Time-to-Event Data (Survival Analysis) based Modelling of Maneuver Occurrence of Non-Cooperative Satellites

Shivshankar S.
PhD Student, Department of Aerospace Engineering, Indian Institute of Science
Debasish Ghose
Professor,Department of Aerospace Engineering, Indian Institute of Science

ABSTRACT

Modelling pattern-of-life of non-cooperative space objects is an essential requirement of Space Situational Awareness (SSA). Maneuvers of non-cooperative satellites is an important event of interest in their life pattern. Ground-based sensors are predominantly used to collect orbital data of a non-cooperative satellite based on which the maneuver occurrence is detected. However, the active satellites to be monitored is huge in number and the available ground surveillance sensors are limited as well as costly. Due to this limitation, a non-cooperative satellite may go unobserved by ground based sensors. Therefore, there exists a problem of gaps in the available orbital data of non-cooperative satellites. Moreover, the satellite maneuver (event of interest) occurrence information of some samples may be lost, due to noise in the ground sensor observations or due to observation window limits or losing tracks. Therefore, a need arises for detection of maneuvers from incomplete data as well as to devise a methodology for optimal use of ground sensors to collect data. If the probability of maneuver occurrence of a non-cooperative satellite at each time point and estimation of time until occurrence of a maneuver are available, then the ground sensors can be more effectively scheduled.

Conventional machine learning regression methods are not suited to be able to include both the event and time aspects as the outcome. The conventional models are also are not equipped to handle censored examples (incomplete data due to non-observability). Therefore, in this work, we introduce a solution methodology by applying Time-to-Event data analysis (survival analysis) techniques to assess whether a satellite maneuvered, that is whether the event of interest occurred or not, and also estimate when the next maneuver would occur. Time-to-Event analysis is a branch of statistics concerned with analyzing temporal data and predicting the probability of occurrence of an event and has an inherent capability to accommodate censored data samples.

We have explored a variety of approaches including Cox proportional hazards model, Weibull distribution model, Kaplan-Meier model, Nelson-Aalen model, Random survival forest, Survival Support Vector Machines, Gradient boosted survival analysis and Deep learning based survival analysis. Detailed experimental results based on real life satellite orbital datasets and discussions on the results are presented.

1. INTRODUCTION

1.1 Problem Statement

The Space based capabilities and space power has taken centre stage in deciding the overall power of a Nation and therefore the space is becoming an increasingly contested domain. Space Situational Awareness (SSA) or Space Domain Awareness (SDA) is therefore essential for a Nation to safeguard its own assets. SSA is defined as the comprehensive knowledge of Resident Space Objects [RSO] which may include satellites, rocket bodies, debris etc. and the ability to understand and predict their behaviour and future location with required accuracy. SSA solutions should provide a quantifiable and timely technical evidence of behavioural attributes of specific space domain threats, hazards and its implications.

L.Chen et al., [2] brings out that Situational analysis based on space catalogue data is realized by modelling and analyzing the orbital data to understand all space situations. Orbital Maneuvers are an essential part of a life pattern of a satellite. The orbital maneuvers can be benign, for example a maneuver essential for housekeeping of a satellite or a maneuver essential for maintaining a specific mission objective of the satellite in orbit. The types of maneuvers performed by Low Earth orbit (LEO) and Geostationary (GEO) satellites are briefly explained below.

1.1.1 Types of Satellite Maneuvers

The following are the types of maneuvers performed by the LEO satellites

Inclination Adjustment Maneuver (IAM) : An IAM is commonly referred to as a Delta-Inclination (Delta-I) maneuver. This out-of-plane maneuver is performed in the cross-track direction (i.e. perpendicular to the direction, the spacecraft is moving.) An IAM changes the angle of the equatorial plane to the orbital plane. This type of maneuver is performed periodically to maintain the mission's Mean Local Time (MLTAN).

Drag Makeup Maneuver (DMM) : This type of maneuver is also referred to as a delta-velocity (Delta-V) maneuver (or an orbit altitude adjustment maneuver). It is used to raise (or lower) the orbit's semi-major axis and is an in-plane maneuver. A DMM is a specific type of Delta-V (positive) which increases the orbital velocity thus increasing the orbital altitude and is used to counteract the effects of atmospheric drag on the spacecraft and maintain orbit circulation as well as maintain the strict constraints on the projected spacecraft ground track. (A negative "DMM" is a retrograde maneuver to lower the altitude).

Risk Mitigation Maneuver (RMM) : This type of maneuver is also a Delta-V maneuver to change the orbit altitude. An RMM is executed to avoid orbital debris and may be either a velocity increase (prograde maneuver -semi-major axis increase) or a velocity decrease (retrograde maneuver - semi-major axis decrease). An RMM is only performed if the flight operations team determines that the Probability of Collision (P_c) meets certain thresholds as determined by complex conjunction assessments.

Similarly for the Geostationary (GEO) orbital regime, besides the dominant 2 body acceleration, various other accelerations affect the motion of a satellite in GEO orbital regime. Due to this, the satellite location drifts and to bring it back to its designated location, station keeping maneuvers are required. The station keeping maneuvers can be classified into East-West stationkeeping maneuvers and North-South stationkeeping maneuvers. The third body gravitational effects from the Sun and the Moon affect the Inclination and RAAN orbital parameters of a GEO satellite and a North-South stationkeeping maneuver corrects it. These maneuvers require larger delta-V on the order of 1 m/s. Whereas the solar radiation pressure and higher order gravity terms in the Earth's gravitational field affect the Semi-major axis and eccentricity parameters and a East-West stationkeeping maneuver corrects it. These maneuvers require delta-V of the order of 0.01 to 0.1 m/s.

1.1.2 Anomalous and Threatening Maneuvers

A non-cooperative space object is defined as a non-friendly object in space and can be perceived as a threat if it performs anomalous maneuvers in space. A non-cooperative space object does not share information like operational mission objectives, orbital maneuvers, special orbital events, and station keeping maneuvers, unlike civilian satellite missions which publish orbital maneuver details in the open domain. Hall,Z. and Singla,P. [6] elegantly bring out that an important aspect of space object monitoring is detection and tracking of non-cooperative maneuvering space objects in a data-sparse environment.They also bring out that although there has been a lot of work in terms of maneuvering targets, relatively a little of this work is applicable to non-cooperative space objects. Some satellites may perform maneuvers which may be anomalous and not benign. For example, a maneuver by a non-cooperative chaser satellite to pursue another satellite, a maneuver to perform Rendezvous and Proximity Operations (RPO), etc. To identify the threatening maneuvers, the non-cooperative space objects needs to be tracked and observed for collecting orbital position and velocity data. The foremost challenge of SSA is the huge rate of increase of RSOs in recent years. Therefore the space objects to be monitored is huge in number and the available ground surveillance sensors are very limited as well as costly. This is depicted by the authors in Figure 1 below where $S_1, S_2, ..., S_n$ are satellites to be observed and $T_1, R_2, ..., T_m$ are ground based Telescope and Radar sensors.

The constraint in a real world scenario is that the space objects are very very larger in number as compared to the available ground based sensors $[n >> m]$. Due to this limitation, a non-cooperative satellite may go unobserved by ground sensors and may result in missing or incomplete observed data. Satellite maneuvers of non-cooperative space objects with adversarial intent needs to be known with such limited data available. Such a real life scenario is explained in Figure 2 below. A non-cooperative space object is defined as a non-friendly object in space and can be perceived as a threat if it performs anomalous maneuvers in space. A non-cooperative space object does not share information like operational mission objectives, orbital maneuvers, special orbital events, and station keeping maneuvers, unlike

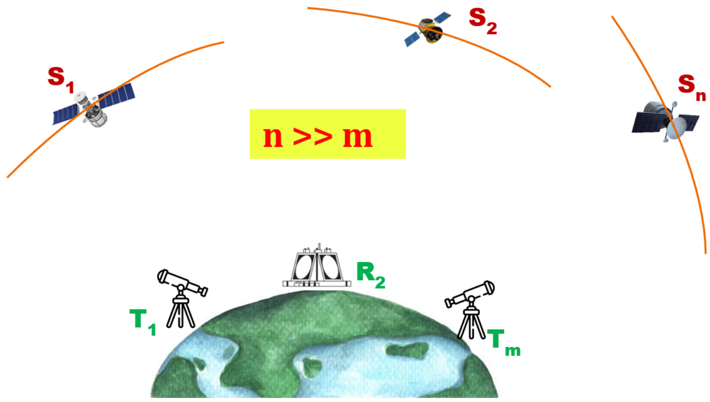

Fig. 1: Depiction of Lack of surveillance resources (Image by authors)

civilian satellite missions which publish orbital maneuver details in the open domain. Understanding the actions of a non-cooperative object in space is an essential requirement for SSA. Such information of non-cooperative space objects must be derived from the orbital data obtained through ground-based tracking radars, telescopes, etc.

Therefore there is a need to develop solution methods through which we can estimate the maneuver occurrence time of non-cooperative satellites. This will help in scheduling the ground sensors to track the non-cooperative objects only during their respective estimated maneuver time period window. The result would be an optimal sensor schedule tasking to track the large number of resident space objects with very limited sensors.

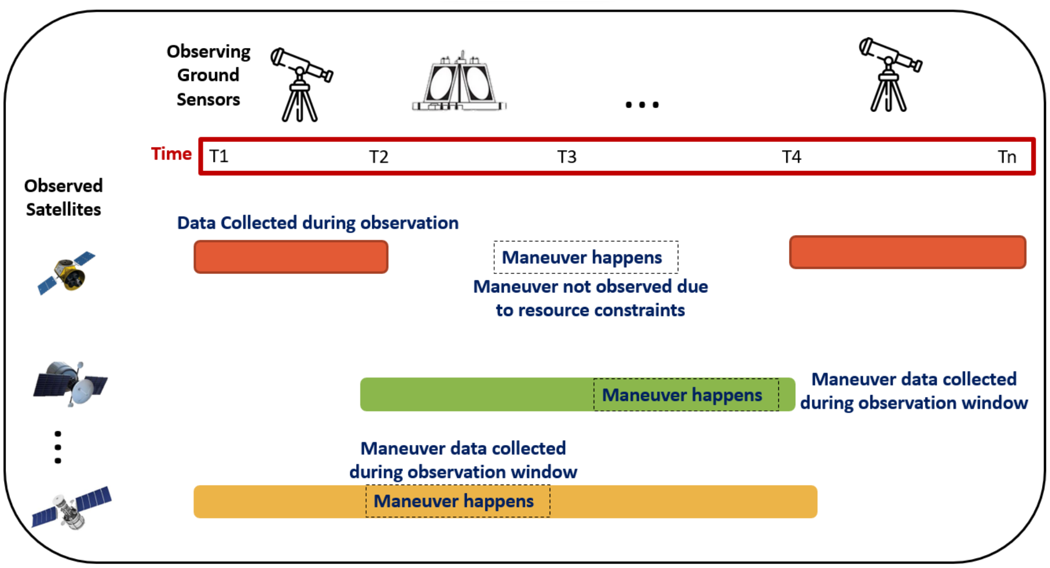

Fig. 2: Depiction of Unobservable scenario due to lack of sensors (Image by authors)

1.2 Conventional Machine Learning and Statistical Methods

Incomplete or censored data is a prevalent problem in many research studies [16]. Unfortunately, unlike missing data, which is very easy to spot, censored data is obscure and can easily evade detection. If this problem of censored data is not appropriately addressed, it can lead to biases and inefficient estimation that can impact the conclusions of the study. In conventional statistical concepts, these incomplete data will be disregarded. The use of standard methods to analyze censored data will generate results that, in a way, has some level of biases because some important information would be left out. Conventional machine learning methods like logistic and linear regression are not suited to be able to include both the event and time aspects as the outcome in the Time-to-Event estimation model. Traditional regression methods are also not equipped to handle censored examples in the data.

The authors propose a solution methodology to tackle this problem using Time-to-Event data analysis (also called survival analysis) which has been an active research topic due to its impactful applications in a variety of disciplines as found in literature survey. To the best knowledge of the authors, Time-to-Event data analysis techniques have not yet been explored for SSA applications.

2. SOLUTION METHODS

There are two main streams of survival analysis (refer Figure 3 and Figure 4) [1]. The first view is based on traditional statistics scattering in three categories [13]

(i) Non-parametric methods including Kaplan-Meier estimator [8] and Nelson-Aalen estimator [1] are solely based on counting statistics, which is too coarse-grained to perform personalized modeling.

(ii) Semi-parametric methods such as Cox proportional hazard model [14] and its variants Lasso-Cox [15] assumes some base distribution functions with the scaling coefficients for fine-tuning the final survival rate prediction.

(iii) Parametric models assume that the survival time or its logarithm result follows a particular theoretical distribution such as Exponential distribution [9] and Weibull distribution [12].

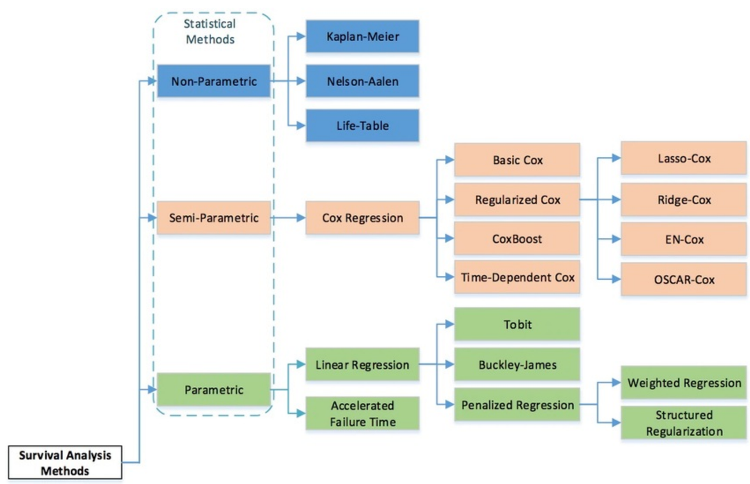

Fig. 3: Statistical Survival Analysis Methods

These methods are either based on statistical counting information or pre-assume distributional forms for the survival rate function, which generalizes not very well in real-world situations.The goal of Time-to-Event data analysis (survival analysis) is to estimate the time until occurrence of the particular event of interest, which can be regarded as a regression problem [2] [19]). It can also be viewed as to predict the probability of the event occurring over the whole timeline [9]. Specifically, given the information of the observing object, survival analysis would predict the probability of the event occurrence at each time point.

Time-to-Event data analysis is a learning framework and a set of techniques that can be used to estimate the time it takes for an event of interest to occur based on observations. In the research problem at hand, the event of interest may be considered as the maneuver of a non-cooperative satellite.

In the context of the figure. 3 below, our solutions methodology attempts to answer some of the follwoing pertinent questions

[1] Figure reference https://humboldt-wi.github.io/blog/research/information systems 1920/group2 survivalanalysis/

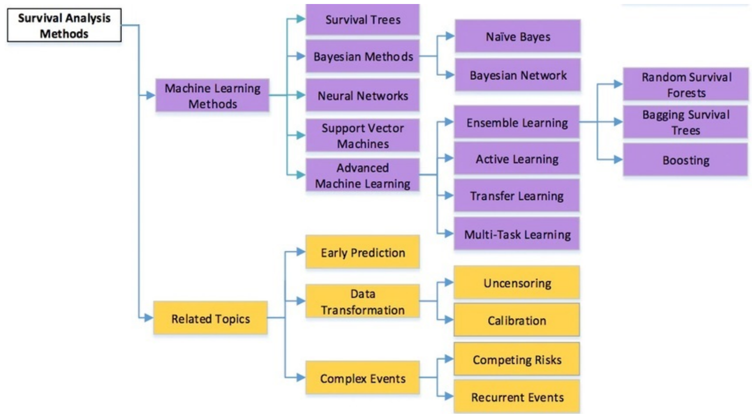

Fig. 4: Machine Learning Methods for Survival Analysis

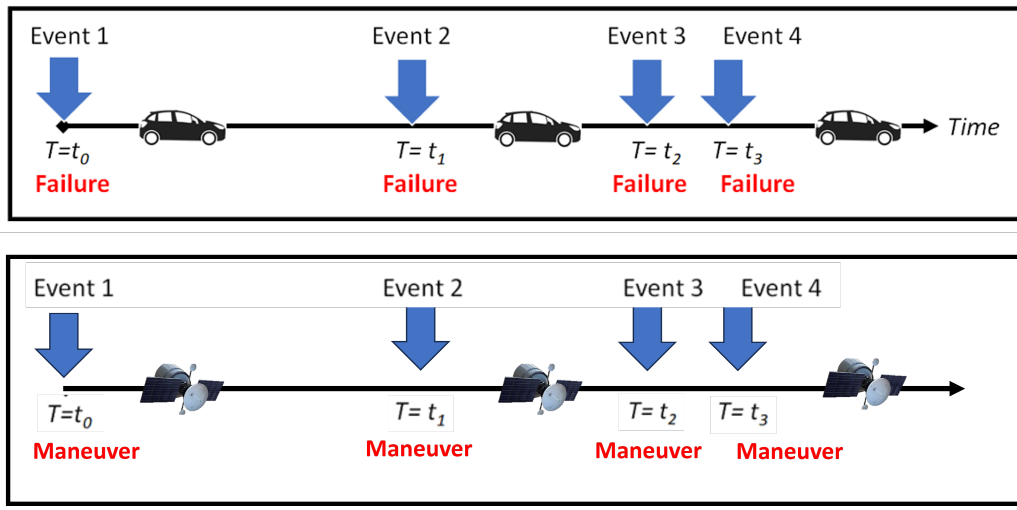

Fig. 5: Depiction of Analogy between car failure event and satellite maneuver event

(i) What is the probability of occurrence of event 2 under different combinations of event 1?

(ii) What is the relative risk of occurrence of a satellite maneuver in a specific time period?

(iii) What are the factors influencing the duration between events 1 and 2?

(iv) What is the probability that event 2 will occur before, or at, or after a certain time conditioned upon the influencing covariates?

Time-to-Event estimation model not only helps to assess whether or not a satellite maneuver (event of interest) occurred, but also when that event occurred. Survival analysis is used to analyze or predict when an event is likely to happen. It originated in medical research, but its use has greatly expanded to many different fields.The real strength of Survival Analysis is its capacity to handle situations when the event has not happened yet. Predicting the probability of an event occurring is good and to be able to predict the time remaining before an event occurs is even better. Time-to-Event data analysis is also called survival analysis, reliability analysis, duration modelling and event history analysis. Many popular ideas of machine learning such as gradient boosting, random forests and support vector machines have been adapted from Time-to-Event analysis.

Survival Random Forests, Survival SVM and Bayesian models are some of the techniques available in the second stream of survival analysis which is based on machine learning perspective. Recently, deep learning, i.e., deep neural network has gained traction as the third stream of survival analysis in many tasks (Ranganath et al. 2016; Grob et al. 2018; Lee et al. 2018).

Time-to-Event data is data on how long it takes until some event of interest we care about happens. Time-to-Event data is unique , because the outcome of interest is *"whether event occurred or not"* and also *"when that event occurred"*. The unknown or incomplete data is taken into account in Time-to-Event data analysis and is regarded as an important component. Survival models are able to take censoring into account and incorporate the uncertainty, so that instead of predicting the time of an event, we predict the probability that an event happens at a particular time.

For example, consider a factory containing various machines running in a production line. Using Time-to-Event data analysis, it is possible to predict with great degree of certainty when a machine will fail. The Data Science team could predict the machines survival function every day, so that 1 or 2 weeks before the machine is supposed to fail, the factory manager can be notified so that the necessary actions can be taken. Similarly, in the case of our problem at hand, the maneuver occurrence time of non-cooperative space objects can be predicted or estimated, so that we can anticipate and prioritize ground sensor tasking more effectively and optimally to track the non-cooperative space objects and avoid any imminent threat.

2.1 Key Components of Time-to-Event Data Analysis

Let T be a non-negative random variable denoting the time to event of interest (survival time/event time/failure time). The distribution of T could be discrete, continuous or a mixture of both. We will focus on the continuous distribution. The distribution of a random variable $T \geq 0$ can be characterized by its probability density function (pdf) and cumulative distribution function (cdf). However, in Time-to-Event data analysis, we often focus on the following three functions

2.1.1 Survival Function

Let *S(t)* be the survival probability, the probability that an event has NOT occurred until time *'t'*. Let *F(t)* be the failure probability, the probability that the event occurred by time *'t'*. *S(t)* and *F(t)* can thus be represented mathematically as

$$\begin{aligned} S(t) &= P(T > t) \\ F(t) &= P(T \leq t) \\ S(t) &= 1 - F(t) \end{aligned} \tag{1}$$

Survival function reflects the cumulative non-occurrence of an event of interest. The theoretical survival function and a typical survival function in practice is shown in Figure 6 below [1] . T is the random lifetime taken from the population under study and cannot be negative. The survival function *S(t)* outputs values between *0* and *1* and is a non-increasing function of *t*. In theory the survival function is smooth, in practice the events are observed on a concrete time scale, e.g. days, weeks, months, etc., such that the graph of the survival function is like a step function.

2.1.2 Hazard Function

Hazard rate is the instantaneous rate or probability that an event has occurred during a very small-time interval Δt between t and Δt, given that the individual did not have an event until *'t'*. It describes the instantaneous potential per unit time for the event to occur

$$\text{h(t)} = \lim_{\Delta t \to 0} \left(\frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} \right) \tag{2}$$

Therefore the hazard function models which periods have the highest or lowest chances of an event. In contrast to the survival function, the hazard function does not have to start at 1 and go down to 0. The hazard rate usually changes over time. It can start anywhere and go up and down over time. An example hazard function is shown in Figure 7 below [2]

[1]Figure reference https://www.slideshare.net/zhe1/kaplan-meier-survival-curves-and-the-logrank-test
[2]https://www.statisticshowto.datasciencecentral.com/hazard-function/

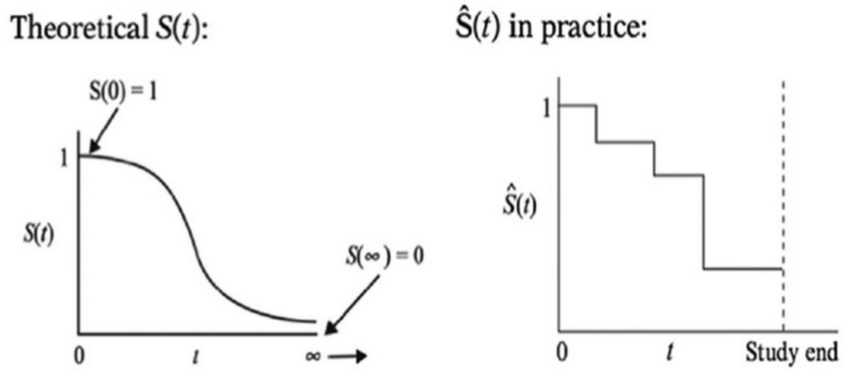

Fig. 6: Survival Function

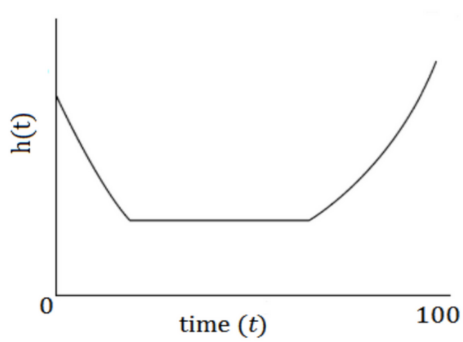

Fig. 7: Hazard Function

The hazard function $h(t)$ is NOT the probability that the event (such as death) occurs at time t or before time *t*. $h(t)\Delta t$ is approximately the conditional probability that the event occurs within the interval $(t, t + \Delta t]$ given that the event has not occurred before time *t* for small $\Delta t > 0$.

2.1.3 Cumulative Hazard Function

The cumulative hazard function is given by

$$H(t) = \int h(t) \cdot d(t) \tag{3}$$

Hazard function is useful in finding the periods that are the safest or riskiest with respect to the occurrence of the event of interest.It has the advantage in analyzing censored data, analytic simplification and modelling sensibility.

2.1.4 Relation between Survival and Hazard functions

Since we need to estimate the probability of occurrence of event of interest, lets consider [1]

$$P(t < T \leq t + \Delta t \mid T > t) \tag{4}$$

Using Bayes theorem

[1] https://towardsdatascience.com/the-mathematical-relationship-between-the-survival-function-and-hazard-function-74559bb6cc34

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0 \tag{5}$$

$$\text{h(t)} = \lim_{\Delta t \to 0} \left(\frac{P(t < T \leq t + \Delta t) \cap P(T > t)}{P(T > t) \cdot \Delta t} \right) \tag{6}$$

The probability of hazard occurring $P(t < T \leq t + \Delta t)$and $P(T > t)$ is actually $P(t < T \leq t + \Delta t)$. Therefore,

$$\text{h(t)} = \lim_{\Delta t \to 0} \left(\frac{P(t < T \leq t + \Delta t)}{s(t) \cdot \Delta t} \right) \tag{7}$$

Rearranging equation (7) gives the following

$$\text{h(t)} = \lim_{\Delta t \to 0} \left(\frac{P(T \leq t + \Delta t) - P(T \leq t)}{S(t) \cdot \Delta t} \right) \tag{8}$$

$$\text{h(t)} = \lim_{\Delta t \to 0} \left(\frac{F(t + \Delta t) - F(t)}{\Delta t} \right) \cdot \frac{1}{S(t)} \tag{9}$$

$$h(t) = \frac{dF(t)}{dt} \cdot \frac{1}{S(t)} \tag{10}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{11}$$

Rewriting equation (11) into something of an equation containing only *h(t)* and *S(t)* by applying the chain rule of differentiation.

$$\begin{aligned} h(t) &= \frac{d}{dt}(1 - S(t)) \cdot \frac{1}{S(t)} \\ h(t) &= -\frac{d}{dt} S(t) \cdot \frac{1}{S(t)} \end{aligned} \tag{12}$$

Therefore, hazard rate is simply the negative natural logarithm of survival rate (survival probability) differentiated over the time

$$h(t) = -\frac{d}{dt} \ln(S(t)) \tag{13}$$

and the cumulative hazard rate (cumulative hazard function) at time *t* is the negative logarithm of survival rate at time *t*

$$H(t) = -\ln(S(t)) \tag{14}$$

2.1.5 Censoring

As explained above, censoring occurs when we don't know the exact time-to-event of an included observation. In our problem at hand, censoring can occur due to sensor constraints of not able to observe the non-cooperative space object during a time period in between observations. Censoring can be among any of the following types

(i) Left Censoring : Occurs when the event is observed, but exact event time is unknown.

(ii) Right Censoring : If the events occur beyond the end of the study, then the data is right-censored.

(iii) Interval Censoring : Occurs when the event is observed, but subjects come in and out of observation, so the exact event time is unknown.

(iv) Random Censoring : A combination of the above three types.

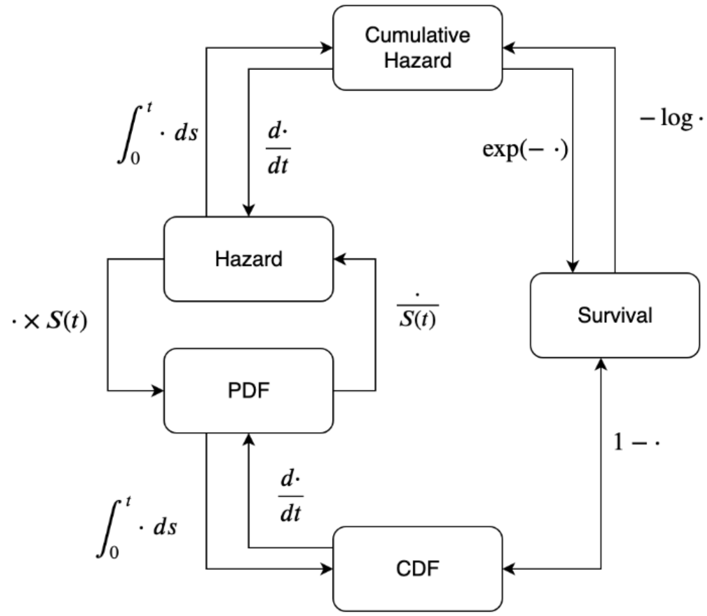

Fig. 8: Depiction of relation between survival function , hazard function , cumulative hazard (source : Lifelines Python library documentation)

2.2 Time-to-Event Regression Analysis

Time-to-Event or Survival regression can be used to estimate the conditional probability of an event occurring within a specific time period or event-horizon. A time-to-event estimation problem thus reduces to estimating the conditional distribution of survival :

$$E[1\{T > t\} \mid X = x] = P(T > t \mid X = x) = 1 - P(T \le t \mid X = x)] \tag{15}$$

Note that X is a set of covariates, and T refers to the distribution of censored survival time $T = \min(T^*, C)$ where T^* is the distribution of the true time-to-event and C is the distribution of the censoring time. Assuming conditional independence between T and C (that is $T \perp C \mid X$) allows identification of the distribution of $P(T|X)$. Survival regression naturally allows accounting for censored data. In the case of survival regression, the likelihood ℓ under censoring is given as

$$\ell(\{x, t, \delta\}) \propto P(T = t \mid X = x)^{\delta} P(T > t \mid X = x)^{1-\delta} \tag{16}$$

Here $x \in R^d$ are the covariates, $t \in R^+$is the event or censoring time and $\delta \in \{0, 1\}$ is a binary indicator denoting if the individual was censored. For the censored individuals, the likelihood corresponds to the probability that the event takes place beyond the time horizon, $t, P(T > t \mid X = x)$ also known as the survival function.

The Cox Proportional Hazards is a Time-to-Event semi parametric regression model. It describes the relation between the event incidence (as expressed by the hazard function) and a set of covariates. The Cox model is written as follows

$$h(t \mid X) = \underbrace{h_{\text{hazard rate}}(t)}_{\text{base}} \times \underbrace{\exp(X_1 \beta) \times \ldots \times \exp(X_m \beta)}_{\text{proportional hazards}} \tag{17}$$

The first part in the above equation describes how the occurrence of event of interest evolves through time and the second part models the effects of the explanatory covariates on the occurrence of event. So, with Cox regression, we are interested in how long it takes for something to happen. In Cox regression, the goal is to find the most probable parameters $\beta = (\beta_1, ..., \beta_m)$ and the most probable base hazard function $h_0(t)$.

We can choose survival analysis regression when we need to predict the time until a specific event, especially in situations with censored data and when the focus is on event occurrence. We can choose simple time series forecasting when we are interested in predicting continuous variables over time, and the timing of specific events isn't our main concern.

2.3 Random Survival Forests

Another feasible machine learning approach which can be used to avoid the proportional constraint of the Cox proportional hazards model is a random survival forest (RSF). The random survival forest is defined as a tree method that constructs an ensemble estimate for the cumulative hazard function. Early applications of random forests (RF) focused on regression and classification problems. Random survival forests (RSF) [7] was introduced to extend RF to the setting of right-censored survival data. Implementation of RSF follows the same general principles as RF:

(i) Survival trees are grown using bootstrapped data

(ii) Random feature selection is used when splitting tree nodes

(iii) Trees are generally grown deeply

(iv) The survival forest ensemble is calculated by averaging terminal node statistics

Constructing the ensembles from base learners, such as trees, can substantially improve the prediction performance. Basically, RSF computes a random forest using the log-rank test as the splitting criterion. It calculates the cumulative hazards of the leaf nodes in each tree and averages them in following ensemble. The tree is grown to full size under the condition that each terminal node have no less than a prespecified number of deaths. The out-of-bag samples are then used to compute the prediction error of the ensemble cumulative hazard function. The presence of censoring is a unique feature of survival data that complicates certain aspects of implementing RSF compared to RF for regression and classification. The technical implementation is based on scikit-survival package, which was built on top of scikit-learn: that allows the implementation of survival analysis while utilizing the power of scikit-learn.

Compared with regression based approaches, random survival forest has several advantages. First, it is completely data driven and thus independent of model assumptions [18]. Second, it seeks a model that best explains the data and thus represents a suitable tool for exploratory analysis where prior information of the survival data is limited. Third, in case of high dimensional data, limitations of univariate regression approaches such as overfitting, unreliable estimation of regression coefficients, inflated standard errors or convergence problems do not apply to random survival forest [4]. Fourth, similar to survival trees, random survival forest is robust to outliers in the covariate space.

2.4 Survival Support Vector Machine

Survival Support Vector Machines is an extension of the standard Support Vector Machine to right-censored time-to-event data. Its main advantage is that it can account for complex, non-linear relationships between features and survival via the so-called kernel trick. A kernel function implicitly maps the input features into high-dimensional feature spaces where survival can be described by a hyperplane. This makes Survival Support Vector Machines extremely versatile and applicable to a wide a range of data. A popular example for such a kernel function is the radial basis function. Instead of modeling the probability that an event will occur, we could look at survival analysis as a Ranking Problem in the survival SVM modelling.

Survival models based on support vector machines (SVM) [17] are able to incorporate non-linearities in an automatic way and using non-additive kernels, interactions are automatically incorporated. These methods use an approach which is different from the standard statistical approach. SVM-based models do not assume a true underlying function for which the parameters need to be estimated. Instead the empirical risk of misranking two instances with regard to their failure time, is minimized. The survival problem was therefore reformulated as a ranking problem.

2.5 Deep Learning for Survival Analysis

The initial adaptation of survival analysis to meet neural networks (Farragi and Simon, 1995) was based on generalization of the Cox proportional hazards model with only a single hidden layer. The main focus of the initial model was to learn relationships between primary covariates and the corresponding hazard risk function. Following development of the neural network architecture with Cox regression proved that in real-world large datasets with non-linear interactions between variables it is rather complicated to keep the main proportionality assumption of Cox regression model.

However, Farragi and Simon's network extended this non-linearity quality. A few years ago, the more sophisticated deep learning architecture, DeepSurv, was proposed by J.L. Katzman et al. as an addition to Simon-Farragi's network. It showed improvements of the CoxPH model and the performance metrics when dealing with non-linear data . This architecture was able to handle the main proportional hazards constraint. In addition to that, while estimating the log-risk function h(X) with the CoxPH model we used the linear combination of static features from given data X and the baseline hazards. With DeepSurv we can also drop this assumption out.

2.6 Recurrent events survival analysis

Recurrent events data have two main characteristics viz. within subject correlation and time varying covariates [20]. Recurrent event within subject are very unlikely independent, they are related and this phenomenon is known as within-subject correlation and there are two possible sources this within subject correlation viz. within-subject. Correlation due to event dependency and with subject correlation due to heterogeneity. Within-subject correlation due to event dependency refers to a situation where an event itself accelerates or decelerates the rate of subsequent event.

Another important concern related to recurrent event analysis is how to deal with time-varying covariates. In many studies there are some covariates which are subject to change over time. The event of interest , maneuvers, in the life of a satellite are recurrent events and the authors have paid due attention to deal it in its perspective.

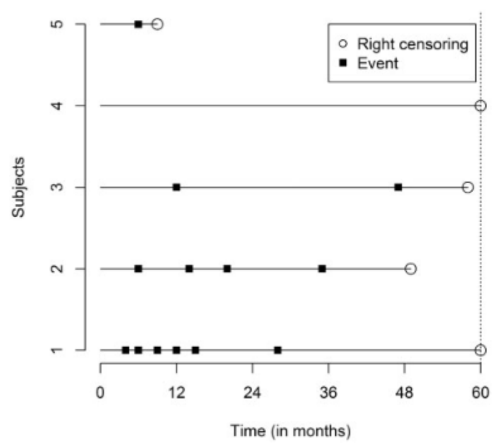

Fig. 9: Depiction of right censoring and recurrent events among five subjects

3. DATA COLLECTION

The maneuver histories of some of the civilian satellites are maintained by both the International Laser Ranging Service (ILRS) housed at NASA Goddard and the International DORIS Service (IDS) formed by the French Space Agency (CNES). The Orbital data history of ENVISAT, AQUA, SENTINEL-3A, JASON, SPOT, LANDSAT-7 satellites (that is, the NORAD GP element sets current data) for the respective mission life of each satellite was collected[1] along with the satellite orbital maneuver history[2] and used as the event of interest column in the dataset. The JSON format of the NORAD GP element sets current data of LANDSAT-7 satellite is shown in Table.III for illustration. The covariates used in the Time-to-Event regression analysis were the orbital parameters of the respective satellites, that is Semi-Major axis (derived from mean motion), Eccentricity, Inclination, Argument of Pericenter.

4. RESULTS AND DISCUSSION

4.1 Models Tested

The following models were tested on both simulated and real satellite datasets with various covariates and parameters. The simulated dataset was created to contain 100 satellites with Semi-major axis, Eccentricity and Inclination as co-variates. The observation time duration in weeks and a event of interest (maneuver) occurrence column with censoring

[1] https://celestrak.org/

[2] https://ids-doris.org/doris-system/system-monitoring/table-of-all-events.html, https://airs.jpl.nasa.gov/data/outages/anomalies

OBJECT_NAME	”LANDSAT 7”
OBJECT_ID	”1999-020A”
EPOCH	”2022-07-27T22:34:44.879
MEAN_MOTION	14.59777691
ECCENTRICITY	0.001356
INCLINATION	97.9764
ARG_OF_PERICENTER	71.8598
RA_OF_ASC_NODE	255.4786
ARG_OF_PERICENTER	71.8598
MEAN_ANOMALY	288.2735
NORAD_CAT_ID	25682
REV_AT_EPOCH	23847
BSTAR	0.000058342
MEAN_MOTION_DOT	0.00000235

Table 1: Illustration of LANDSAT-7 Orbital Elements in JSON Data

was also generated in the simulation. The simulated dataset contains fifty (50) active remote sensing (LEO) and fifty (50) passive remote sensing (LEO) satellites to test the dependency of maneuver occurrence on this parameter. For the real life datasets the historical orbital parameters of satellites like ENVISAT, LANDSAT, SENTINEL-3A, CRYOSAT, HAIYANG were used along with the corresponding maneuver event information (ground truth for reference) from the ILRS website.

Non-Parametric	Semi-Parametric	Parametric	Regression Analysis
Kaplan-Meier	Cox Propotional Hazards	Exponential Distribution Fitter	Random Survival Forests
Nelson-Aalen		Weibull Distribution Fitter	Gradient Boosting Survival Analysis
Aalen-Additive			Fast Survival SVM
			Fast Kernel Survival SVM

Table 2: Models explored in this work

4.2 Python libraries used

The following python libraries available in the open domain for survival analysis applications were used in the experiments.

(i) Lifelines [3]

(ii) PySurvival [5]

(iii) Scikit-Survival [11]

(iv) Auton-Survival [10]

(v) Deephit

4.3 Metrics for evaluation

When there are multiple models which are available for the study, we can identify the better model by gauging the model’s performance with evaluation metrics. A good model can predict the maneuver occurrence time more accurately. The following metrics were used for comapring the various models

(i) Concordance Index (c-index) - This is the most commonly used metric for time-to-event analyses. This is a rank correlation score between predicted risk values with observed time. A c-index of 0.5 is no better at predicting an outcome than random chance. A c-index around 0.7 indicates a good model and a c-index greater than 0.8 indicates a strong model.

(ii) Akaike Information Criterion (AIC) - It is an estimate of the prediction error of the model. A lower AIC score corresponds to a better model. The AIC is used when we evaluate model fit with the within-sample validation. For out-of-sample validation, log-likelihood values are used.

(iii) Survival function - Estimated as explained above in section 2.1.1

In addition in some models, the coefficients and p-values provide more insights. The hazard ration for each covariate is equivalent to e to the power of the covariate's coefficient (e^{coef}).

4.4 Results

The Figure 10 above compares various parametric and non-parametric models tested on the simulated dataset using the estimated survival function.

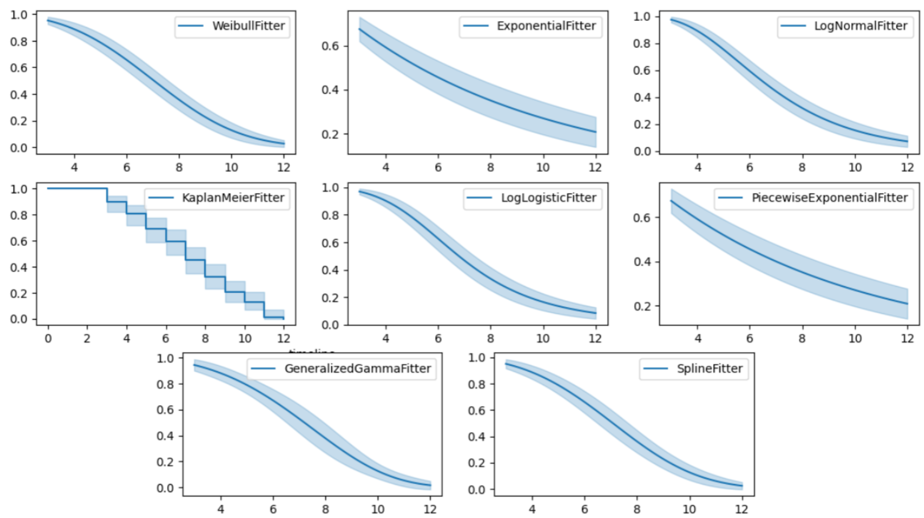

Fig. 10: Comparison of parametric and non-parametric models

The Figure 11 (a), (b), (c) below shows the comparison of estimated survival function between active and passive remote sensing satellites obtained from Kaplan-Meier fitter, the cumulative hazard estimated using Nelson-Aalen fitter and the comparison between Cox propotional hazards model, Weibull AFT model and Aalen Addditive models respectively.

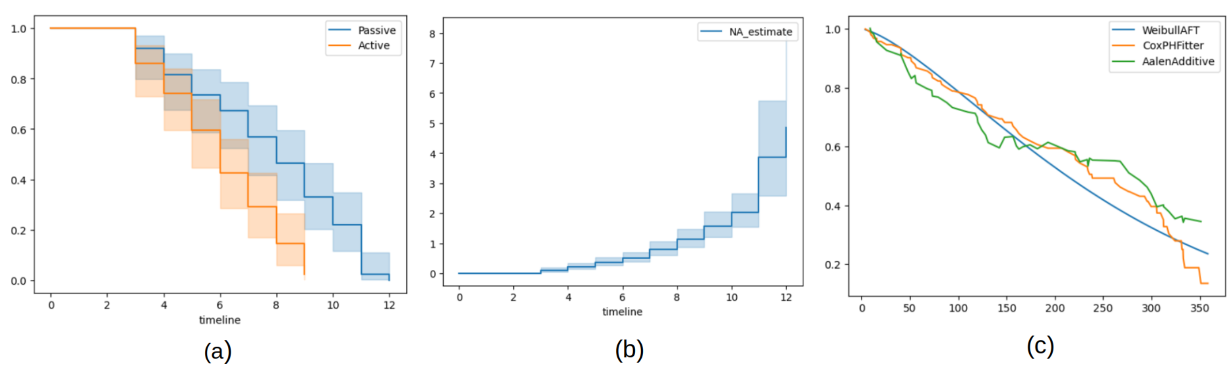

Fig. 11: Comparison of estimated survival function

The above table compares the Cox-PH and Aalen Additive models tested on some of the reallife datasets of satellites. The Cox-PH came out as a better model as compared to the Aalen Additive model in all the satellite datasets except Cryosat where both models gave similar performance. Also it was noticed that both the models performed slightly poorer when it was fitted on the Aqua satellite dataset as it had lesser amount of data as compared to other satellite datasets.

Satellites and Models	Cox PH Model		Aalen Additive Model
	c-index	**AIC**	**c-index**
ENVISAT	0.70	1159.24	0.52
AQUA	0.69	4771.55	0.42
CRYOSAT	0.78	1840.89	0.77
HAIYANG-2	0.97	946.78	0.89
LANDSAT-7	0.89	1283.52	0.77
SENTINEL-3A	0.85	330.32	0.50

Table 3: Table comparing Cox PH and Aalen Additive Models on reallife Satellite datasets

Cox's proportional hazard's model is often an appealing model, because its coefficients can be interpreted in terms of hazard ratio, which often provides valuable insight. The Figure 12 below explains the effect of covariates on the probability of occurrence of event. For example, the logarithm of hazard ratio with 95% confidence interval obtained in the Cox-PH model for ENVISAT satellite dataset is shown below on the left. As explained above, the coefficients in a Cox regression relate to hazard; In the context of medical applications of patient survival analysis, a positive coefficient indicates a worse prognosis and a negative coefficient indicates a protective effect of the variable with which it is associated. In our problem at hand, the semi-major axis and eccentricity orbital parameters of a satellite are more dominating covariates in the probable occurrence of maneuver. Since the satellites taken for experiment are all Low Earth Orbit (LEO) satellites and it is known that these satellites in general perform In-Plane maneuvers to adjust for drag and do not perform an out-of-plane maneuver which is influenced by Inclination orbital parameter. The impact of covariates on the coefficients is shown on the right.

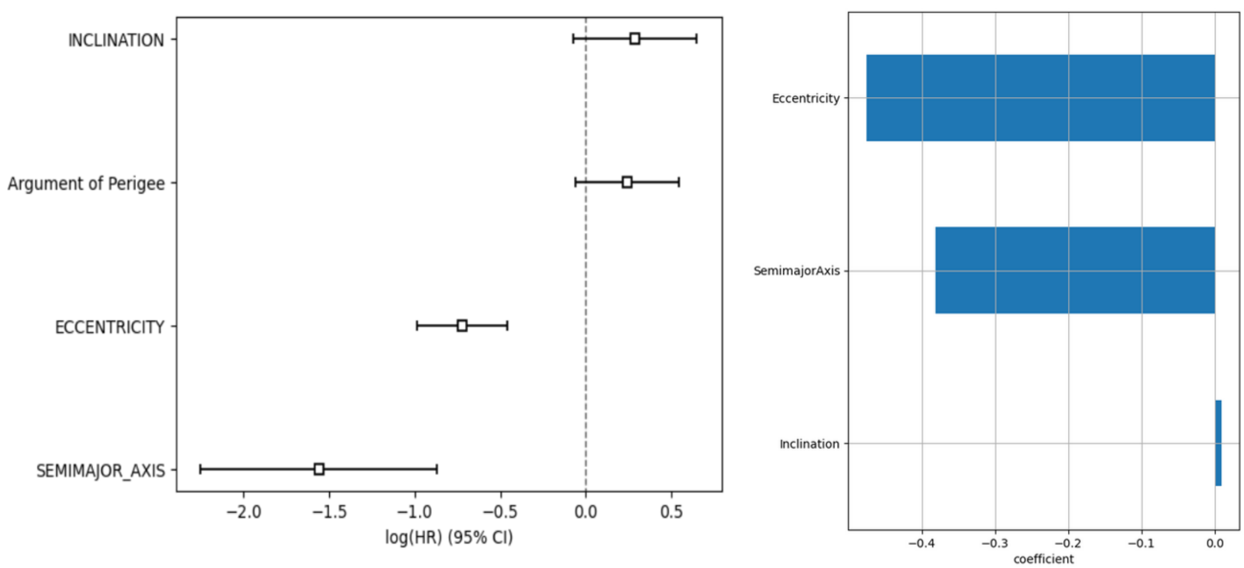

Fig. 12: Effect of covariates on the probability of occurrence of event

In the Figure 13 below, alpha in the X-axis denotes the hyper-parameter that controls the amount of shrinkage of the β coefficients of the Cox model to zero. The Y-axis on the right is the scale of β coefficients and Y-axis on the left corresponds to each co-variate. If the penalty has a large weight (to the right), all coefficients are shrunk almost to zero. As the penalty's weight is decreased, the coefficients' value increases.

The table 4 below lists the Cox-PH model coefficients, e^{coef} and p-values for the ENVISAT satellite dataset. The p-value indicates which covariates have a significant effect on the survival duration. Given the coefficient's small p-values (0.005) , it can be established that the semi-major axis and eccentricity orbital parameters are statistically significant predictors for determining the duration for event of interest.

In the context of survival analysis and the Cox PH model, the coefficients provide insights into how changes in the covariates affect the hazard rate (risk of an event occurring) for the subjects. The coefficient value with respect to the

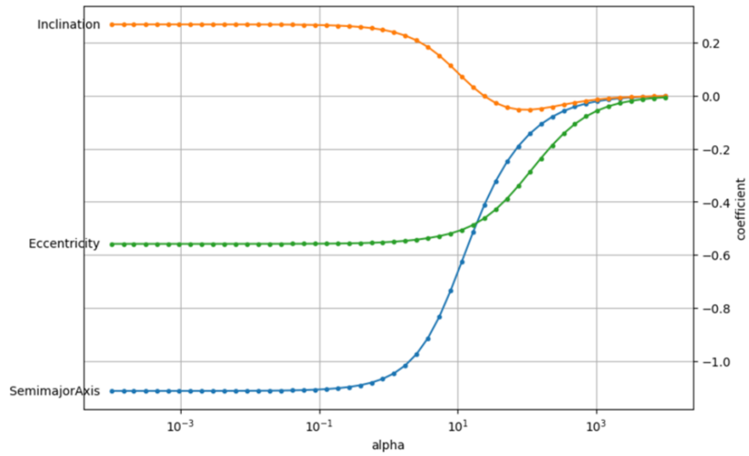

Fig. 13: Relation between hyperparameter and CoxPH model coefficients

covariate is explained in detail below

(i) SemimajorAxis Coefficient (-0.95): A negative coefficient indicates that a decrease in the semimajor axis is associated with a higher hazard rate (increased risk of an event). In the case of a satellite, this could suggest that satellites with lower semimajor axes are more likely to experience the event (e.g., a maneuver).

(ii) Eccentricity Coefficient (-0.49): A negative coefficient suggests that a higher eccentricity is associated with a lower hazard rate (decreased risk of an event). This means that satellites with higher eccentricities are less likely to experience the event compared to those with lower eccentricities.

(iii) Inclination Coefficient (0.20): A positive coefficient indicates that an increase in inclination is associated with a higher hazard rate (increased risk of an event). This suggests that satellites with higher inclinations are more likely to experience the event.

	coefficients	exp(coef)	p-value
Semimajor Axis	-0.95	0.39	0.005
Eccentricity	-0.49	0.11	0.005
Inclination	0.20	1.23	0.18
Argument of Perigee	0.34	1.40	0.02

Table 4: Cox-PH model coefficients and p-values

The table 5 below compares the c-index obtained by fitting various survival regression models on the real-life satellite datasets. The overall performance of across multiple models and multiple satellites was observed good as the c-index score was around 0.7. However, the reasons for poor performance of Fast-kernel survival SVM on the HAIYANG-2 and LANDSAT-7 satellites are still being explored.

Satellites and Models	ENVISAT	SENTINEL-3A	CRYOSAT	HAIYANG-2	LANDSAT-7
Random Survival Forest	0.714	0.908	0.940	0.944	0.998
Gradient Boosted Survival Analysis	0.676	0.910	0.818	0.887	0.919
Fast Survival SVM	0.628	0.892	0.748	0.959	0.858
Fast Kernel Survival SVM	0.63	0.893	0.738	0.664	0.656

Table 5: c-index scores obtained from various regression models for various satellites

The box plots shown below are obtained through GridSearchCV method. GridSearchCV is the process of performing hyperparameter tuning in order to determine the optimal values for a given model. GridSearchCV can leverage multiple cores by evaluating multiple parameter settings concurrently. The grid search provided by GridSearchCV exhaustively generates candidates from a grid of parameter values. We can observe that the regression models seem to be relative robust with respect to the choice of α for the datasets used.

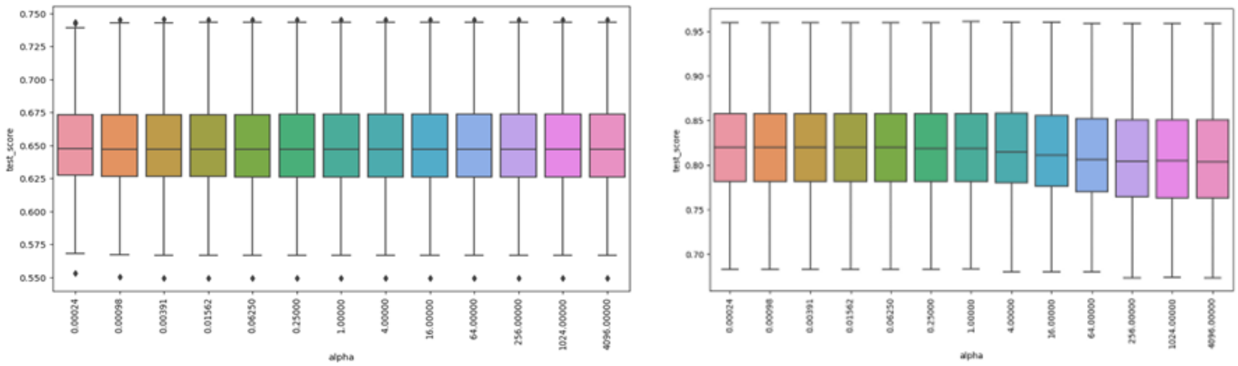

Fig. 14: Depiction of robustness of regression models

The comparison of concordance index for varying hyperparameter is shown for the four regression models in the Figure 15 below

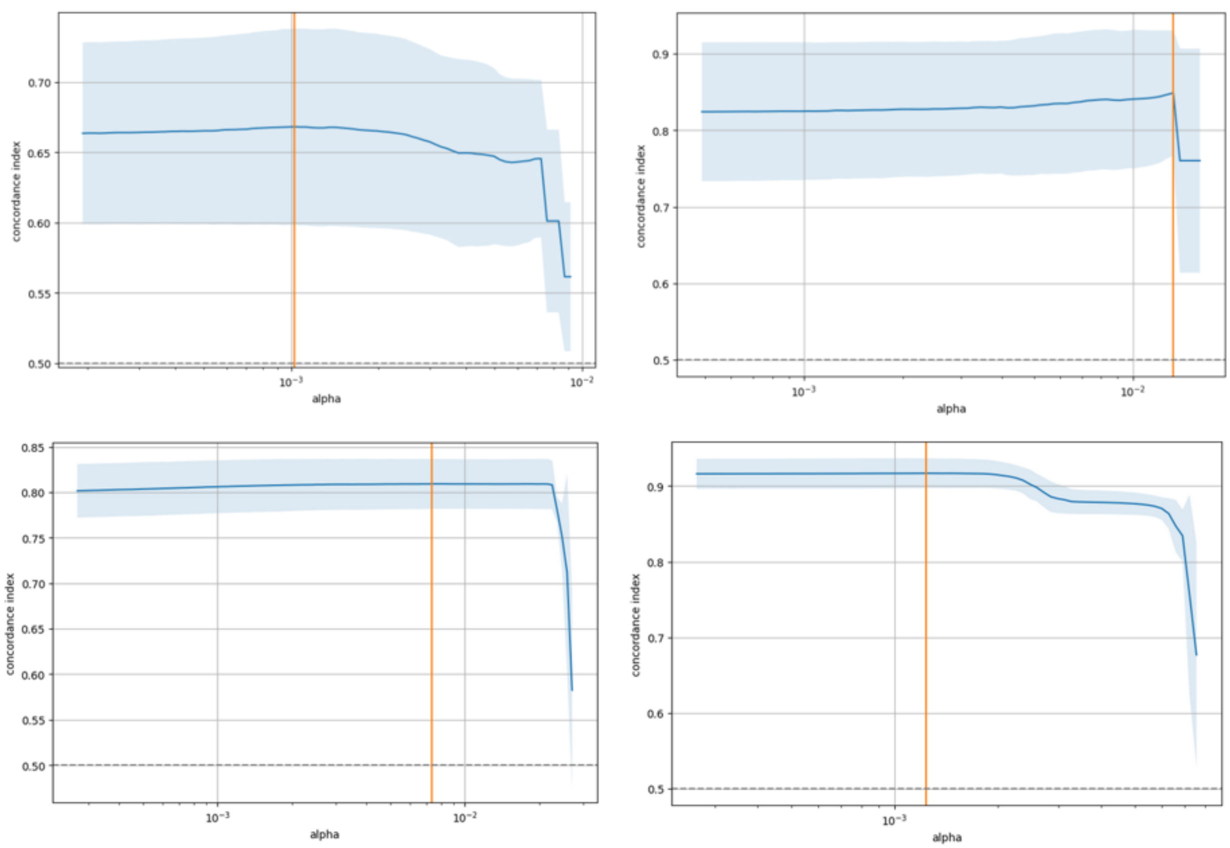

Fig. 15: Variation of c-index with hyperparameter of model

Deep Learning models were trained on the real-life satellite datasets using the DeepHit Python library. The training and validation loss curves are shown on the left in the Figure 16 below and the estimated survival function is shown on the right. The Brier score and Negative Binomial Log Likelihood [NBLL] was used to evaluate the deeplearning experiments. The Brier score is used to evaluate the accuracy of a predicted survival function at a given time t ; it represents the average squared distances between the observed survival status and the predicted survival probability and is always a number between 0 and 1, with 0 being the best possible value. The negative binomial log-likelihood at t measures how close the survival probability is to 1 if the given data survived at t and how close the survival probability is to 0 if the given data failed before t. The integrated Brier score and NBLL was 0.0061 and 0.022 respectively.

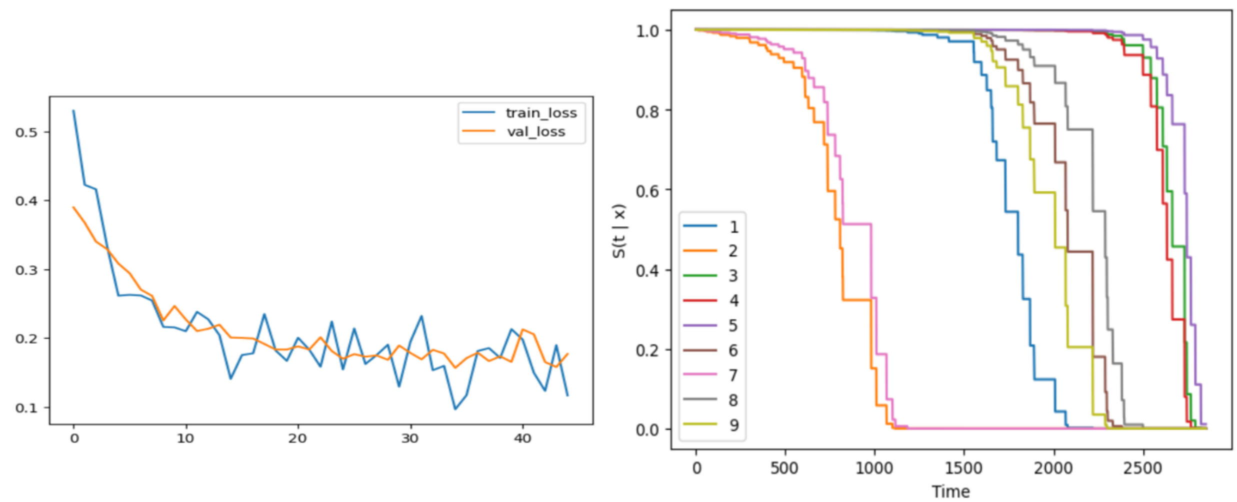

Fig. 16: Results from Deeplearning trials

5. CONCLUSION AND FUTURE WORK

Time-to-event or survival analysis helps us explain how long it will take until some event of interest we care about happens. In the context of a surveillance resource constrained perspective, the ground surveillance sensors are overburdened due to the huge quantum of space objects available for monitoring. Due to the threat perception, in the case of non-cooperative space objects, we cannot afford to miss the observation to collect data during its maneuver. Therefore, estimating the probability of maneuver occurrence and time until the maneuver will greatly enhance the optimal tasking of sparse ground sensors. Conventional statistical , regression and time series forecasting models fail to accommodate censored data. Ignoring censoring will bias results and ignores much of information collected. We have presented the novel solution methodology of using survival analysis for modelling maneuver occurrence of non-cooperative satellites. Different statistical, machine learning and deep learning models have been compared by experimenting on real-life satellite datasets.

The benefit of Time-to-Event analysis techniques is that it can incorporate data from multiple time points across various satellites. The data of satellites which have not maneuvered till a time instant or data of satellites unavailable during a time window can still contribute to the Time-to-Event analysis. Therefore, the behaviour modelling of non-cooperative satellites can also be done using historical orbital data of cooperative satellites operating in similar orbital characteristics and operating for similar mission objectives. This is being considered to be explored as part of future work. The expected behaviour of a benign active satellite can be modelled based on orbital regime in space in which the satellite is operating (that is LEO, GEO, etc.,) and the mission objectives for which the satellite is operating.

6. REFERENCES

[1] Per Kragh Andersen, Ronald B Geskus, Theo de Witte, and Hein Putter. Competing risks in epidemiology: possibilities and pitfalls. *International journal of epidemiology*, 41(3):861–870, 2012.

[2] Lei Chen, Xian-Zong Bai, Yan-Gang Liang, and Kebo Li. *Orbital Data Applications for Space Objects*. 01 2017.

[3] Cameron Davidson-Pilon. lifelines: survival analysis in python. *Journal of Open Source Software*, 4(40):1317, 2019.

[4] Stefan Dietrich, Anna Floegel, Martina Troll, Tilman Kühn, Wolfgang Rathmann, Anette Peters, Disorn Sookthai, Martin Von Bergen, Rudolf Kaaks, Jerzy Adamski, et al. Random survival forest in practice: a method for modelling complex metabolomics data in time to event analysis. *International journal of epidemiology*, 45(5):1406–1420, 2016.
[5] Stephane Fotso et al. PySurvival: Open source package for survival analysis modeling, 2019–.
[6] Zachary Hall and Puneet Singla. *Reachability Analysis Based Tracking: Applications to Non-cooperative Space Object Tracking*, pages 200–207. 11 2020.
[7] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. 2008.
[8] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.
[9] Elisa T Lee and John Wang. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.
[10] Chirag Nagpal, Willa Potosnak, and Artur Dubrawski. auton-survival: an open-source package for regression, counterfactual estimation, evaluation and phenotyping with censored time-to-event data, 2022.
[11] Sebastian Pölsterl. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *Journal of Machine Learning Research*, 21(212):1–6, 2020.
[12] Rajesh Ranganath, Adler Perotte, Noémie Elhadad, and David Blei. Deep survival analysis. In *Machine Learning for Healthcare Conference*, pages 101–114. PMLR, 2016.
[13] Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang, Weinan Zhang, Lin Qiu, and Yong Yu. Deep recurrent survival analysis, 2018.
[14] Michael Schemper. Cox analysis of survival data with non-proportional hazard functions. *Journal of the Royal Statistical Society Series D: The Statistician*, 41(4):455–465, 1992.
[15] Robert Tibshirani. The lasso method for variable selection in the cox model. *Statistics in medicine*, 16(4):385–395, 1997.
[16] Anthony Turkson, Francis ayiah mensah, and Vivian Nimoh. Handling censoring and censored data in survival analysis: A standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences*, 2021:1–16, 09 2021.
[17] Vanya Van Belle, Kristiaan Pelckmans, Johan AK Suykens, and Sabine Van Huffel. Learning transformation models for ranking and survival analysis. *Journal of machine learning research*, 12(3), 2011.
[18] Hong Wang and Gang Li. A selective review on random survival forests for high dimensional data. *Quantitative bio-science*, 36(2):85, 2017.
[19] Xin-Sen Xu, Wei Chen, Run-Chen Miao, Yan-Yan Zhou, Zhi-Xin Wang, Ling-Qiang Zhang, Kai Qu, Qing Pang, Rui-Tao Wang, and Chang Liu. Survival analysis of hepatocellular carcinoma: A comparison between young patients and aged patients. *Chinese Medical Journal*, 128(13):1793–1800, 2015.
[20] Chander Prakash Yadav, V Sreenivas, MA Khan, and RM Pandey. Epidemiology: Open access. 2018.