JOURNAL OF AIRCRAFT
Vol. 42, No. 3, May–June 2005

Text Mining the Global Abrupt-Wing-Stall Literature

Ronald N. Kostoff*
Office of Naval Research, Arlington, Virginia 22217
Gabriel Karpouzian†
U. S. Naval Academy, Annapolis, Maryland 21402
and
Guido Malpohl‡
University of Karlsruhe, 76128 Karlsruhe, Germany

Text mining was used to derive technical intelligence from an abrupt-wing-stall database derived from the Science Citation Index database. Both concept and document clustering were used to provide the structural taxonomy of the global abrupt-wing-stall literature and the estimated relative levels of effort in the major subcategories. Bibliometric analysis of the abrupt-wing-stall literature generated author/journal/institution publication and citation data.

I. Background

ABRUPT wing stall (AWS) reflects uncommanded lateral motions of high performance aircraft. In military operations, AWS can lead to loss of target tracking. AWS variants have occurred on almost every high-performance aircraft for the past five decades.[1] In the late 1990s, the U.S. Navy, U.S. Air Force, and NASA undertook a joint research program to understand the fundamental causes of AWS. Prior to program initiation, a historical literature survey of AWS was commissioned, and a report was issued by Chambers in 1999 (Ref. 1).

The present study was initiated to complement the Chambers report. The main study objective was to identify global science and technology (S&T) that had both direct and indirect relations to abrupt wing stall. One subobjective was to estimate the overall relative level of global effort in abrupt wing stall S&T, as reflected by the emphases in the published literature. Another subobjective was to identify the linkage between the historical references from the modern published literature and the historical references in the comprehensive literature survey conducted by Chambers.

II. Approach

To execute the study reported in this paper, a database of relevant abrupt-wing-stall articles was extracted from the Science Citation Index (SCI),[2] using the iterative search approach of simulated nucleation.[3] The Web version of the SCI accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research). The extracted database represents a fraction of the available abrupt-wing-stall (mainly research) literature, which in turn represents a fraction of the abrupt-wing-stall S&T actually performed globally. It does not include the large body of classified literature or company proprietary technology literature. It does not include technical reports or books or patents on abrupt wing stall. It covers a finite slice of time (1991–mid-2002). The database used represents the bulk of the peer-reviewed high-quality abrupt-wing-stall research literature.

The database was analyzed to produce the following characteristics and key features of the AWS global literature: 1) recent prolific AWS authors; 2) journals that contain numerous AWS papers; 3) institutions that produce numerous AWS papers; 4) keywords most frequently specified by the AWS authors; 5) authors, papers, and journals cited most frequently; and 6) the technical structure (taxonomy) of the AWS literature.

Because of space limitations, only summary results are presented here. The readers interested in the detailed bibliometric and clustering results are advised to obtain Ref. 4, which also contains in-depth descriptions of the text mining processes used.

The bibliometric portion of the study was based on counts of authors, journals, institutions, countries, and citations. Once the relevant AWS records were obtained, the bibliometrics tabulation was straightforward.

The computational linguistics (clustering/taxonomy generation) portion of the study was more complicated computationally. Two methods of clustering were used: concept clustering[5,6] and document clustering.[7,8] In concept clustering, related words/ phrases are grouped into clusters, based on their cooccurrence within some computational domain (e.g., journal paper abstract). In document clustering, documents are grouped into clusters based on some measure of their similarity. Both clustering approaches offer different perspectives on the structure of the technical literature. Neither is inherently superior.

III. Results

A. Bibliometrics: Comparison of Cited Papers in Present Study with Chambers' Study References

The initial Chambers literature survey report[1] contained 37 separate references, mostly organizational reports. Only one reference linked to the peer-reviewed published journal literature. An updated historical survey was published in 2003 (Ref. 9). The new survey contained 61 separate references, of which 24 were in common with the 1999 survey. Seventeen of the 61 references in the new study were to papers resulting from the joint NASA/Navy/Air Force Abrupt Wing Stall program. All of these 17 papers (plus Chambers' updated study) were presented in three special abrupt-wing-stall sessions at the 41st Aerospace Sciences Meeting and Exhibit in Reno, Nevada, in January 2003. In the updated survey, three references linked to the peer-reviewed published journal literature, compared to one in the initial survey.

One initial purpose of the present text mining study was to examine the relatively recent peer-reviewed published journal literature to both complement the 1999 Chambers' study and to ascertain the connectivity between the references in the abrupt-wing-stall published

Received 2 June 2003; revision received 24 October 2003; accepted for publication 24 October 2003. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States. Copies of this paper may be made for personal or internal use, on condition that the copier pay the $10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 0021-8669/05 $10.00 in correspondence with the CCC.

*Physical Science Administrator, 800 North Quincy Street; kostofr@onr.navy.mil.
†Professor, Aerospace Engineering Department; karpouzi@usna.edu.
‡Doctoral Candidate, Department of Informatics, Postfach 6980; Malpohl@ipd.uka.de.

The views in this paper are solely those of the authors, and do not represent the views of the Department of the Navy or any of its components, or the University of Karlsruhe.

literature and the references in Chambers' 1999 report. (This goal was updated to include Chambers' 2003 report as well.) To accomplish this expanded goal, all of the references cited in the present study were searched to see which of the separate references in the Chambers first and second reports were also contained in the present study.

From Chambers 1999 study,[1] only three of Chambers' 37 references are in common with the 40,245 references in the present study. Agreement between the 2003 Chambers' study[9] and the present study is moderately better (seven out of Chambers' 61 references in common with the present study). For a relatively focused topic such as abrupt wing stall, one would have expected a more substantial overlap. Many of Chambers' references appear to be quite applied, with focus on tests of specific aircraft. In contrast, the SCI papers tend to be more fundamental, concentrating on the analysis of basic flowfield phenomena and physics. In addition, authors tend to cite even more fundamental references, and so the combination of fundamental papers from the SCI retrieval coupled with their even more fundamental references identifies a fundamental set of cited papers that complements the more applied set in Chambers' study.

These results offer further insight into one of the main findings of Chambers' studies, namely, that abrupt-wing-stall-related problems have occurred on almost all high-performance aircraft since the early 1950s. As Chambers' studies show, many of the abrupt-wing-stall-related problems on specific aircraft tend to be published in contractor reports or other organizational reports not easily accessed to the degree of the SCI, if they are published at all. Neither the sponsors, vendors, or researchers and developers are motivated to report widely such problems on aircraft that they have funded or developed. Again, as Chambers' reports show, the developers generate "band-aid" approaches to resolve the problem for a particular aircraft. However, the absence of widely available and easily accessible documentation leads to repetition of the problem on future aircraft. This history emphasizes the need for a nationally mandated policy to document all federally funded research and development (whose results are both positive and negative) widely in the readily accessible published literature, so that mistakes in the future can be avoided by lessons from the past.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly cited papers listed in Ref. 4 and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

B. Computational Linguistics Results

Before the elemental clusters and the resultant taxonomy are described, the organic differences between AWS and technical disciplines text mined previously require some discussion. Most previous text mining studies (performed by the first author) using concept or document clustering focused on basic physical science disciplines (e.g., analytical chemistry,[6] electrochemical power sources,[5] nonlinear dynamics,[10] and fractals[11]). The focus tended to be on basic phenomena, rather than including systems or technologies. A strong disciplinary thread at a consistent descriptive level throughout the data allowed division of thematic categories into relatively crisp and complementary subcategories. At any hierarchical level in the taxonomy, the categories were sharply defined and complementary (e.g., nonlinear phenomena could subdivide into chaotic nonlinear and nonchaotic nonlinear).

For text mining studies that did not focus exclusively on a single descriptive level, such as fundamental phenomena, but focused on multiple descriptive levels, such as fundamental phenomena, technologies, and systems, the taxonomy categories had a different type of structure. In those cases, there was a competition in the algorithm among phenomenon subdivision, technology subdivision, and systems subdivision. One technology could contain multiple phenomena, and one system could contain multiple technologies. Sometimes the phenomena subdivision dominated, sometimes the technology subdivision dominated, and sometimes the system subdivision dominated.

Abrupt wing stall is an engineering science discipline, rather than a physical science discipline. It is not a single descriptive level discipline in the sense of chemistry or physics, but rather is a multi-meta-level descriptive discipline. Many abrupt-wing-stall papers integrate the phenomena with the technology, or the technology with the system, or all three. The consequence for clustering is that the larger cluster themes are not thematically pure in every case. In particular, some of the smaller clusters under the larger computational cluster have both computational and experimental components, although different from the surface-pressure measurements characteristic of the other major cluster.

The concept clustering taxonomy results showed that the central abrupt-wing-stall theme is understanding and controlling the main factors that lead to vortex breakdown, flow separation, and eventual stall. The highest taxonomy level can be divided into two categories, each relating to how vortex breakdown and separation are induced. The first category covers angle-of-attack induced vortex breakdown and flow separation leading to stall. It focuses on vortical flows over delta wings at high angles of attack, emphasizing surface-pressure distributions and vortex breakdown locations. The second category covers other causes of vortex breakdown and flow separation leading to stall, mainly shock-induced and aeroelastic coupling induced. It focuses on numerical computations of transonic shock-turbulent boundary-layer flows, as well as aeroelastic coupling between the flowfield and the body.

The next highest taxonomy level can be divided into four categories. The first category focuses on vortex breakdown on high-angle-of-attack delta wings and its subsequent impact on flight dynamics, supplemented by laboratory experiments that emphasize surface (especially pressure) measurements. Its five component subcategories are as follows: 1) vortical flows over delta wings at angles of attack, concentrating on measuring surface-pressure distributions, determining kinematics, dynamics, and breakdown location of the vortices emanating from the leading edges, and examining vortex shedding from the trailing edges; 2) airfoil flow separation and subsequent stall over a range of Reynolds Numbers and freestream Mach numbers; 3) delta-canard configurations, emphasizing vortex structure, dynamics, and breakdown, and their impact on stall; 4) temporal amplitude and frequency of unsteady pitching and rolling motions, especially the impact of the rolling forces and moments on the aircraft flight controls and dynamic stall, and methods to dampen the rolling moments for improved flight dynamics control; and 5) visualization of flow phenomena over swept wings in wind and water tunnels, for the purpose of identifying vortex bursting and breakdown, using root chord and apex to nondimensionalize the results.

The second category focuses on maximum lift/drag augmentation by leading-edge flap deflections that increase camber, smooth the flow to reduce suction, retard vortex breakdown, and delay separation and stall. It cannot be subdivided further.

The third category focuses on numerical solutions of the Navier–Stokes and Euler equations for transonic shock waves, mainly focused on shock interactions with the turbulent boundary layer that lead to vortex breakdown and flow separation from the shock-driven adverse pressure gradients and including mass bleeding to control separation. Its five component subcategories are as follows:

1) The first subcategory uses numerical solutions of the compressible Navier–Stokes and Euler inviscid equations, emphasizing implicit algorithms with finite volume mesh integration, to simulate shock-wave interactions in transonic flow and generate complex three-dimensional supersonic and subsonic flowfield structures.

2) The second subcategory uses accurate spatial discretization finite volume codes with high-order upwind smoothing operators to solve coupled Navier–Stokes/turbulence transport equations, emphasizing the one-equation Spallart–Almaras flow turbulence model.

3) The third subcategory includes turbulent boundary-layer separation related to pressure gradients, especially to adverse pressure gradients caused by shock-induced pressure increases. Emphasized are the laminar-turbulent transition, separation bubble growth, and subsequent flow reattachment. Adequacy of kinematic eddy

viscosity closure and k–ϵ wall region models is examined, and adequacy of measurements of turbulent shear stress, mean turbulence energy, skin friction, and velocity profiles in different layers is studied.

4) The fourth subcategory uses direct rapid numerical simulation of incompressible Navier–Stokes equations for the inner viscous region, coupled to an integral formulation for obtaining inviscid external potential flows.

5) The fifth subcategory includes normal and oblique shock-wave/boundary-layer interactions. Emphasized are normal and slanted bleed slots upstream, downstream, and across the shock impingement point to retard separation and delay stall. Boundary-layer displacement and momentum thickness distributions are provided.

The fourth category focuses on oscillatory aeroelastic interactions affecting flight dynamics and methods to control the vortex breakdown locations and shedding that drive these unsteady oscillations. Its six component subcategories are as follows:

1) The first subcategory is the use of high-image-density particle image velocimetry to study vortex breakdown. Emphasized are instantaneous three-dimensional vorticity structures and streamline patterns, including crossflow plane variations, and especially velocity field structures and vorticity contours.

2) The second subcategory is vortex kinematics and dynamics leading to breakdown and eventual stall, concentrating on the axial vorticity and axial-to-crossflow energy evolution in the primary and secondary vortex cores, as the vortex cores evolve in the streamwise direction, and on the parameters that can delay or retard the transient onset of vortex breakdown. Emphasized are velocity field data at different incidence angles at different chordwise locations above 75-deg delta wings and their relation to conical flow behavior.

3) The third subcategory is the simulation of aircraft forebody tangential jets blowing along the leeward side to delay the onset of leading-edge vortex breakdown and subsequent stall.

4) The fourth subcategory is linear and nonlinear aeroelastic responses to excitations and disturbances, emphasizing the evolution from multifrequency responses at low speed through single-frequency responses at critical speed (signifying the onset of flutter) to limit-cycle oscillations accompanying wing rock at higher speed.

5) The fifth subcategory is dominant mechanisms of characteristic flow modes of vortex formation, surface evolution, and shedding in the wake of NACA airfoils, emphasizing the correlation between the different evolution processes of the surface flow and the wake instabilities, as well as the evolution to stable vortex shedding in the wake.

6) The sixth subcategory is laser-driven flow visualization around fighter aircraft, with emphasis on buffeting of vertical tail/fin (especially on the F/A-18 aircraft), and feedback from the tail location and dynamics on the vortex breakdown location and eventual stall. Rigid and flexible fins are studied, including root bending of the flexible fins to obtain the buffeting response, with emphasis on spectral power fluctuations, especially focusing on peak power spectra fluctuations.

The document clustering results showed that about 58% of the documents in the database can be classified as angle-of-attack induced stall, whereas the remaining 42% can be viewed as shock and aeroelastic coupling induced stall, subject to the uncertainties caused by phenomena-technology systems and experimental-theoretical mixing discussed earlier. Angle-of-attack induced stall subdivides into slightly over $\frac{5}{6}$ angle-of-attack variation and slightly under $\frac{1}{6}$ lift augmentation because of effectively changing angle of attack through flap modifications. Shock-and-aeroelastic-coupling induced stall subdivides into slightly under $\frac{4}{5}$ shock-induced stall because of boundary-layer separation, and slightly over $\frac{1}{5}$ aeroelastic-coupling induced stall because of change in vortex breakdown location from aeroelastic feedback.

IV. Conclusions

The open literature references were almost disjoint with the contractor-report oriented references. Both studies served to complement each other. Future literature surveys should integrate the Chambers' report traditional methodology with the present text mining approach.

Concept clustering was appropriate for identifying the technical structure of the AWS literature. Document clustering provided a similar technical structure for the AWS literature and provided estimates of relative levels of effort as well.

Appendix: Text Mining Bibliometrics Overview

The combination of text mining and bibliometrics is being developed by different researchers for these and many other applications. Its component capabilities are as follows.

Science and technology (S&T) text mining[12–15] is a process for extracting useful information from large volumes of technical text, based mainly on the mechanics of computational linguistics. It identifies pervasive technical themes in large databases from frequently occurring technical phrases. It also identifies relationships among these themes by grouping (clustering) these phrases (or their parent documents) on the basis of similarity. Text mining can be used for the following: 1) enhancing information retrieval and increasing awareness of the global technical literature[3,16,*]; 2) potential discovery and innovation based on merging common linkages between very disparate literatures[17–20]; 3) uncovering unexpected asymmetries from the technical literature[21,22]; 4) estimating global levels of effort in S&T subdisciplines[23,24]; 5) helping authors potentially increase their citation statistics by improving access to their published papers and thereby potentially helping journals to increase their impact factors[10,11]; and 6) tracking myriad research impacts across time and applications areas.[25,26]

A typical text mining study of the published literature develops a query for comprehensive information retrieval, processes the database using computational linguistics and bibliometrics, and integrates the processed information.

Evaluative bibliometrics[27–29] uses counts of publications, patents, citations, and other potentially informative items to develop science and technology performance indicators. Its validity is based on the premises that 1) counts of patents and papers provide valid indicators of research and design activity in the subject areas of those patents or papers; 2) the number of times those patents or papers are cited in subsequent patents or papers provides valid indicators of the impact or importance of the cited patents and papers; and 3) the citations from papers to papers, from patents to patents, and from patents to papers provide indicators of intellectual linkages between the organizations that are producing the patents and papers, and knowledge linkage between their subject areas.[30] Evaluative bibliometrics can be used to do the following: 1) identify the infrastructure (authors, journals, institutions) of a technical domain, 2) identify experts for innovation-enhancing technical workshops and review panels; 3) develop site visitation strategies for assessment of prolific organizations globally; and 4) identify impacts (literature citations) of individuals, research units, organizations, and countries.

References

[1] Chambers, J. R., "Historical Review: Perspectives on Experiences with Uncommanded Lateral Motions at High-Subsonic and Transonic Speeds," Ball Aerospace and Technologies Corp., Aerospace Systems Div., Boulder, CO, Feb. 1999.

[2] *Science Citation Index*, Inst. for Scientific Information, Philadelphia, 2002.

[3] Kostoff, R. N., Eberhart, H. J., and Toothman, D. R., "Database Tomography for Information Retrieval," *Journal of Information Science*, Vol. 23, No. 4, 1997, pp. 301–311.

[4] Kostoff, R. N., Karpouzian, G., and Malpohl, G., "Abrupt Wing Stall Roadmaps Using Database Tomography and Bibliometrics," Naval Air Warfare Center, Aircraft Div. TR NAWCAD PAX/RTR-2003/164, Patuxent River, MD, June 2003.

[5] Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A., "Electrochemical Power Source Roadmaps Using Bibliometrics and Database Tomography," *Journal of Power Sources*, Vol. 110, No. 1, 2002, pp. 163–176.

*Data available online at http://trec.nist.gov/.

[6]Kostoff, R. N., and DeMarco, R. A., "Science and Technology Text Mining," *Analytical Chemistry*, Vol. 73, No. 13, 1 July 2001, pp. 370–378A.

[7]Willet, P., "Recent Trends In Hierarchical Document Clustering: A Critical Review," *Information Processing and Management*, Vol. 24, 1988, pp. 577–597.

[8]Rasmussen, E., "Clustering Algorithms," *Information Retrieval Data Structures and Algorithms*, edited by W. B. Frakes and R. Baeza-Yates, Prentice–Hall, Upper Saddle River, NJ, 1992, pp. 419–442.

[9]Chambers, J. R., and Hall, R. M., "Historical Review of Uncommanded Lateral-Directional Motions at Transonic Conditions," AIAA Paper 20034-0590, Jan. 2003.

[10]Kostoff, R. N., Shlesinger, M., and Tshiteya, R., "Nonlinear Dynamics Roadmaps Using Bibliometrics and Database Tomography," *International Journal of Bifurcation and Chaos*, Vol. 14, No. 1, Jan. 2004.

[11]Kostoff, R. N., Shlesinger, M., and Malpohl, G., "Fractals Roadmaps Using Bibliometrics and Database Tomography," (to be published).

[12]Kostoff, R. N., "Text Mining for Global Technology Watch," *Encyclopedia of Library and Information Science*, 2nd ed., edited by M., Drake, Vol. 4, Marcel Dekker, New York, 2003, pp. 2789–2799.

[13]Hearst, M. A., "Untangling Text Data Mining," *Proceedings of ACL 99, the 37th Annual Meeting of the Association for Computational Linguistics*, Univ. of maryland, June 20–26, 1999, pp. 3–10.

[14]Zhu, D. H, and Porter, A. L., "Automated Extraction and Visualization of Information for Technological Intelligence and Forecasting," *Technological Forecasting and Social Change*, Vol. 69, No 5, 2002, pp. 495–506.

[15]Losiewicz, P., Oard, D., and Kostoff, R. N., "Textual Data Mining to Support Science and Technology Management," *Journal of Intelligent Information Systems*, Vol. 15, No. 2, 2000, pp. 99–119.

[16]Greengrass, E., "Information Retrieval: An Overview," National Security Agency, TR-R52-02-96, 1997.

[17]Swanson, D. R., "Fish Oil, Raynauds Syndrome, and Undiscovered Public Knowledge," *Perspectives in Biology and Medicine*, Vol. 30, No. 1, 1986, pp. 7–18.

[18]Swanson, D. R., and Smalheiser, N. R., "An Interactive System for Finding Complementary Literatures: A Stimulus to Scientific Discovery," *Artificial Intelligence*, Vol. 91, No. 2, 1997, pp. 183–203.

[19]Kostoff, R. N., "Stimulating Innovation," *International Handbook of Innovation*, edited by Larisa V. Shavinina, Elsevier, Oxford, England, UK, Pt. 6, 2003, Chap. 5, pp. 388–401.

[20]Gordon, M. D., and Dumais, S., "Using Latent Semantic Indexing for Literature Based Discovery," *Journal of the American Society for Information Science*, Vol. 49, No. 8, 1998, pp. 674–685.

[21]Goldman, J. A., Chu, W. W., Parker, D. S., and Goldman, R. M., "Term Domain Distribution Analysis: A Data Mining Tool for Text Databases," *Methods of Information in Medicine*, Vol. 38, No. 2, 1999, pp. 96–101.

[22]Kostoff, R. N., "Bilateral Asymmetry Prediction," *Medical Hypotheses*, Vol. 61, No. 2, 2003, pp. 265–266.

[23]Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A., "Database Tomography Applied to an Aircraft Science and Technology Investment Strategy," *Journal of Aircraft*, Vol. 37, No. 4, 2000, pp. 727–730.

[24]Viator, J. A., and Pestorius, F. M., "Investigating Trends In Acoustics Research from 1970–1999," *Journal of the Acoustical Society of America*, Vol. 109, No. 5, Pt. 1. 2001, pp. 1779–1783.

[25]Davidse, R. J., and Van Raan, A. F. J., "Out of Particles: Impact of CERN, DESY, and SLAC Research to Fields Other Than Physics," *Scientometrics*, Vol. 40, No. 2, 1997, pp. 171–193.

[26]Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A., "Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling," *Journal of the American Society for Information Science and Technology*, Vol. 52, No. 13, 2001, pp. 1148–1156.

[27]Narin, F., "Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity (monograph)," National Science Foundation, NSF C-637, Contract NSF C-627, NTIS Accession No. PB252339/AS, Computer Horizons, Inc., Chery Hill, NJ, 1976.

[28]Garfield, E., "History of Citation Indexes for Chemistry—A Brief Review," *Journal of Chemical Information and Computer Sciences*, Vol. 25, No. 3, 1985, pp. 170–174.

[29]Schubert, A., Glanzel, W., and Braun, T., "Subject Field Characteristic Citation Scores and Scales for Assessing Research Performance," *Scientometrics*, Vol. 12, Nos. 5–6, 1987, pp. 267–291.

[30]Narin, F., Olivastro, D., and Stevens, K. A., "Bibliometrics Theory, Practice and Problems," *Evaluation Review*, Vol. 18, No. 1, 1994, pp. 65–76.