Science, traffic density, loss of separation events.**

I. Introduction

The idea of extracting knowledge from a set of data is not recent, but has existed, in its manual form, since the beginning of civilization. For instance, one may think in the *Kabbalah*, a school of thoughts inside Judaism, born in the 12th- to 13th-century Southern France and Spain, which uses different methods to analyze hidden meanings and messages inside the Torah [1]; or the analysis of the astronomical observations performed by Johannes Kepler in the 16th century [2].

After this beginning, a dramatic increase in data collection, storage, and manipulation abilities has been witnessed only in the last decade, mainly due to the proliferation, ubiquity and increasing power of computer technology. Many fields of research have benefit from these advances. Among others, astronomy, both in the form of digital repositories, as the Sloan Digital Sky Survey [3], or the distributed analysis of massive data sets [4]; physics, through the automated scanning of particle collision images [5]; and social sciences, with the characterization of large-scale networks of human interactions [6], [7].

The analysis of such large collections of data presents several challenges, which were not even hypothesized 10 years ago. These include the fusion of a diversity of data sources, with different refresh rates and formats; manage the concurrency of the data sources; understand how knowledge could be extracted from unstructured data; and deal with high volumes of data, possibly developing algorithms capable of reducing the computation time if data streams are to be integrated. These challenges have resulted in the birth of *Data Science*, defined as a set of fundamental principles that support and guide the extraction of information and knowledge from data, merging techniques from knowledge discovery, statistics and computer science [8], [9].

In spite of the above, the application of Data Science to air transport and ATM has been of limited scope; among the few examples that can be found, of special relevance is the characterization of safety occurrences, like separation losses between aircraft [10], [11], [12], [13]. The main reasons behind this lack of applications are two. Firstly, the problem of the diversity and heterogeneity of data sources: already a limiting factor in other fields of research, it is of utmost importance in ATM, as data usually come from different institutions, are recorded with different technical standards for different aims, and even comply non-homogeneous legislations. Secondly, most of the information that may have value from a Data Science perspective, also has a high value from the commercial point of view, or is considered sensitive for safety and security: thus, accessing such data sets is usually a challenge on its own. Even if some important problems have to be overcome, the added value that a Data Science analysis may yield to ATM is extremely high. Specifically, analyzing historical data sets may provide insights about the dynamics of the system, that cannot be easily discovered just by dint of manual analysis, or by relying on expert judgement.

The hypothesis behind this contribution is then the following: the comprehension of even very simple and well established operational concepts may be enhanced, or *fine tuned*, by means of proper Data Science analyses. In order to demonstrate this point, a comparison is developed between two different approaches for evaluating the traffic density within a sector, namely a standard aircraft count, and an assessment based on counting flights within concentric circles. Both approaches are compared, by means of Data Science techniques, against their capacity for explaining and forecasting the appearance of safety events, *i.e.* losses of separation. Results indicate that the latter method has a higher predictive capacity, which in turns reflects a higher capacity for explaining the mechanisms behind the appearance of safety events. Beyond the specific measures here compared, the relevance of this work resides in the general framework presented, *i.e.* the use of Data Science to improve the knowledge of an operational aspect of ATM.

This contribution is organized as follows. Section II introduces the criterium used to compare both ways of defining traffic density, *i.e.* the capacity of forecasting the appearance of a given type of safety events. Section III and IV respectively presents the two methods and their performance. Finally, Section V draws some conclusions.

II. Identification of safety events in real data

In order to compare the two aircraft density measures that will be presented in the following sections, it is firstly necessary to define a way of assessing their significance. In this work, the approach presented in Ref. [14] is followed, which envisions the use of the score of a classification task as a proxy of the quantity of information codified by each measure. In other words, the two traffic measures, that will be described in the next Section, are used to classify events into two categories, solved and unsolved safety events (as described below): a high classification score would then indicate the presence of valuable information for understanding the dynamics of the system, specifically the appearance of such events.

In the sake of simplicity, basic events will be considered, yet with important operational consequences: potential separation losses. These *iEvents* were identified by projecting the intentions of each aircraft (that is, the future route according to the filed flight plan) starting from a given position (obtained by the radar trajectory), and by detecting if two aircraft may break the separation minima in the near future. Notice that this is equivalent to the surveillance task performed by any Air Traffic Controller. Following this definition, iEvents include both events that may result in a safety-related condition (*e.g.*, a reduction of the separation between aircrafts), and situations that might have resulted in similar conditions, but in which the intervention of the controllers (or of the pilots) solved the problem before its appearance. By analyzing the real evolution of both flights, all iEvents have been classified in these two groups, called “unsolved iEvents” and “solved iEvents” respectively.

Trajectories data have been extracted from the ALL_FT+ data set, collected by the EUROCONTROL PRISME group. This includes information about planned, regulated and executed trajectories for all flights crossing the European airspace. The data set covers the period from 1^{st} March to the 31^{st} December 2011, including a total of 10.3 million flights. After some pre-processing, aimed at filter those trajectories that were suspected to not correspond to real flights, 100.032 iEvents were detected, 4.316 of which have been classified as unsolved. The interested reader may refer to Ref. [15] for further detail on iEvents definition and detection.

III. Standard approach: traffic density within sectors

Much work can be found in the Literature about analyzing the factors leading to safety-related events, with a special attention devoted to the concept of *traffic density* [16], [17]. This metric assesses the number of aircraft in a given sector: the higher the density, the more potential trajectory crossings are expected to appear, and thus the higher the risk of separation losses. While this metric does not include information about the structure of flows created by aircraft [18], it is still a principal indicator of the controller workload, which has a direct impact on the controller’s ability to successfully manage adverse operational conditions [19]. Thus, extremely high (but even extremely low [20]) workloads are expected to correlate with the appearance of unsolved iEvents.

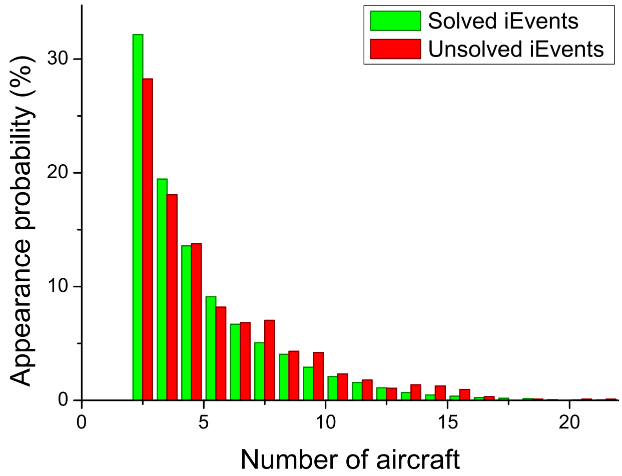

Fig. 1. *Traffic density within sectors.* Histograms of the traffic density in sectors corresponding to solved (green bars) and unsolved (red bars) iEvents.

Fig. 1 depicts the relationship between traffic density and appearance of unsolved iEvents, as obtained in the ALL_FT+ data set. Specifically, green (red) bars represent how solved (unsolved) iEvents are distributed as a function of the number of aircraft found in the sector of interest at the moment of the detected event. It can be appreciated how red bars slightly dominate in the right part, thus confirming that unsolved events are more likely to appear in high workload conditions. This is confirmed by an analysis of the statistical properties of both distributions: their means (4.504 and 4.854 for solved and unsolved iEvents, respectively) are different in a statistically meaningful way (two-sample T-test, p-value $< 10^{-5}$, significance level of 0.01).

Such small difference between both probability distributions can, in principle, be used to forecast if a given event is going to evolve into an unsafe situation. To this end, a *Naïve Bayes* classification algorithm (see [21], [22]) has been trained using this information, obtaining a classification score of 51.101% (significant at 0.01 in a binomial test). While this score is quite low, it is still relevant, as the high number of factors influencing the outcome of a safety-related event makes any prediction an extremely difficult task.

IV. Alternative approach: traffic density in the neighborhood of an event

As an alternative approach, in this Section we define the traffic density as the number of aircraft located within a circle centered on the event of interest. Fig. 2 depicts the calculation process; all blue aircraft, *i.e.* those located within the blue dashed circle, are included in the calculation, irrespectively on the sector they lay within. On the other hand, other aircraft

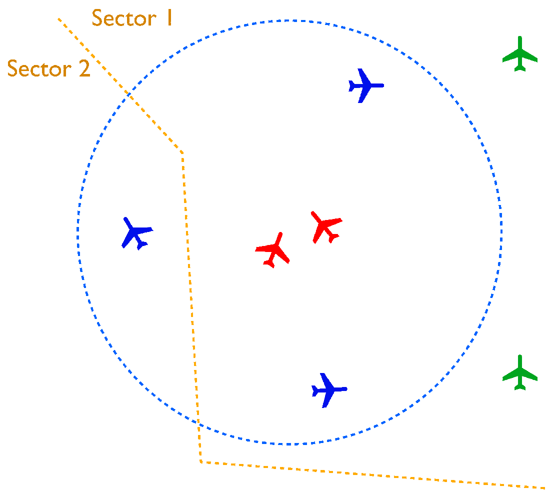

Fig. 2. *Alternative traffic density calculation*. Example of the calculation of the alternative traffic density measure. Flights are counted only when within a given distance from the event (blue dashed circle). Notice that aircraft within the sector (depicted by the dashed orange line), but outside this radius, are not considered (*e.g.*, the two green aircraft on the right part).

(in green), even if in the same sector, are not accounted for.

This approach implies two elements of flexibility not present in the classical case: *(i)* the circle radius can assume arbitrary values, thus including more or less traffic, and *(ii)* different radii can be combined together, thus yielding information integrating different spatial scales. While beneficial towards a better understanding of the system, such flexibility also results in the need for an objective way of optimizing the representation, *i.e.* of selecting the best number of circles and the best radii.

Before presenting the results corresponding to this approach, the next subsection depicts how such obstacle can be overcome by means of Data Science techniques.

A. Optimizing the representation

The selection of the best radii can be performed by means of a Data Science analysis, following the recipe described in Ref. [14]. By starting from an external classification as ground truth, in this case the designation of each iEvent as solved or unsolved, it is possible to use the output of a data mining classification task as a proxy for the relevance of the radii under study. Specifically, a high classification score indicates that the considered radius is correctly representing the structural differences between the two classes of iEvents; this yields criteria for an optimal traffic representation with respect to the considered problem.

According to this approach, instead of applying a single pre-determined radius r, in what follows a set of different radii $R = \{r_1, r_2, \ldots\}$ are considered, spanning from very small to large circles. For each one of these radii r_i, the traffic density associated to each iEvent is extracted and used to feed a classification algorithm; results are then presented as the score (in percentage) of this task.

In what follows, a *Naïve Bayes* algorithm ([21], [22]) has been used in all classification tasks; in order to validate these results, other classification algorithms have been also tried, *e.g. Decision Trees* and *K-Neightbors*, with qualitatively equivalent findings.

The information provided by the classification task is also complemented by the *Bhattacharyya distance*. Generally speaking, this distance measures the similarity of two discrete or continuous probability distributions, by assessing the amount of overlap between them [23], [24]. In this application, the Bhattacharyya distance is used to estimate how dissimilar are the traffic density distributions corresponding to solved and unsolved iEvents, in a way similar to what presented in Fig. 1. The higher the Bhattacharyya distance, the more different are the considered samples, thus the easier is to discriminate both types of events.

B. Single radius

Before moving to more complex information representations, here the case of a single radius is considered; the reader should notice that this is equivalent to assuming that the system may be characterized by a single spatial scale, clearly a poor approximation of the reality of ATM. Fig. 3 depicts the evolution of the Bhattacharyya distance and of the classification score as a function of the considered radius. Both functions have a maximum between 20 and 40 NM, indicating that the structural differences between solved and unsolved iEvents should be searched within this window.

The best classification score (54.04%) corresponds to 38.4 NM. As discussed in Section III, while such result is only slight above the score expected in a random classification, *e.g.* 50%, this is statistically meaningful (at $\alpha = 0.01$), and is not unexpected, due to the complexity of ATM, and to the high number of factors affecting the appearance of separation losses.

C. Multiple radii

Figs. 4 and 5 respectively represent the Bhattacharyya distance and the classification score corresponding to the si-

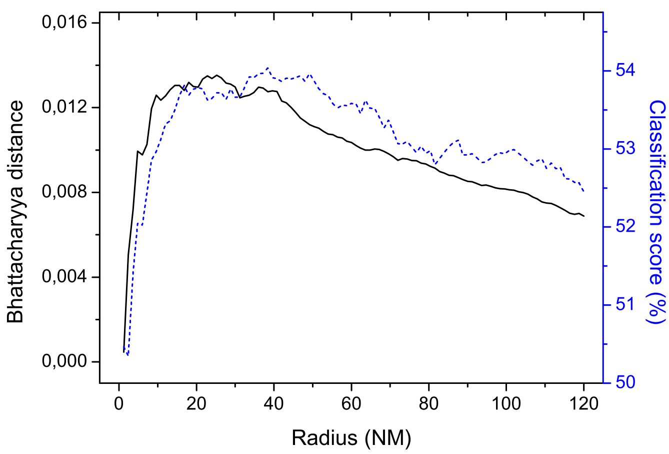

Fig. 3. *Analysis of a single radius*. Representation of the Bhattacharyya distance (left scale, black solid line) and of the classification score (right scale, dashed blue line) as a function of the radius size.

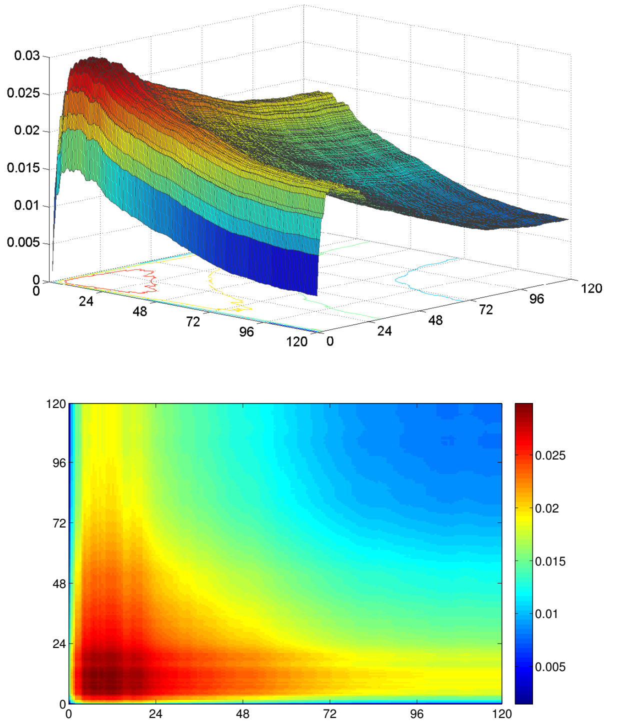

Fig. 4. *Bhattacharyya distance for two radii.* 3D plot (upper graph) and 2D color map (bottom graph) of the Bhattacharyya distance corresponding to the two probability distributions (for solved and unsolved iEvents, respectively) obtained by using two radii - see text for details. Axes values are expressed in NM.

multaneous use of two radii. A maximum in the Bhattacharyya distance is clearly visible between 5 and 20 NM. As with respect to the classification score, the best result is obtained for a combination of short and long range densities, respectively counting aircraft within 12 and 48 NM; the score reached, 54.13%, *i.e.* three times higher than the one obtained in Section III (51.101%) with respect to the random classification baseline (50%), indicates that this novel measure is able to extract significantly more information about the appearance of unsolved safety events.

This latter point deserves more attention. Fig. 6 depicts the morphospace [25] associated to this problem, namely a phenotype space where traffic densities are defined as the axes (corresponding to 12 and 48 NM respectively), and in which iEvents are located. In such morphospace, it is possible to define regions where one type of event is more frequent than the other; in other words, green points (red points) indicate combinations of aircraft densities for which safe (unsafe) iEvents are more frequent than unsafe (safe) ones, in a statistically significant way. Where no dot is plotted, there is no clear predominance of one of the two groups according to a binomial test (p-value of 0.01, significance level of 0.01). It can be appreciated that two main regions appear: one internal region, with a predominance of unsolved iEvents, and a second external one, where solved events are more frequent. The interpretation of this outcome is quite interesting: given an aircraft density for the large radius, unsolved iEvents appear less frequentlyy when most of the aircraft are located in the external part of such radius, *i.e.* their density is described by a *torus*, or when they are mostly concentrated within the small radius. Therefore, unsolved iEvents mostly appear when the traffic is homogeneously distributed in the airspace - which probably implies several attention spots for the controller, and thus a reduction in his / her situation awareness.

V. Conclusions

This contribution presented an analysis centered on the hypothesis that Data Science techniques can be used to improve the understanding of air transport and ATM operational elements. Instead of just relying on manual assessments or expert judgements, the scrutiny of historical data sets provides methods and techniques for selecting the most important features for the understanding of a given problem, in an objective and quantitative fashion. As a simple case study, the problem of quantifying traffic density has been here tackled. Data Science techniques can be used to compare different approaches, specifically for assessing which one is able to codify most information about the system under study. In particular, it has been found that a measure based on concentric circles yields significantly more knowledge than the classical sector-based approach. Furthermore, additional insights can be

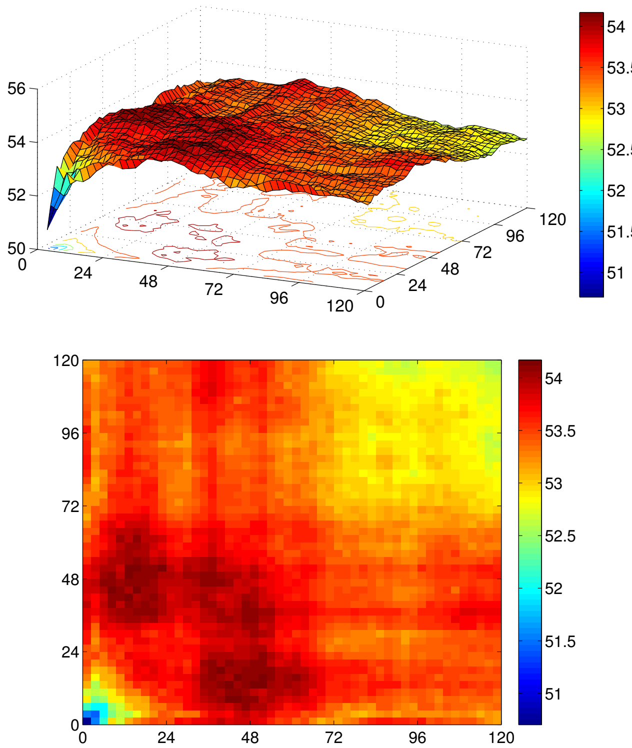

Fig. 5. *Classification score for two radii.* 3D plot (upper graph) and 2D color map (bottom graph) of the classification score (in $\%$) obtained by using two radii - see text for details. Axes values are expressed in NM.

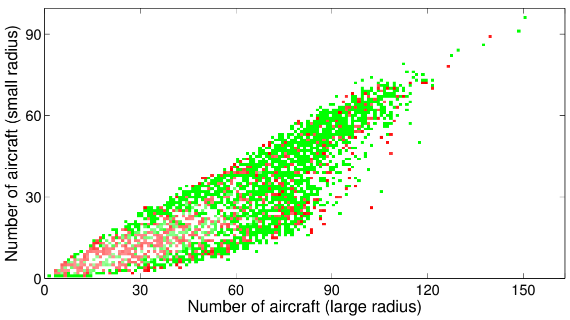

Fig. 6. *Characterization of iEvents in the aircraft density space*. Green (red) points indicate combinations of aircraft densities (for small and large radii of 12 and 48 NM respectively) for which solved (unsolved) iEvents are significantly more frequent - see text for details.

obtained by looking at the characteristics of iEvents, as for instance the importance of flight spatial distribution for the appearance of unsolved events.

In resume, this contribution stresses the importance of a very simple concept, successfully applied in other fields of research, but not yet fully taken into account in air transport and ATM: *listening to the data*.

References

[1] J. Dan, *Kabbalah: A very short introduction.* Addison-Wesley, England, 2005.

[2] R. Taton and C. Wilson, *Planetary Astronomy from the Renaissance to the Rise of Astrophysics.* Cambridge University Press, 1989.

[3] K. N. Abazajian *et al.*, *The seventh data release of the Sloan Digital Sky Survey.* The Astrophysical Journal Supplement Series 182, pp. 543, 2009.

[4] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky and D. Werthimer, *SETI @home: an experiment in public-resource computing.* Communications of the ACM 45 (11), pp. 56–61, 2002.

[5] M. Lamanna, *The LHC computing grid project at CERN.* Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 534 (1), pp. 1–6, 2004.

[6] N. B. Ellison, C. Steinfield and C. Lampe, *The benefits of Facebook "friends:" Social capital and college students' use of online social network sites.* Journal of Computer-Mediated Communication 12 (4), pp. 1143–1168, 2007.

[7] A. Java, X. Song, T. Finin and B. Tseng, *Why we twitter: understanding microblogging usage and communities.* In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007.

[8] W. S. Cleveland, *Data science: an action plan for expanding the technical areas of the field of statistics.* International statistical review 69 (1), pp. 21–26, 2001.

[9] S. Iwata, *Editor's Note: Scientific "Agenda" of Data Science.* Data Science Journal 7 (0), pp. 54–56, 2008.

[10] S. Alam, K. Shafi, H. A. Abbass and M. Barlow, *An ensemble approach for conflict detection in Free Flight by data mining.* Transportation Research Part C 17 (3), pp. 298–317, 2009.

[11] Z. Nazeri, *Data Mining of Air Traffic Control Operational Errors.* In Proceedings of the Conference on Data Mining, pp. 321–324, 2006.

[12] S. Pozzi, G. Lotti, G. Matrella and L. Save, *Turning information into knowledge. The case of automatic safety data gathering.* EUROCONTROL Annual Safety R&D Seminar, Southampton, UK, 2008.

[13] M. Zanin, *Synchronization Likelihood in Aircraft Trajectories.* In Proceedings of the Tenth USA/Europe Air Traffic Management Research and Development Seminar, Chicago, USA, 2013.

[14] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. García-Prieto, F. del Pozo, E. Menasalvas and S. Boccaletti, *Optimizing functional network representation of multivariate time series.* Scientific reports 2, 00630, 2012.

[15] M. Zanin and J. Loadman, *COMPASS Scenario Definition Report.* COMPASS project, E.02.12-220719-C6, 2012.

[16] I. V. Laudeman, S. G. Shelden, R. Branstrom and C. L. Brasil, *Dynamic density: An air traffic management metric.* No. NAS 1.15: 112226, NASA/TM-, 1998.

[17] A. J. Masalonis, M. B. Callaham and C. R. Wanke, C. R., *Dynamic density and complexity metrics for realtime traffic flow management.* In Proceedings of the 5th USA/Europe Air Traffic Management R&D Seminar, 2003.

[18] M. Zanin, *Measuring Aircraft Flows Complexity.* In Proceedings of the 27th International Congress of the Aeronautical Sciences, 2010.

[19] G. B. Chatterji and B. Sridhar, *Measures for air traffic controller workload prediction.* In Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum, 2001.

[20] R. M. Yerkes and J. D. Dodson, *The relation of strength of stimulus to rapidity of habit-formation.* Journal of Comparative Neurology and Psychology 18, pp. 459–482, 1908.

[21] T. Hastie, R. Tibshirani and J. J. H. Friedman, *The elements of statistical learning.* Springer, 2001.

[22] J. Han, M. Kamber and J. Pei, *Data mining: concepts and techniques.* Morgan Kaufmann, 2006.

[23] A. Bhattacharyya, *On a measure of divergence between two multinomial populations.* The Indian Journal of Statistics 7, pp. 401–406, 1946.

[24] T. Kailath, *The divergence and Bhattacharyya distance measures in signal selection.* IEEE Trans. on Comm. Tech. 15, pp. 50–60, 1967.

[25] G. R. McGhee, *Theoretical morphology: the concept and its applications.* Columbia University Press, 1999.