Executive summary:

This report provides a summary of the SVETLANA project: an EC – Russia cooperation project in the field of aviation safety and maintenance improvement. The project is aimed at improving the capabilities of flight data monitoring programmes for civil aviation by using the EC-Russian synergy. In front of the growing volume of air traffic and higher requirements from safety management systems, the aviation industry is challenged to even further improve the aviation safety record. The volume of recorded flight data has increased so significantly that analysis tools are challenged to release the full potential of all the data available.

In order to provide answers to these challenges, the SVETLANA project developed a new analysis cycle process based on a 2 steps approach: first an automated detection process based on data-mining techniques and then a human expertise feedback loop to assess smart maintenance and flight safety improvement. In the frame of the project, a common basis for the processing of all available flight data and anomaly detection in real-time has been designed. Several data-mining algorithms (2 patents applied) were developed, based on the most recent research outcome in the field of anomaly detection, they are fully data-driven and not base on threshold exceedence anymore.

The core of SVETLANA system, the algorithmic toolset, is embedded in a completely new architecture, extensible, compatible with additional data sources through common formats, distributed and highly relying on widely adopted aeronautical standards. This new flight data analysis environment has then been assessed with both real and simulated data, focusing on currently operated fleets in order to target short term exploitability. Finally, to support the project's developments and progress, an End User group has been involved to provide requirements and to assess operational benefits.

Several building blocks of the architecture and results have been successfully assessed in a representative environment. Combined with the robust and novel improved flight phase detection techniques (patent applied), the algorithms provide seamlessly a better accuracy, revealing unexpected anomalies; the research aiming at reducing the rate of false events removes almost all flaws in the raw data, anomaly detection algorithms developed during SVELANA have been successfully utilized in the medical environment to develop a real-time patient monitoring and prediction system. It is also necessary to highlight the fact that all algorithms and methods issued from SVETLANA are independent of the type of aircraft as well as of the type and number of parameters analyzed.

These results leaded to short term and direct exploitation plans in legacy systems despite some limitations issued from features not yet mature enough. We observe still a high sensibility to data quality and to flight dynamics, which can raise many minor or identical abnormalities. In addition, the correct classification of statistically detected abnormalities remains a challenge.

Finally, to anticipate a future implementation, the consortium of the project recommends that enough efforts are allocated to the definition of almost independent sub-systems for an efficient analysis and a reliable classification of the anomaly. Also, we expect the promotion of common format and preprocessing methods to offer additional services or

functionalities and ease flight safety related data exchanges among stakeholders.

Project Context and Objectives:

Air transport is a very safe mode of transportation. Due to design improvements and safety programmes the aviation accident rate has been significantly reduced since the 1960s, although the number of departures has been steadily increasing. From 1960 to 1970 the accident rate, as measured in terms of accidents per million departures, fell drastically. However, the accident rate has remained approximately unaltered over the last decades whilst air transport has continued to grow.

As air traffic is projected to grow in the future and since accidents are becoming more costly (among others due the impact on the aviation travel caused by world-wide coverage of instantaneous accessible media) the aviation industry realizes that an increase in the number of accidents is not acceptable. This means that the accident rate needs to be reduced in the future. At the same time the airline community faces difficult times due to the effects of the financial crisis in combination with the need to invest heavily in green solutions.

The EU has set the objective to achieve a five-fold reduction in the average accident rate of global commercial aviation operators as expressed in European Aeronautics Vision for 2020. In order to achieve this there is a high interest to extend EC and international law to neighbouring countries and particularly to the Russian Federation in order to create a European-wide aviation area that should lead to equally high standards in terms of aviation safety across the European continent. Around 75% of all Russian passenger traffic is directed towards European directions and this is forecast to grow annually at a rate of 5.8% in the next years. For the EU, the Russian federation is currently the fourth largest foreign aviation market. However different approaches to safety and environmental impact are a constant source for potential misunderstandings in EU/Russia aviations relations and increase risks of accidents. There are substantial economic and political reasons for the European Community to define a coherent and common aviation policy with its major trading partners, notably Russia, with whom the EU is developing a close economic and political partnership under the framework of the Partnership and Cooperation Agreement (PCA)

Russia airlines still operate a significant number of old aircraft, which are non-compliant with international and European safety and environmental standards. The growing intensity of flight safety concerns in Russia has caused not only members of the aviation community, but also the society at large to search for new ways of increasing the level of air transport safety. In tackling flight safety problems in the Russian Federation it must take into account the rapid growth of air traffic. Solutions that can be rapidly implemented are urgently requested.

Russian Federation engaged to improve aviation safety in Russia by promoting international standards Air Transport Association (IATA) and International Advisory Committee (IAC) Agreement - 19 April 2009). Expert exchange in international experience and best practices is vitally important for the development of complex measures and solutions of flight safety issues. A great interest exists to develop common standards with EU in order to extend the market opportunities for Russian Federation In the EU area.

The EU and Russia are working towards new Partnership and Cooperation Agreement (PCA) with the EU. As a result the four "Common Spaces" were defined at the EU-Russia Summit in St. Petersburg in May 2003. In particular the Common Economic Space (CES) was seen as a important milestone as this agreement represents the common EU/Russia economic and political interests.

Regulatory authorities as well as airlines are aiming to develop EU-Russia aviation relations in-line with international norms. If they get it right this has the potential to yield huge economic benefits. Working together, the two aviation communities can generate improved market opportunities, ensure compliance with each other's laws, promote convergence of aviation laws, and establish joint mechanisms for co-operation on security, safety and environmental standards.

Despite the high interest in Flight Data Monitoring and the promising results, current systems are not used to their full potential. The efficiency of the Flight Data Analysis process (FDA) is limited due to:
- the lack of harmonized methods between the different actors. FDA is currently "event based" and not strictly "data based". Due to the fact that each airline has its own event detection processes, there is no common methodology to analyse flight data.
- the fact that the recorded data is only analysed to a limited extend. One reason is because of the high workload for expert staff to operate the FDA process in a systematic approach. Often data needs to be analysed with intense understanding of the context of the operation. Currently, this involves a lot of labour, making it difficult to explore the full potential of the data.
- the huge amount of currently recorded data that is not suited for systematic validation. Currently the FDA process is often limited to an individual aircraft or an individual flight and it does rarely allow the combination of data from different aircraft or flights. It has also the limitations in the ability to detect complex events and unforeseen events.
- the coexistence of various aviation standards and regulations worldwide. The need for common standards or at least of common methodology to compare FDM analysis results has been identified together with the need of a common approach towards risk identification and then enhance flight safety
- The lack of interoperability of the FDM systems also highlights the variety of the solutions in use today.

Both, in the field of aviation safety and maintenance planning, improvements can be made by more intelligent use of the available flight data. In order to achieve this, important challenges need to be addressed.

The first challenge will be to be able to handle the data volume from various sources (different aircraft from different airlines) and to conduct automatic and systematic data analysis on a daily basis for all flights. The improved flight safety evaluation and advanced maintenance scheduling will depend on the ability to process extensively large amounts of flight. This requires the development of appropriate data mining techniques in order to extract from this huge amount of data the information that the expert can handle for safety or maintenance recommendation purpose.

The second challenge will be to quantify the detection level of flight minor event for such innovative data mining technique.

The third challenge shall be to design a new analysis process economically viable. Handling these FDA processes will always need the involvement of domain experts. However, in order to reduce costs the new process shall depend as much as possible on computing power and limit the dependence on expert resources as the investment in advanced flight data analysis technology, methods and procedures needs to be compensated by significant cost reduction in flight data management programmes or maintenance activities.

The last challenge is to create common standards in the EU and Russia. The viability of an automated process will rely on its ability to process data from different standards with limited human effort: the new processes shall be applicable for airlines in EU and Russia. In order to be able to handle this data in automated processes with limited effort a certain level of standardisation will be needed in FDM solutions (which are currently "event based", and events very specific to each airline). This would also improve comparison of processing results to compare different standards with the highest levels.

In response to these challenges the SVETLANA project has been initiated with the first objective to develop an advanced flight data analysis philosophy. The analysis process is conducted in a 2-step approach:

1. In a first step, innovative data mining techniques are applied automatically on all flight data to identify flight "singularities". Such technique is applied on all available flight data without any hypothesis on the expected "singular" data.
2. In a second step, limited to the automatically detected flight "singularities", flight operation and maintenance experts provide their recommendation regarding the detected flight event (or analyse it as a non-event)

A second main objective of SVETLANA project is to involve important EU and Russian Federation industries and research organisations from the aviation area into a common ambitious effort to enhance flight operational safety and initiate smart maintenance through intelligent and automated Flight Data Monitoring (FDM). More specifically, SVETLANA aims to provide an advanced smart update of flight safety by:

- Deploying an automated and standardised flight data management cycle that is capable of processing routinely large amounts of data to allow operators to examine all data from every flight deeply using advanced sophisticated algorithms,

- Propose a common standard methodology for flight data analysis using singular points automation based on data from various systems and operators to be combined and processed by advanced algorithms. Analysis on a larger statistical footprint with data from various aircraft types or operators would become comparable and could be combined,

- Identifying, detecting and correcting potentially unsafe trends before they manifest themselves in an incident using self-learning adapted methodologies,

- Performing a study of potential parameters or assessment techniques that can help in prediction of the condition or failure of hardware on the basis of current FDR data and can reduce the amount of corrective maintenance and provide better insight for predictive maintenance cycles,

- Limit the need for specialist involvement in the FDM cycle only to the decision making process,

- Providing insight into abnormal events in order to adjust training, maintenance and procedures to prevent re-occurrence,

- Establishing isolation criteria templates for well-known conditions and situations from large simulated datasets, allowing a broader and earlier identification of abnormal conditions,

- Validate and assess the advanced FDM Methodology in a simulated/synthetic environment,

- Provide a common standardised solution that has acceptance in both the EU and the Russian Federation.

Beyond all these objectives, the SVETLANA project expects to provide several benefits, among them:
- To improve flight safety: SVETLANA shall design an automated and standardised flight data management cycle that is capable of processing routinely large amounts of data from various sources in order to improve flight safety evaluation.
- To improve maintenance support: the customization of the FDM cycle expects additional benefits to support smart maintenance processes, as airline operations and maintenance departments will be informed about a potentially unsafe situation and how to detect it.
- To analyse more flight data to a more extensive level: this will lead to the reduction of expert involvement by qualifying the singularities detected automatically (and not preselected by experts). As a result human intervention will be used only to validate critical decisions for flight safety and maintenance enhancements.
- To provide feedback to inform stakeholders and improve the FDM process: SVETLANA shall refine the search of singularities that have not been selected for immediate risk mitigation. The most appropriate parameters to monitor will be selected and, when possible, confirm the event.
- To update FDM process to enhance the analysis process: this method is a real innovative approach as it develops the FDM process to adapt to new trends.

Project Results:

Requirements and FDM definitions / standards

Research activities initially focused on the definition of requirements specifications for the improvement of the current flight data analysis process and prepared the development of several modules for a demonstrator of an implementation of SVETLANA. A study has been realized to propose several improvements to current FDM limitations: "historical" event based strategy, new strategies of analysis (data mining), exploitation of environmental data, data processing automation…

The comparison of Flight Data Analysis systems developed in RF and EU has been conducted and no essential difference have been found out since both systems rely on the same concept of exceedence and threshold based anomaly detection. New deep data analysis algorithms have then been designed during the project, based on a data mining fully data driven approach. Nevertheless, using the data mining algorithms separately from present-day algorithms requires to implement additional experts assessments of data mining algorithms results and comparisons between aircraft documentation (It can decrease the adaptation capability). Separate SVETLANA algorithms usage (without existing systems) is difficult for users in the near future, the concept of simultaneous usage of SVETLANA algorithms and existing flight data analysis systems was then proposed.

As a result, the first proposals for the operational concept of SVETLANA were issued and shared with End Users through dedicated interviews.

During the study on how the limitation of actual FDM could be solved, the reflexion has been conducted along several lines:

1- Improvement of the current "historical" event-based strategy for FDM

In relation to the improvement of the current strategy for FDM, it has been proposed to enhance the current standards event sets. This set has not been modified since 2005 and does not take into account the number of parameters recorded for modern aircraft. This idea shall be further developed in the work on enhanced self-inspection. Another idea is to implement a set standard snapshots, this point should be developed in the same work and also with the development of algorithms.

A pre-study about workload linked to FDM (especially on event invalidation) has been performed. This study intends to prove that improvement of current algorithms is necessary (flight decoding, flight phase identification and event detection algorithms). It also proposes to build a KPI for FDM system automation. Further studies about this subject should be conducted with EUG.

In order to improve current FDM, a study about event detection algorithms detection should be done, in order to prove the lack of harmonization between operators (e.g.: unstabilized approach for several ATR, Boeing and Airbus operators)

2- Development of new strategies for FDM such as data mining methods

Another way to improve FDM would be to develop and integrate data-mining algorithms in the analysis loop. The pre-study about "unstabilized approach" event patterns clustering for a B737 operator has been used to judge the potential of such types of data mining analysis.

3- Adding of external and contextual information to FDM data

The potential of external data sources that should be correlated to FDM has been studied. It appears that the following external data could be integrated in the analysis process:
- weather data (METAR),
- astronomical data (aeronautical day/night),
- operational flight plan data (fuel data / trajectory data / loading data),
- crew status information,
- crew roistering.

During end-user interviews it was noted that the option to include data from other sources would be welcomed significantly. However, the introduction of these options will need adjustment of the FDM software as operators lack the knowledge and access rights to adjust their FDM software to include external and company databases in the FDM process.

4- Improvement of automation for FDM data processing

Another way to improve FDM efficiency is to automate all the processing that doesn't really require a human understanding. This aspect has been treated in the work which defines the architecture of the automatic flight data analysis system.

Finally, since the beginning of the project, the quality of the input data and the pre-processing steps that shall be used were questioned. In this context, research activities where conducted to improve flight phase detection in terms of accuracy, of automation, and maybe the most important aspect of aircraft independence. In this context,
- new flight phases definitions were developed from the study of the standardized ICAO flight phase definitions,
- definition of robust and generic new algorithms for automatic flight phases determination.

This work leaded to a patent application around a "Flight State Machine", a novel and robust method to cut raw data in flights and flights in phases.

Flight Data Management Cycle Process

The architecture of SVETLANA analysis system has been achieved and the constraints on this system (modularity, reusability, and existing FDA tool in the core) have been satisfied. The main innovation brought is the definition of a common format for both flight data and analysis results:
- having a specified common format for flight data allows the use of flight data from an unlimited type of source, as long as a "transcoding" function is developed in the system providing the source.
- having a specified common format for analysis results opens a new field of possibilities by allowing the comparison of results from different types of analysis. In the case of SVETLANA, the different types of analysis are: classical western FDM analysis, classical Russian FDM analysis, and data-mining oriented analysis.

The main results of the work above are:

- the definition of a common data format for flight data, available for Russian and European legacy systems,
- the definition of a common data format for analysis results, as a preparation to a dual exploitation with currently used analysis tools,
- the definition of a platform that can integrate several types of analysis, and is opened to other types of analysis,
- a platform allowing analysis results comparison.

Another innovation brought by the architecture of SVETLANA analysis system is the possibility to integrate external data in the process. An important work on data formats has been realised, which is essential for standardisation of a common analysis process between Europe and Russia.

Several points of convergence have been studied and the commonalities identified so far enable a high reusability of the mathematical analysis component, which is essential for an assessment into different process cycle or systems. The overall architecture of SVETLANA is highly modular and can be deployed as a single independent system as well as in a distributed environment. This architecture also takes into account existing flight data analysis systems thanks to the elaboration of common flight data formats at both input and output and based on engineering and semantics values of parameters. Clear benefits of the proposed architecture are: a smooth integration into existing environment, the capability to compare with existing systems, bringing new insights into data streams, a potential modular extension with new functionalities or features and finally the compatibility with existing systems and processes.

The system is composed of 7 functional modules, which are:
- the inbox module
- the "convert and import" module
- the data repository module
- the SKAT classical analysis module
- the AGS classical analysis module
- the SVETLANA data mining analysis module
- the results display module

Flight data (real or simulated) enter the system through the inbox module, and are then imported into the data repository.

The different analysis algorithms can access the flight data of the data repository, use these data, and send back their analysis results thanks to the inbox module.

Flight data and analysis results are available to partners of the project via the results display module.

This architecture corresponds to some dataflow that has been defined during architecture specifications.

Here is the dataflow for flight data analysis:

The input for the flight data analysis dataflow is raw data, in various formats depending of the way it has been generated.

The first step of the process is to transfer these data to the SVETLANA analysis platform, through the network. Once transferred on the platform, the different types of data are converted to the common flight data forma (which is composed of an XML descriptive file and binary files for engineering values)

Then different types of data are converted to the common format, they are imported into the data repository. Data information is stored in a database, and engineering values are stored in a folder structure.

The content of this data repository is used by the different algorithms to perform advanced analysis:
- Classical SKAT analysis
- Classical AGS analysis
- Data mining analysis

These different types of algorithms produce output in a standardized format. This format is pure XML, and is referred as the common analysis results format.

All the results produced by the different types of analysis in common format are then imported into the results database of the data repository

Here is the dataflow for results display:

As said before, all the information about flight data and analysis results are stored in the data repository. The location of this data repository is variable depending on the retained architecture (for example it can be located at an airline).

All the content of this data repository is used by a web application to display an overview and the results of:
- Flight list
- Events list (classical and data-mining analysis)
- Recorder reprogramming status
- Any information generated from the content of the data repository

The web application generates web pages so that users can watch the results through a standard web browser.

For the structured storage of the flight data it is necessary to develop a database. For compatibility with the proposed architecture it is necessary to divide the database into a data repository and a data mining database. The original flight data should be stored in the repository with a common format and the flight data processing results using algorithms of the express analysis. The repository database provides access to the stored flight data and processing results. The data upload is performed entirely for each flight. To perform the upload it is only necessary to store links to files with the flight data and the results. It is important to ensure the processing of queries is logarithmic in time depending on the number of records uploaded off the database.

Information used by the developed data mining algorithms should be kept in a separate database. The database will provide access to any recorded parameters separately upon a given query. All information is stored not in files but directly in the tables. For the uploading of flight data that has a number of samples comparable with the flight duration, the number of records reaches 200 thousand. In such circumstances, the effective data search is becoming increasingly difficult. It is necessary to use the database indexing, which leads to a marked improvement in performance.

The database should keep flight data in tabular form and in the original format. Nevertheless, flight data are written to the on board recorder with different frequency and for the normal operation of algorithms, the parameters shall be brought to a single frequency. Then, the database implementation must provide means for storing of accessing the normalized flight data (with equal frequency). The flight data stored in such a form can be used directly in the processing algorithms.

The data mining database should also store the computed parameters of the algorithms and the results of the flight data processing, allowing to quickly provide the flight data in-depth processing results of a specific aircraft and information about possible failures.

To implement the database corresponding to the requirements presented, several options were considered for DBMS (database management system).

Among the criteria for choosing a database that were searched, are:
- The architecture and functionality: mobility, scalability, distributivity, network capabilities, …
- The system operation monitoring
- Features of the application development:: design tools, multilanguage support, ability to develop web-based applications, supported programming languages, …
- Performance: rating TPC, ability of a parallel architecture, ability to optimize queries, …
- Reliability: recovery after failures, backup copying, rolling back changes, multi-layer protection, …
- Requirements for the operating medium
- Mixed criteria: localization, model of the value chain, …

In the items above, were considered criteria to be taken into account when choosing a DBMS. The structure of the developed database corresponds to the relational model of data representation. Based on these DBMS options the developed database structure can be implemented. Comparing the performance of the DBMS, it should be noted that 'Oracle' and 'Cache' provide better query performance in comparison with the 'Microsoft SQL Server' and 'IBM DB2'. However, the performance of the DBMS considered is sufficient for performing the tasks and is not a limiting factor when developing the database in question. Based on the application of the developed database, the maximum intensity of queries in these DBMS is unattainable not only in research studies but also in commercial operation.

The essential factor is difficulty in developing and supporting a database. The best choice at this point is 'MS SQL Server', as this DBMS requires a minimum qualification for administering when comparing with the alternative DBMS. Moreover, the license price of this DBMS is minimal. Thus, the implementation of the developed database was based on the 'MS SQL Server'.

Knowledge extraction from data streams

The third phase of the project is the central part of the new process and consists in the development of advanced algorithms, including the data base needed for the processing itself. A number of methods were proposed and tested on real flight data: both Russian and European.

A summary of what has been achieved during the preliminary work of mathematical candidate tools selection is below:
- Algorithms prototypes have been developed which mainly conveyed techniques such as recursive density estimate (RDE) and clustering analysis. These techniques were further tuned to per data vector analysis and per parameter analysis. In addition, RDE was tested globally on single pass of the whole data and locally on data groups obtained through clustering analysis which in had its own share of different approaches such as clustering based on Euclidean distance, Mahalonobis distance and K means, eClustering, and ARD-EM mixture Gaussian.
- The result of the RDE analysis was applied to classifiers which, again, varied between Ada-Boost, Quadratic Fisher Discriminant, Linear Fisher Discriminant.
- Using data obtained through the SVETLANA project, the format of these data were digitized, standardized and then used to test several algorithms prototypes.

Results so far proves that using ARD-EM algorithms with a number of clusters determined through evidence maximization (which adds up to 14) and input vectors obtained by Mahalonobis density, which then is classified to: good or bad through Fisher Linear Discriminant prove to be the good bet. Using a test data obtained through the SVETLANA project the overall error rate between the proposed algorithm and expert analysis in detecting anomalies was: 9.98%, whereas if the square of the variables is used, the overall rate of error increases slightly to 10.84%. This is mainly because Fisher LD techniques have the capability of avoiding over-fitting.

Although, the first results obtained prove the feasibly of the algorithm, it's continually been improved. The steps that have been followed are:
1- Digitized all express-analysis results for Russian flight data;
2- Implemented preliminary analysis of the data, using techniques of RDE and eClustering, validating that data was converted correctly;
3- Implemented number of clustering techniques on Russian data:
a. K means,
b. eClustering,
c. ARD-EM mixture Gaussians;
4- Different types of RDE analysis was implemented:
a. per vector,
b. per parameter;
c. global,
d. per cluster\phase,
e. Euclidian distance,
f. Mahalonobis distance.

The main reason for calculating and estimating densities is that following the variation of the values of the parameters analyzed along the time scale, when any anomaly arises, the density will decrease in comparison with the values it has during the "normal" operation/flight.

The plot below illustrated the density fall:
- x-axis represents the time of the flight,
- blue line represents the density value (per-vector using Mahalonobis distance),
- red line indicates the problems.

To provide more robust analysis and to avoid increasing the number of false-positives (FP) and false-negatives (FN), different types of estimated density could be used in final decision rule.

5- Also obtained mixture-gaussian pdf analysis was implemented;
6- Different classification techniques were applied over the obtained density values:
a. Ada-Boost,
b. Quadratic Fisher Discriminant,
c. Linear Fisher Discriminant.
d. Sigma-value criteria
7- Online versions of the algorithms and respective recursive calculation formulas were investigated (patent application including these results has been submitted to Lancaster University).
8- Per-phase, per-A/C RDE analysis was implemented over the EU flight data.

In the search of behaviour predictive capabilities, long- and short- term prognostic algorithms were developed and applied. Short-term algorithms demonstrate the capability to perform precise prediction in case of stable parameter behaviour. However, in case of drastic parameter change all tested techniques demonstrate a "delay" of correct prediction, comparable with prognosis horizon.

The following approaches for behavior prediction were investigated, developed and tested:

- Anomaly prediction based on "recursive density estimation" (RDE) function prognostic. This approach has advantage that it is computationally low-cost, often RDE has partial local trends, which is a convenient for application of prognostic models fact. Disadvantage is that such type of prognostic does not provide enough information for predicted anomaly diagnostics. Determining of the causal parameters requires additional analysis without prognostic information for every registered value.

- Anomaly prediction and diagnostic based on RDE partial derivatives (or other types of deviation measurements for every parameter). Advantage of this approach is automatic diagnostic of the predicted anomaly. This fact provides capability to build expert system, that allows to prevent some events in advance. RDE property of local trends is essential for the prognostic, because it cause the same property for its derivatives. Disadvantage of currently used approaches is high-computational resources and memory requirements. Computation of every derivative involves O(n2) operations. So, to calculate partial derivative for every parameter it is required to use O(n3). This fact is a significant issue for on-board or online application of this technique.

- Prediction of parameter values and following RDE estimation. This approach is convenient for the reason that such information is useful not only for statistical analysis method, but also for legacy FDA systems. This approach allows following diagnostic application. For that reason this approach is effective for early prevention of the event. As disadvantage of the approach is that parameter value prediction is hard-to-solve problem, because most of the parameters do not usually have a clear trend as well as required computational and memory resources are significantly high.

Focusing on prediction capabilities, long-term analysis tends to be more informative, as it is possible to apply simple algorithms for prediction, trends could be detected and high-level outlier detection is possible. The newly proposed and developed approach is less prone to over-fitting that the well known Ada-boost approach and on benchmark data sets performs extremely well. The main advantage is that a simple decision rule provided to this algorithm will take into account only the general dependencies in the data set, and will ignore misleading dependencies in the observed data set.

The developed algorithms were again mostly based on the RDE (patented) approach comparison with legacy system in blue, SVETLANA detection matches and is in red) and allow:
1- building a model of a normal functionality of an aircraft system,
2- using expert knowledge for training,
3- updating parameters for a better fitting of algorithms to the specified aircraft,
4- determining abnormalities and problems in functioning,
5- detecting new functioning modes of an aircraft that are not abnormal.

The results are very important to allow developing algorithms of problem source identification, i.e. parameters and parameter subsystems that cause abnormalities, as well as prognostication of problems. Among objectives that were achieved:
1- A number of data-mining technologies were investigated, developed and tested.
2- Algorithm for anomaly identification was investigated and tested (online and offline).
3- Algorithm demonstrator was developed (with reinforcement\online learning capability)
4- Additional software tools for inter-module communications are developed.

Algorithm is able to use expert knowledge and can also process flight data per subsets of parameters, the results have been compared with the legacy system analysis.

Enhanced self inspection

The fourth step of the project work on the objective to design a dedicated feedback loop, aiming at further improving the detection capabilities at each new analysis.

The study of recording systems and ACMS designed in EU and RF was performed. EU systems provide ample opportunities for re-programming, but unfortunately, Russian systems have little opportunities for re-programming and all tools for it is available for aircraft and analysis system designers only. Onboard adaptive algorithms were then adapted from the previous stage of the project.

A scalable process for reconfiguration of the acquisition configuration has also been designed allowing early detection of abnormalities if ACMS functionality is available on board. Most of the steps of the SVETLANA processes could be automated but the level of automation depends on technical and certification availability (certified bi-directional wireless, on ground updated databases, already carried out learning phase).

In overall, the performances of the automated process will depend on:
- The performance of diagnostics algorithms (identification of parameters involved in singularities)
- The quality of correlation knowledge database.

For each step of the generic process, an analysis has been carried out to identify:
- current state of art,
- capabilities to be automated and to be embedded on board,
- human and operational impacts.

The process can be decomposed in the following steps:
1- Acquisition of (a first set of) aircraft parameters
2- Events identification
3- Events diagnostic
4- Subsystem parameters selection
5- Acquisition function customization
6- Acquisition of new aircraft parameters

Some functions required the development of dedicated algorithms : events identification, events diagnostic. Nevertheless, early identification of events, potentially on board is subject to a high dependence to data quality (see requirements section) which leaded to specific studies on flight data pre-processing and more precisely filtering. Several techniques were tested with real and simulated data: Kalman filter, Fast Fourier Transform, Wavelet transform, but only the latter proved to be technically acceptable (not deforming the input signal beyond unrecoverable alterations).

After the implementation and validation of the proposed algorithms, Wavelet filtering approach was identified as most robust and algorithm with highest performance. All other approaches were considered as unacceptable due to high level of data corruption, despite the fact of successful data spikes removing. The final algorithm still has several drawbacks, as it does not filter some spikes with low amplitude, but it increase the processing quality significantly. Achievement of better performance is possible in case of taking into account the "shape" and standard behavior of the spikes.

Below is an example of application of the wavelet filtering;

However, still there are some drawbacks in case of low-amplitude spikes. Example is provided as follows:

Despite these problems, the wavelet filter was able to find all instances of anomalies with the data. The following algorithm was developed:

1- Filtering by a hard filter using wavelet transform at level 5 (best value in the experiments) and find maximum absolute derivatives of the resulting data. This is used to detect high-amplitude signals.
2- Mark as a wrong source the data point where the derivative is greater than the max derivative.
3- Merger intervals for the data with slightly unusual values:
4- If the resulting length of the interval is higher than some constant, set all point of the interval as 'normal';
5- Else set all the point of the resulting interval as 'abnormal.
6- Interpolate the 'abnormal' values

Integration and Assessment

A complete flight data analysis system has been developed in order to implement all the steps of the flight data analysis cycle. Initially planned to be an extension of a version of AGS, it finally has been decided to develop a new analysis chain due to some technical constraints.

Among the features of the data flow, as it contributes to standardization efforts, the common formats for data input and output constitute a major achievement. Such data formats foster interoperability between systems and between FDM stakeholders. This common format is also the link between all the modules of the architecture above, including low level layers of database and higher levels with a completely new HMI.

An important work has been done to handle various data sources, as a wide input of safety related pieces of information, not restricted to flight data:

- Real flight data have been provided to the project from Tu204 and B737 aircraft. Both Russian and Western Europe developments were then supported by these datasets which were also essential to study specific behaviour of algorithms by providing different usage and regulation frames.
- A large set of flight data based on Fokker 100 flight dynamics were delivered to the project. This contribution has a high added value for the developments of SVETLANA but also for the elaboration of a methodology to maturate statistical algorithms dedicated to the analysis of flight data. This precious dataset shall be reused extensively to further develop and validate algorithms, even in different environment or activities.
- BADA database (courtesy of Eurocontrol) has been used during the review of the state of art in anomaly detection algorithms.
- A hybrid database prototype with Russian and EU flight data has also been developed and is a backbone of the system : algorithms depend on efficient data access as well as on a dedicated data model to ensure real-time and/or on-line data processing.
- A thorough study, based on End Users requirements, of contextual information available in other information systems has been conducted, and the prototype demonstrator make available to final users of SVETLANA much more information when compared to existing systems, among others weather information, crew details, flight plans.

Among noticeable results:
- Flight data pre-processing: which includes flight phases identification and detection improvements, new parameter validity technique, handling of data flaws (gaps, blanks, …), sub-system definitions, …
- As mentioned above, the whole flight data workflow, with common data formats,
- Aircraft independent and automated event detection rules
- Data aggregation from various sources, including ACMS
- New generation HMIs to combine the analysis from various sources
- Reusable components of the prototype architecture : data converters, data base structure, data loaders, …
- Novel methods, algorithms, software for real-time analysis of data streams, successfully reused in other applications (medical environment)

Since a full demonstrator has been developed, below are a few screenshots of the dedicated HMI that has been developed, showing sample data:

For the last stage of the project, the overall assessment has been conducted with both the simulated environment and real flight data sources. During the evaluation the filtering techniques developed by for removal of erroneous data have been assessed. The testing that used the simulated data sets had to remove spikes from the data and accommodate any blanks that might exist. It is shown that the performance of the updated wavelet filtering algorithm has improved performance. In general the filtering technique is consistently removing evident spikes in the data. However the filtering algorithms are still sensitive to signals with rapid changes, such the roll and angle of attack, and improvements are still required.

The assessment of the simulated flights with abnormal events showed that the automated algorithms had trouble in correctly identifying and classifying abnormal events. In all cases the false positive rate was still high. In the cases where the abnormal events were identified correctly the classification of the event proved to be challenging.
In addition T5.1 evaluated the analysis of the SVETLANA algorithms on the flights with pre-programmed abnormal events. Here it was shown that the detection in combination with correct classification of abnormal events is still challenging for automated routines working without pre-knowledge.

For the data handling SAGEM produced two versions of a new flight data analysis interface. The first version was assessed to be limited in functionality but the second and final version however showed the full potential and succeeded in meeting quite significantly the demands as noted by the EUG during the project initialisation. For example the new interface allows additional information sources to be integrated into the flight data monitoring cycle, this includes weather data, flight scheduling and daylight conditions. Also the interface support more enhanced data handling and improved assessment tools using graphical representations.

The assessment with real data has been performed with the integrated simulator and by processing a set of flight data provided by a European Boeing 737-800 operator. The analysis has been launched for a first small subset of 13 basic flight parameters and a second subset of 47 fight parameters. Classical FDM analysis (event detection) was also performed on this data set so as to be able to compare classical FDM event detection and data mining abnormalities detection. The Ground Data Replay and Analysis Station configuration chosen for the classical FDM analysis was the one chosen for daily FDM routine operations of the airline. This configuration enables to monitor 101 potential events related to flight safety and aircraft airworthiness.

The full comparison between the data mining and classical methods applied to this reduced set of data showed that most of the classical FDM events were detected by the data mining algorithms, and the related parameters were also identified by the data mining algorithms. As the figure here below shows, a flight with a High Vertical Speed Event followed by a Go-Around was also tagged by the data mining algorithms as containing an "IVV (vertical speed) too large" abnormality and a "Flight phase too large" abnormality.

The study made to decrease the rate of invalid events showed that even if the detection algorithm has its part of responsibility, parameter invalidity is often responsible of false detection. Invalidity patterns of parameters have been identified and algorithms detecting these patterns have been developed. An event detection algorithm can then use the validity of the parameter it monitors in order to bypass event detection.

Here is an example of CAS (computed air speed) invalidation and correction:

Accurate and "flawless" flight phase detection is a first step for better event detection. Indeed the massive test of the algorithms showed that some events where not detected with the previous algorithms, whereas some other where false positive. But this improvement alone would not be sufficient without correction of event detection algorithms themselves, which need a powerful checking of parameter validity.

As a conclusion we can say that these three improvement on classical algorithms (flight phases – parameters invalidation – better events detection) contribute to increase flight safety by detecting events more precisely and by reducing the ratio of false events, thus growing confidence of the users (pilots, safety officer, airline management) for FDM systems.

The work performed introduced also a new type of maintenance events, when the parameters or event the whole recording has a too high invalidity ratio. This gives the maintenance team a better feedback of the quality of the data which are analysed. It avoids according too much trust in results when the input data are really bad, and even trigger an action for the replacement of a media or a recording unit.

Interaction with End User Group, IPR and Exploitation

Finally, the support of End Users has been essential: input have been collected from end-users on the SVETLANA concept and the potential introduction of standards. In addition dissemination tools have been developed: a project identity set, a brochure, a public website etc... Presentations were held at MAKS 2011 , at the SAGEM AGS user conference in China,at the Avionics Europe 2012 in Munich and the annual meeting of European Air Safety Investigators (ESASI 2012). Additional dissemination presentations are expected beyond the duration of the project in 2013.

In addition SVETLANA was given the third place for the "Best International Aviation Cooperation Project 2012"by Russian Union of Aviation Industrialists under leadership of General Director Evgeny A. Gorbunov (see http://www.aviationunion.org/ online) A SVETLANA representation was in Moscow during the final price giving event on December 5th at the conference hall of the Cathedral of Christ the Savior in Moscow. For information on the award see:
http://www.aviationunion.ru/competitions.php?conam=

Potential Impact:
SVETLANA project is focused on a common EU/Russia approach for flight data analysis and aims to promote:
- a new common analysis process based on the existing set of recorded flight data, from different operators (EU and Russian Federation), following regulations convergence,
- the adoption of advanced data mining techniques without any hypothesis on the data modelling or on the event search patterns ; Allowing automated analysis and behaviour predictions with a minimal parameterization will foster the standardization of processes

The implementation of SVETLANA aims to bring further technical innovations by the training of algorithms with simulated data and systematic analysis of all available data. The development of the capability to detect sensors failures by algorithms is also the integration of a major innovation to promote focused and proactive on-time maintenance.

The new FDM methodology and its implementation as extensions of existing tools will then increase the insight that an airline can have on its flight data thanks to a wider coverage of data analysis related to operational safety.

The development of SVETLANA automated FDM shall provide improvements of airlines' operating procedures by having in line of sight flight safety and the reduction of serious incidents and accidents. Among all possible areas of enhancements, maintenance preventive actions process, fuel consumption, pilot training programs, availability of aircraft are clearly where direct benefits are expected.

Finally, SVETLANA aims at providing short return on investments and visible improvements on airlines' cost control, such as reduced insurance costs or efficient allocation of maintenance tools, test equipments and technicians. The reduction of unnecessary maintenance and repairs, for example, can be improved at several scales:
- On a micro scale, early warnings of deteriorating equipment tied to tail numbers or engine serial numbers;
- On a macro scale, the resultant data can more accurately predict future maintenance requirements;
- On both scales, FDM allows maintenance planners to more proactively and accurately plan and position spare parts;

Expected final results

An immediate benefit in a short term time frame is expected to come from the industrialisation and integration of the foundation components of SVETLANA architecture into legacy FDM tools. The combination of different analysis based on opposite hypothesis shall contribute to the maturation of algorithms developed so far. The capability to detect the unforeseen without a priori and with the help of additional sources of data shall support a better understanding of the root causes in non trivial situations. This mutual enrichment of legacy and new system is then expected to reduce efforts to maintain legacy algorithms and to ensure a higher robustness.

Consequently, the new developments that can potentially support a standardisation (common formats, HMI) shall enable a seamless integration of additional capabilities along with existing features of legacy systems. It will then be possible to offer data mining based tools to the existing anomaly detection tool, leading to a better acceptance of these new tools, reducing training efforts because the analysis environment will not change.

In addition, a better automation is also expected by implementing the more robust parameter validity techniques and the flight and flight phase identification algorithms. The project has clearly demonstrated the benefits obtained from a better initial flight data pre-processing, which includes advanced filtering techniques. A higher quality of the data analysed by data mining statistical algorithms is a key improvement in order to obtain stability in algorithms output, and as a consequence less efforts from end users and analysts required to discard events/abnormalities.

Beyond the improvement of data quality at each stage of the analysis process, it is expected that an improved data fusion from various sources will also provide a better understanding of the root causes in case of events or abnormalities. This widely shared requirement among stakeholders shall pave the path to initiate a new framework to exchange flight safety related data more effectively.

Finally, thanks to the development of an automatic advanced flight data analysis platform integrated in legacy FDM tools, we expect:
- A capacity to offer additional flight data monitoring services to airlines, and thus to enhance the potential market of such FDM tools and services,
- An environment suitable to assess advanced detection capability brought by the competition of several algorithms working on the same input data,
- A standardisation of all in-house tools, and an opened capability to communicate with end-users' information systems.

Intentions for use and impact

Several outcomes of the project are already planned to be exploited in evolutions of existing legacy systems. But beyond the technical integration of new features, prototypes and literature can be used as guidelines to prepare a step by step implementation of SVETLANA new FDM cycle.

In a real environment, the results of SVETLANA are expected to support training and to offer new flight data analysis based services.

Among other outcomes candidates for a short-term exploitation, one can highlight the definition of high level requirements for flight safety improvement, based on existing standards and procedures which can be reused to extend the existing or candidate for new standards.

Also the common format for flight data and analysis results can have an impact on other projects than SVETLANA. As it allows using multiple sources for analysis, and comparison of different types of analysis , it's opening new possibilities for collaboration between heterogeneous systems. Future flight safety studies will benefit from a standardization of singular points to define the characteristic points in a flight, this knowledge can also be explored in other aviation safety studies.

The architecture and all contributing software components can be used to implement a distributed flight data analysis system, widely opened to various data processings and standard compliance between Europe and Russia.

The definition of high level requirements for flight safety improvement is based on existing standards and procedures and can be reused to extend the existing or candidate for new standards.

In addition, the data sources are expected to be linked to stakeholders' systems in order the integrate contextual information such as weather conditions, flight plans, in-house safety management systems…

Dissemination and exploitation of results

Input has been collected from end-users on the SVETLANA concept and the potential introduction of standards. In addition dissemination tools have been developed: a project identity set, a brochure, a public website etc... Presentations were held at MAKS 2011, at the SAGEM AGS user conference in China and at the Avionics Europe 2012 in Munich additional dissemination presentations are expected.

Most of these results can be exploited as guidelines and as prototype tools (software and algorithms) to be incorporated in the existing tools (such as AGS, SKAT etc.) or as stand-alone FD analyser. To support this, several channels of dissemination were used during the project, and the main noticeable results would be (full details are given in WP6 deliverables):

- 3 patent applications were filed: 1 SAG + 2 LANC
- 2 publications were submitted in a scientific journal
- Members of the consortium attended aviation events, such as MAKS 2011, AGS User Conference (2011 and 2013), ESASI 2012 and Avionics Europe 2012
- RTD collaboration with End Users conducted to valuable achievements for the implementation (e.g. parameter validity methodology). Future collaborations are being set-up, such as a participation in a working group managed by EASA concerning the standardisation of flight safety related data.

Finally, the whole consortium has already identified exploitation plans built of with results of the SVETLANA project:

Sagem will pursue 3 lines of exploitation in parallel:
- From the Industrial point of view: Sagem will Implement the architecture and the full processing chain. Due to the design of SVETLANA with multiple databases, it will be possible to implement as separate information systems the algorithmic part (and its own database) and the infrastructure part (hosting all available flight data and the basic low level management services). Then a complete product-line will handle the whole processing chain, data mining algorithms will constitute one of the black boxes of this chain (middleware). The final expected system will then be able to use several algorithmic boxes on demand.
- From the operational point of view: Sagem will propose new analysis tools and services in parallel of the existing chain of analysis. End-user willing to go deeper in the analysis of the data shall have access to new data mining features on equivalent commercial basis as current AGS offers. The new HMIs will be the enabler of this access to different results from analysis tools (at least for legacy analysis and data mining analysis). In addition, aircraft independent and automated event detection rules shall now be integrated in AGS products and offer a more

efficient "default configuration". Specific, per customer algorithms will still remain available, but a harmonization and a reduction of per customer specificities is expected.
- From the dissemination point of view: Sagem will continue researches through other collaborations, including with PhD students to build a collection of algorithmic toolboxes. At a first level, there shall coexist different mathematical tools to address maintenance and flight safety issues. In addition, Sagem will continue to disseminate the results of SVETLANA and future developments in at least 2 forthcoming conferences in the first half of 2013: EOFDM and AGS User Conference.

NLR has on-going activities in the field of Flight Data Monitoring. Due to its aviation research nature NLR is especially active in the field of developing and testing new and improved FDM processes and applications for flight data (management) to support the safety and efficiency of flight operations of their customers. The knowledge gained during the SVETLANA project will be used to further enhance the potential of FDM. Especially the work on standards and criteria has provided insights to further future research. Also the inside knowledge accrued form the automated data processing and future concepts will be valuable to further advise customers on their developments in the future. The simulated flight data set has been put to good use in the SVETLANA project. However this data will also remain valuable in testing other FDM techniques in the future.

In addition to activities defined by the SVETLANA project, researchers at Lancaster University were investigating whether tools and algorithms developed for SVETLANA applications can be used in other applications, in particular to help prevent complications after surgery.

For this purpose, a team from Lancaster University, lead by Prof.G.Markarian partnered with the team from the Academic Surgery Unit at University Hospital of South Manchester led by Professor Charles McCollum aiming to develop new technology for risk reduction in cardiac intensive care based on algorithms designed to monitor and predict "health of their aircraft".

Lancaster University Aviation Security Team are utilizing SVETLANA results to develop a real-time patient monitoring and risk prediction system, similar to those used by airlines to monitor the safety of their aircraft. The basic principle behind the work is the fact that there are a lot of parallels between flying an aircraft and observing a critically ill patient. Both the surgeon and the pilot are dealing with a lot of information coming from a variety of sensors. They both need to know not only what is happening now but what might happen in the future and safety is absolutely critical. When a patient is critically ill or recovering from surgery, doctors monitor the patient's blood pressure, temperature, pulse and other vital signs very closely but have to rely on their experience to predict what is likely to happen next. Pilots have the additional benefit of tools to help them do that. The new tool developed by SVETLANA and adapted for cardiac applications has the potential to give doctors an extra layer of intelligence to draw upon.

The new tool is being designed to make sense of a diverse range of patient data to provide health care professionals with a clearer indication of what might happen to their patients in the near future; buying them precious time to take preventative action. Doctors can then potentially access this information at any time, even from home on their laptop or phone. The tool is in the early stages of development but once up and running it is hoped that it will have applications in a number of different healthcare settings.

Dr Stuart Grant, Research Fellow in Surgery, who works on the project, said: "There are vast amounts of clinical data currently collected which is not analysed in any meaningful way. This tool has the potential to identify subtle early signs of complications from real time data. If the aviation technology can be successfully transferred to healthcare it has the potential to provide doctors with information which could improve outcomes for patients."

- UAC will cooperate with Russian airlines operating Russian and foreign aircrafts to integrate SVETLANA concept and modules of the system into existing FDA tools.

- UAC will take into account SVETLANA requirement in the course of new aircraft design as existing on-board equipment imposes limitation on SVETLANA algorithm implementation. It is necessary to change acquisition hardware and user access permissions for a list of parameters for wide spread of FDA tools based on SVETLANA.

- UAC will continue research to implement SVETLANA concept as on-board system. It's important for pilots to have short time on-line on-board predictions. Prediction based on deep data mining analysis is one of the key benefits of SVETLANA.

- UAC will improve algorithms to make user-friendly interface. It will require more cooperation with airlines, more tests on real FD during long period of aircraft operation.

List of Websites:

http://www.svetlanaproject.eu