A proposed Digital Twin concept for the smart utilization of Sustainable Aviation Fuels

B. Enderle*, B. Rauch, C. Hall and U. Bauder
German Aerospace Center (DLR)
Institute of Combustion Technology
Pfaffenwaldring 38-40, 70569 Stuttgart, Germany

In the efforts to reduce both CO_2 emissions and non-CO_2 climate impacts in global aviation, encouraging the take-up of *Sustainable Aviation Fuels* (SAF) is seen as a key element. Although there are currently seven ASTM-approved SAF production pathways, availability of the required feedstock and increased costs compared to conventional jet fuels pose major challenges for a fast market uptake. It is therefore advisable to use the currently available capacities in a *smart* way, i.e. taking into consideration the maximum potential for mitigating climate impact for a given scenario as well as a cost-optimal and economic efficient allocation in the operational context. For such complex tasks, all information regarding the system under consideration should be made available over the entire lifecycle. For this purpose, a Digital Twin concept for SAF is proposed and demonstrated, linking together information from models and data to form a virtual representation of the fuel and mirror the fuel during its lifecycle. Special emphasis is placed on the representation of uncertainties of the involved model predictions and the traceability of the provenance of all information within the Digital Twin. The methodology is tested and demonstrated for a decision making problem typical for the utilization of SAF under the current regulatory constraints, namely the blending of conventional fuel with SAF to a drop-in SAF blend. The fact that the physical properties of the conventional jet fuel might vary within its specification requires a careful selection of the blend ratio and the candidate SAF to achieve an optimal blend given associated cost functions and constraints. It is identified that due to the defined constraints and the cost function, not always the SAF with the highest possible blend ratio is selected as the best choice. It is shown that the inclusion of uncertainties narrows the range of possible blend ratios and hence affects the decision process and the optimal decision. However, this approach reflects a risk-informed decision since uncertainties in the model prediction are considered in the decision process.

I. Introduction

In the efforts to reduce both CO_2 and non-CO_2 emissions in global aviation, encouraging the take-up of *Sustainable Aviation Fuels* (SAF) is seen as a key element [2, 3]. While the general term *Aviation Alternative Fuels* (AAFs) refers to fuels produced from feedstock other than crude oil, SAFs additionally require strict compliance with sustainability criteria along the production cycle and a substantial reduction of lifecycle CO_2 emissions compared to a fossil fuel [4]. This broad definition results in a variety of possible production pathways for SAFs [5] involving a diverse range of bio-based feedstock. Although there are currently seven ASTM-approved SAFs [6], regulatory policies as well as technological limitations only allow the use of SAFs in blends with conventional jet fuels in a range between 10% and 50% SAF content. Thus, depending on the blend ratio and the chosen feedstock, the reduction potential for CO_2 and non-CO_2 emissions of a specific SAF varies. The fact that the price of SAF today is higher than petroleum-based Jet A and SAF availability is still limited poses another major challenge in its market uptake.

It is therefore advisable to use the currently available volumes in a *smart* way, i.e. taking into consideration the maximum potential for mitigating climate impact for a given scenario, as well as a cost-optimal and economic efficient allocation in the operational context. This requires the consistent collection and tracking of all available data and information along the lifecycle of a SAF with respect to the considered scenario. Only on the basis of this knowledge optimal decisions regarding SAF deployment can be made and strategies for future climate mitigation effective SAF utilization can be inferred. Furthermore, there is an increasing need for airlines to precisely track the CO_2 footprint of the utilized fuel in order to profit from accounting and offsetting schemes like CORSIA [7] or EU-ETS [8] when using

*Post-doctoral researcher, DLR Institute of Combustion Technology, Stuttgart, Germany

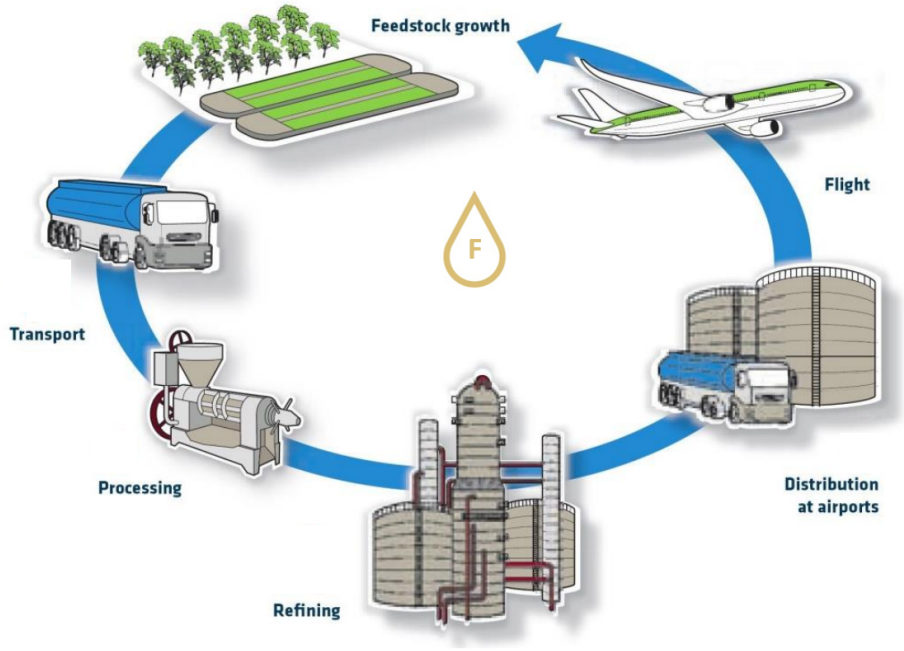


Fig. 1 Sustainable Aviation Fuel lifecycle. [1]

SAF. From a technical perspective, the quantitative impact of SAF utilization on Maintenance, Repair and Operation (MRO) is not clear. Point measurements showed potentially reduced maintenance and repair costs, as well as operational advantages due to lower aromatics content and higher thermal stability of SAF. Systematic tracking of fuel impacts on MRO could potentially guide future decision making.

In addition to the data collected, predictions from computer models enable the evaluation of SAF usage over the life cycle prior to the actual deployment in an aviation mission. These models cover, inter alia, the prediction of unknown fuel properties [9], the sensitivity of the combustion process on varying fuel composition in the engine [10] or the expected climate impact, e.g. from the resulting soot emissions. Connecting these models and augmenting them with the gathered real-world data yields a holistic virtual representation of the fuel's lifecycle and its impact on the various stakeholders, commonly referred to as a *Digital Thread* [11]. As an integral part of this Digital Thread, a *Digital Twin* (DT) embodies a digital representation to closely mirror or emulate the life of an entity along the Digital Thread, in this particular case the fuel under consideration. It is important to note that a single Digital Thread might incorporate multiple Digital Twins interacting with each other as for example in the case of multiple candidate fuels which are compared and selected for a specific scenario.

Digital Twins are gaining increasing interest for their potential of a unified virtual representation in fields such as manufacturing, aerospace, healthcare and medicine. Proposed applications range from smart manufacturing of components [12] up to full scale model representations of physical systems [13] for the purpose of a streamlined certification process. As the latter involves high consequence decisions, uncertainties in the models and data contained in the Digital Twin must be taken into consideration. Clearly, this requires the underlying architecture of the Digital Twin to cope with uncertain data, commonly implemented in a probabilistic approach [14]. Combining methods and concepts for Digital Twins and uncertainty quantification finally enables risk-informed decision making on the basis of a comprehensive and consistent source of information.

Due to the various sources of data, information and model predictions contributing to a Digital Twin, keeping track of *data provenance* is another critical issue in the application of Digital Twin concepts. The broad term provenance thereby refers to understanding where a piece of data originates from and how it is connected to other datasets and models [15]. Therefore, data provenance modeling of the Digital Twins can be seen as a connecting backbone within a Digital Thread system.

Although a growing body of literature on Digital Twin concepts exists [16, 17], state-of-the art Digital Twins are largely the result of custom implementations that require considerable deployment resources and a high level of expertise in the specific field of application [14]. This paper therefore aims at developing a Digital Twin for Sustainable

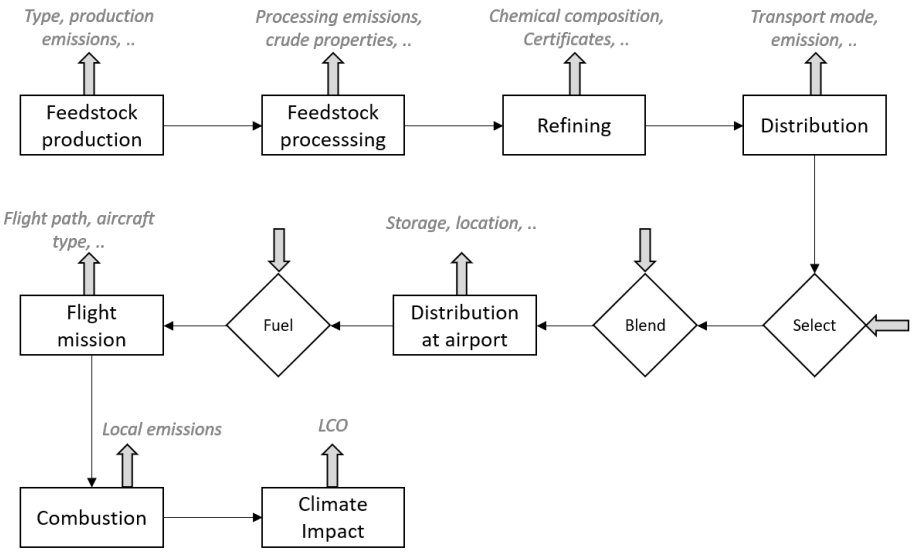


Fig. 2 Flowchart of the SAF lifecycle involving operations (boxes), decisions (diamonds) and data and information streams (gray arrows).

Aviation Fuels that is in conjunction with a superordinate Digital Thread concept. Special emphasis is placed on the representation of uncertainties of the involved model predictions and the traceability of the provenance of all information within the system.

II. Problem formulation

This section describes the problem considered in this paper and identifies and defines the involved elements which must be represented in the Digital Twin system.

A. Sustainable Aviation Fuel lifecycle

Figure 1 summarizes the typical lifecycle of a bio-based SAF from feedstock growth to actual utilization in a flight mission. It involves the production of an arbitrary feedstock (e.g. algae, biowaste, used cooking oil), its transport and pre-processing, the conversion into a SAF through a refining process as well as the distribution at the airport. Although not shown here, all of these high level elements might incorporate several additional processes which are specific to the chosen production pathway and utilization strategy. This *feedstock-to-wake* lifecycle can be seen as an extension of the *well-to-wake* lifecycle concept for conventional jet fuels [18]. In contrast to the latter, several SAF specific challenges exist:

- Varying lifecycle emissions depending on the SAF production pathway
- Increased number of stakeholders (e.g. feedstock producers) compared to conventional jet fuels
- Increased level of process complexity over the whole lifecycle
- Additional regulatory policies, for example regarding the allowed blend ratio with a conventional fuel

On a more abstract level, this lifecycle can be exemplified in a flowchart as depicted in Fig. 2. It interprets the SAF lifecycle as a sequence of operations, decisions and information streams. Operations (indicated as black boxes) can be seen as a high level summary of events connected to the fuel. Again, only the most relevant ones are shown for brevity. Amid all of these operations information and data regarding the fuel are generated, ranging from the CO_2 emission during the production over the chemical composition until the local emissions during the flight. On the basis of these information, decisions (shown as diamonds) regarding the fuel and its utilization must be made. Exemplary shown are the decision which SAF to select, how to blend with a conventional fuel and how much fuel to lift on the wings of the aircraft for a specific mission. Naturally, all of these decisions influence the downstream part of the lifecycle. Furthermore, for an optimal decision, all upstream information and data should be available. Thus, collecting all of the data along the lifecycle and providing a basis for the associated decisions is a key task for a Digital Twin in this scenario.

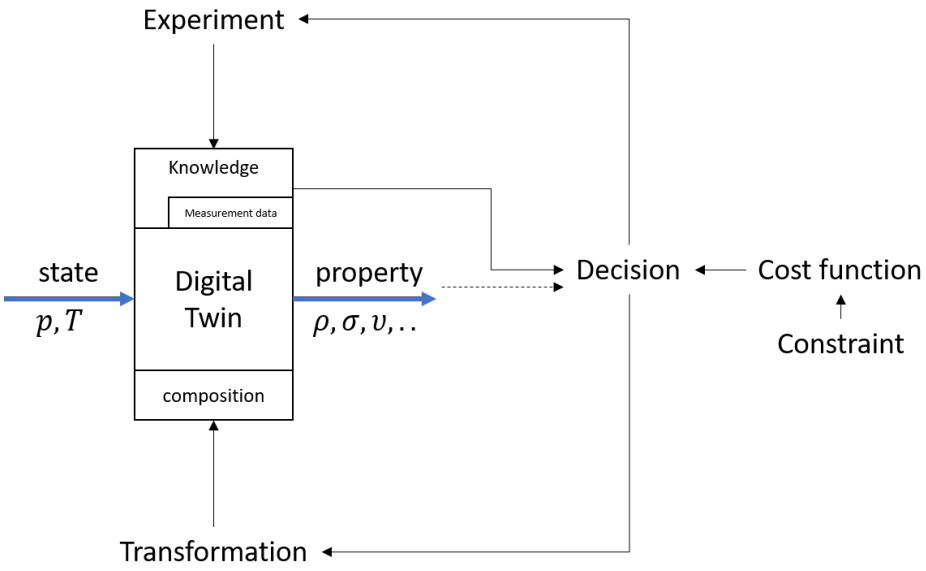


Fig. 3 System abstraction of a Digital Twin -based decision.

Ultimately, information and decisions from consecutive runs of the Digital Twin serve as a feedback loop to optimize the usage of the fuel in the operational context, hence enabling a *smart utilization*.

Note that the work at hand specifically focuses on the last elements of the lifecycle in which the deployment and utilization of a produced SAF at the airport is considered. Therefore, upstream processes from the SAF production are not explicitly taken into consideration but are represented in terms of resulting data, for example for the chemical composition of the SAF.

B. System abstraction

For the definition of key terms and concepts, we take a closer look at the abstraction of a Digital Twin-based decision as summarized in Fig. 3. The Digital Twin, i.e. the virtual representation of the fuel, forms the core element. In the DT, *knowledge* regarding the real world entity is collected and processed. Here, knowledge entails all information characterizing the Twin *as it is* as well as *how it could be*. While the former refers to data collected during the runtime of the DT, the latter comprises for example results and findings from model based predictions regarding downstream operations in the lifecycle. A specific example of a knowledge element is given by measurement data as a result of an experimental characterization of the real word entity. The chemical composition of the fuel represents the most fundamental piece of measurement data, as it provides the basis for the characterization of the fuel.

At any point within the lifecycle a *state* is imposed on the DT by the surrounding. In the case of a liquid fuel, this state is determined by the surrounding pressure and temperature. From the imposed state and the knowledge regarding the Twin, physical properties can be deduced, for example the density or viscosity. In the present DT concept, this state/property mapping is realized using advanced models for each of the required fuel properties (see Sec. III.A and III.A.1).

Inherent characteristics of the DT (e.g. the chemical composition) are changed by a *transformation*. The most prominent example is the blending of the fuel within the DT with an other fuel (change of composition). Operations feeding data and knowledge into the Digital Twin are summarized as *experiments* and are not limited to physical experiments but can include auxiliary models which use the DT, providing insight into the entity and hence additional knowledge.

Finally, transformations or experiments can be affected by a *decision*. For example, a decision regarding a specific blend ratio causes a transformation of the chemical composition of the DT, thus changing the inherent characteristics of the DT. Decisions are driven by a *cost function* which is an arbitrary function that must be optimized to achieve an optimal decision. Thus, this cost function is not limited to economic quantities but might incorporate fuel properties and knowledge calculate and contained in the DT. Typically, such cost functions are subject to constraints and policies

restricting the domain over which the cost is optimized. An important example for such a constraint is the maximum allowed blend ratio of 50% volume for synthetic fuels as stated in ASTM D1655.

III. Methodology

During the distribution and utilization process as detailed in the previous section, different stakeholders require different data regarding the SAF, e.g. fuel properties for blending or CO_2 production footprint for reporting and accounting. Preferably, this data is collected, stored and provided through a unified virtual representation of the fuel under consideration. For this purpose, a Digital Twin concept for SAF is proposed.

A. Digital Twin

In this approach, the fuel and its properties from the physical space are represented in the virtual space using property models and collected data. A schematic of this concept is given in Fig. 4. A fuel in the physical space is determined by its chemical composition C and a state $S = \{p, T\}$ imposed by the environment. p and T denote the pressure and temperature, respectively. In dependance of the state and the composition, physical properties Π of the fuel can be measured. Π might include, inter alia, density ρ, surface tension σ or kinematic viscosity ν. For the Digital Twin, this system is abstracted in the virtual space through suitable property models and data storage and data fusion methods. Hence, for all states imposed on the fuel in the physical space, the corresponding fuel properties are returned by the Digital Twin in the virtual space without the need of further measurements or experiments.

1. Fuel property models

The fact that conventional and alternative jet fuels both are highly complex mixtures consisting of hundreds of possible hydrocarbon combinations still poses major challenges to the accurate model-based prediction of their respective physical properties. Physical properties of the fuel are therefore calculated within the Digital Twin using probabilistic Machine Learning models. From this broad class of models, a probabilistic model based on the Mean Quantitative-Structure Property Relationship (MQSPR) method [19] is selected for the work at hand. In this approach, fuels are represented as one mean pseudo-structure that is inferred by a weighted average over structures of 1887 molecules that could be present in the individual fuel. The MQSPR method assumes a relationship between pseudo-structures and physical properties of compounds as well as similar chemo-physical properties for compounds with similar structure and substructures [20]. The resulting pseudo-structure is then correlated with the physical properties of the jet fuel.

A Deep Neural Network algorithm with the Monte-Carlo Dropout technique (MCNN) [21] is applied for the correlation of the fuel properties with the pseudo-structures. During the prediction of a dataset, network neurons are deactivated randomly and the prediction is repeated multiple times, resulting in varying outputs and finally a distribution of the predictions. This distribution thus characterizes the model uncertainty. The models are trained using a broad database of XX conventional fuels and YY synthetic fuels. An example of the distribution of property predictions for a conventional jet fuel is given in Fig. 5. The density of the fuel is evaluated at $T = 15°$ C.

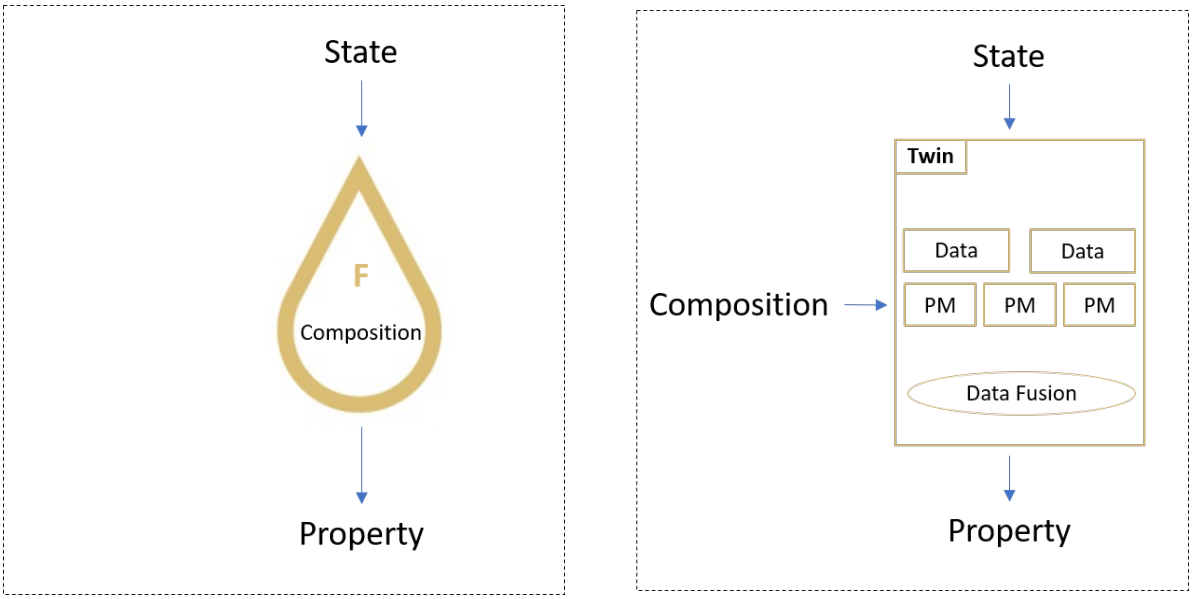
Chemical composition C from GCxGC characterization and the state S are required as inputs to the MQSPR model. Thus, fuel properties are calculated from the corresponding MQSPR model as

$$\Pi = \mathcal{M}_{MQSPR}(S, C). \tag{1}$$

Further details on the used fuel property models are given in the work of Hall et al. [19].

2. Data provenance

With each interaction of the Digital Twin, data and information contained in the Digital Twin might be altered or updated, for example when a new state of the fuel is examined and hence the fuel properties are recalculated. In order to keep track of such changes, information regarding the provenance of data must be captured and stored within the Digital Twin. Several approaches for provenance models are available from the literature [22], among which the W3C PROV Recommendation [23] is a widely accepted standard. In this standard, data provenance is defined as information about *entities*, *activities*, and *agents* involved in the generation of a piece of data and their relationship to each other. The core concept involving these three main components and their relationship is shown in Fig. 6: An *entity* is a generic thing that one wants to describe the provenance of. An *activity* is a process that takes place in a specific setting and timing, changing or using the *entity*. Finally, an *agent* is responsible for an *activity*. In the context of the fuel Digital Twin,



(a) Physical space **(b) Virtual space**

Fig. 4 Abstraction of a fuel through a Digital Twin (PM=Property Model).

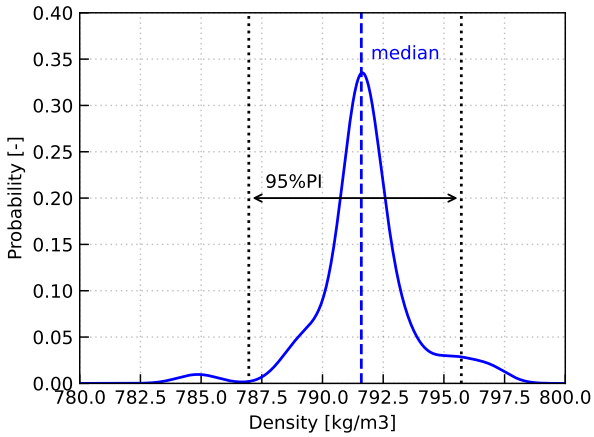


Fig. 5 Distribution of density predictions from the MQSPR model for a conventional jet fuel at $T = 15°$ C.

properties of the fuel would correspond to *entities*, while a model calculation within the Digital Twin can be seen as an *activity*. Lastly, the subordinated process in which the Digital Twin is used acts as an *agent* in the provenance model.

All operations concerning the Digital Twin are tracked and described during the runtime using these definitions. Provenance information is made available in terms of a graph of all recorded provenance.

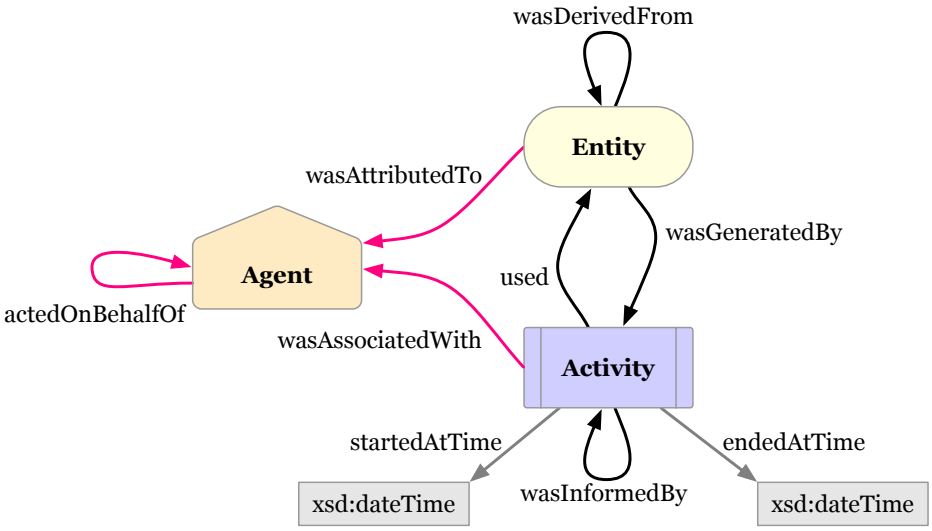


Fig. 6 Core concept of the W3C PROV model [24].

3. Uncertainty management

To enable risk-informed decision making using the Digital Twin, uncertainties in the model prediction must be considered. The fact that probabilistic Machine Learning models are utilized for the prediction of fuel properties within the Digital Twin provides access to the uncertainties in the model predictions. Consequently, respective prediction intervals are provided and stored for each fuel property. With this information at hand, non-deterministic simulations can be conducted using the Digital Twin. The full distribution of the property models is reduced to the median of the distribution as well as the boundaries of the interval covering 95% of the predictions (95% prediction interval/PI). Both values are indicated in Fig. 5 for an exemplar density prediction.

IV. Example case study

In order to demonstrated the proposed Digital Twin framework, an example case involving the blending of a conventional jet fuel with one of three candidate SAFs is studied. In this scenario, three different types of SAF are available at a blending location (e.g. an airport). The fact that the physical properties of the conventional jet fuel might vary within its specification and all three SAFs differ in their respective physical properties requires a careful selection of the blend ratio and the candidate SAF to achieve an optimal blend. Figure 7 summarizes the resulting decision problem and the corresponding DT workflow according to the concept from Fig. 3. In a first step, a virtual blending study is conducted involving each of the three SAFs. Hence, this operation describes an *experiment* increasing the information stored with in the DT. Based on the acquired information, a *decision* which of the three fuels to select for blending and which blend ratio to use is made. This decision involves *constraints* from the regulatory limits regarding SAF blending (e.g. maximum blend ratio, minimum density of the resulting blend) as well as a *cost function* rating the available choices in the decision. Finally, the conventional jet fuel is blended with the SAF DT from the decision.

The inclusion of uncertainties are demonstrated through the example of prediction intervals of the underlying fuel property models in the DT (see Sec. III.A.1). A simplified cost function based on the estimated soot reduction potential is utilized. Through the DT concept, data provenance as described in Sec. III.A.2 is tracked using the W3C model.

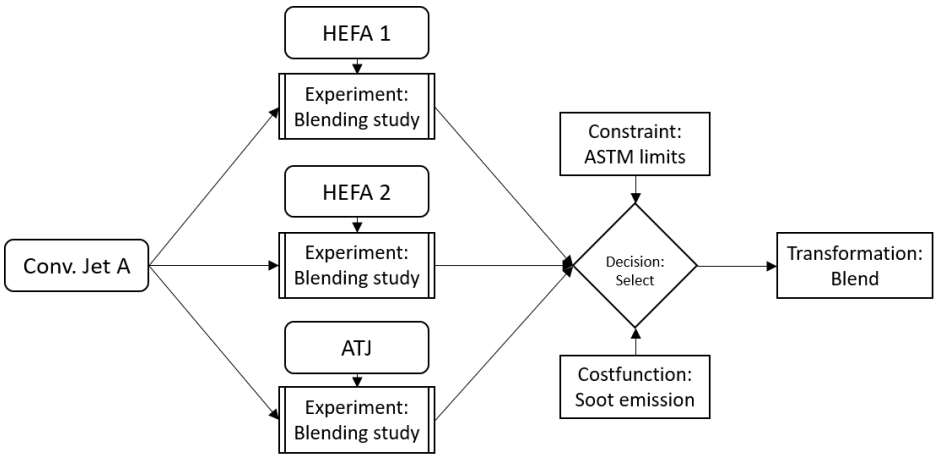


Fig. 7 Decision problem and DT workflow for SAF blending.

A. Fuels

Results from the 2006 CRC world fuel survey [25] provide the data basis for the conventional jet fuels in the present study. From this survey, detailed data including physical properties as well as chemical composition (from GCxGC measurements) is available for 28 different Jet A compliant jet fuels. These 28 fuels are used to mimic the variation in physical properties over conventional fuels that might be available at the blending location. This is illustrated in Fig. 8 for four key properties, namely aromatics content Y_{arom}, density ρ_{15} at 15 deg celsius, hydrogen content Y_h and net specific energy E_{net}. Boxplots portray the variation of these quantities over the 28 fuels. Data was extracted from respective experimental measurements reported in the CRC world fuel survey.

As SAFs, two fuels from a HEFA (Hydro-processed Esters and Fatty Acids) production pathway and one from an ATJ (Alcohol to Jet) production route are considered. While the HEFA pathway has the highest technological maturity the ATJ production routes currently do not have much technological development but might gain importance in the future.

Especially in the HEFA case, the chemical composition and thus the performance of the fuel highly depends on the feedstock type. For this reason, two different HEFA fuels are compared. HEFA 1 was characterized in depth during the European Unions JETSCREEN research project. From this project, an extensive collection of measurement data regarding chemical and physical properties is available. HEFA 2 is a fuel currently under investigation at the DLR Institute of Combustion Technology and varies slightly in the chemical composition compared to HEFA 1.

The ATJ fuel under consideration was also characterized as part of the JETSCREEN project and is composed of 99.9% of iso-paraffins. A comparison of the key properties of the three SAFs is given in Fig. 8. All three SAFs have an aromatics content below 0.5% and a hydrogen content and specific energy above the conventional fuels from the CRC survey. Thus, resulting blends will have a higher hydrogen content then the neat conventional fuel but a reduced aromatics content and density.

B. Cost function and constraints

A simplified correlation for the soot emission of a fuel based on its hydrogen content is utilized as a cost function. This reflects the fact that through the use of SAF blends not only the CO_2 emissions are reduced but also non-CO_2 emissions such as soot can be reduced. Through this approach, the SAF blend with the highest soot reduction potential is selected.

The correlation is derived from experimental data. During the EU JETSCREEN project [26], the effect of different conventional and alternative jet fuels on pollutant emissions, principally non-volatile particles, but also NOx was measured in a detailed experimental measurement campaign. Various combustor configurations were tested under laboratory conditions, going from academic ones (for better understanding), to more representative ones (APU, tubular combustor and injection system representative of a real engine). Results from the APU measurements [27] are considered in the following since this configuration and the conditions (power setting etc.) closest resemble the conditions in a real aircraft engine. Experimental data for the relative black carbon mass concentration in dependence of the fuel hydrogen content is given in Fig. 9. They demonstrate an exponential decay in concentration with increasing hydrogen content.

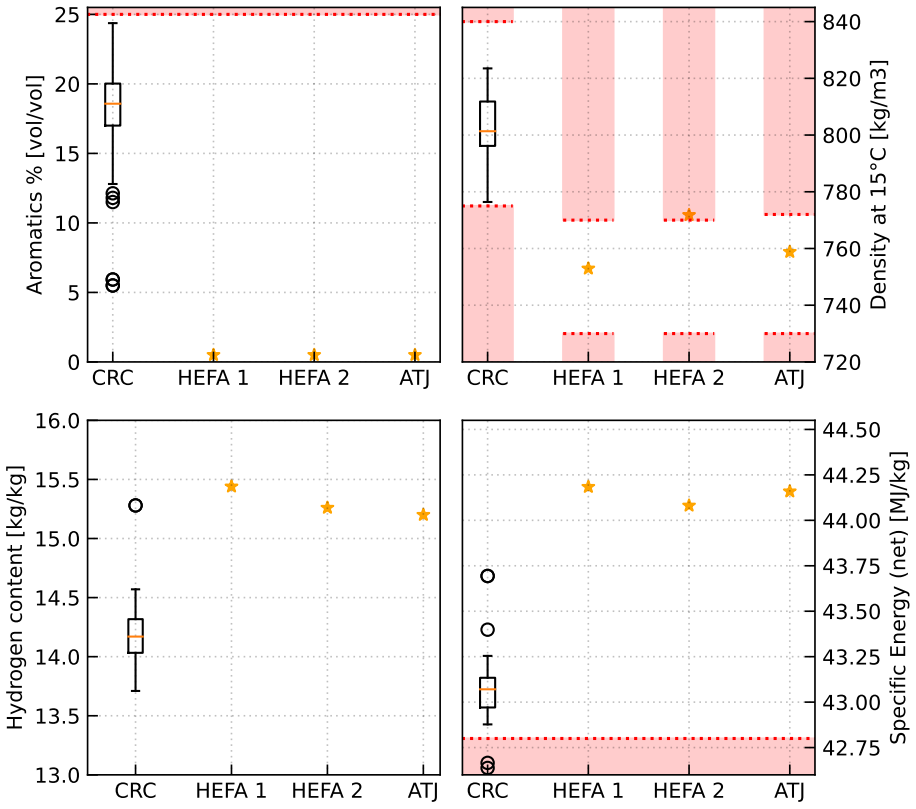


Fig. 8 Variation in fuel properties for conventional jet fuels (CRC) and the three SAFs.

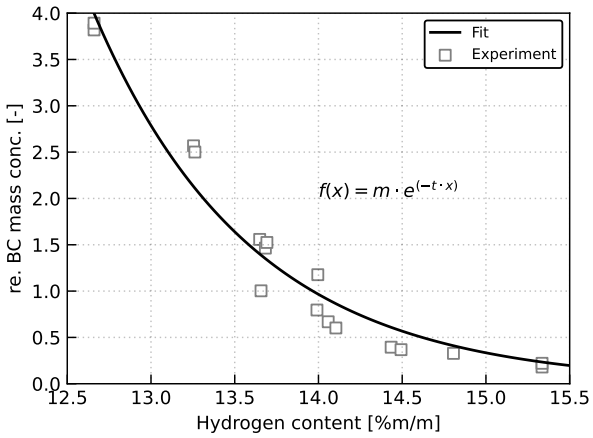


Fig. 9 Cost function based on hydrogen content and black carbon mass concentration (soot).

Thus, an exponential function can be fitted against the experimental results (least squares fit with correlation coefficient $R^2 > 0.9$).

Hydrogen content of the SAF blends is calculated based on the resulting chemical composition. Thus, the cost function $\mathrm{Soot}(Y_H)$ can be transformed to $\mathrm{Soot}(x_Bl)$, evaluating the soot reduction in dependence of the blend ratio.

Constraints for the selection problem are determined by the standard specification for jet fuels containing synthesized hydrocarbons under ASTM D7566-20c. For brevity, only the constraints regarding density ρ_{Bl} and aromatics content $Y_{arom,Bl}$ of the blend are considered for the present example. Both are known to be challenging to match when aiming for higher blend ratio. Furthermore, for both ATJ and HEFA blends, the maximum allowed blend ratio in ASTM D7566 is 50% volume w.r.t the synthetic compound. Since the density ρ_{Bl} for the standard specification must be evaluated at a liquid temperature of $T = 15$ °C, all Digital Twins operate at a state of $\Pi = \{T, p\} = \{15\ °\mathrm{C}, 1\ \mathrm{bar}\}$. Hence, the resulting set of state, cost function and constraints reads:

State:

$$\{T, p\} = \{15\ °\mathrm{C}, 1\ \mathrm{bar}\}$$

Cost function:

$$\mathrm{Soot}(Y_H) = \mathrm{Soot}(x_{Bl})$$

Constraints:

$$x_{Bl} \in [0, 0.5]$$
$$\rho_{Bl} \in [775, 840]$$
$$Y_{arom} \in [8.0, 25]$$

C. Results

The decision problem from Fig. 7 is evaluated for each of the 28 conventional fuels from the CRC survey (see Sec. IV.A). In the following, results from a single evaluation of fuel *CRC334 POSF4385* are discussed at first. Then, results from all evaluations are compared.

1. Single conventional fuel evaluation

Figure 10 displays the provenance graph of the conventional fuel *CRC334 POSF4385* before blending and hence before the decision problem. Three operations (orange boxes) were already conducted: the DT was created followed by an experimental characterization of the chemical composition through GCxGC analysis. This data was read into the DT, determining the composition. With this information, all physical properties were recomputed. Note that for brevity, only density, viscosity and surface tension are included in the graph. Furthermore, additional data from physical experiments within the CRC survey were provided to the DT, indicated by an additional entity (yellow element *MeasurementData:density*). Thus, this graph provides traceable information regarding the history of the DT and its

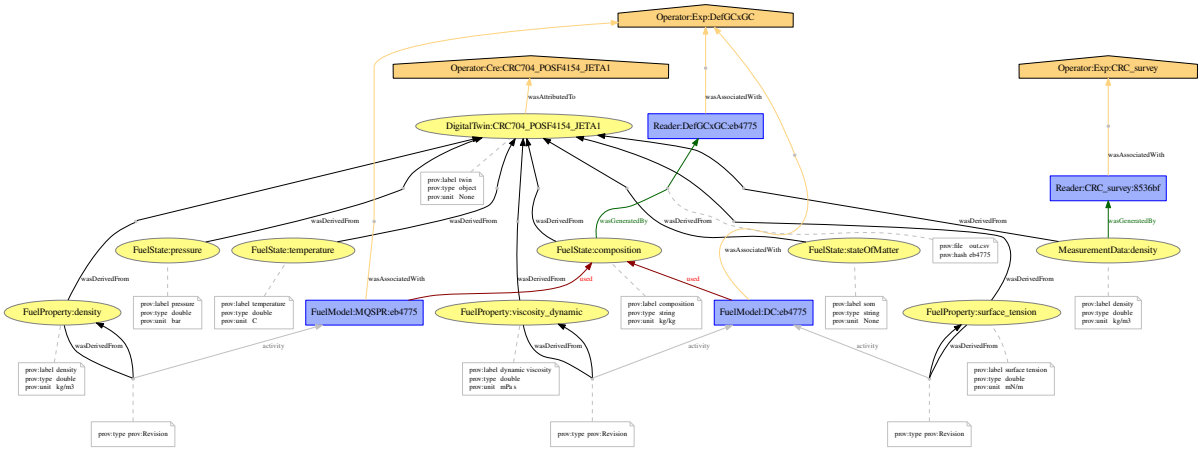


Fig. 10 Provenance graph of a conventional fuel DT before blending.

contained data.

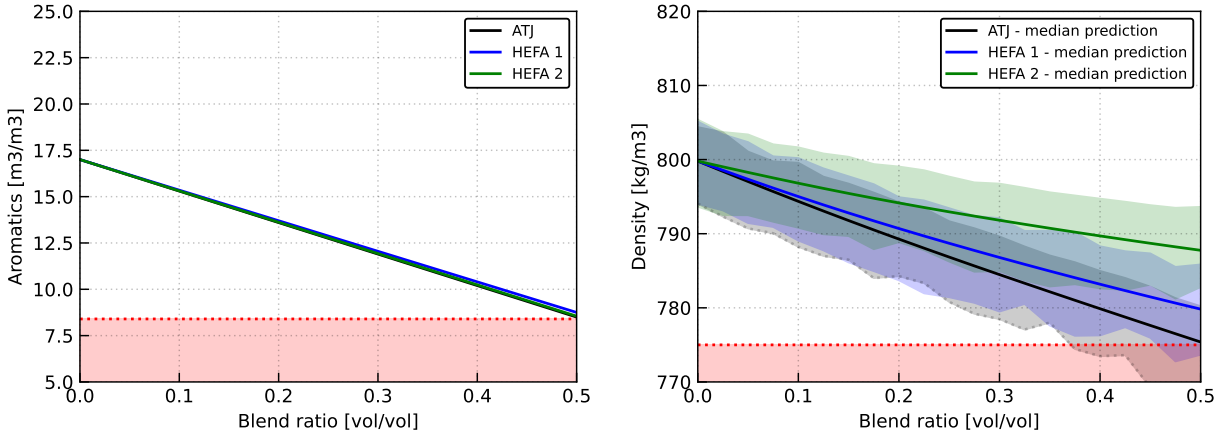
In the next step, a blending study is conducted for each of the three SAFs. For this purpose, the conventional fuel is gradually mixed with each of the three SAFs taking into consideration the constraint for the blend ratio. For each mixture, physical properties are evaluated. As aforementioned, only aromatics content and density are considered at this point. While the density is computed using the MQSPR model detailed in Sec. III.A.1, aromatics content is calculated using a linear mixing rule:

$$Y_{arom,bl} = \sum_i r_i Y_{arom,i} = x_{bl} \cdot Y_{arom,conv.} + (x_{bl} - 1) \cdot Y_{arom,SAF}, \tag{2}$$

where $r_{rel,i}$ is the relative volume in the mixture. For the conventional fuel, $r_i = x_{bl}$ while for the SAF $r_i = (x_{bl} - 1)$. Results from the blending study for fuel *CRC334 POSF4385* are given in Fig. 11. Figure 11a shows the aromatics content in the blended fuel over the blend ratio. Values outside the constraint are indicated by a red shade. Since the aromatics content is computed from a mixing rule, only deterministic results, i.e. a single line for each SAF, is given. For all three SAFs, the blend is within the constrained area regarding the blend ratio. In contrast, the blends' density is predicted by calculation the new composition of the blend based on the GCxGC data of the neat fuels and the blend ratio. This new composition is then used as an input to the probabilistic ML model and hence the 95% prediction interval (colored shades) can be taken into consideration alongside the median prediction (solid lines). This reveals uncertainties around the median prediction in the magnitude of $\pm 5 \mathrm{kg/m^3}$ and highlights the importance of the clear portrayal of uncertainties: When considering only the median values, all blends from all three SAFs would stay within the constraint regarding minimum density. However, if it is demanded that the whole 95% prediction interval must be above the minimum constraint, ATJ blends with a blend ratio $x_{bl} > 0.375$ as well as HEFA1 blends with $x_{bl} > 0.375$ drop below the minimum density.

Due to this requirement, the cost functions for ATJ and HEFA1 are limited to maximum blends of $x_{bl} > 0.375$ and $x_{bl} > 0.375$ as shown in Fig. 12. For this case, HEFA1 represents the best choice, i.e. the SAF for which the minimum value of the cost function is achieved, corresponding to the highest possible reduction in soot emission. It is important to point out that this does not relate to the maximum possible blend ratio for fuel *CRC334 POSF4385* which could be achieved using HEFA2. Hence, a higher impact on the ultimate objective (here: reducing soot emissions) is possible without the maximum possible blend ratio among the three SAFs.

Finally, Fig. 13 shows the essential elements of the provenance graph of fuel *CRC334 POSF4385* after the decision and the blending. Again, all operations are given as orange boxes, namely the three blending studies, the decision and the the blending with the best choice. Interaction with other DTs (for the three SAFs) is indicated by gray entities. It can be traced back that the blending operation changed the composition of the DT and triggered the recomputation of the physical properties of the fuel.



(a) Aromatics content **(b) Density**

Fig. 11 Calculated physical properties of the blends for fuel CRC334 POSF4385. Shaded areas around median prediction indicate 95% prediction interval.

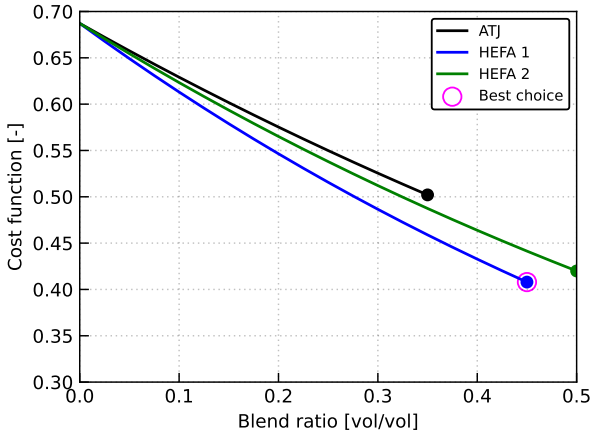


Fig. 12 Constrained cost function for fuel CRC334 POSF4385.

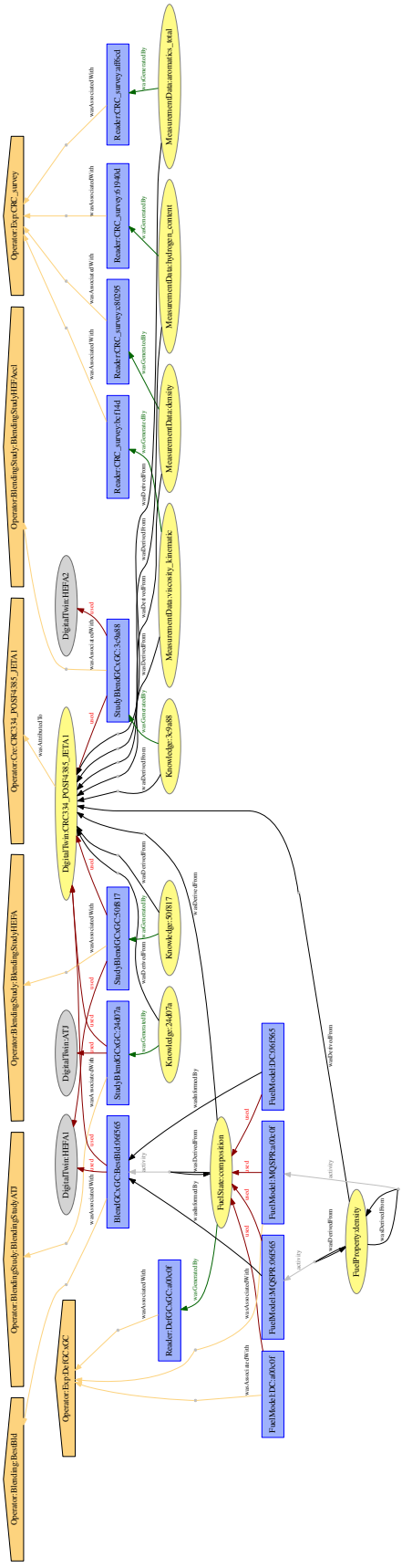


Fig. 13 Provenance graph of a conventional fuel DT after blending..

2. Overall evaluation of the CRC fuels

The decision discussed in the previous section is now repeated for all of the 28 conventional fuels from the CRC survey to identify the optimal SAF and blend ratio for each fuel. As the CRC fuels vary in their composition (see Fig. 8), optimal choices differ between the fuels. A summary of the 28 decisions is given in Fig. 14. Cases exist for which all three fuels can be utilized up to the maximum allowed blend ratio of $x_{bl} = 0.5$ (e.g. *CRC334 POSF4385*). In these cases, HEFA1 is selected as best choice, since it has the highest potential for soot reduction in the cost function.

As already identified in the previous section, there are conventional fuels for which the SAF with the highest potential blend ratio is not the optimal choice under the given constraints and cost function. This highlights the *smart* utilization of SAF as discussed in the introduction, meaning that the optimal utilization of a given SAF cannot only be determined by a single objective (e.g. high blend ratio) but should take multiple targets into consideration .

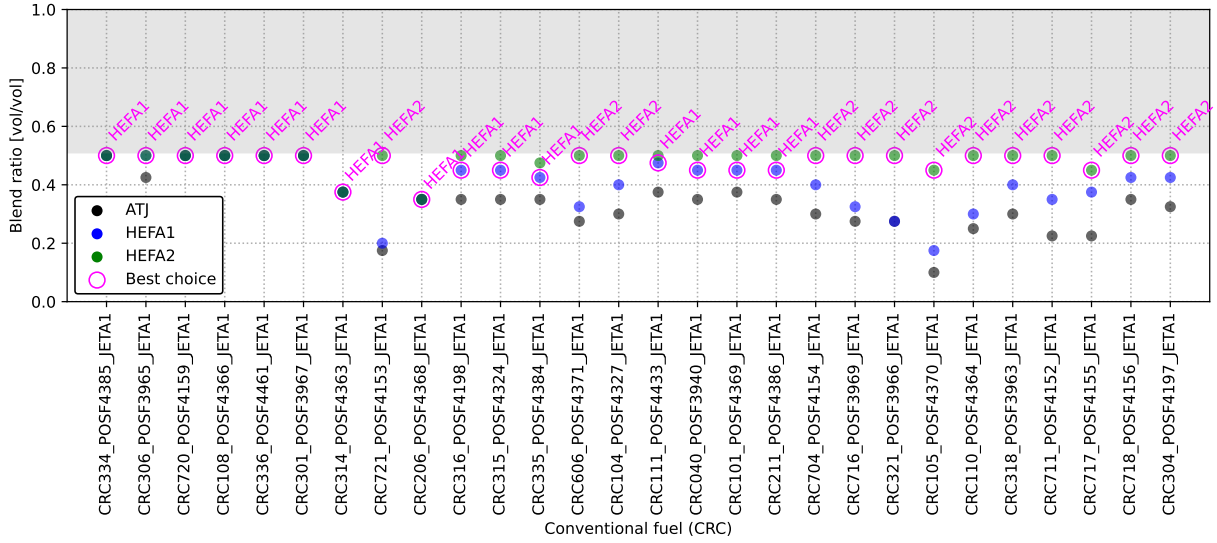


Fig. 14 Summary of options and decisions for all CRC fuels.

3. 50 – 100% SAF evaluation considering drop-in fuel property constraints

Both ATJ and HEFA SAF are currently limited to 50% volume blend ratio due the corresponding specifications. However, there are efforts to advance these limits towards higher blend ratio and ultimately 100% SAF utilization in order to maximize the CO_2 and non-CO_2 emission reduction potential. In order to shed light on the effects on fuel blending, the presented case study is repeated and adjusted towards possible 100% SAF utilization. For this purpose, the constraint on the blend ratio x_{bl} is altered to

$$x_{Bl} \in [0, 1.0] \tag{3}$$

while the constraints on density and aromatics content remain unchanged. The latter ensures that the drop-in characteristic of the resulting fuel blend is maintained.

Resulting blends and best choices are summarized in Fig. 15. Although the blending limit was increased to a $x_{bl,max} = 1.0$, all possible blends stay below $x_{bl,max} = 0.65$ due to the constraints on density and aromatics content. Obviously, the system is more sensitive to these constraints then to the maximum allowed blend ratio.

There are fuels for which the best choice of SAF changes when changing the constraint on maximum blend ratio (indicated by orange labels). For example, for fuel *CRC040 POSF3940* or *CRC101 POSF4369* the best choice is HEFA1 under the constraint of $x_{Bl} \in [0, 0.5]$ but this shifts to HEFA2 under the constraint of $x_{Bl} \in [0, 1.0]$. This is due to the fact that the blend ratio for HEFA2 is primarily constrained by the maximum allowed blend ratio and increases when changing this constraint. Hence, the associated cost function is able to drop below the HEFA1 cost function, which is primarily constrained by minimum density and aromatics content.

These examples again emphasize that each decision depends not only on the fuels under consideration, but also on the constraints placed on the system.

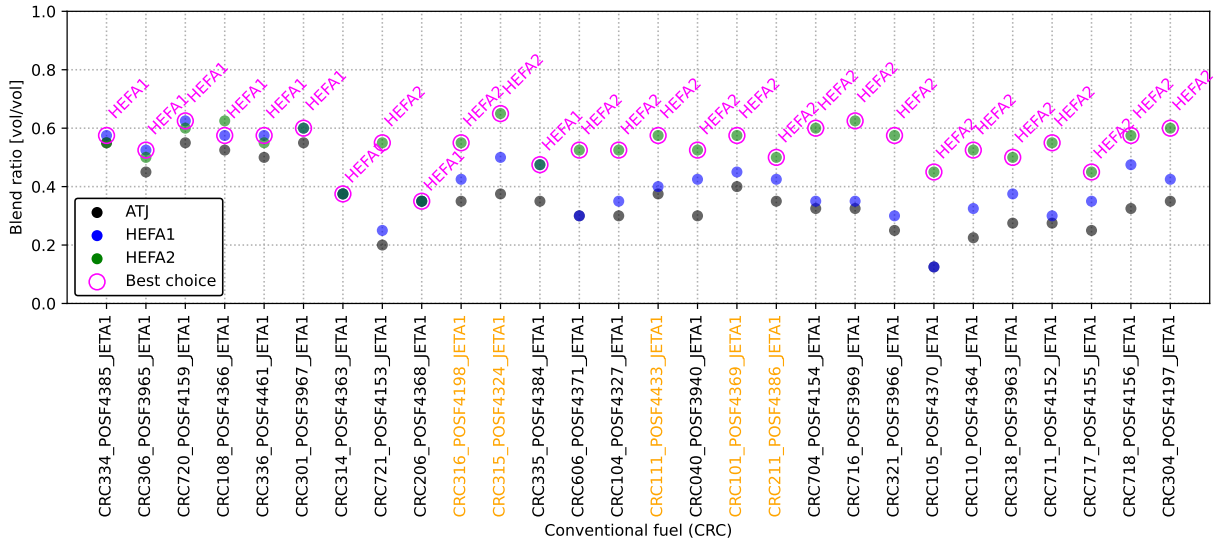


Fig. 15 Summary of options and decisions for all CRC fuels for $x_{bl,max} = 1.0$. Orange labels indicate fuels for which the best choice of SAF changes when changing the constraint on maximum blend ratio.

V. Summary and outlook

In order to track aviation fuels over their lifecycle and support decision making using the collected data and information, a Digital Twin concept based on a virtual representation of the fuel was presented. In this concept, the state-property relationship of fuels is represented through probabilistic machine learning models for each physical property under consideration. This enables fast modeling of the real world behavior of the entity over multiple operations within the lifecycle, e.g. production, blending or combustion of the fuel. Essential elements occurring in all of these operations - experiments, transformations, decisions - were identified and interpreted in the Digital Twin context. Furthermore, a strict tracking of the provenance of each piece of data and information within the Digital Twin through the W3C provenance model is adopted. The fact that probabilistic Machine Learning models are utilized for the prediction of fuel properties within the Digital Twin provides access to the uncertainties in the model predictions and builds the key element for uncertainty management in the proposed concept.

The methodology was successfully demonstrated for a decision making problem typical for the utilization of SAF under the current regulatory constraints, namely the blending of conventional fuel with SAF to a SAF blend. It involved three different SAFs that should be blended with a conventional jet fuel. The fact that the physical properties of the conventional jet fuel might vary within its specification required a careful selection of the blend ratio and the candidate SAF to achieve an optimal blend given associated cost functions. For this purpose, the expected soot emission of the blend was utilized as a cost function. It was shown that the inclusion of uncertainties narrowed the range of possible blend ratios and hence affected the decision process and the optimal decision. However, this approach reflects a risk-informed decision since uncertainties in the model prediction are considered in the decision process. It was identified that due to the defined constraints and the cost function, not always the SAF with the highest possible blend ratio was selected as the best choices. In a last step, the constraint on the maximum blend ratio was adjusted towards a possible 100% SAF utilization. Although the blending limit was increased to 1.0, all possible blends stayed below $x_{bl,max} = 0.65$ due to the constraints on density and aromatics content, which were kept constant in this study. Furthermore, conventional fuels were identified for which the best choice of SAF changes when changing the constraint on maximum blend ratio.

The methodology was successfully demonstrated, yet with a limited focus on fuel blending and a simplified cost function. In further studies, the following issues should be addressed:

- Additional quantities should be included in the cost function to fully reflect the complexity of the decision problem. This should include, inter alia, the economic cost and availability of the SAF, the CO_2 reduction potential and the lifecycle CO_2 emissions. Further critical properties for blending should be included in the constraint of the cost function to fully reflect the requirements for drop-in blends as summarized in ASTM7566 under Table 1 and 2.

For example, this should include the distillation curve as pointed out by Kaltschmitt [28]. As already identified, including the uncertainties in all of this data and predictions is a must for a robust decision. Furthermore, the scenario in the work at hand should be extended to cover further operations and decisions from the SAF lifecycle as summarized in Fig. 2. This could include models and data from the SAF feedstock, production and deployment and would aid in identifying dependencies between information and decisions over the lifecycle.

- For the presented case study, data from existing resources was provided to the Digital Twin. However, in order to mimic a real world/real time entity through the DT, interfaces capturing data from real-time resources, e.g. sensors, must be incorporated. For example, data from sensors measuring physical properties of the fuel before and after blending could be used to update the model predictions and provide a feedback loop. Furthermore, coupling sensors with the Digital Twin would enable *smart sensing*. This means that only a few, easily measurable quantities are measured by sensors while others are predicted using the Digital Twin.

- Although a single ML model was utilized for each physical property in the Digital Twin, the predictive capabilities of the models might differ depending on the application domain [9]. Thus, it would be advisable to dynamically select adequate models for each physical property. Methods for automatic model selection or model fusion should therefore be included into the DT concept.

- The cost function in the example at hand preferred fuels and blends with high reduction capabilities for soot emissions. As demonstrated in Sec. IV.C.1, this resulted in cases in which not the SAF with the highest possible blend ratio was selected, but the SAF with the highest soot reduction. However, this would mean that although the maximum soot reduction was achieved, there was a potential for a higher CO_2 emission reduction through a higher blend ratio. Clearly, there are cases with a trade-off between CO_2 and non-CO_2 reduction, depending on the available SAF. Since there are implications in recent research that the climate impacts of non-CO_2 effects are highly sensible to flight mission parameters such as aircraft trajectory and ambient atmospheric conditions [29], these information must be included in the decision. This would result in a truly *smart* utilization of SAF.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 957824 - *Copenhagen Airport: a Lighthouse for the introduction of sustainable aviation solutions for the future (ALIGHT)*.

References

[1] IATA, "Sustainable Aviation Fuels - Fact Sheet 5," Tech. rep., 2018.

[2] Association, I. A. T., et al., "IATA sustainable aviation fuel roadmap," , 2015.

[3] Colket, M., Heyne, J., Rumizen, M., Gupta, M., Edwards, T., Roquemore, W. M., Andac, G., Boehm, R., Lovett, J., Williams, R., et al., "Overview of the national jet fuels combustion program," *AiAA Journal*, Vol. 55, No. 4, 2017, pp. 1087–1104.

[4] EASA, "Sustainable Aviation Fuel Monitoring System," Tech. rep., 2015.

[5] Mawhood, R., Gazis, E., de Jong, S., Hoefnagels, R., and Slade, R., "Production pathways for renewable jet fuel: a review of commercialization status and future prospects," *Biofuels, Bioproducts and Biorefining*, Vol. 10, No. 4, 2016, pp. 462–484.

[6] ASTM, "D7566 - 20c Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons," Tech. rep., ????

[7] Maertens, S., Grimme, W., and Scheelhaase, J., "ICAO's new CORSIA scheme at a glance–a milestone towards greener aviation," *Aviation and Climate Change: Economic Perspectives on Greenhouse Gas Reduction Policies*, 2020.

[8] Ares, E., *EU ETS and aviation*, House of Commons Library London, 2012.

[9] Hall, C., Rauch, B., Bauder, U., Le Clercq, P., and Aigner, M., "Predictive Capability Assessment of Probabilistic Machine Learning Models for Density Prediction of Conventional and Synthetic Jet Fuels," *Energy & Fuels*, Vol. 35, No. 3, 2021, pp. 2520–2530.

[10] Ruoff, S. J., Rauch, B., Le Clercq, P., and Aigner, M., "Assessment of the comparability of droplet evaporation fuel sensitivities between a unit test case and an aviation gas turbine combustor," *AIAA Scitech 2019 Forum*, 2019, p. 0727.

[11] Kraft, E. M., "The air force digital thread/digital twin-life cycle integration and use of computational and experimental knowledge," *54th AIAA aerospace sciences meeting*, 2016, p. 0897.

[12] Qi, Q., Tao, F., Zuo, Y., and Zhao, D., "Digital twin service towards smart manufacturing," *Procedia Cirp*, Vol. 72, 2018, pp. 237–242.

[13] Aydemir, H., Zengin, U., and Durak, U., "The Digital Twin Paradigm for Aircraft Review and Outlook," *AIAA Scitech 2020 Forum*, 2020, p. 0553.

[14] Kapteyn, M. G., Pretorius, J. V., and Willcox, K. E., "A Probabilistic Graphical Model Foundation for Enabling Predictive Digital Twins at Scale," *arXiv preprint arXiv:2012.05841*, 2020.

[15] Reitenbach, S., Vieweg, M., Hollmann, C., and Becker, R.-G., "Usage of Data Provenance Models in Collaborative Multi-Disciplinary Aero-Engine Design," *Turbo Expo: Power for Land, Sea, and Air*, Vol. 84096, American Society of Mechanical Engineers, 2020, p. V02DT38A001.

[16] Barricelli, B. R., Casiraghi, E., and Fogli, D., "A survey on digital twin: Definitions, characteristics, applications, and design implications," *IEEE Access*, Vol. 7, 2019, pp. 167653–167671.

[17] Jones, D., Snider, C., Nassehi, A., Yon, J., and Hicks, B., "Characterising the Digital Twin: A systematic literature review," *CIRP Journal of Manufacturing Science and Technology*, Vol. 29, 2020, pp. 36–52.

[18] De Jong, S., Antonissen, K., Hoefnagels, R., Lonza, L., Wang, M., Faaij, A., and Junginger, M., "Life-cycle analysis of greenhouse gas emissions from renewable jet fuel production," *Biotechnology for biofuels*, Vol. 10, No. 1, 2017, pp. 1–18.

[19] Hall, C., Creton, B., Rauch, B., Bauder, U., and Aigner, M., "Probabilistic Mean Quantitative StructureProperty Relationship modelling of Jet Fuels Properties," *submitted to Energy & Fuels*, 2022.

[20] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J.-C., Ungerer, P., Rousseau, B., and Adamo, C., "A general guidebook for the theoretical prediction of physicochemical properties of chemicals for regulatory purposes," *Chemical reviews*, Vol. 115, No. 24, 2015, pp. 13093–13164.

[21] Gal, Y., and Ghahramani, Z., "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *international conference on machine learning*, PMLR, 2016, pp. 1050–1059.

[22] Simmhan, Y. L., Plale, B., and Gannon, D., "A survey of data provenance techniques," *Computer Science Department, Indiana University, Bloomington IN*, Vol. 47405, 2005, p. 69.

[23] Missier, P., Belhajjame, K., and Cheney, J., "The W3C PROV family of specifications for modelling provenance metadata," *Proceedings of the 16th International Conference on Extending Database Technology*, 2013, pp. 773–776.

[24] Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., et al., "Prov-DM: The prov data model," *W3C Recommendation*, 2013.

[25] Hadaller, O., and Johnson, J., "World fuel sampling program," *Alpharetta: Coordinating Research Council*, 2006.

[26] JETSCREEN, "D8.7 – Report on accuracy assessment of screening tool, Fuel compositions with added value," Tech. rep., 2021.

[27] JETSCREEN, "D7.3 - Report and analysis of APU emission measurements," Tech. rep., 2017.

[28] Kaltschmitt, M., and Neuling, U., *Biokerosene: Status and prospects*, Springer, 2017.

[29] Dahlmann, K., Grewe, V., Matthes, S., and Yamashita, H., "Climate assessment of single flights: Deduction of route specific equivalent CO2 emissions," *International Journal of Sustainable Transportation*, 2021, pp. 1–12.