The **introduction** highlights the growing availability of condition-monitoring data for components/systems and the development of data-driven Remaining Useful Life (RUL) prognostics in recent years. However, the authors point out that most studies focus on point RUL prognostics without providing insights into the uncertainty associated with these estimates, which limits their applicability in maintenance planning.

The authors discuss the limitations of existing studies, which either focus solely on developing RUL prognostics or propose predictive maintenance planning models that rely on simple, generic probability distributions for the time-to-failure/RUL. They emphasize the need for component-specific data-driven RUL prognostics that capture the unique degradation patterns exhibited by the sensor measurements.

The authors also mention that few studies integrate data-driven RUL prognostics into maintenance planning, and those that do often focus on single-component systems or have other limitations, such as relying on point RUL estimates or requiring high-fidelity physics-based models.

To address these gaps, the authors propose an end-to-end framework for multi-component predictive maintenance, starting from raw sensor measurements, progressing to probabilistic RUL prognostics, and finally to maintenance planning. They highlight the main contributions of their paper, which include:

1. Obtaining reliable data-driven probabilistic RUL prognostics that enable uncertainty-aware maintenance planning.
2. Specifying a renewal-reward process for RUL-integrated single-component maintenance optimization.
3. Proposing a multi-component maintenance planning model that incorporates RUL prognostics and considers resource constraints.
4. Demonstrating the approach on a case study and analyzing the performance relative to benchmarks.

**Section 2,** titled "Probabilistic RUL prognostics for turbofan engines using convolutional neural networks", focuses on developing a method to predict the remaining useful life (RUL) of aircraft turbofan engines using sensor data. The section is divided into three subsections.

**Subsection 2.1**, "Description of the dataset", introduces the C-MAPSS dataset used in the study. The dataset consists of four subsets (FD001, FD002, FD003, and FD004), each containing sensor measurements from 21 sensors for a set of engines. The data is further divided into training and test sets. The training sets include complete run-to-failure data, while the test sets contain partial sensor data up to a certain point before failure. The goal is to predict the RUL at that point. The authors preprocess the data by selecting 14 out of the 21 sensors with non-constant measurements and normalizing the sensor data using min-max normalization.

**Subsection 2.2,** "Architecture of the Convolutional Neural Network", describes the proposed CNN architecture for RUL prediction. The input to the CNN is a sliding window of the last N cycles of normalized sensor measurements. The CNN consists of L convolutional layers, each with K filters and a kernel size of 1×S. The convolutional layers are followed by a single convolutional layer with one filter and a kernel size of 1×S'. Finally, two fully connected layers are used to predict the RUL based on the extracted features from the convolutional layers.

**Subsection 2.3,** "Monte Carlo dropout", explains the technique used to obtain probabilistic RUL predictions. The authors apply dropout with a rate ρ to each layer of the CNN, except for the input layer. During inference, M forward passes are performed for each test sample, with different randomly selected neurons being dropped in each pass. This process approximates a deep Gaussian process, allowing the estimation of the probability density function (PDF) of the RUL. The authors provide a brief overview of the theoretical background behind this approach, explaining how the posterior distribution of the RUL is approximated using variational inference.

**Section 3,** titled "Results - Probabilistic RUL prognostics for aircraft turbofan engines", presents the results of applying the proposed CNN with Monte Carlo dropout method to the C-MAPSS dataset. This section is divided into three subsections.

**Subsection 3.1,** "Hyperparameter tuning", discusses the optimization of the CNN hyperparameters. The authors list the chosen hyperparameters in Table 2, which includes:

* Window size N: The number of past flight cycles included in each input sample (set to 30).
* Convolutional layers L: The number of convolutional layers in the CNN (set to 5).
* Number of filters K: The number of filters in each convolutional layer (set to 10).
* Kernel size S: The size of the one-dimensional kernels in the convolutional layers (set to 10).
* Kernel size S' in the last convolutional layer: The size of the one-dimensional kernel in the last convolutional layer (set to 3).
* Number of nodes in the fully connected layer: The number of nodes in the first fully connected layer (set to 100).
* Monte Carlo dropout rate ρ: The dropout rate applied to each layer except for the input layer (set to 0.5).
* Number of passes M: The number of forward passes performed for each test sample during inference (set to 1000).
* Rearly: The threshold for the piece-wise linear RUL target function (set to 125 flight cycles).

The authors also mention the use of zero padding for test instances in subsets FD002 and FD004 that have fewer than 30 historical flight cycles.

**Subsection 3.2,** "Mean RUL prognostics", evaluates the performance of the mean RUL predictions using the Root Mean Square Error (RMSE) metric. The authors compare their results with existing studies and show that their approach achieves comparable performance, especially for the more challenging subsets FD002 and FD004.

**Subsection 3.3**, "PDF of the RUL prognostics", assesses the reliability of the probabilistic RUL predictions using the α-coverage metric and the reliability diagram. The α-coverage measures the proportion of true RUL values falling within the α-confidence interval of the predicted RUL distribution. The authors report the α-coverage for α ∈ {0.5, 0.9, 0.95} and find that the values are close to the expected α, indicating reliable RUL prognostics. However, they also note that the mean widths of the confidence intervals are quite large, suggesting high uncertainty in the predictions. The reliability diagram, which compares the predicted probabilities with the observed frequencies, further confirms that the RUL prognostics are reliable, with the reliability curves being close to the ideal diagonal line.

**Section 4,** titled "Maintenance scheduling", focuses on integrating the probabilistic RUL prognostics obtained from Section 2 into maintenance planning for single and multiple components. This section is divided into two subsections.

**Subsection 4.1,** "Single-component replacement using probabilistic RUL prognostics and renewal-reward processes", formulates the problem of determining the optimal replacement time for a single component as a renewal-reward process. The objective is to minimize the long-term average cost per unit of time, considering the expected replacement cost and the expected lifetime of the component.

The authors introduce the following notation:

* 𝑘: The current time step (i.e., the number of time steps the component has been used).
* 𝑡𝑘: The number of time steps from the current moment until the scheduled preventive replacement.
* 𝑐r: The cost incurred for a preventive replacement.
* 𝑐f: The cost incurred for a failure replacement, with 𝑐f > 𝑐r.
* 𝜙𝑘(𝑖): The probability that the component has a remaining useful life of exactly 𝑖 time steps, given that it has been used for 𝑘 time steps.

The optimal replacement time 𝑡∗𝑘 is determined by minimizing the ratio of the expected replacement cost to the expected lifetime of the component:

E[𝐶(𝑘, 𝑡𝑘)] / E[𝐿(𝑘, 𝑡𝑘)]

where E[𝐶(𝑘, 𝑡𝑘)] and E[𝐿(𝑘, 𝑡𝑘)] are the expected replacement cost and the expected lifetime of the component, respectively, given the current usage 𝑘 and the scheduled replacement time 𝑘 + 𝑡𝑘.

Subsection 4.2, "Multi-component replacement using probabilistic RUL prognostics", extends the single-component problem to a multi-component setting. The authors consider a set of components 𝑉 and introduce additional notation:

* 𝑑p: The present day.
* 𝑑v0: The installation day of component 𝑣 ∈ 𝑉.
* 𝑘p: The usage time of component 𝑣 ∈ 𝑉 at the present day 𝑑p.
* 𝑆v: The set of maintenance slots available for component 𝑣 ∈ 𝑉 in the period [𝑑p, 𝑑p + 𝑙], where maintenance slots are known 𝑙 days in advance.
* 𝑆g: The set of generic maintenance slots available in the period [𝑑p, 𝑑p + 𝑙], where any component can be replaced at an additional cost 𝑐g.
* ℎ: The maximum number of components that can be replaced per day due to limited resources.

The authors propose an integer linear programming (ILP) model to plan component replacements at the present day 𝑑p for the time window [𝑑p, 𝑑p + 𝑙). The objective function minimizes the total expected costs over the expected lifetime of all components. The model considers costs related to preventive replacements, failure replacements, and the use of generic maintenance slots.

The multi-component maintenance planning problem is solved using a rolling horizon approach, where the planning is performed for a period of 𝑇 days. The maintenance decisions for the first 𝜏 days (𝜏 < 𝑙) are executed, and the process is repeated by updating the present day and the RUL prognostics until the entire planning horizon of 𝑇 days is covered.

**Section 5,** titled "Results - Maintenance planning for turbofan engines", demonstrates the application of the maintenance planning methodology proposed in Sections 2 and 4 to the C-MAPSS dataset. This section is divided into four subsections.

**Subsection 5.1,** "Results - Probabilistic RUL prognostics for the maintenance planning of turbofan engines", describes the dataset used for maintenance planning. The authors randomly select 80% of the engines from each C-MAPSS training set (568 engines in total) to train the CNNs and use the remaining 20% (141 engines) to generate RUL prognostics for maintenance planning. The subsection also presents examples of the predicted RUL distributions for specific engines at different time points and analyzes the performance of the RUL prognostics using metrics such as RMSE and α-coverage.

**Subsection 5.2,** "Results - Single-engine maintenance planning", discusses the optimal replacement times obtained using the renewal-reward process described in Section 4.1. The authors present examples of the optimal replacement times for four engines at different points in their lifetimes. They observe that the optimal replacement time tends to be close to the lower bound of the 99% confidence interval of the predicted RUL distribution. The subsection also highlights the relationship between the uncertainty in the RUL predictions and the timing of preventive replacements.

**Subsection 5.3,** "Results - Multi-engine maintenance planning", presents the results of the multi-component maintenance planning approach described in Section 4.2. The authors consider a fleet of 50 engines and set up the problem with specific parameters, such as maintenance slot availability, replacement costs, and planning horizon. They illustrate the maintenance planning process using examples at different time points and discuss how the planned replacement times can deviate from the single-component optimal times due to resource constraints.

**Subsection 5.4,** "Long-term performance: Maintenance with probabilistic RUL prognostics vs maintenance with perfect RUL prognostics and time-based maintenance", compares the long-term performance of three maintenance strategies: the proposed approach using probabilistic RUL prognostics, a strategy with perfect RUL prognostics, and a time-based maintenance strategy. The authors conduct a Monte Carlo simulation with 10,000 runs, each covering a 10-year period. They evaluate the strategies based on metrics such as the expected number of replacements, failures, and maintenance costs. The results show that the proposed approach leads to significant cost savings and failure reductions compared to time-based maintenance, and performs only slightly worse than the perfect prognostics scenario.

In the conclusions section, the authors summarize their work on developing an end-to-end framework for predictive maintenance of complex components/systems. They highlight the key aspects of their approach, which include:

1. Data-driven probabilistic RUL prognostics using Convolutional Neural Networks (CNNs) with Monte Carlo dropout.
2. Integration of the obtained probabilistic RUL prognostics into maintenance planning for both single-component and multi-component systems.

The authors reiterate the results of their case study on turbofan engines, emphasizing that the proposed approach leads to:

1. Reliable RUL prognostics with high α-coverage values.
2. Significant cost savings and failure reductions compared to time-based maintenance strategies.
3. Performance close to the perfect prognostics scenario, with only a slight increase in costs and failure occurrences.

The authors conclude by highlighting the potential benefits of their data-driven predictive maintenance approach in terms of cost reduction and reliability improvement. They also mention their plans for future work, which include further analyzing the impact of cost choices on maintenance planning results and improving the RUL prognostics by incorporating additional features, such as attention mechanisms, into the neural network architecture.