aerospace

MDPI

Systematic Review

Aircraft Trajectory Prediction: A Systematic Review of System-Level Organization, Methodologies, and Performance Evaluation

Chengzhi Zhu [1,2], Benkui Zhang [1,2], Chenguang Yang [3], Huanyu Bian [1,2], Ying Chang [1,2], Lei Wang [1,2,4], Dongyu Li [5] and Ning Wang [6,*]

1 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; wanglei002931@aircas.ac.cn (L.W.)
2 Key Laboratory of Target Cognition and Application Technology, Chinese Academy of Sciences, Beijing 100190, China
3 Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, China; cyang@ieee.org
4 School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China
5 School of Cyber Science and Technology, Beihang University, Beijing 100191, China
6 School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield S1 2NU, UK
* Correspondence: wang.ning@ieee.org

Abstract

Aircraft trajectory prediction supports trajectory-based operations and air traffic management by estimating future aircraft motion. This systematic review synthesizes studies on trajectory, position, and flight state prediction for civil, manned, fixed-wing aircraft during airborne operation from takeoff to touchdown. Following evidence assessment, 198 studies were retained as primary evidence, of which 55 met the predefined criteria for formal performance synthesis. The review uses a Preparation, Prediction, Update, and Output (PPUO) framework and organizes the evidence around modeling methodology, task formulation, and performance evaluation. The findings show that physics-based, state estimation, data-driven, and hybrid methods have different modeling assumptions and operating conditions, and the available evidence does not support any one paradigm as universally superior across tasks and operating scenarios. Reported performance is jointly influenced by prediction horizon, target cardinality, interaction context, and evaluation conditions. The evidence is concentrated on single-target prediction; quantitative evidence for joint multi-aircraft prediction and probabilistic uncertainty evaluation remains limited, while heterogeneity in experimental conditions restricts direct cross-study comparison. Future research should establish reproducible benchmarks with complete reporting of experimental conditions and conduct cross-scenario validation while evaluating trajectory accuracy, uncertainty quality, computational efficiency, and operational assurance together.

Keywords: aircraft trajectory prediction; systematic review; civil aviation; trajectory-based operations; prediction horizon; multi-aircraft prediction

Check for updates

Academic Editors: Xavier Olive and Michael Schultz

Received: 26 September 2026
Revised: 25 October 2026
Accepted: 27 October 2026
Published: 3 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Growing traffic demand and operational complexity continue to challenge airspace safety, efficiency, and capacity [1]. Modern air traffic management (ATM) programs, including the Single European Sky ATM Research (SESAR) initiative and the U.S. Next Generation Air Transportation System (NextGen), increasingly organize planning and decision support

around trajectory-based operations (TBO) [2–4]. In TBO-oriented environments, predicted aircraft trajectories provide a common information basis for conflict monitoring, sequencing and timing, demand–capacity assessment, and related decision-support functions [5–7]. The accuracy and timeliness of trajectory prediction therefore affect how future aircraft motion can be incorporated into ATM planning and control.

Aircraft trajectory prediction uses information available at prediction time to estimate future aircraft motion. Depending on the task and output formulation, the prediction may comprise a sequence of future positions, a four-dimensional (4D) trajectory, future flight states, waypoint or arrival time information, or a probabilistic representation of plausible future motion. Predictor inputs may include recent surveillance states, historical trajectories, flight plans and intent, aircraft performance information, meteorological conditions, and surrounding traffic or airspace context [7,8]. This review focuses on future trajectory, position, or flight state prediction for civil, manned, fixed-wing aircraft during airborne operation from takeoff to touchdown, including commercial transport, business aviation, and manned general aviation. The scope covers departure, climb, cruise, descent, and airborne terminal operations.

Trajectory prediction is challenging because future aircraft motion depends on interacting sources of information and uncertainty. Recent position, velocity, heading, and vertical motion constrain near-term evolution, whereas flight intent, procedures, weather, aircraft performance characteristics, and traffic context become more influential as the lookahead extends across future maneuvers or operational changes. Surveillance data may also contain measurement noise, irregular sampling, or missing observations. The same observed history may therefore support several plausible future trajectories when intent or interaction outcomes remain uncertain. The requirements of trajectory prediction vary with prediction horizon, information availability, and the unit of prediction [9–11].

These requirements have produced a diverse methodological landscape. Physics-based predictors propagate aircraft motion using explicit dynamical, performance, environmental, and intent assumptions. State estimation methods infer continuous aircraft states, discrete motion regimes, or combinations of the two from available observations and propagate those estimates into the future. Data-driven approaches encompass classical statistical and machine learning methods that rely on historical trajectories and engineered contextual variables, as well as deep-learning architectures that learn representations of temporal, spatial, relational, and multimodal information. Hybrid formulations combine mechanisms from multiple modeling families, including learned or data-derived components, physical models, recursive estimators, and explicit prior or constraint mechanisms. The suitability of these approaches depends on the prediction task, available information, prediction horizon, and evaluation setting.

Previous reviews provide complementary perspectives on aircraft trajectory prediction. Mondoloni and Rozen examine the trajectory prediction process and synchronization requirements for ATM and TBO applications [7]. Zeng et al. review civil aviation 4D trajectory prediction through a process description, three broad method categories, evaluation indices, and data sources [8]. Xie et al. compare physical, classical machine learning, and deep-learning approaches with emphasis on computational efficiency and application conditions [12]. Kwak and Lee conduct a structured review focused on deep-learning models, including model taxonomies, datasets, evaluation metrics, and practical deployment considerations [13]. These reviews differ in scope, review method, process organization, and performance evidence, as summarized in Table 1. Taken together, these reviews do not explicitly integrate system-level function, modeling paradigm, task formulation, and performance evaluation within a common evidence framework for civil aircraft trajectory prediction.

Table 1. Comparison of prior reviews and the present systematic review.

Review	Study Subject	Included Studies	Method Classification	Task Dimensions	Performance Evaluation
Mondoloni and Rozen [7]	Commercial aircraft trajectory prediction and synchronization	Not reported	Process framework	Strategic/tactical prediction; synchronization	Spatial, temporal, stability, and computational measures
Zeng et al. [8]	Civil aviation 4D trajectory prediction	Not reported	State estimation, kinetic models, and machine learning	10 min boundary; deterministic or probabilistic output	EE, ATE, CTE, and AE
Xie et al. [12]	Aircraft trajectory prediction, including autonomous-flight settings	Not reported	Physical, classical machine learning, deep learning, and hybrid methods	Strategic, pre-tactical, and tactical stages	Error, similarity, complexity, and uncertainty
Kwak and Lee [13]	Aircraft trajectory prediction from 2020 to June 2025	46	Deep learning	Long-horizon prediction; multi-aircraft interaction	MAE, RMSE, ADE, FDE, NLL, and CRPS
This review	**Civil, manned, fixed-wing aircraft during airborne operation from takeoff to touchdown**	**198 (55 for performance synthesis)**	**Physics-based, state estimation, data-driven, and hybrid methods**	**Prediction horizon; target cardinality; interaction context**	**Accuracy, uncertainty, and engineering requirements**

Accordingly, this review provides a systematic and multi-level synthesis of aircraft trajectory prediction for civil, manned, fixed-wing airborne operations. It connects evidence identification and screening with a system-level functional organization, a taxonomy of modeling paradigms, two cross-cutting task dimensions, and a structured framework for performance evaluation. The review is guided by four research questions:

1. **RQ1:** What functions, information flows, and outputs recur in aircraft trajectory prediction, and how can they be organized at the system level?
2. **RQ2:** How do physics-based, state estimation, data-driven, and hybrid approaches differ in their assumptions, applications, advantages, and limitations?
3. **RQ3:** How do prediction horizon, target cardinality, and interaction context influence task formulation, method selection, and output representation?
4. **RQ4:** Which accuracy, uncertainty, engineering, and operational measures are used to evaluate aircraft trajectory prediction, and what performance patterns and evidence gaps can be identified across the reported studies?

The answers to these questions are developed in Sections 3–8 and consolidated in Section 9. The main contributions are summarized as follows.

- **Systematic and traceable evidence foundation.** The review applies a structured workflow for literature identification, deduplication, screening, full-text assessment, evidence-role classification, and assessment of eligibility for performance synthesis. Restricting the core scope to civil, manned, fixed-wing aircraft during airborne operation provides a consistent basis for the methodological and quantitative synthesis.
- **System-level organization of the trajectory prediction process.** The Preparation, Prediction, Update, and Output (PPUO) view organizes how heterogeneous observations and contextual information are converted into predictor-ready inputs, future-motion

estimates, revised predictions, and outputs for downstream use. It provides a system-level analytical framework applicable across different predictor architectures.

- **Unified modeling and task taxonomy.** The methodological synthesis covers physics-based models, state estimation models, data-driven approaches, and hybrid formulations. Prediction horizon and target cardinality are treated as cross-cutting task dimensions, while the use of inter-aircraft interaction information is distinguished from joint multi-aircraft prediction.
- **Structured performance evaluation synthesis.** Performance evidence is stratified by dataset or scenario, prediction horizon, sampling interval, metric family, output formulation, and target cardinality. Accuracy, uncertainty, and engineering or operational considerations are treated as complementary evaluation dimensions.
- **Integrated synthesis of challenges and future directions.** The review identifies recurring challenges in benchmark reproducibility, cross-scenario generalization, uncertainty and dependency representation, interaction modeling, computational efficiency, and operational assurance, and uses these evidence gaps to organize future research priorities.

Section 2 describes the systematic review methodology, and Section 3 presents the PPUO system-level organization of trajectory prediction. Sections 4 and 5 review traditional, data-driven, and hybrid prediction methods, while Sections 6 and 7 examine prediction horizon and target cardinality as cross-cutting task dimensions. Section 8 presents the evaluation framework and experimental context, compares quantitative performance evidence across modeling categories, and synthesizes cross-cutting performance findings. Finally, Section 9 integrates the main evidence patterns and discusses remaining challenges and future research directions.

2. Systematic Review Methodology

The reporting of this systematic review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [14]. This section defines the review scope and describes the procedures used to identify, screen, classify, and synthesize the literature on aircraft trajectory prediction. The resulting evidence base supports the system-level, methodological, task-oriented, and performance analyses developed in the subsequent sections.

2.1. Review Scope and Eligibility Criteria

The review focuses on the prediction of future trajectories, positions, or flight states of civil, manned, fixed-wing aircraft during airborne operation from takeoff to touchdown. Both single-target and joint multi-aircraft prediction studies are considered across the modeling paradigms covered in this review, while interaction-aware prediction is treated as a complementary task dimension. The prespecified eligibility criteria are summarized in Table 2. These criteria define eligibility for the systematic-review evidence base; the additional criteria used for formal performance synthesis are specified separately in Section 2.6.

2.2. Information Sources and Search Strategy

The systematic literature search was conducted on 29–30 July 2026 in Scopus, Web of Science Core Collection, Engineering Village–Compendex, and IEEE Xplore through institutional subscriptions. No publication-year or language limits were applied during database searching. Two complementary query groups were used. Q1 targeted terminology explicitly associated with aircraft trajectory prediction and forecasting, whereas Q2 broadened the search to related trajectory estimation, generation, computation, propagation, reconstruction, and flight intent terminology.

Table 2. Prespecified eligibility criteria of the systematic review.

Dimension	Inclusion Scope
Aircraft scope	Civil, manned, fixed-wing aircraft, including commercial transport, business aviation, and manned general aviation
Operational phase	Airborne operation from takeoff to touchdown
Prediction task	Future trajectory, position, flight state, flight path, 4D trajectory, or probabilistic future-motion prediction
Modeling paradigm	Physics-based, state estimation, classical statistical and machine learning, deep learning, and hybrid approaches
Target setting	Single-target and joint multi-aircraft trajectory prediction
Report type	Journal articles, conference papers/articles, Article in Press, and Early Access research reports

The common Q1 concept block was:

("trajectory prediction" OR "trajectory predictor" OR "trajectory forecasting" OR "trajectory forecast" OR "flight path prediction" OR "4D trajectory prediction" OR "four-dimensional trajectory prediction" OR "four dimensional trajectory prediction" OR "4D track prediction") AND (aircraft OR airplane OR aeroplane OR airliner OR aviation OR "air traffic" OR "civil aviation" OR "commercial aviation").

The common Q2 concept block was:

("aircraft trajectory estimation" OR "flight trajectory estimation" OR "aircraft trajectory generation" OR "flight trajectory generation" OR "aircraft trajectory computation" OR "flight trajectory computation" OR "aircraft trajectory propagation" OR "flight trajectory propagation" OR "aircraft trajectory reconstruction" OR "flight trajectory reconstruction") OR (("flight intent prediction" OR "aircraft intent prediction" OR "flight intent modeling" OR "flight intent modelling") AND (aircraft OR airliner OR "air traffic" OR "civil aviation" OR "commercial aviation")).

The database-specific search strings and associated search details are provided in Supplementary File S1. The eight database-query combinations yielded 3771 exported records, which were subsequently merged and deduplicated before screening. Table 3 summarizes the database-level export counts.

Table 3. Database search summary.

Database	Q1	Q2	Total
Scopus	1279	115	1394
Web of Science Core Collection	552	59	611
Engineering Village–Compendex	1085	98	1183
IEEE Xplore	535	48	583
Total	**3451**	**320**	**3771**

2.3. Record Management, Study Selection, and Full-Text Assessment

The 3771 exported records were merged into a unified bibliographic dataset and deduplicated using standardized DOI, title, author, year, and other record-level information. Records with identical DOIs were treated as the same study. For records without a DOI, matching was performed using the normalized title, authors, and publication year; records were merged only when the title similarity was high and the author and year information

was compatible. Records with conflicting non-empty DOIs were not merged directly. Database source, search group, source file, source row, and database accession number were retained for traceability. Citation counts were retained as descriptive metadata and were not used for inclusion, exclusion, ranking, or weighting.

After 2072 duplicate or related-version records were removed or merged, 1699 unique study-level records remained for title and abstract screening. Titles and abstracts were screened by one reviewer according to the predefined eligibility criteria specified in Section 2.1. Records that could not be classified reliably from the title and abstract were retained for full-text assessment. A total of 1168 records were excluded during title and abstract screening, and seven additional records were separated as background evidence, leaving 524 studies for full-text assessment.

All 524 candidate reports underwent full-text assessment by the same reviewer according to the predefined eligibility criteria. Of these, 226 were excluded for failing to meet one or more prespecified eligibility criteria, leaving 298 studies for detailed evidence assessment. Of these 298 studies, 198 were retained as primary evidence, 81 as secondary/supporting evidence, and 19 were excluded after the final evidence assessment. Figure 1 summarizes the overall process of record identification, merging and deduplication, screening, full-text assessment, and evidence classification. The 198 primary studies formed the core evidence base for the subsequent methodological and task-oriented synthesis; studies meeting the additional criteria for formal performance synthesis were analyzed as a subset of this group.

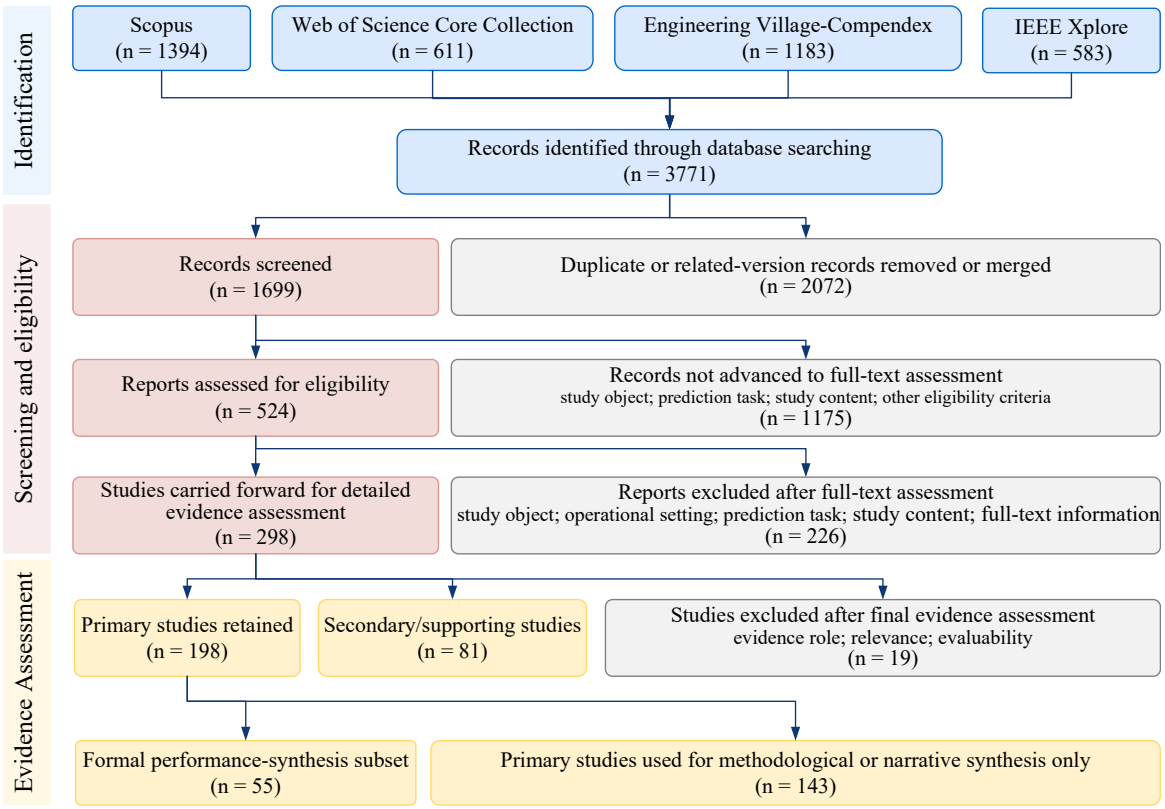


Figure 1. Literature identification, screening, evidence classification, and performance synthesis workflow.

During manuscript revision, a second author reviewed all 55 studies included in the formal performance synthesis, focusing on study eligibility, methodological quality, and evidence confidence. Separately, a third author rechecked the exclusion decisions for all 226 studies excluded during full-text assessment and all 19 studies excluded during the final evidence assessment against the predefined eligibility and evidence classification criteria.

2.4. Evidence Classification and Data Extraction

The retained evidence was assigned one of two roles. Primary evidence comprised original studies that proposed, implemented, or empirically evaluated qualifying future-trajectory prediction methods. Secondary/supporting evidence comprised studies that directly informed system-level, methodological, or contextual aspects of trajectory prediction but were not treated as core prediction-method studies.

Other contextual references, including standards, prior reviews, datasets, simulation platforms, and foundational methods, were cited where necessary to support definitions, background, or methodological context but were not included in the systematic-review study counts.

Data were extracted by one reviewer using a standardized extraction framework. For the retained studies, standardized information on modeling approach, prediction setting, data source, and evaluation outcomes was extracted. Study investigators were not contacted for additional or missing information. The extracted data supported the subsequent methodological and task-oriented synthesis, while quantitative performance information from eligible primary studies formed the basis of the formal performance analysis described below.

For the mutually exclusive modeling categories used in the formal performance synthesis, each study was assigned to one primary category according to its dominant prediction mechanism. Auxiliary contextual inputs, transition probabilities, or uncertainty representations did not by themselves trigger hybrid coding; hybrid coding was used when distinct mechanisms from more than one principal modeling family materially participated in trajectory generation or uncertainty propagation.

2.5. Risk of Bias and Certainty of Evidence

To address potential bias associated with the original study selection and data extraction process, the second author conducted a structured qualitative assessment of the 55 studies included in the formal performance synthesis. The assessment considered five aspects: (1) study scope and eligibility, (2) data selection and experimental design, (3) comparison and evaluation methods, (4) completeness and transparency of result reporting, and (5) confidence in the supporting evidence.

Particular attention was paid to whether the study population and prediction tasks satisfied the predefined eligibility criteria, whether experimental designs introduced potential selection or information-leakage bias, whether baseline comparisons and evaluation metrics were appropriate, and whether the reported results adequately supported the stated findings. Any differences in judgment were discussed between the two reviewers until consensus was reached.

Given the substantial heterogeneity in modeling paradigms, datasets, prediction horizons, and evaluation protocols, the assessment was conducted using a structured qualitative approach. Confidence in the main findings of the review was considered in relation to the methodological limitations, consistency, directness, and reporting completeness of the supporting studies. The assessment was intended to inform the interpretation of the synthesized evidence.

2.6. Performance Synthesis Criteria

Formal performance synthesis was restricted to a subset of the 198 primary studies. Eligibility required (1) an identifiable civil aircraft dataset or experimental scenario, (2) an explicit prediction horizon, (3) a defined evaluation metric, and (4) at least one quantitative result that either directly evaluates the eligible trajectory prediction or evaluates a task-specific outcome explicitly computed from that prediction. These criteria were applied to

define a quantitatively interpretable subset of the primary evidence; they did not change the eligibility of studies for the systematic-review evidence base. Application of the criteria yielded 55 studies for formal performance synthesis. The remaining 143 primary studies were retained for methodological or narrative synthesis.

The criteria favor studies that report sufficiently complete evaluation conditions, including a defined horizon, metric, and quantitative result. Consequently, studies from some modeling paradigms or experimental settings may be less likely to enter the formal performance subset when their performance information is reported incompletely, even when they remain eligible primary evidence. The 55 studies should therefore be interpreted as a structured quantitative subset of the 198 primary studies rather than as a complete representation of their distribution. The 143 studies not included in the formal performance synthesis continued to inform the methodological or narrative synthesis, and their absence from the quantitative subset does not diminish their methodological relevance.

Because the reviewed studies differ substantially in experimental conditions and evaluation settings, the performance evidence was synthesized in a stratified manner according to modeling paradigm, dataset or scenario, prediction horizon, sampling interval, metric family, target cardinality, and output formulation. This organization supports condition-specific interpretation of reported results. Performance values were interpreted in relation to the datasets, prediction settings, and metric definitions under which they were obtained.

The review was not registered, and no formal review protocol was prepared.

Figure 1 summarizes the overall literature identification, screening, evidence classification, and performance synthesis process.

This section defined the review scope, eligibility criteria, and performance synthesis criteria and described the procedures used to search, merge, deduplicate, screen, assess, and classify the literature across four databases. The resulting evidence base comprised 198 primary studies and 81 secondary or supporting studies. Of the primary studies, 55 met the criteria for formal performance synthesis, while the remaining 143 informed the methodological or narrative synthesis.

3. General Process of Aircraft Trajectory Prediction

Aircraft trajectory prediction forms part of an information processing chain that prepares observations and operational context, infers future aircraft motion, revises predictions as new information becomes available, and delivers prediction results for operational use. For system-level analysis, this review organizes the chain into four connected functional stages: preparation, prediction, update, and output (PPUO). The framework describes functional relationships, while individual predictors may combine several stages or implement some of them implicitly. In trajectory-based operations, predicted trajectory information may also be exchanged or synchronized across operational and decision-support systems [5–7]. Figure 2 summarizes this organization.

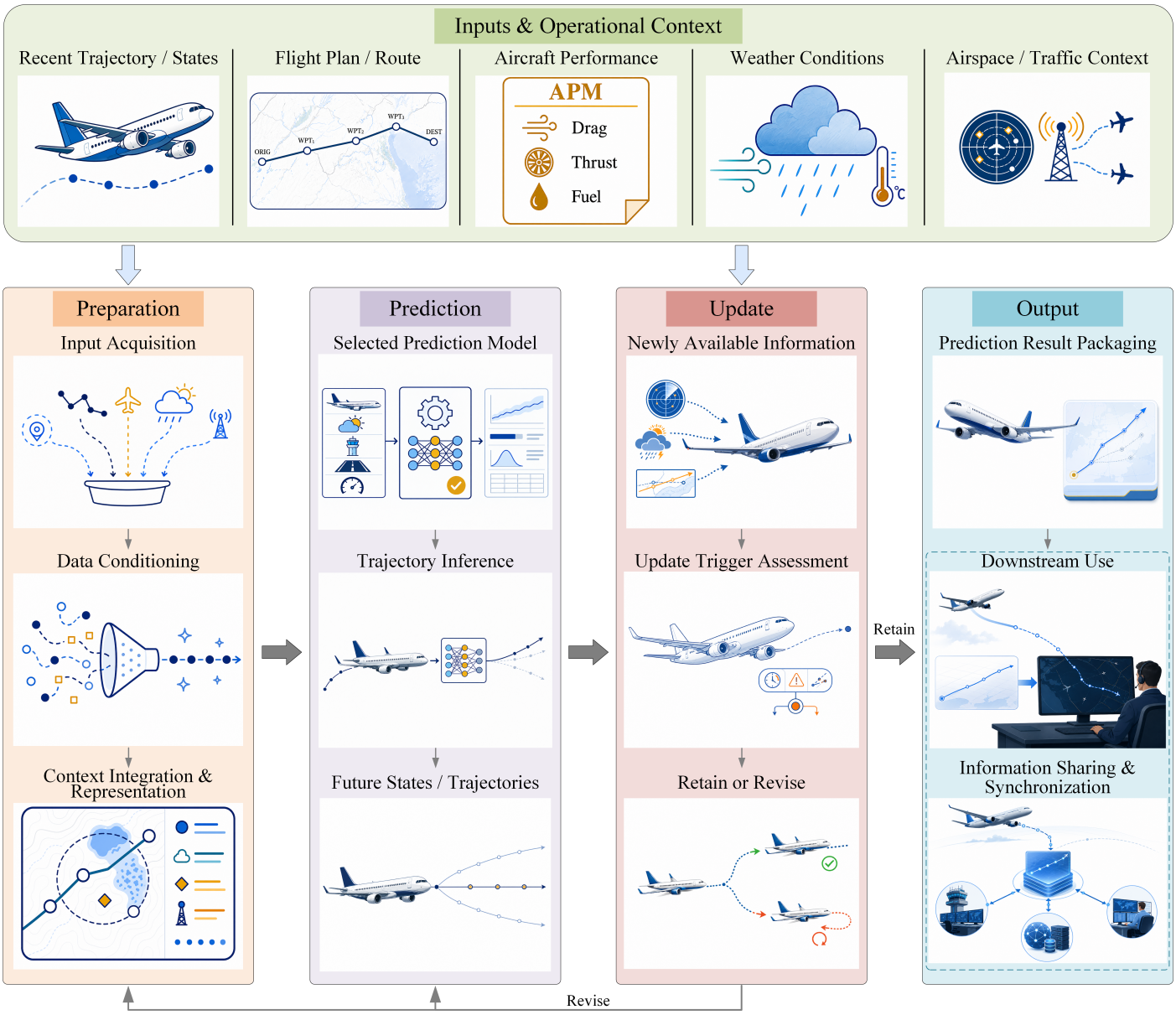


Figure 2. System-level PPUO organization of aircraft trajectory prediction.

3.1. Preparation: Input and Context Processing

The preparation stage transforms heterogeneous observations and operational context into a representation suitable for the selected trajectory predictor. Common inputs include recent aircraft states or trajectory history, flight plans and operational intent, aircraft performance parameters, meteorological conditions, and airspace or traffic context. The relative importance of these information sources varies across prediction methods. Physics-based and intent-informed approaches commonly rely on aircraft performance and operational information, whereas data-driven approaches primarily use historical state sequences and available contextual features [7,15].

3.1.1. Input Acquisition and Data Conditioning

Trajectory prediction inputs originate from heterogeneous sources and may differ in sampling rate, coordinate representation, completeness, and measurement reliability. Preparation therefore includes acquiring and temporally aligning observations, together with data cleaning, normalization, filtering, and other transformations required by the selected predictor. When observations have different or irregular sampling intervals,

resampling or interpolation may be used to place them on a common temporal grid. The selected procedure should be consistent with the temporal resolution and information available at the prediction time.

Input quantities that are unavailable or uncertain may also be estimated or calibrated using historical or in-flight observations. Aircraft performance parameters, for example, can be inferred from observed trajectories and supplied to a trajectory predictor [16,17]. Such parameter estimation procedures support trajectory prediction and form part of the prediction chain when the estimated quantities are subsequently used to generate future aircraft states or trajectories.

3.1.2. Intent and Context Representation

Operational intent describes how an aircraft is expected to evolve under planned or commanded conditions. Depending on data availability, intent-related information may be represented explicitly through flight plans, routes, waypoint sequences, speed or altitude constraints, clearances, and procedure information, or inferred from observed aircraft behavior and surrounding context [18–20]. Traditional intent-based predictors commonly use structured representations, whereas learning-based methods may incorporate intent-related variables as contextual features or infer behaviorally relevant representations from trajectory and operational data [10,21].

More generally, the output of the preparation stage is a predictor-ready representation containing the observations, constraints, and available or inferred context required by the selected prediction mechanism.

3.2. Prediction: Future Trajectory Inference

The prediction stage converts the predictor-ready information produced during Preparation into future aircraft states or trajectories over a specified prediction horizon. According to the mechanism used to generate future states, prediction methods can be grouped into four families: physics-based methods propagate aircraft motion using dynamical, performance, environmental, and operational assumptions; state estimation methods infer and propagate continuous states or discrete motion regimes from observations; data-driven methods learn predictive relationships from historical trajectories and contextual information; and hybrid methods combine prediction mechanisms from multiple modeling families. Detailed formulations are reviewed in Sections 4 and 5.

Prediction outputs may be deterministic, consisting of a single future state sequence or trajectory, or probabilistic, representing uncertainty through predictive distributions, intervals, ensembles, or multiple plausible trajectories. The prediction depends on the information available at prediction time and on the assumptions embedded in the selected model. As new aircraft state, operational intent, meteorological, or traffic context information becomes available, an earlier prediction may require revision through the update stage described next [22,23].

3.3. Update: Online Prediction Revision

As new surveillance observations, operational information, or environmental conditions become available, an existing prediction may diverge from the current aircraft state or operational context. The update stage incorporates these new inputs and supports either retaining the current prediction or revising it through the preparation and prediction stages using updated information. This function may be implemented through explicit recalculation or rolling prediction as new observations arrive [24,25].

According to the triggering mechanism, updates can be classified as time-based, information- or event-based, and conformance-based. Time-based updating revises predictions at fixed intervals. Information- or event-based updating responds to changes in

clearances, operational intent, weather, or state estimates. Conformance-based updating is triggered when observed or estimated aircraft behavior deviates substantially from the previously predicted trajectory or falls outside an associated tolerance region [25–27]. Once an update is triggered, newly available observations or contextual information are fed back to the preparation and prediction stages to support continuous prediction revision.

3.4. Output: Prediction Results and Downstream Use

The output stage provides the current prediction in a form suitable for interpretation and operational use. Deterministic predictions may provide future aircraft state sequences, trajectory points, waypoints, or four-dimensional trajectories, while probabilistic predictions may additionally provide predictive distributions, uncertainty intervals, trajectory ensembles, or multiple plausible future trajectories. Predictive uncertainty representations should be distinguished from retrospective evaluation quantities obtained by comparing predictions with observations, such as realized prediction errors. These evaluation measures are discussed in Section 8.

Predicted trajectories can support conflict detection and monitoring, sequencing and timing, demand–capacity assessment, and other decision-support tasks. These functions use trajectory prediction results and belong to downstream operational applications. In trajectory-based operational architectures, trajectory information may also be exchanged or synchronized across participating systems through shared information infrastructure [5–7]. Distinguishing prediction results from their downstream use helps define the scope of the trajectory prediction problem.

Overall, the PPUO framework characterizes aircraft trajectory prediction as a system-level process comprising the preparation of predictor-ready information, inference of future aircraft motion, online prediction revision, and delivery of results for operational use. The framework distinguishes prediction mechanisms from input preparation, update functions, and downstream applications and provides the system-level basis for the methodological, task-oriented, and performance analyses developed in the following sections.

4. Traditional Trajectory Prediction Methods

Traditional trajectory prediction methods are approaches that generate future aircraft states primarily using explicit motion and aircraft performance models or state transition and state estimation mechanisms. They are organized into two principal methodological families: physics-based models, which propagate aircraft motion using explicit dynamical and performance relations together with operational and environmental information; and state estimation models, which infer continuous aircraft states, discrete motion regimes, or combinations of the two from observations and propagate these estimates into the future. The main distinction between the two families is the mechanism used to generate future states, while state estimation approaches also differ in how aircraft state and motion uncertainty are represented and propagated. Table 4 summarizes the methods and principal subfamilies considered in this section.

4.1. Physics-Based Models

Physics-based trajectory predictors generate future trajectories by propagating a known or estimated initial aircraft state using explicit motion and performance relations. Civil aircraft applications commonly combine simplified point-mass dynamics with aircraft performance information, operational intent, and meteorological conditions. Model fidelity depends on the prediction task, prediction horizon, and availability of aircraft-specific and environmental information [28,29]. Figure 3 illustrates the principal force and angular variables in a point-mass aircraft model.

Table 4. Taxonomy and characteristics of traditional aircraft trajectory prediction methods and subfamilies.

Method Family	Subfamily	Core Prediction Mechanism	Representative Formulations	Output/Uncertainty Representation	Key Characteristics/Limitations
Physics-based	–	Numerical propagation of aircraft motion using explicit dynamics, performance relations, operational intent, and environmental conditions	3-DOF point-mass models; total-energy formulations; APM/BADA-supported trajectory integration; intent-informed propagation	Often deterministic; input and model uncertainty may be propagated through parameter distributions, Monte Carlo simulation, or trajectory ensembles	Physically interpretable and able to incorporate performance and operational constraints; performance depends on the fidelity of aircraft performance, intent, and weather information, and model mismatch may accumulate with prediction horizon
State estimation	Discrete-state and Markov	Probabilistic propagation of discrete locations, latent states, maneuver regimes, or mode transitions to infer future motion	Hidden Markov models; Markov chain prediction; HMMs combined with trajectory similarity or mode transition models	Future-state probabilities and probabilistic trajectory hypotheses derived from transition and observation models	Represents discrete-mode uncertainty explicitly and supports probabilistic multi-step prediction; performance depends on state-space design and transition model coverage, while coarse discretization or sparse transitions may limit extrapolation
State estimation	Recursive Bayesian continuous-state	Recursive estimation of continuous aircraft states followed by propagation beyond the latest observation	KF, EKF, and UKF; particle filters; sequential Monte Carlo methods	Gaussian state distributions for Kalman-family methods; particle-based posterior approximations for nonlinear or non-Gaussian formulations	Provides principled treatment of measurement and process uncertainty with recursive updating; performance depends on transition and measurement models, Gaussian approximations may be restrictive, and particle methods can be computationally demanding
State estimation	Multiple-model	Parallel propagation of candidate motion models with probabilistic mode updating and fusion of mode-conditioned state estimates	Interacting multiple model (IMM); multiple-model Kalman filtering; state-dependent mode transition variants	Mode probabilities combined with mode-conditioned state distributions or trajectory hypotheses	Adapts to changes among predefined motion regimes and retains uncertainty over competing hypotheses; performance depends on model set coverage and transition probability specification, while complexity increases with the number and sophistication of candidate models

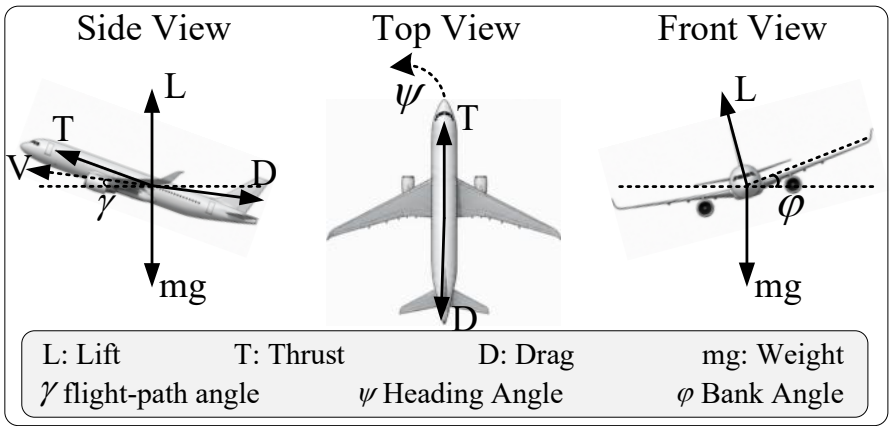


Figure 3. Three-view illustration of forces and angular variables in a point-mass aircraft formulation.

In operational prediction systems, motion propagation is combined with aircraft performance models, operational intent, and environmental information. Aircraft performance models provide thrust, drag, and related parameters as functions of aircraft type, configuration, and flight condition. Operational intent describes planned or commanded conditions, while the initial aircraft state and meteorological information provide the state and environmental conditions required for propagation. BADA is a representative aircraft performance model used in trajectory computation infrastructures [30,31]. Physics-based studies therefore commonly address model input determination, performance parameter calibration, and online trajectory updating.

One group of retained studies estimates aircraft mass, thrust, or related performance parameters from observed motion and supplies the estimates to point-mass or performance-based predictors [16,32]. Other studies infer or update aircraft intent from flight plans, observed behavior, or new surveillance information and recompute future trajectories for online four-dimensional prediction [20,24]. Related work examines the effects of wind and other meteorological information on propagated trajectories [33,34].

Overall, physics-based studies support future-trajectory prediction by refining motion propagation, calibrating aircraft performance parameters, incorporating operational intent, and updating environmental information. The reviewed applications indicate that the treatment of aircraft performance information, operational intent, and meteorological information affects reported prediction results, with effects that vary across prediction horizons and operational settings. These structured prediction mechanisms can also inform the state estimation, data-driven, and hybrid formulations discussed in the following sections.

4.2. State Estimation Models

State estimation models infer aircraft states or motion regimes from available observations and propagate the resulting estimates into the future. The estimated state may be continuous, discrete, or conditioned on a motion mode, while uncertainty may be represented by state distributions, particle samples, discrete-state probabilities, or probabilities assigned to multiple candidate motion models. For trajectory prediction, state estimation must extend beyond the latest observation time to produce future state sequences, trajectory hypotheses, or predictive trajectory distributions. Figure 4 summarizes the relationship between observations, state updates, and future-state propagation.

4.2.1. Discrete-State and Markov Models

Discrete-state models describe aircraft motion through a finite set of locations, motion regimes, flight phases, or other latent states. Hidden Markov models use observations to infer the current latent state and then apply state transition probabilities to predict

future states, which can subsequently be mapped to future positions, waypoint transitions, or trajectory segments.

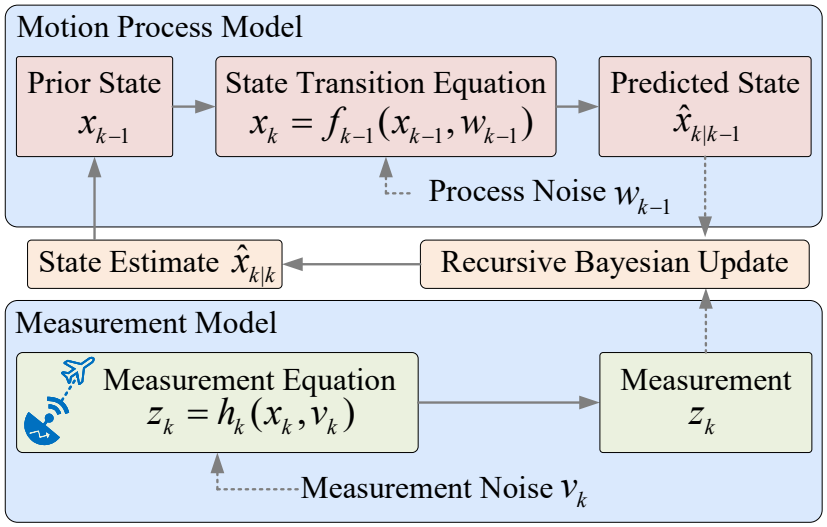


Figure 4. Recursive Bayesian state estimation and future-state prediction workflow.

In the retained evidence, Pan et al. model aircraft locations as hidden states and use weather-related observations to generate minute-by-minute future positions, which are then concatenated to form the remaining mid-flight trajectory [35]. Lin et al. combine hidden Markov state inference with trajectory similarity to predict future waypoint-related information and update the prediction as flight information changes [36]. These approaches represent uncertainty about discrete future states, but their performance depends on the construction and coverage of the state space and transition structure. Coarse state definitions, sparse transitions, or operating conditions that are insufficiently represented can limit prediction beyond previously observed patterns.

4.2.2. Recursive Bayesian State Prediction

For continuous aircraft states, Kalman-filter variants recursively update state estimates under different assumptions about the transition and measurement models. The Kalman filter is formulated for linear-Gaussian systems, the extended Kalman filter applies local linearization to nonlinear models, and the unscented Kalman filter propagates deterministically selected sigma points through nonlinear transformations [37]. In trajectory prediction applications, estimated position, velocity, acceleration, or maneuver state provides the initial condition for future extrapolation. Short-term civil aircraft studies illustrate this use, with performance depending on whether the transition model, measurement model, and uncertainty assumptions represent aircraft motion over the intended lookahead horizon [38].

When aircraft motion exhibits substantial nonlinearity or non-Gaussian uncertainty, particle filters approximate the posterior state distribution with weighted particles and propagate them into the future. This representation can retain complex uncertainty structure, but particle propagation and resampling increase computational demand [39]. Retained studies use adaptive particle filtering and sequential Monte Carlo for future trajectory prediction, including multi-aircraft prediction in air traffic management settings [40,41]. Their applications show the value of explicit posterior-sample propagation, while also indicating that computational requirements grow with particle count, state dimension, and transition model complexity.

4.2.3. Multiple-Model State Estimation

A single transition model may be inadequate when an aircraft changes among distinct motion regimes. Multiple-model estimators maintain several candidate dynamic models and update their probabilities as new observations become available. In trajectory prediction, the state associated with each candidate model can be propagated beyond the current observation time, and the resulting trajectory hypotheses can be retained separately or combined according to their model probabilities.

Civil aircraft studies use multiple-model estimation to represent uncertainty over alternative motion regimes. Intent-guided IMM prediction uses spatial and temporal intent information to adjust candidate model probabilities and further refines the trajectory with intent-derived pseudo-measurements [42]. Other studies combine IMM with wind modeling to predict future trajectories of light aircraft or incorporate mode-switching dynamics into probabilistic four-dimensional trajectory prediction [43,44]. Related multiple-model Kalman-filter formulations follow the same principle of maintaining and combining several mode-conditioned estimators.

Multiple-model methods reduce reliance on a single motion hypothesis, but their coverage is bounded by the candidate model set and transition structure. Missing maneuvers or poorly specified transition probabilities can leave model mismatch unresolved, while additional or more complex candidate models increase computational demand.

State estimation methods generate trajectory predictions through observation updates and future propagation, making them suitable for tasks that require continuous assimilation of new information and updates to state uncertainty estimates. Their recursive estimation, probabilistic propagation, and multiple-model mechanisms can also be integrated with data-driven modules to provide state estimates and uncertainty representations in hybrid formulations, as discussed in Section 5.

5. Data-Driven and Hybrid Trajectory Prediction Methods

Data-driven trajectory prediction uses historical flight trajectories and available contextual information to estimate future aircraft states or trajectories through statistical relationships or learned input–output mappings. In this review, data-driven methods comprise classical statistical and machine learning approaches as well as deep learning models. Hybrid models are defined as explicit cross-paradigm formulations in which distinct mechanisms from different modeling families participate in the same prediction chain. These mechanisms may include data-driven modules, aircraft performance or dynamics models, state estimation structures, flight intent, and physical or operational constraints. Table 5 summarizes these method families and their principal characteristics.

5.1. Data-Driven Models

Data-driven models learn mappings from observed trajectory histories and available contextual information to future aircraft states or trajectories. Inputs may include position, speed, heading, altitude, their rates of change, flight plans, operational intent, meteorological conditions, and surrounding traffic information. Depending on the output formulation, a data-driven model may produce a deterministic future trajectory, a predictive distribution, prediction intervals, or multiple plausible future trajectories.

Table 5. Taxonomy and characteristics of data-driven and hybrid aircraft trajectory prediction methods.

Method Family	Subfamily	Core Prediction Mechanism	Representative Formulations	Output/Uncertainty Representation	Key Characteristics/Limitations
Data driven	Classical statistical and machine learning	Engineered trajectory and contextual representations are mapped to future aircraft states or trajectory quantities using non-deep statistical or machine learning models	Linear and functional regression; Gaussian process and kernel regression; support vector regression; random forests; gradient boosting and related ensembles	Usually point estimates; probabilistic regression and ensemble formulations may additionally provide predictive variance, intervals, or empirical uncertainty estimates	Explicit feature definitions facilitate interpretation and moderate model complexity; performance depends on feature design, training-data coverage, and generalization to unseen operating conditions
Data driven	Deep learning	Trajectory and contextual representations are learned jointly with the future-trajectory mapping using multilayer neural architectures	Recurrent and convolutional sequence models; attention and Transformer models; graph-based models; deep probabilistic and generative models; composite deep architectures	Deterministic multi-step trajectories, predictive distributions, or multiple trajectory samples depending on the output formulation	Flexible nonlinear representation learning supports temporal, spatial, contextual, and relational information; training and inference cost, interpretability, calibration, and robustness depend on architecture and training design
Hybrid	Parameter and dynamics augmentation	Data-driven components estimate parameters, dynamics terms, or correction quantities used by an explicit structured propagation model	Learned parameter or residual dynamics augmentation; Neural ODE and analytical dynamics coupling	Deterministic or probabilistic according to the coupled propagation model	Combines learned adaptation with explicit propagation; performance depends on the learned–structured interface and consistency of the coupled model
Hybrid	State estimation and probabilistic fusion	Learned predictions, pseudo-measurements, corrections, or auxiliary forecasts are incorporated into recursive state estimation and future propagation	Estimator–neural fusion; learned pseudo-measurements or corrections; residual mean IMM and LSTM–IMM formulations	Deterministic or probabilistic according to the estimator and coupled learned component; uncertainty may be represented through state or mode distributions	Combines learned trajectory information with recursive estimation and uncertainty handling; requires consistent fusion and estimator design
Hybrid	Constraint- and operational prior-informed learning	Explicit physical, procedural, or operational priors or constraints directly regularize, restrict, or shape training, prediction, or decoding	Prior- or constraint-informed learning; constrained decoding; feasibility correction or projection	Deterministic or probabilistic according to the underlying predictor; constraints may act during training, prediction, decoding, or feasibility enforcement	Supports structural or operational consistency but requires valid priors or constraints and careful weighting or enforcement

5.1.1. Classical Statistical and Machine Learning Models

Classical statistical and machine learning methods construct explicit features from observed trajectories and operating conditions, and predict future aircraft states or trajectory quantities using regression, kernel, tree-based, or ensemble methods. Stochastic predictive analytics has been applied to aircraft trajectory prediction in air traffic management [45], while functional regression has used historical radar tracks to forecast aircraft positions over short and medium horizons [46]. Figure 5 summarizes the general workflow.

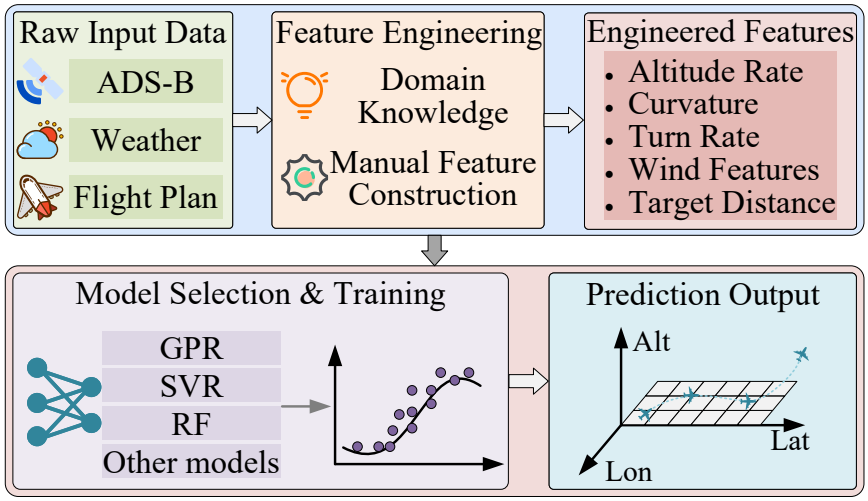


Figure 5. General workflow of classical statistical and machine learning trajectory prediction.

Regression, kernel, and trajectory statistics methods have been applied to final-approach prediction, historical trajectory extrapolation, and stochastic trajectory prediction. Related studies use operational and meteorological variables for final-approach regression [47], local functional regression in Sobolev space for short- and medium-term position prediction [46], and Gaussian process regression to produce predictive means together with variances [48]. Trajectory data-mining methods extract typical trajectories from historical observations and use them to update transition probabilities and refine nominal trajectory propagation [49].

Tree-based and ensemble methods are useful when prediction relies on heterogeneous engineered features and nonlinear interactions. Ensemble meta-estimators have been developed using flight plan, surveillance, and meteorological variables [50]. Reported performance remains dependent on the informativeness of the feature representation and the coverage of the training data. Missing routes, weather conditions, or operating states in historical data can limit generalization.

5.1.2. Deep Learning Models

Deep learning models jointly learn trajectory representations and future-trajectory mappings from sequential observations and contextual information. They may produce deterministic multi-step trajectories, predictive distributions, or multiple trajectory samples. Figure 6 illustrates the general workflow, and Table 6 summarizes the principal mechanisms recurring in the retained studies. These mechanisms may be combined within one predictor.

Table 6. Representative modeling mechanisms in deep learning-based aircraft trajectory prediction.

Modeling Mechanism	Temporal/Spatial Representation	Interaction Representation	Output/Uncertainty Formulation	Representative Architectures	Structural Limitations
Recurrent and convolutional sequence	Recurrent states represent temporal evolution, while convolutional operators encode local trajectory, weather, or spatial patterns; the mechanisms can be used separately or jointly	Interactions are typically introduced through shared contextual features, pooling, fusion, or additional relational modules when required	Commonly deterministic sequence prediction; probabilistic heads or sampling-based extensions can also be incorporated	RNN, LSTM, GRU, CNN, ConvLSTM, recurrent encoder–decoder models	Long temporal dependencies may remain difficult for recurrent propagation; convolution alone does not represent long-range temporal structure; recursive decoding can accumulate error
Attention and Transformer	Self-attention relates trajectory tokens, state variables, or contextual representations across the observation sequence and supports direct modeling of long-range dependencies	Dependencies may be represented through self-attention, cross-attention, token construction, or additional relational modules	Deterministic multi-step output is common, while probabilistic heads, latent-variable modules, or generative decoders can be integrated	Transformer, inverted Transformer, attention-based encoder–decoder, non-autoregressive Transformer-like predictors	Standard attention cost increases with sequence or token length; performance depends on representation design, positional information, context construction, and training-data coverage
Graph and interaction-aware	Aircraft, trajectory attributes, or spatiotemporal states are represented as nodes, with relational information propagated through neighborhood aggregation or graph attention and often combined with temporal encoders	Explicit edges or learned adjacency structures encode relative position, velocity, proximity, correlation, or other relational information	May produce deterministic trajectories or probabilistic outputs depending on the associated prediction head	GCN, GAT, ST-GCN, dynamic graph networks, graph–attention sequence models	Graph construction and neighborhood definition introduce sensitivity to the selected relational structure; computational and memory requirements increase with graph size and temporal resolution
Deep probabilistic and generative	A conditional distribution over future trajectories is modeled from historical and contextual representations; generation may use Bayesian approximations, latent variables, adversarial learning, diffusion, or related generative mechanisms	Context and interaction information enter through the conditioning representation and may itself be encoded recurrently, through attention, or through graph-based mechanisms	Predictive distributions, uncertainty intervals, or multiple future-trajectory samples	Bayesian deep models, GAN/CGAN, diffusion models, and related conditional generative formulations	Calibration and mode coverage require explicit evaluation; some formulations introduce substantial training or sampling cost and require distribution-sensitive evaluation metrics
Composite deep architectures	Several learned modules are connected serially, in parallel, or through fusion mechanisms to combine complementary temporal, spatial, and contextual representations	Interaction information is inherited from constituent modules or incorporated through attention, pooling, graph, or branch-fusion mechanisms	Depends on the prediction head and may be deterministic or probabilistic	CNN–LSTM/GRU, TCN–BiGRU, Transformer–BiGRU, multi-branch and multi-scale feature-fusion networks	Additional components increase architectural complexity and can make attribution of performance gains to individual mechanisms more difficult; redundancy and tuning requirements may also increase

Note. The rows describe recurring modeling mechanisms rather than mutually exclusive deep learning categories. Attention, graph-based interaction modeling, probabilistic generation, and composite architectures may coexist within the same trajectory predictor.

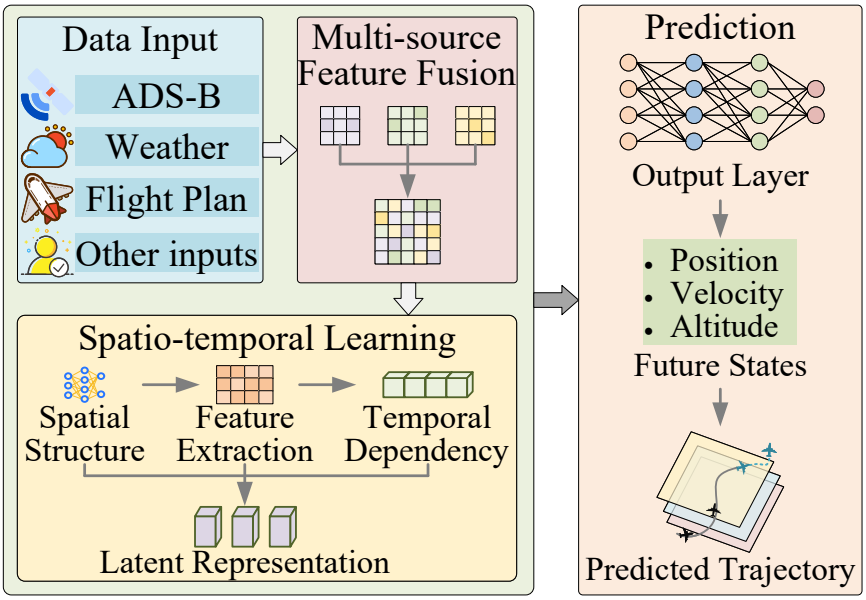


Figure 6. General deep learning workflow for aircraft trajectory prediction.

Recurrent and Convolutional Sequence Models

LSTM, GRU, ConvLSTM, and recurrent encoder–decoder models have been used for aircraft trajectory prediction. Applications include direct LSTM prediction [51], online GRU updating with newly observed trajectories [52], convolutional encoding of convective weather and planned flight information [53], and multi-step four-dimensional prediction in terminal airspace [54]. Recursive decoding can nevertheless accumulate error over successive prediction steps.

Attention- and Transformer-Based Models

FlightBERT combines binary encoding with a stacked Transformer backbone [55], while FlightBERT++ introduces non-autoregressive multi-horizon prediction with horizon-aware context generation and differential decoding [56]. Improved Transformer models have been evaluated for long-term civil aircraft prediction [57], and an inverted Transformer represents individual trajectory variables as variate tokens to model multivariate flight state dependencies [58]. These studies use attention mechanisms to model long-range dependencies, support multiple prediction horizons, and capture multivariate trajectory relationships.

Graph- and Interaction-Aware Models

Reviewed studies use attribute graphs, spatiotemporal graphs, and social pooling structures to represent relations among trajectory variables, aircraft, or evolving aircraft states. Attribute graph methods use graph convolution to extract local and global structure among trajectory variables [59]; spatiotemporal graph models combine graph modules with temporal prediction for four-dimensional trajectories [60,61]; and other studies use social pooling to incorporate neighboring aircraft hidden states [62] or flight pattern guidance [63]. Graph structures specify how relational information enters the predictor, while single-target and joint multi-aircraft outputs are treated separately in Section 7.

Deep Probabilistic and Generative Models

Dropout-based Bayesian approximate inference has been used for trajectory prediction under uncertain weather conditions and can provide uncertainty intervals for predicted longitude and latitude [64]. Deep Gaussian process models have also been used to predict nominal trajectories and associated uncertainty information [65].

Generative models produce a conditional trajectory distribution or multiple plausible futures. Conditional GANs have been applied to weather-conditioned aircraft trajectory prediction [66], and convolutional or recurrent GAN formulations have been used for long-term four-dimensional prediction [67]. Diffusion models provide another mechanism for conditional trajectory generation through iterative denoising [68]. Recent studies either condition diffusion models on historical behavior, social interaction, and operational context [11] or estimate an intention distribution before generating trajectories with a Transformer-based diffusion model [69]. The main evaluation concerns for these methods are uncertainty calibration, coverage of plausible modes, and training or sampling cost.

Composite Deep Architectures

Many predictors combine neural modules to process complementary input information. CNN–LSTM models combine convolutional feature extraction with recurrent temporal modeling [70]; TCN–BiGRU models combine temporal convolution, bidirectional recurrent processing, and attention [71]; and convolution–GRU encoder–decoders incorporate tactical intent into trajectory prediction [21]. Other work evaluates recurrent architectures on data spanning multiple routes and dates to examine generalization [72]. These architectures remain within the data-driven family when they combine only learned representation or prediction modules. They are classified as hybrid only when physical models, state estimation mechanisms, or explicit priors and constraints materially participate in trajectory generation or uncertainty propagation.

5.2. Hybrid Models

In this review, hybrid trajectory prediction models are defined by explicit coupling between different modeling paradigms. Distinct mechanisms from two or more principal modeling families participate materially in the same prediction chain. These mechanisms may include data-driven modules, aircraft performance or dynamics models, recursive state estimators, and explicit priors or constraints derived from flight intent, physical principles, or operational requirements. This definition distinguishes hybrid models from composite deep architectures consisting only of neural modules and also applies to state estimation formulations involving explicit cross-paradigm coupling.

5.2.1. Parameter and Dynamics Augmentation

One hybrid pattern uses a data-driven module to estimate quantities required by an explicit trajectory-propagation model. NODE-FDM combines analytical kinematic relations with Neural ODE components learned from quick access recorder (QAR) data to model the temporal evolution of aircraft flight dynamics quantities [73]. The learning–dynamics interface and the consistency of the coupled components over the operating conditions of interest determine the effectiveness of this pattern.

5.2.2. State Estimation and Probabilistic Fusion

A second hybrid pattern introduces learned trajectory information into a recursive state estimation process through a pseudo-measurement, prediction error correction, auxiliary forecast, or related data-derived input. This arrangement combines trajectory patterns learned from historical data with the current observed state and its uncertainty.

A terminal area framework learns collective trajectory behavior from surveillance data and supplies the data-driven prediction to a residual mean IMM as a pseudo-measurement, allowing the estimator to reconcile learned future behavior with the current aircraft state [74]. An LSTM–IMM formulation uses an LSTM-based component to predict time-related information and compensate for errors in longitude, latitude, and altitude associated with IMM propagation [75]. The reported formulations differ in the information exchanged

across the learning–estimation interface, while their Bayesian and multiple-model estimation mechanisms follow the state estimation processes described in Section 4.2.

5.2.3. Constraint- and Operational Prior-Informed Learning

A third hybrid pattern uses explicit physical, procedural, or operational priors or constraints to shape model training, prediction, or decoding. Such information may include waypoints, route structure, flight phase, operational intent, aircraft performance, and runway conditions. A model is classified as hybrid when this information is explicitly used to regularize training, correct infeasible predictions, or constrain decoding.

Constrained LSTM prediction provides one example in which top-of-climb, waypoint, and runway direction constraints are associated with climb, cruise, and descent/approach phases, respectively [76]. When route, phase, intent, performance, or other operational information is supplied only as an input feature, the model remains in the data-driven category. When the same information directly restricts trajectory generation or enforces feasibility, it forms an explicit prior or constraint within the hybrid model.

The reviewed hybrid methods combine different modeling families through parameter or dynamics augmentation, state estimation fusion, or the incorporation of explicit operational priors and constraints. Prediction horizon and target cardinality are examined separately in the following sections.

6. Trajectory Prediction Under Different Time Horizons

Prediction horizon is a cross-cutting task dimension applicable to different aircraft trajectory prediction paradigms. Physics-based, state estimation, data-driven, and hybrid models can all be used over different lookahead intervals, but the relative roles of recent motion, operational intent, environmental context, uncertainty representation, and sequence generation change as the horizon extends. The reviewed studies describe prediction extent in various forms, including explicit seconds or minutes, numbers of future steps, author-defined short- or long-term tasks, flight phases, full-flight prediction, and distance-based lookahead.

This review uses 10 min as a boundary for time-normalized analytical stratification. Ten-minute prediction tasks or intervals that include 10 min recur in the retained evidence, providing a common timescale for part of the synthesis [10,16,49]. The boundary also provides a practical separation between settings dominated by recent state continuity and settings in which flight intent, weather, aircraft performance assumptions, and operational changes become more influential. This boundary serves as an analytical convention for organizing the evidence, while operational definitions may vary among studies. Studies are assigned to the time-normalized strata only when their reported information permits a reliable conversion to time. Step-defined, phase-specific, full-flight, and distance-defined horizons are retained in their original form when such a conversion is unavailable. Table 7 summarizes the principal horizon-dependent considerations.

Table 7. Horizon-dependent considerations in aircraft trajectory prediction.

Horizon-Sensitive Dimension	Shorter-Horizon Analytical Stratum	Longer-Horizon Analytical Stratum
Review horizon stratum	Time-normalized prediction horizons of up to and including 10 min	Time-normalized prediction horizons exceeding 10 min
Dependence on recent observations	The latest aircraft state and recent motion history provide relatively strong constraints on the near future, although local maneuvers can still change the trajectory rapidly	The latest observed state constrains a progressively smaller portion of the future evolution as additional maneuvers, operational changes, and environmental effects may occur

Table 7. *Cont.*

Horizon-Sensitive Dimension	Shorter-Horizon Analytical Stratum	Longer-Horizon Analytical Stratum
Contextual information	Current kinematics and recent trajectory history remain central; local weather, traffic, intent, and procedural information may provide additional predictive context	Flight intent, route and procedure information, weather forecasts, aircraft performance assumptions, and anticipated operational changes can become increasingly consequential
Uncertainty sources	Observation noise, missing or irregular samples, near-term maneuver or mode ambiguity, and local environmental disturbances	Intent or plan changes, procedure or ATC changes, weather forecast uncertainty, aircraft performance uncertainty, and accumulated prediction uncertainty over a longer future sequence
Sequence generation implications	Recursive and direct multi-step formulations are both applicable; recursive prediction may accumulate error over successive future steps	Longer output sequences increase the importance of controlling rollout drift, preserving trajectory and endpoint consistency, and representing alternative future evolutions when required
Updating and recomputation	New observations can revise near-future predictions through rolling state updates, parameter adaptation, or repeated inference when supported by the prediction system	Recomputation becomes particularly relevant when intent, flight plan, weather, procedural, or constraint information changes during the extended lookahead interval
Evaluation implications	Horizon-specific state or position errors retain the relationship between reported accuracy and the corresponding lookahead interval	Horizon-dependent errors, sequence or endpoint consistency, and uncertainty information become increasingly important for interpreting performance across extended prediction intervals

Note. The 10 min boundary is used only as a pragmatic analytical stratification in this review for studies whose prediction horizons can be reliably normalized to time. Reported prediction intervals beginning at 10 min and extending beyond it are treated as longer-horizon evidence. Step-defined, phase-specific, full-flight, and distance-defined horizons are retained in their reported form unless a reliable time mapping is available.

6.1. Shorter-Horizon Prediction

At shorter lookahead intervals, the latest aircraft state and recent motion history generally provide strong constraints on future states. Choi et al. combine a data-driven trajectory pattern model with a residual mean interacting multiple-model estimator and evaluate a 2 min lookahead in terminal airspace [74]. Thipphavong et al. adapt an aircraft-weight model using observed track data and report a 5 min lookahead for climbing aircraft [32]. These studies use different prediction mechanisms, but both show how recent observations support near-term prediction.

Shorter-horizon prediction remains sensitive to uncertainty in the current aircraft state and local operating conditions. Measurement noise, missing or irregular surveillance samples, maneuver ambiguity, and short-range weather variation can affect the inferred future motion. Feng et al. combine historical trajectories and meteorological information with LSTM and Bayesian neural modeling to represent uncertainty in short-term prediction [77]. GooDFlight is described by its authors as a long-term intention-guided framework, but its TrajAir evaluation predicts 120 s ahead and therefore belongs to the shorter-horizon stratum used here [69]. These tasks can benefit from recent state assimilation and online updating, while evaluation should report prediction errors at their corresponding lookahead intervals.

The generation of multiple future states also affects shorter-horizon prediction. Recursive prediction can propagate errors through successive steps, whereas direct multi-step or multi-horizon formulations can generate several future states from the observed history. TPGAN formulates an author-defined short-term multi-horizon task as a direct conditional generative problem to reduce iterative error accumulation [78]. FlightBERT evaluates 1, 3,

9, and 15 future steps, mapped by the source to approximately 20 s, 1 min, 3 min, and 5 min, respectively [55]. Step-defined outputs can be assigned to a time-normalized stratum only when the temporal spacing is explicitly reported.

New observations can support rolling revision of shorter-horizon predictions. A two-stage GRU framework combines offline initialization with online adjustment from newly observed trajectories [52]. Trajectory data-mining methods use typical historical trajectories to update transition probabilities, with evaluation extending to a 10 min lookahead [49]. Multimodal prediction uses newly available operational information, including spoken air traffic control instructions, to condition predictions over horizons from 1 to 10 min [10]. These examples show how recent states and context can be incorporated when the prediction system supports online revision, without implying a fixed update frequency for all shorter-horizon tasks.

6.2. Longer-Horizon Prediction

As the prediction horizon increases, a larger part of the future trajectory cannot be determined from recent motion continuity alone. Future maneuvers, wind evolution, aircraft performance assumptions, route adherence, and operational decisions become more influential. Sequential Monte Carlo prediction has been evaluated over a 20 min multi-aircraft horizon while estimating wind forecast errors from radar observations [41]. Adaptive particle filtering has examined horizons up to 40 min and used sequential observations to estimate uncertain model parameters and weather effects before future propagation [40]. Local functional regression has evaluated future aircraft positions over a 10–30 min interval [46]. These studies show that recent observations remain useful at longer horizons, but increasingly as a basis for estimating states, parameters, or trajectory patterns whose effects extend across the prediction interval.

Operational intent becomes more important when the prediction interval spans future maneuvers or mode changes that are not determined by the current aircraft state. Intent-based probabilistic prediction combines aircraft-dynamics models with verified or inferred intent and has been evaluated at horizons up to 20 min [79]. Route, waypoint, procedural, weather, and aircraft performance information can constrain future evolution that recent kinematics alone cannot determine. Their value therefore depends both on elapsed prediction time and on whether the horizon crosses operational decisions, flight phase transitions, or environmental changes.

Longer sequences also increase the importance of the sequence generation strategy. Recursive decoding may propagate early errors through more future steps, whereas direct sequence generation reduces dependence on intermediate predicted states. A civil aviation Transformer model for long-term prediction replaces serial dynamic decoding with direct sequence generation [57]. A TCN–Transformer model evaluates 96 future steps at 15 s intervals, corresponding to 1440 s or 24 min [80]. Another long-term Transformer uses positional encoding and multi-head attention to process extended trajectory sequences [81]. These studies illustrate alternative generation strategies rather than establishing one architecture as universally preferable.

As the prediction interval expands, several future evolutions may remain compatible with the information available at prediction time. Probabilistic and generative models can represent predictive intervals or alternative trajectories when required by the task. A deep Gaussian process framework predicts nominal trajectories with confidence information [65], while a GAN-based formulation addresses long-term four-dimensional trajectory prediction using historical route data [67]. Such results should be interpreted using horizon-specific accuracy and uncertainty measures that distinguish performance across different forecast lead times.

Extended lookahead also increases the possibility that prediction conditions will change before the end of the forecast. Flight plan amendments, revised intent, weather updates, procedural changes, and observed deviations may therefore trigger recomputation. Online four-dimensional prediction has been designed to recompute trajectories when observed position or speed deviates beyond predefined thresholds and to revise horizontal or vertical intent accordingly [24]. Recomputation can therefore be triggered by changes in available information across different prediction horizons.

6.3. Cross-Horizon and Alternative Horizon Definitions

Aircraft trajectory prediction studies do not use a single standardized definition of prediction extent. In addition to explicit seconds or minutes, studies may report future steps, flight phases, complete routes or flights, distance-based lookahead, or author-defined short- and long-term tasks. These definitions should not be converted to a common temporal scale unless the source provides sufficient information for a reliable mapping. Identical labels such as "short term" or "long term" may refer to substantially different temporal or operational extents.

Step-defined prediction is the clearest case in which conversion may be possible. If both the number of future steps and their temporal spacing are reported, the lookahead time can be derived directly. The TCN–Transformer study predicts 96 future steps at 15 s spacing, yielding a 24 min horizon [80]. By contrast, describing a model as multi-step or multi-horizon does not establish a particular temporal extent. FlightBERT++ directly predicts multiple future horizons, but this does not imply a lookahead greater than 10 min [56]. Multi-step and multi-horizon therefore describe output structure, whereas assignment to a time-normalized stratum requires an explicit temporal mapping.

Phase-specific prediction characterizes the operational segment covered by a task, with duration varying across operating conditions. Online four-dimensional prediction has been evaluated with intent updating during climb [24]; constrained LSTM prediction uses different constraints for climb, cruise, and descent or approach [76]; and sequence-to-sequence prediction in terminal airspace models the future trajectory from a selected start through the terminal segment without reporting a universal duration [54]. These studies are retained as phase-specific evidence unless the source provides a defensible time conversion.

Full-flight and route-based prediction are treated in the same way. They describe the operational extent of the predicted trajectory, while actual flight duration varies across routes and operating conditions. An HMM and trajectory similarity framework predicts waypoint flyover time and altitude before departure and revises the prediction after departure using similarity between the executed and historical trajectories [36]. The horizon is therefore retained as a full-flight or route-based descriptor because a reliable time interval cannot be established from the reported information.

Distance-defined lookahead is also retained in its original form unless sufficient information is reported for a reliable conversion to time. This avoids introducing artificial temporal comparability into tasks defined by operational range, spatial extent, or sequence structure.

Prediction horizon influences the relative roles of recent state information, contextual information, uncertainty, sequence generation, and prediction updating across modeling paradigms. Identical horizon labels can correspond to different temporal or operational extents. Prediction horizon should therefore be interpreted jointly with the reported task definition, input information, and evaluation conditions.

7. Trajectory Prediction by Target Cardinality

Target cardinality is determined by the number of aircraft represented in the predicted output. This review distinguishes single-target prediction from joint multi-aircraft prediction and treats interaction context as a complementary dimension. A dataset containing many aircraft does not imply joint prediction if each future trajectory is generated separately, and the use of neighboring aircraft information does not imply joint output when only one focal aircraft is predicted. Joint multi-aircraft prediction requires the future states or trajectories of several aircraft to be represented within a coupled prediction problem. Table 8 summarizes these relationships.

Table 8. Relationship between target cardinality and inter-aircraft interaction context in aircraft trajectory prediction.

Analytical Dimension	Single-Target Output	Interaction-Aware Focal/Per-Aircraft Prediction	Joint Multi-Aircraft Output
Prediction unit and output structure	One target aircraft is predicted per prediction instance; the output may be deterministic, probabilistic, or multimodal	A focal or per-aircraft trajectory is predicted while neighboring aircraft information conditions the prediction	Future states or trajectories of multiple aircraft are represented within one coupled prediction problem
Interaction context	Neighboring aircraft information is not required	Neighbor states, relative geometry, traffic structure, or learned inter-aircraft relations explicitly condition prediction	Inter-aircraft information is incorporated within the joint state, shared representation, or coupled prediction process
Coupling across predicted trajectories	No cross-target output coupling is required	Interaction context may alter individual predictions while outputs remain separately generated	Multiple aircraft futures are explicitly coupled through the state, representation, or output formulation
Uncertainty representation	Uncertainty concerns alternative futures of one target aircraft	Neighboring aircraft uncertainty may affect the focal prediction without establishing a joint predictive distribution	Per-aircraft or joint uncertainty may be represented; cross-aircraft dependence requires explicit joint modeling

Note. Interaction awareness is a contextual dimension applicable to both single-target and joint multi-aircraft prediction.

7.1. Single-Target Prediction

Single-target prediction produces the future trajectory or state sequence of one aircraft per prediction instance. Choi et al. used a data-driven trajectory pattern model as a pseudo-measurement in an RM-IMM estimator for an individual incoming flight [74]. FlightBERT maps the observed attributes of one flight to its future states with a Transformer-based predictor [55], while a deep Gaussian process formulation predicts a nominal trajectory with uncertainty for the same target aircraft [65]. These examples illustrate deterministic and probabilistic single-target prediction. More generally, multimodal outputs can also be generated for a single target aircraft.

Single-target prediction can be repeated across a dataset containing many aircraft without becoming joint multi-aircraft prediction. It can also use flight plan, weather, performance, or neighboring aircraft information, provided that the predicted output remains associated with one target aircraft. Likewise, comparing separately predicted trajectories in a downstream conflict or separation analysis does not change the target cardinality of the underlying predictors.

7.2. Multi-Aircraft Interaction Context

Interaction-aware prediction uses surrounding aircraft information to condition the prediction of a focal aircraft or a set of predicted aircraft. Relevant information includes relative states, proximity, sequencing or merging relationships, and local traffic structure. This context can influence input construction, neighborhood selection, relational representation, and updating, but it does not by itself determine the number of aircraft in the predicted output.

The reviewed civil aircraft studies include spatiotemporal attention for close-range terminal airspace interactions [82], learnable adjacency representations for spatial correlations among aircraft [60], and conditional diffusion prediction using historical behavior, social interaction, and operational context [11]. GooDFlight also incorporates interaction-aware goal information into stochastic trajectory generation [69]. These studies illustrate different ways in which neighboring aircraft information can be incorporated into trajectory prediction. Such information can support either single-target prediction or coupled multi-aircraft prediction, depending on the output formulation.

7.3. Joint Multi-Aircraft Prediction

Joint multi-aircraft prediction represents the future states or trajectories of several aircraft within one coupled prediction problem. Lymperopoulos and Lygeros formulate multi-aircraft prediction as a high-dimensional state estimation problem and develop a sequential Monte Carlo method that exploits the structure of the coupled system [41]. TrajAirNet provides a data-driven formulation that predicts trajectories for the agents represented in a multi-aircraft scenario while incorporating agent interactions and environmental context [83]. The flight pattern-guided Social LSTM of Chu et al. addresses joint trajectory prediction [63], while the Bayesian Social LSTM of Zhong et al. provides a related formulation incorporating uncertainty representation [84].

Joint output does not automatically imply a joint probability distribution across aircraft. A model may produce several deterministic trajectories or separate per-aircraft uncertainty estimates without representing cross-aircraft dependence. Claims about correlated or joint predictive uncertainty should therefore be limited to formulations that explicitly model this dependence.

Target cardinality also affects computational structure. Repeated single-target prediction treats different aircraft as separate prediction targets, while interaction-aware formulations incorporate neighboring aircraft information and joint formulations couple the predictions of multiple aircraft. Joint formulations can increase state or output dimensionality and the number of represented relationships as the number of targets grows. The resulting cost depends on neighborhood selection, relational sparsity, model architecture, update frequency, and implementation strategy.

For evaluation, per-aircraft trajectory errors remain applicable to both single-target and joint formulations. Interaction-aware or joint output studies may additionally evaluate relative-position consistency, separation-related quantities, or joint uncertainty when these quantities are explicitly represented. Conflict probability, loss of separation, and downstream decision-support performance should be reported as operational outcomes distinct from trajectory prediction accuracy.

Target cardinality specifies the number of aircraft represented in the predicted output, while the output formulation indicates whether their future states are coupled. Interaction context describes the use of surrounding aircraft information in either single-target or joint multi-aircraft prediction. These distinctions inform the selection of evaluation measures, including per-aircraft errors, relational consistency, and joint uncertainty.

8. Performance Evaluation and Comparative Analysis

Performance evaluation in aircraft trajectory prediction is influenced by the prediction method, dataset or operational scenario, prediction horizon, sampling interval, target cardinality, output formulation, and evaluation metric. Following the performance synthesis criteria described in Section 2.6, 55 of the 198 primary studies entered the formal performance synthesis. These studies comprise 9 physics-based, 6 state estimation, 11 classical statistical and machine learning, 24 deep learning, and 5 hybrid studies. This 55-study set is a performance-evidence subset with an identifiable horizon, evaluation metric, and quantitative result. Its category distribution describes the composition of the quantifiable evidence and should not be interpreted as the distribution of the 198 primary studies or of the field as a whole.

The quantitative evidence is synthesized in a stratified manner using the distribution of study counts, within-study baseline comparisons, changes in error across prediction horizons, and the reported distributions of target cardinality, output formulation, and metric families. Because datasets, scenarios, sampling intervals, coordinate representations, prediction horizons, and metric definitions differ across studies, cross-study numerical results are interpreted to identify condition-specific performance patterns, with direct comparisons restricted to sufficiently comparable experimental settings. Studies with incomplete performance reporting remain part of the methodological and task-oriented synthesis but are not included in the formal quantitative synthesis presented here.

This section first summarizes the evaluation metrics and experimental data conditions represented in the literature, then organizes quantitative performance evidence by the five modeling paradigms, and finally identifies cross-paradigm findings concerning prediction horizon, target cardinality and interaction context, deterministic and probabilistic output, and engineering performance.

8.1. Evaluation Metrics

Evaluation metrics determine which aspects of trajectory prediction performance are observed and therefore affect the interpretation of reported results. The metrics identified in the reviewed studies are organized into three complementary families: point-wise and trajectory-level error metrics, probabilistic and uncertainty metrics, and engineering or task-specific metrics. Point or state errors quantify deviations from observed aircraft states, trajectory-level metrics describe errors over a predicted sequence or at its endpoint, probabilistic metrics assess the accuracy, calibration, and sharpness of predictive distributions, and engineering or task-specific metrics characterize operational consequences. Metric values must be interpreted in relation to the prediction horizon, coordinate representation, output formulation, and experimental setting.

8.1.1. Error Metrics

MAE and RMSE are commonly reported for longitude, latitude, horizontal position, altitude, velocity, and arrival time errors. MAE summarizes the average absolute deviation, whereas RMSE gives greater weight to larger deviations and is more sensitive to occasional large errors [85]. For multi-step prediction, ADE summarizes displacement over the predicted sequence and FDE emphasizes the final predicted point [86]. These measures can therefore lead to different interpretations when the application prioritizes trajectory-wide accuracy or endpoint accuracy.

ATE and CTE describe deviations parallel and perpendicular to a reference route and are useful in route-following and terminal area analyses [8]. MAPE and DTW occur as complementary measures in specific settings, respectively describing relative error and sequence similarity; MAPE becomes difficult to interpret when the reference quantity

approaches zero [87,88]. For multimodal or generative prediction, minADE and minFDE report the minimum ADE and FDE, respectively, among K generated trajectories [89]. These best-of-K errors characterize the accuracy of the closest candidates but do not establish the reliability or calibration of the complete predictive distribution.

8.1.2. Probabilistic and Uncertainty Metrics

Probabilistic trajectory prediction methods represent future motion through predictive distributions, confidence or prediction intervals, covariance structures, ensembles, or multiple trajectory hypotheses. Their evaluation should distinguish trajectory accuracy from the reliability, calibration, and concentration of the represented uncertainty.

When an explicit predictive density is available, NLL evaluates the negative log predictive density assigned to the realized outcome, while CRPS measures the discrepancy between a predictive cumulative distribution and the observed outcome [90,91]. Conventional CRPS is defined for a univariate predictive distribution. When it is applied separately to longitude, latitude, altitude, or another state component, it provides component-wise evidence. If a study reports a joint distribution for future states or a complete trajectory, it should specify the multivariate scoring rule used, including any explicitly defined multivariate extension of CRPS. Component-wise CRPS values should not be treated as equivalent to a score for the full joint trajectory distribution.

Prediction-interval coverage measures the proportion of realized values contained within a specified interval, while interval width characterizes its sharpness [92]. These quantities should be interpreted together because high coverage may be obtained with excessively broad intervals, while narrow intervals are uninformative if realized trajectories frequently fall outside them. Reliability curves, calibration errors, and related calibration analyses provide additional evidence on the agreement between predicted probabilities and observed frequencies.

Some studies characterize uncertainty using covariance matrices, confidence regions, Monte Carlo envelopes, empirical error distributions, or sample-based trajectory sets. These representations provide useful quantitative evidence but are not numerically interchangeable with standardized probabilistic scores. For generative prediction, best-of-K errors, sample diversity, and distributional similarity measures describe candidate coverage or sample characteristics; they do not replace NLL, CRPS, coverage, or calibration assessment.

8.1.3. Engineering and Task-Specific Metrics

Trajectory accuracy does not fully characterize the computational or operational properties of a prediction method. Some studies report inference latency, throughput, execution time, parameter count, or computational complexity, which are relevant when predictions must be updated repeatedly in real time. These quantities depend on the hardware platform, software implementation, input size, output horizon, and model configuration [93].

Task-specific measures depend on the operational function associated with trajectory prediction. Arrival-oriented studies may report estimated-time-of-arrival (ETA) error or waypoint time error, whereas conflict-oriented studies may report conflict probability, loss-of-separation indicators, false alarms, or missed detections. These measures describe downstream operational effects and should be reported separately from geometric trajectory accuracy.

Table 9 summarizes the principal metric families encountered across the reviewed trajectory prediction settings.

Table 9. Evaluation metric families for different aircraft trajectory prediction settings.

Prediction Setting	Primary Evaluation Metrics	Complementary Metrics
Shorter-horizon prediction	MAE, RMSE, ATE/CTE, position error	Inference latency, covariance or confidence bounds
Longer-horizon or full-flight prediction	ADE, FDE, horizon-specific MAE/RMSE, ETA error	ATE/CTE, endpoint error, uncertainty measures
Probabilistic or multimodal prediction	NLL, CRPS, coverage	Interval width, minADE/minFDE, calibration or reliability measures
Interaction-aware or joint multi-aircraft prediction	Per-aircraft ADE/FDE, position error	Joint or dependency measures when cross-aircraft dependence is explicitly modeled; downstream conflict or separation measures
Phase- or procedure-specific prediction	Position or altitude error, ATE/CTE, ETA or waypoint time error	Phase-specific uncertainty and operational constraint measures

Deterministic trajectory accuracy, probabilistic reliability, and engineering or task-specific performance represent distinct but complementary evaluation dimensions. Probabilistic or generative outputs require attention to accuracy, candidate coverage, distributional reliability, and calibration. For interaction-aware or joint multi-aircraft prediction, per-aircraft errors may be supplemented by relational, dependency, or joint-uncertainty measures when these structures are explicitly represented. The choice of metrics should therefore remain consistent with the prediction task, output formulation, horizon, and operational setting.

8.2. Evaluation Data and Experimental Context

The data used for model development and evaluation determine the information available to the predictor and affect the applicability and comparability of reported performance. The formal performance evidence uses different types of data and contextual information, including historical trajectory observations, environmental and flight intent information, and simulated or synthetic data. These sources differ in spatial coverage, temporal resolution, measurement quality, flight phase, traffic environment, and availability of auxiliary operational information. Table 10 summarizes their representative sources, typical information, and roles in trajectory prediction.

Table 10. Principal data and contextual information used in aircraft trajectory prediction.

Data/Context Category	Representative Sources	Typical Information	Main Characteristics	Typical Roles in Trajectory Prediction
Historical trajectory data	ADS-B, OpenSky, surveillance radar, operational ATC records	Position, altitude, speed, heading, vertical rate, time stamp	Direct observations of aircraft motion; may contain noise, missing reports, irregular sampling, and regional coverage differences	Model development and evaluation, state estimation, shorter- and longer-horizon prediction
Environmental information	Numerical weather prediction products, wind fields, meteorological observations	Wind, temperature, pressure, convective weather, and related atmospheric variables	Describes exogenous conditions that influence aircraft motion and trajectory deviation	Weather-aware prediction, uncertainty analysis, and extended-horizon prediction

Table 10. *Cont.*

Data/Context Category	Representative Sources	Typical Information	Main Characteristics	Typical Roles in Trajectory Prediction
Flight intent information	Flight plans, waypoints, SID/STAR procedures, target flight levels, ATC instructions	Planned route, procedural constraints, target conditions, and control intent	Represents intended future behavior that may not be recoverable from recent motion alone	Intent-aware, route-constrained, arrival/departure, and full-flight prediction
Simulated or synthetic data	BlueSky, AirTOp, aircraft performance-based simulation, and other civil aviation simulation environments	Configurable trajectories, aircraft states, traffic scenarios, environmental and operational variables	Controllable and reproducible but dependent on the fidelity of simulated dynamics and operations	Controlled evaluation, sensitivity analysis, rare scenario generation, and uncertainty experiments

8.2.1. Historical Trajectory Data

Historical trajectory data provide the main empirical basis for aircraft trajectory prediction. Common sources include Automatic Dependent Surveillance–Broadcast (ADS-B), surveillance radar, and operational air traffic control records. ADS-B can provide aircraft identification, position, altitude, ground speed, vertical rate, ground-track or heading information where available, and timestamps for reconstructing aircraft trajectories. The OpenSky Network provides a representative open infrastructure for large-scale ADS-B research data [94].

Surveillance data commonly require filtering, interpolation, resampling, anomaly handling, and trajectory segmentation to address missing reports, irregular sampling, duplicated messages, and measurement noise [95]. These operations affect the effective sampling interval, input sequence, and reported evaluation results. Radar data are frequently used in terminal area, shorter-horizon, and state estimation studies and may provide comparatively controlled surveillance coverage, although operational radar data are generally less accessible than open ADS-B records [96].

Datasets may cover terminal arrivals or departures, en-route sectors, particular city pairs, specific procedures, or complete flights. Differences in route regularity, flight phase composition, traffic density, aircraft type, geographical region, and observation period affect the scope of model applicability and performance comparisons.

8.2.2. Environmental and Flight Intent Information

Observed aircraft states alone do not uniquely determine future motion. Meteorological and operational information can provide additional constraints and influence both model inputs and reported performance.

Meteorological inputs include wind speed and direction, temperature, pressure, convective activity, and related atmospheric variables. Wind affects the relationship between air-relative motion and ground trajectory, and errors in wind information can lead to substantial along-track or arrival time prediction errors over longer horizons. Such information may come from numerical weather prediction products, gridded meteorological fields, or operational weather databases and may be used in physics-based, statistical, or learning-based predictors [97,98].

Flight intent information includes flight plans, waypoint sequences, Standard Instrument Departures (SIDs), Standard Terminal Arrival Routes (STARs), target flight levels, speed or altitude constraints, and air traffic control instructions. These inputs can constrain trajectory propagation, provide variables for statistical models, or condition learned trajectory representations [20,21,24,99]. Dynamic operational information may also capture

changes not reflected in a filed plan or recent motion; for example, spoken controller instructions have been combined with observed trajectories to condition future prediction [10]. Performance comparisons should therefore account for the type, availability, and timing of contextual information.

8.2.3. Simulated and Synthetic Data

Simulated and synthetic trajectories provide controllable experimental conditions for sensitivity analysis, rare scenario generation, and evaluation of prediction methods under specified operating conditions. Aircraft type, route, traffic demand, weather, maneuver patterns, and operational constraints can be configured to support sensitivity analysis, rare scenario generation, and controlled evaluation of probabilistic or generative models.

Civil air traffic simulation platforms such as BlueSky can generate trajectories and reproduce traffic scenarios [100], while AirTOp-based environments can be used to evaluate air traffic management concepts and trajectory-related procedures [101]. Aircraft performance models can also be combined with flight plans and meteorological information to construct controlled trajectory and uncertainty experiments. Simulated data are controllable and reproducible, but their validity depends on the fidelity of the modeled dynamics, traffic, and environmental processes. Results based on simulated trajectories should therefore be distinguished from evaluations based on operational flights.

Table 10 summarizes representative data sources and contextual information used to interpret the formal performance evidence. Dataset source, preprocessing, contextual inputs, and operational scenario should be interpreted jointly with prediction horizon, output formulation, and evaluation metric. The comparative analysis retains these conditions to distinguish differences associated with the prediction method from differences associated with data and scenario.

8.3. Comparative Performance Analysis

The formal performance synthesis subset contains 55 studies: 9 physics-based, 6 state estimation, 11 classical statistical and machine learning, 24 deep learning, and 5 hybrid studies. Each included study provides an identifiable civil aircraft prediction scenario, an explicit prediction horizon, a defined evaluation metric, and at least one aircraft-specific quantitative result. This is the quantifiable evidence subset of the 198 primary studies, and its category counts do not represent the distribution of the full evidence base or of the field as a whole. Sampling information is retained when reported.

The synthesis is organized by modeling paradigm while preserving the experimental context of individual results. Target cardinality is coded as single-target (S) or joint multi-aircraft (J), consistent with Section 7; interaction-aware prediction does not by itself change this coding. Output formulation is coded as deterministic (D) or probabilistic/generative (P). The tables provide the study-level evidence, while the accompanying text emphasizes category-level patterns and within-study comparisons. Given these differences, the analysis emphasizes within-study comparisons and performance patterns under specific experimental conditions.

8.3.1. Performance of Physics-Based Models

The nine physics-based studies in the formal performance subset use single-target deterministic prediction. Their evaluation settings cover climb- and descent-specific prediction and extended lookahead scenarios. Reported outcomes are dominated by altitude, along-track, position, and time-related errors, while aircraft performance parameters, weather information, and operational intent recur as major experimental factors. The corresponding quantitative evidence is summarized in Table 11.

Table 11. Stratified quantitative performance evidence for physics-based models.

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Alligier et al. [16] PMM with learned mass/thrust	EUROCONTROL/BADA, radar and Mode-S climb data	10 min; 15 s	S	D	Point/position	At 10 min, altitude RMSE was reduced by about 40% relative to the standard BADA setting; reductions approaching 50% were reported against the reduced-power reference.
Yepes et al. [20] Intent inference trajectory predictor	Civil surveillance, flight plan and intent data	14 min reported result; sampling NR	S	D	Point/position	At a 14 min lookahead, the intent-based predictor reduced along-track error by about 20% relative to the comparison predictor.
Warren and Ebrahimi [102] Vertical path trajectory predictor	B737-400 climb/FMS-based ATM scenario	20 min; sampling NR	S	D	Point/position; ETA/time	A 5% thrust profile difference produced about 1 NM along-track divergence, roughly 2000 ft maximum altitude divergence, and about 2.0 min difference in time to top of climb.
Zammit and Zammit-Mangion [103] Idle-descent physics model	A320-231 descent scenario with aircraft performance data	about 2 min (16 NM); sampling NR	S	D	Point/position; ETA/time	Absolute energy error remained below about 100 ft (approximately 2%); maximum time-of-arrival error was about 9 s (approximately 8%).
Bronsvoort et al. [33] Descent predictor with longitudinal intent	237 civil descents from US, European and Australian operations	Descent phase; sampling NR	S	D	Point/position; ETA/time	Reported altitude error IQRs were about 6900 ft (Europe), 3100 ft (US), and 1900 ft (Australia), illustrating the effect of missing longitudinal intent.
Cásek et al. [104] FMS-generated trajectory predictor	FMS-generated trajectories with radar/surveillance comparison	5–15 min; 20 s	S	D	ETA/time; Point/position	For 10–15 min lookahead, average time error was typically below 5 s and the reported confidence interval was below 10 s.
Legrand et al. [34] Wind-networked trajectory predictor	French airspace, approximately 8000 flights with networked wind observations	5–45 min; 10 s	S	D	ETA/time; Point/position	At 30 min, time error decreased from about 20.3 s to 2.2 s after the networked wind update; at 45 min it decreased from about 22.4 s to 2.7 s.
Paglione et al. [105] RTA descent trajectory prediction experiment	FMS/surveillance simulation, 24 RTA descent scenarios	up to 20 min; 30 s	S	D	Point/position; ETA/time	Under altered RTA conditions, meter-fix altitude errors were approximately 500–550 ft on average in the reported descent experiments.
Thipphavong et al. [32] Adaptive climb PMM	Cleveland Center departures with NASA/EUROCONTROL/ BADA inputs	5 min; sampling NR	S	D	Point/position; ETA/time	At 5 min, altitude RMSE was 1545 ft (about 20% lower) and top-of-climb time RMSE was 112 s (about 19% lower).

Notes: T denotes target cardinality (S: single-target); O denotes output formulation (D: deterministic). Interaction-aware single-target prediction remains coded S. Approximate values identified as figure-read were extracted from aircraft-specific figures when no scalar table value was reported.

Within-study comparisons indicate that refining the information used to initialize and constrain trajectory propagation can improve prediction performance. Learning aircraft mass and thrust reduced 10 min altitude RMSE by about 40%, intent inference reduced a reported 14 min along-track error by about 20%, and updated wind information substantially reduced time error at longer horizons [16,20,34]. The evidence therefore links physics-based performance to the fidelity of aircraft performance, meteorological, and operational intent information.

8.3.2. Performance of State Estimation Models

The six state estimation studies include five single-target formulations and one joint multi-aircraft formulation. Two produce deterministic outputs and four use probabilistic or stochastic representations. Evaluated horizons range from seconds to approximately 40 min, with position or state error and uncertainty propagation forming the principal quantitative outcomes. The corresponding quantitative evidence is summarized in Table 12.

Table 12. Stratified quantitative performance evidence for state estimation models.

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Baumgartner and Maeder [43] Low-cost collision-avoidance state estimator	Real glider/aircraft flight dataset (about 600 flights; >1700 h)	0–20 s; sampling NR	S	D	Point/position	From the reported RMS error curves, the proposed predictor is about 100 m at 20 s in circling flight, versus roughly 145–285 m for the legacy predictor depending on wind condition.
Chatterji [38] Short-term KF/hybrid estimator	NASA/NOAA/ radar/GPS surveillance data	1–6 min; 12 s update	S	D	Point/position	Ground-speed RMS error was 15.10 kt, track-angle RMS error 6.00 degrees, and turn prediction RMS error about 1.5–2.5 NM.
Liu and Hwang [44] Probabilistic 4D trajectory predictor	Civil aircraft conflict detection simulation with radar and flight plan information	120 s; 5 s	S	P	Point/position; Safety/conflict	From the aircraft-specific RMS error curve, at about 60 s prediction depth the intent-based method is approximately 550 m versus about 1050 m without intent.
Lymperopoulos and Lygeros [40] Adaptive particle filter	Radar-based aircraft trajectories with NASA/ EUROCONTROL/ BADA inputs	40 min; 30 s	S	P	Point/position; Probabilistic/uncertainty	Along-track uncertainty decreased from about ±3 km to ±0.3 km in climb and from about ±40 km to ±2 km in cruise; vertical error decreased from about ±200 m to ±50 m.
Pan et al. [35] Online stochastic trajectory predictor	OpenSky ADS-B trajectories over the USA with NOAA information	20 min; sampling NR	S	P	Point/position	The reported 20 min evaluation improved horizontal error by 26.3% and vertical error by 31.8% relative to the comparison setting.
Lymperopoulos and Lygeros [41] Structured conditional particle filter	Multi-aircraft simulation using NASA/ EUROCONTROL/ BADA and weather inputs	20 min; 30 s	J	P	Point/position; Safety/conflict	The SCPF reported aircraft state RMS error of about 60 m for the 20 min joint multi-aircraft prediction experiment.

Notes: T denotes target cardinality (S: single-target; J: joint multi-aircraft); O denotes output formulation (D: deterministic; P: probabilistic/generative). Interaction-aware single-target prediction remains coded S. Approximate values identified as figure-read were extracted from aircraft-specific figures when no scalar table value was reported.

The reviewed studies span short-horizon recursive estimation based on repeated surveillance updates and longer-horizon particle filtering with explicit uncertainty propagation. Adaptive filtering substantially narrowed reported along-track uncertainty in both climb and cruise, while the structured conditional particle filter provided a joint multi-aircraft state estimate over a 20 min experiment [40,41]. State estimation performance is therefore characterized by the interaction between update frequency, state uncertainty, and prediction horizon.

8.3.3. Performance of Classical Statistical and Machine Learning Models

The 11 classical statistical and machine learning studies use single-target prediction in the formal performance subset. Nine produce deterministic outputs and two use probabilistic formulations. Their experimental settings include final approach, minute-based prediction horizons, flight phase-specific forecasting, distance-defined arrival prediction, and strategic ATM tasks. The corresponding quantitative evidence is summarized in Table 13.

Table 13. Stratified quantitative performance evidence for classical statistical and machine learning models.

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Gong and Sadovsky [47] Final-approach regression model	NASA/METAR final-approach data	45–120 s; sampling NR	S	D	Point/position; ETA/time	At 120 s lookahead, the mean inter-arrival separation prediction error was about 0.2 NM, corresponding to roughly 33% improvement over dead reckoning.
Ramchandra et al. [106] kNN trajectory regressor	OpenSky DFW event data (2557 flights; 23,314 windows)	30 s; sampling NR	S	D	Point/position	With the selected 18 s input window, the 30 s forecast reported mean position error of 92.41 m.
Christien et al. [107] Supervised conflict-free trajectory predictor	Recorded Maastricht UAC traffic and flight plan data	5 min; sampling NR	S	D	Point/position	Absolute model deviation remained below about 1.3 NM horizontally and about 540 ft vertically, compared with roughly 2760 ft vertical error for the baseline.
de Leege et al. [108] Supervised arrival time predictor	Operational arrival trajectories with wind and surveillance information	45 NM; sampling NR	S	D	ETA/time	At the runway threshold, the full-input model reported MAE 18.3 s and RMSE 23.4 s, corresponding to a 43% MAE reduction.
Song et al. [49] Probabilistic trajectory data mining	North China ATM Bureau radar and flight plan data	10 min; 15 s	S	P	Point/position	Across nine 10 min prediction sets, average prediction accuracy improved by about 7.7%, with a maximum improvement of 10.2% (about 1.6 NM at 10 min).
Tastambekov et al. [46] Local functional regression	EUROCONTROL radar trajectories	10–30 min; 10 s	S	D	Point/position	Relative prediction error ranged from about 2% to 16%, with a final altitude error of 72 ft in the reported experiment.
Ayhan and Samet [45] Predictive analytics for descent	EUROCONTROL/ NOAA radar and flight plan data	Descent phase; 60 s	S	D	Point/position	Mean cross-track error was approximately 12.601 km on the evaluated real-trajectory dataset.
Hernández et al. [50] Ensemble meta-estimators	ADS-B/ EUROCONTROL/ BADA climb data with weather variables	10 min lookahead task; 1 Hz	S	D	ETA/time; Point/position	The selected gradient boosting model reported an estimated-time RMSE of 9.96 s; latitude/longitude RMSE values were 0.0220/0.0367 in the reported coordinate representation.
Petrou et al. [109] Route-pattern mining	Madrid/Barcelona aviation trajectories	1, 2, 4, 8, 16, 32 min; 10 s	S	D	Point/position	From the reported aviation error boxplots, mean horizontal error rises from about 3 km at 1 min to about 35 km at 32 min.
Rocha and De [110] Probabilistic terminal airspace trajectory model	ADS-B/Flightradar24 terminal arrival trajectories	60 NM arrival horizon; sampling NR	S	P	ETA/time; Point/position	Landing time MAE decreased from 82.3 s to 66.4 s, while horizontal error decreased from 4.1 NM to 3.3 NM at 60 NM.
Arnaldo et al. [111] Strategic AI regression	Historical Madrid ACC strategic ATM data	Strategic/sector transit time; sampling NR	S	D	ETA/time	For time-to-sector prediction, MAE was 3.24, RMSE 5.80, relative error 0.0446, and $R^2 = 0.9956$.

Notes: T denotes target cardinality (S: single-target); O denotes output formulation (D: deterministic; P: probabilistic/generative). Interaction-aware single-target prediction remains coded S. Approximate values identified as figure-read were extracted from aircraft-specific figures when no scalar table value was reported.

Within-study comparisons indicate improvements when trajectory, weather, route, or procedural information is incorporated, including gains in altitude, position, arrival time, and terminal airspace prediction. The reported outcomes are concentrated on position

and time variables, while multimodal output and joint multi-aircraft prediction remain sparsely represented. This category therefore provides evidence for feature-informed single-target prediction across several operational settings, with limited coverage of richer output structures.

8.3.4. Performance of Deep Learning Models

Deep learning constitutes the largest category in the formal performance subset, with 24 studies. Under the output-cardinality definition adopted in Section 7, 21 use single-target output and three provide confirmed joint multi-aircraft prediction. Sixteen produce deterministic outputs and eight use probabilistic or generative formulations. The studies cover a broad range of prediction horizons, interaction settings, and evaluation measures. The corresponding quantitative evidence is summarized in Table 14.

Table 14. Stratified quantitative performance evidence for deep learning models.

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Guo et al. [55] FlightBERT	Real Chinese ATC trajectories	20 s/1/3/5 min; 20 s	S	D	Point/position	Longitude/latitude MAE was 0.0039/0.0033 degrees at one step and 0.0208/0.0187 degrees at 15 steps (5 min).
Guo et al. [56] FlightBERT++ non-autoregressive predictor	Historical civil flight trajectories	20 s/1/3/5 min; 20 s	S	D	Point/position	At the 5 min horizon, FlightBERT++ reported mean deviation error of 1.97 NM.
Guo et al. [112] Wavelet multi-horizon predictor (WTFTP+)	Civil trajectory data used for multi-horizon FlightBERT-family evaluation	20 s/1/3/5 min; 20 s	S	D	Point/position	At the 5 min horizon, WTFTP+ reported mean deviation error of 1.06 NM.
Guo et al. [10] Controller-instruction-aware predictor	Civil trajectories with spoken controller instructions	1–10 min (5 min reported result); 20 s	S	D	Point/position	At 5 min, the instruction-aware model reduced mean deviation error by 20.25% relative to the LSTM+Attention baseline.
Ma et al. [113] CNN-BiGRU terminal area predictor	Real ADS-B terminal area trajectories	1 step; sampling NR	S	D	Point/position	CNN-BiGRU reported RMSE 1.004679 and MAPE 0.028814, compared with GRU RMSE 2.889301 and LSTM RMSE 4.165763.
Zeng et al. [54] Seq2Seq terminal area predictor	Guangzhou terminal area ADS-B trajectories	90 s departure/150 s arrival; sampling NR	S	D	Point/position	Reported endpoint errors were 206.1 m for departure and 330.1 m for arrival; average errors were 33.6 m and 45.3 m, respectively.
Lu et al. [114] Seq2Seq-Attention terminal predictor	Terminal area ADS-B trajectories	1/3/5/10 min; sampling NR	S	D	Point/position	At 10 min, MAE reductions for altitude/longitude/latitude were 66.30%/54.62%/36.59%; RMSE reductions were 65.45%/38.16%/20.57%.
Tran et al. [21] Encoder-decoder with enriched intent	ADS-B trajectories with flight plan/ intent information	1–10 min; sampling NR	S	D	Point/position	At 10 min, enriched intent RMSE was about 2.6 NM versus about 4.9 NM for ordinary intent; 50% of cross-track errors were below 1 NM.
Kuang et al. [60] MTDA-STGCN	Multi-aircraft ADS-B trajectories with flight plan information	60/90/120 s; 10 s	S	P	ADE/FDE; Probabilis-tic/uncertainty	At 60 s, ADE decreased by about 13% and FDE by about 16% relative to Social-STGCNN.
Yoon and Lee [58] Inverted Transformer	Incheon arrival ADS-B trajectories	120 s; 6 s	S	D	Point/position	Compared with the standard Transformer, average MAE reductions were 7.7%, 4.9%, and 3.1% in the east, north, and up components, respectively.
Yoon and Lee [115] Multi-Agent Inverted Transformer (MAIFormer)	Incheon terminal area ADS-B trajectories	2 min (20 steps); 6 s	J	D	Point/position	MAIFormer reduced MAPE by 16.67%, 3.66%, and 22.08% for latitude, longitude, and altitude, respectively.

Table 14. *Cont.*

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Yang et al. [69] GooDFlight goal-oriented diffusion	TrajAir, Pittsburgh-Butler non-towered terminal airspace	120 s; sampling NR	S	P	ADE/FDE; Probabilistic/ uncertainty	GooDFlight reported ADE/FDE/GLeV of approximately 0.29/0.39/0.08 km for the 120 s prediction task.
Yin et al. [11] Context-aware diffusion	Singapore Changi arrival ADS-B data	120 s; 5 s	S	P	ADE/FDE; Probabilistic/ uncertainty	Best-of-5 3D ADE/FDE was approximately 0.528/1.003 km for the 120 s prediction task.
Chu et al. [63] Flight pattern-guided Social LSTM	Real terminal area multi-aircraft trajectories	up to 6 min; 4 s	J	D	Point/position	For the 6 min horizon, horizontal prediction accuracy improved by 14.5% and height prediction accuracy by 29.5% on average.
Zhang et al. [116] SWIFT dual cross-attention	OpenSky Europe and US multi-aircraft datasets	20 s/1/3/5 min; 10 s	S	D	Point/position	At 5 min, reported composite MAE/RMSE/SMAPE was 0.1178/0.2476/25.21% for Europe and 0.1174/0.2181/25.87% for the US.
Wang et al. [117] IDEA-FastTEN real-time predictor	Real ADS-B trajectory dataset (RTFT)	Phase-specific/descent; 20 s	S	D	Point/position; Engineering	Reported trajectory distance error was below 1 km, with inference throughput close to 60 FPS.
Zhang and Liu [118] Improved CTGAN + LSTM under severe weather	Civil 4D trajectories with severe-weather augmentation	5 min; sampling NR	S	D	Point/position	Increasing the minority severe-weather data share from 4.7% to 13% reduced average RMSE by 20.46%; longitude RMSE decreased by about 35.93%.
Zhu et al. [119] Boosted spatiotemporal convective weather ensemble	Sherlock/ADS-B cruise trajectories with convective weather information	500 km/ 1000 km/full route; sampling NR	S	D	Point/position	At 1000 km on the full test set, MAE was 6.96 km and MAPE 11.0%; under convective weather conditions, MAPE at 1000 km was about 14.1%.
Acharya et al. [120] VAE generative predictor	Towered and non-towered airport trajectories	1/5/10 steps; 4 s/12 s	S	P	Point/position	VAE total MAE for towered data was 0.1091/0.3409/0.5714 at 1/5/10 steps; corresponding non-towered values were 0.1546/0.4014/0.7239.
Zhang and Mahadevan [121] Bayesian neural network	FAA SWIM/SFDPS commercial-flight data	60 s (5 future steps); 12 s	S	P	Point/position; Safety/conflict	The integrated Bayesian model reported RMSE reductions of 49% and 58% on the two evaluated response variables relative to the comparison models.
Wu et al. [67] Conv1D-GAN long-term 4D generator	Historical 4D civil trajectories with BADA information	Full 4D trajectory; 4 s	S	P	Probabilistic/ uncertainty; Point/position	Conv1D-GAN reported FID 0.022, MMD 0.200, and MED 0.0006; generation of 1825 trajectories took about 0.192 s.
He et al. [122] FlightDiff diffusion predictor	Civil flight-trajectory and arrival time datasets	10 min FTP; 15–60 min arrival time task; 20 s for FTP	S	P	Point/position; ETA/time	For 10 min trajectory prediction, FlightDiff reported MDE 7.240 km; for the 20–30 min arrival time task, MAE was 3.321 min versus 5.242 min for FlightBERT++.
Ye and Gombolay [123] Conditional flow matching (T-CFM)	OpenSky data, single Cessna, 474 trajectories	up to 30 min; sampling NR	S	P	Point/position; Probabilis-tic/uncertainty	At 30 min, T-CFM longitude/latitude/altitude MAE was 0.130 degrees/0.075 degrees/782.6 m; the aircraft experiment reported 35.4% average improvement.
Zhu et al. [124] DST-GAT joint multi-aircraft predictor	Joint multi-aircraft approach trajectories with flight plan information	5 and 10 min reported; 30 s	J	D	Point/position	At 5 min, mean/final 3D errors were about 3.95/5.09 km; at 10 min they were about 5.37/6.98 km.

Notes: T denotes target cardinality (S: single-target; J: joint multi-aircraft); O denotes output formulation (D: deterministic; P: probabilistic/generative). Interaction-aware single-target prediction remains coded S. Approximate values identified as figure-read were extracted from aircraft-specific figures when no scalar table value was reported.

Multi-horizon evaluations indicate that prediction error can increase with the lookahead interval, as illustrated by the FlightBERT results [55]. Additional operational context can improve horizon-specific predictions. For example, incorporating spoken controller instructions reduced the reported 5 min mean deviation error by 20.25% relative to an

LSTM-attention baseline [10]. Interaction-aware and joint output studies provide a second evidence stream: MTDA-STGCN reported reductions of approximately 13% in ADE and 16% in FDE at 60 s, while a flight pattern-guided Social LSTM reported average horizontal and height improvements of 14.5% and 29.5%, respectively, at 6 min [60,63]. These results describe study-specific contributions of operational context or coupled output and should not be interpreted as a common architecture ranking.

Probabilistic and generative studies extend the output beyond a single deterministic path and report ADE/FDE, sample-based, distributional, or task-specific measures over horizons ranging from seconds to approximately 30 min. The deep learning studies cover diverse deterministic and probabilistic output structures, interaction settings, and confirmed joint multi-aircraft prediction, although their heterogeneous evaluation conditions limit direct numerical comparison.

8.3.5. Performance of Hybrid Models

Five studies are classified as hybrid because their prediction chains integrate distinct mechanisms across modeling paradigms, such as learned trajectory representations coupled with recursive estimation, data-driven components combined with physical dynamics, or explicit operational intent integrated with probabilistic state propagation. Under the target-cardinality definition used in this review, all five retain single-target trajectory output, although one study subsequently propagates predicted trajectory uncertainty into a two-aircraft conflict analysis setting. Two studies produce deterministic outputs and three use probabilistic formulations. The corresponding quantitative evidence is summarized in Table 15.

Table 15. Stratified quantitative performance evidence for hybrid models.

Study/Method	Dataset/Scenario	Horizon/Sampling	T	O	Metric Family	Aircraft-Specific Quantitative Result
Choi et al. [74] Machine learning + residual mean IMM	Incheon International Airport arrival surveillance data	120 s; 5 s	S	D	Point/position	Relative to the GMM comparison, horizontal/along-track/cross-track/vertical errors were reduced by 46.8%/48.6%/24.2%/1.2%, respectively.
Jarry et al. [73] Neural ODE flight dynamics model	Operational A320 QAR/BADA descent data	Descent phase; 4 Hz QAR	S	D	Point/position	True airspeed MAE over the evaluated phases was about 1.35 m/s versus 2.73 m/s for BADA; descent altitude MAE was about 182 m versus 560 m for BADA.
Liu et al. [75] LSTM-IMM	Historical climb trajectories	Climb phase; sampling NR	S	P	Point/position; ETA/time	Test RMSE was 0.01041 degrees longitude, 0.01032 degrees latitude, 61.91 m altitude, and 4.54 s time.
Hwang and Seah [79] Intent-based probabilistic trajectory prediction and conflict detection	Civil two-aircraft conflict scenario with flight plan and surveillance information	20 min; 10 s prediction steps	S	P	Safety/conflict; Engineering	The intent-based method identified the future conflict about 4.5 min before occurrence versus about 45 s for the non-intent method; 20 min trajectory/conflict computation took about 0.26 s for the two-aircraft case.
Mueller et al. [125] Physics-data strategic uncertainty model	Civil DFW–EWR MD-80 trajectory simulation	1 h reported trajectory uncertainty; sampling NR	S	P	Point/position; Probabilistic/uncertainty	From the one-hour uncertainty profile, the 1σ along-path/range error grows to about 5 NM at 60 min; climb altitude uncertainty peaks at roughly 1700–1800 ft.

Notes: T denotes target cardinality (S: single-target); O denotes output formulation (D: deterministic; P: probabilistic/generative). Downstream comparison or conflict analysis of separately predicted aircraft does not by itself constitute joint multi-aircraft trajectory prediction. Approximate values identified as figure-read were extracted from aircraft-specific figures when no scalar table value was reported.

The reported gains arise from different forms of cross-paradigm integration. Within-study comparisons indicate improvements from combining learned trajectory information with recursive estimation, learned components with aircraft dynamics, or intent and state-mode estimation with probabilistic propagation [73,74,79]. The five studies cover different

horizons, data sources, output forms, and task-specific outcomes. The available evidence therefore supports the benefits of hybrid coupling under the specific experimental conditions evaluated in these studies.

Table 16 summarizes the evidence profile of the five modeling categories. Physics-based and classical statistical and machine learning studies are dominated by deterministic single-target evaluation, whereas probabilistic output is more common in state estimation, deep learning, and hybrid studies. Confirmed joint multi-aircraft quantitative evidence remains limited and is concentrated in state estimation and deep learning. These distributions describe the formal performance subset and provide the basis for the cross-paradigm findings in the next section.

Table 16. Evidence profile of modeling categories in the formal performance synthesis subset ($n = 55$).

Modeling Category	*n*	Target S/J	Output D/P	Horizon Coverage	Main Metric Families	Evidence Profile
Physics-based	9	9/0	9/0	Flight phase-specific and approximately 2–45 min lookahead settings	Point/position, altitude, along-track, ETA/time	Deterministic structured prediction; aircraft performance, weather, and intent fidelity recur as performance factors.
State estimation	6	5/1	2/4	Seconds to about 40 min	Point/position, uncertainty, safety/conflict	Recursive prediction with explicit stochastic or probabilistic uncertainty propagation in four studies.
Classical statistical and machine learning	11	11/0	9/2	Seconds to strategic or route-based settings	Point/position, ETA/time, probabilistic/uncertainty	Regression, statistical learning, trajectory pattern modeling, and probabilistic terminal prediction.
Deep learning	24	21/3	16/8	Shorter- to longer-horizon and full-trajectory settings	Point/position, ADE/FDE, probabilistic/uncertainty, engineering	Broad coverage of horizons, interaction-aware modeling, joint prediction, and generative outputs.
Hybrid	5	5/0	2/3	Shorter-horizon, phase-specific, and longer-horizon settings	Point/position, ETA/time, uncertainty, safety/conflict	Heterogeneous cross-paradigm coupling of learned or data-derived, physical, state estimation, intent/prior, and uncertainty-propagation components.
Total	**55**	**51/4**	**38/17**	–	–	–

Note. S denotes single-target prediction and J denotes joint multi-aircraft prediction. D denotes deterministic output and P denotes probabilistic or generative output. Interaction-aware prediction is represented separately from target cardinality and therefore does not automatically imply J.

8.4. Cross-Cutting Performance Findings

The paradigm-specific tables reveal several performance patterns that cut across modeling categories. These patterns concern prediction horizon, target cardinality and interaction context, output formulation, and engineering evidence. They provide a more appropriate basis for cross-study synthesis than direct comparison of numerical errors obtained under heterogeneous experimental conditions.

8.4.1. Horizon-Dependent Performance

Several within-study comparisons show increasing prediction errors as the lookahead interval extends. FlightBERT reports longitude/latitude MAE increasing from 0.0039/0.0033 degrees at one prediction step to 0.0208/0.0187 degrees at 5 min [55]; route-pattern prediction reports mean horizontal error increasing from approximately 3 km at 1 min to approximately 35 km at 32 min [109]; and joint multi-aircraft DST-GAT errors increase from approximately 3.95/5.09 km mean/final 3D error at 5 min to approximately 5.37/6.98 km at 10 min [124].

The magnitude of degradation depends on the flight phase, sampling interval, contextual information, model structure, and metric definition. Longer-horizon performance is particularly sensitive to wind, operational intent, aircraft performance parameters, and controller information [10,21,34]. Prediction horizon should therefore be interpreted as an experimental condition that interacts with the available information and model structure. Horizon-dependent performance is most informative when evaluated under comparable experimental conditions.

8.4.2. Target Cardinality and Interaction Context

The formal performance subset is dominated by single-target prediction: 51 of 55 studies (92.7%) use single-target output, whereas four (7.3%) provide confirmed joint multi-aircraft prediction. The joint output evidence consists of one state estimation study and three deep learning studies; no confirmed joint output study occurs in the physics-based, classical statistical and machine learning, or hybrid categories.

The four joint studies provide limited evidence for a global performance comparison with single-target prediction. The reviewed literature nevertheless reveals two distinct prediction settings. Coupled state estimation and neural joint output formulations represent the future trajectories of several aircraft within one prediction problem [41,63,115,124]. Interaction-aware methods such as MTDA-STGCN and context-aware diffusion incorporate surrounding aircraft information while retaining focal or per-aircraft output [11,60]. Reported improvements associated with graph, attention, social pooling, and related interaction mechanisms illustrate the contributions of traffic context under specific experimental conditions.

8.4.3. Deterministic and Probabilistic Output

Across the 55-study formal subset, 38 studies produce deterministic outputs and 17 use probabilistic or generative formulations. Probabilistic or generative output is more common in state estimation, deep learning, and hybrid studies, whereas all physics-based studies in the subset use deterministic output. The reported probabilistic representations include state covariances, prediction intervals, uncertainty envelopes, sampled trajectories, and multimodal candidate sets.

The evaluation of these outputs remains heterogeneous. Best-of-K metrics, such as minADE and minFDE, quantify the minimum corresponding trajectory errors among K generated candidates, while coverage, NLL, CRPS, and related measures evaluate complementary aspects of probabilistic predictive performance. Sample diversity and distributional similarity do not by themselves establish calibration. Consequently, probabilistic performance should be interpreted jointly with the output representation and the metric used to evaluate accuracy, uncertainty, and calibration.

8.4.4. Engineering Evidence

Computational performance is reported much less consistently than trajectory accuracy. Only a limited portion of the formal subset reports inference time, throughput, generation speed, or related implementation properties. Representative results include inference throughput close to 60 FPS for IDEA-FastTEN [117], generation of 1825 trajectories in approximately 0.192 s for Conv1D-GAN [67], and approximately 0.26 s for a 20 min two-aircraft trajectory and conflict computation [79]. These values depend on hardware, software implementation, input size, output horizon, batch configuration, sampling procedure, and the number of aircraft or hypotheses processed. The current evidence is therefore stronger for trajectory accuracy than for computational efficiency and does not support a cross-paradigm efficiency ranking.

The formal performance evidence shows that prediction error varies with horizon and scenario, joint multi-aircraft evidence remains limited, probabilistic outputs require evaluation of both accuracy and uncertainty quality, and engineering results are reported less consistently than trajectory accuracy. These findings highlight the importance of task- and condition-specific performance evaluation and inform the synthesis presented in Section 9.

9. Synthesis and Future Directions

The preceding sections examine aircraft trajectory prediction from system-level, methodological, task-formulation, and performance evaluation perspectives. This section discusses the implications of these findings for method selection and experimental design, identifies limitations and gaps in the available evidence, and outlines priorities for future research.

9.1. Implications of the Review

The PPUO framework provides a system-level organization for trajectory prediction by distinguishing preparation of inputs and operational information, inference of future states or trajectories, online revision as new information arrives, and delivery of results for downstream use. This organization helps identify the functional problem addressed by each study and relate predictor design to air traffic management applications, without prescribing a particular software architecture.

Physics-based, state estimation, data-driven, and hybrid methods have different input requirements, modeling assumptions, and operating conditions. Method selection should therefore account for the prediction horizon, available information, operating scenario, output formulation, and computational constraints. The evidence supports task-specific selection, but does not establish a single method as superior across all scenarios.

Prediction horizon, target cardinality, and interaction context influence task formulation. Short-horizon prediction generally relies more strongly on recent states and motion history, while longer-horizon prediction often places greater importance on flight intent, route and procedure information, meteorological conditions, and aircraft performance information. Interaction awareness describes the use of neighboring aircraft information and can be incorporated into either single-target or joint multi-aircraft prediction. Target cardinality specifies the number of aircraft represented in the predicted output.

Performance results should be interpreted in relation to experimental conditions and output formulation. The formal synthesis covers 55 of the 198 primary evidence studies, and these studies differ in data sources, operating scenarios, flight phases, sampling intervals, prediction horizons, coordinate representations, contextual inputs, and evaluation metrics. Within-study controlled comparisons are therefore more informative for assessing a modeling choice or information source, whereas cross-study synthesis is more useful for identifying recurring patterns and evaluation gaps. Trajectory accuracy, uncertainty quality, computational efficiency, and operational assurance should be treated as complementary evaluation dimensions.

9.2. Evidence Limitations and Research Gaps

Studies differ substantially in trajectory data filtering, segmentation, resampling, observation windows, train–test partitions, prediction horizons, coordinate representations, and metric definitions. Even metrics with the same name may therefore correspond to different experimental tasks. Public data improve accessibility, but do not by themselves ensure comparability. Complete reporting of experimental conditions and evaluation protocols remains necessary for reproducibility and evidence integration.

Many predictors are developed and evaluated within restricted airports, routes, flight phases, aircraft populations, weather regimes, or observation periods. Such results do not establish robustness under distributional change. Further evidence is needed to distinguish transferable aircraft motion structure from features that primarily encode local operational regularities and to quantify the contribution of each contextual information source across scenarios.

Uncertainty evaluation and multimodal prediction also lack a common evidence base. Studies use covariance estimates, prediction intervals, sampled trajectories, or candidate sets, but report likelihood, coverage, distributional similarity, and calibration inconsistently. Displacement error, sample diversity, and probabilistic reliability describe different properties and cannot be represented adequately by one metric. Confirmed joint multi-aircraft quantitative evidence remains limited, and simultaneous output for several aircraft does not by itself represent correlated uncertainty among their futures.

Engineering and operational evidence is similarly incomplete. Inference latency, throughput, missing-input handling, abnormal-state behavior, failure modes, traceability, and the operational effects of erroneous predictions are not reported consistently. These gaps limit judgments about real-time feasibility, computational trade-offs, and operational risk.

9.3. Future Research Directions

Reproducible and context-aware benchmarking. Future benchmarking should report the observation window, prediction horizon, sampling interval, coordinate representation, preprocessing and segmentation procedures, data split, flight phase, available context, target cardinality, output formulation, and metric definitions together with quantitative results. Benchmark suites spanning multiple airports, routes, flight phases, meteorological conditions, and traffic density regimes would support both controlled within-scenario comparison and explicit evaluation of cross-scenario transfer.

Generalizable prediction and explicit context evaluation. Future models should separately assess the contributions of recent motion, flight intent, procedures, meteorological information, controller instructions, aircraft performance variables, and surrounding traffic. These contributions should be tested across airports, routes, weather conditions, and operational regimes to determine which contextual representations transfer and which require scenario-specific adaptation or updating [10,11,21].

Calibrated probabilistic and dependency-aware prediction. Probabilistic prediction should evaluate uncertainty quality together with trajectory accuracy. Evaluations of single-target multimodal prediction should combine displacement-error measures with calibration, coverage, likelihood-based, or other distribution-sensitive scores. Joint multi-aircraft prediction should distinguish per-aircraft uncertainty from cross-aircraft dependence and evaluate the latter when a joint probabilistic structure is explicitly represented [11,63,69,84].

Physics-guided and adaptively updated prediction. Hybrid methods can couple data-driven components with physical dynamics, state estimation mechanisms, or explicit intent-based priors and constraints. Future studies should evaluate geometric accuracy together with physical and procedural consistency, horizon-dependent behavior, robustness to changing inputs, and the ability to revise predictions when model assumptions no longer hold [73,74].

Efficiency and operational assurance. Studies intended for online or operational use should report computational conditions, inference latency, throughput, input sequence length, prediction horizon, target cardinality, and sampling or generation procedures. Robustness experiments should examine missing observations, degraded surveillance quality, changing weather or intent, and other departures from nominal conditions. Uncertainty

communication, failure behavior, traceability, and the effects of erroneous predictions should be evaluated alongside geometric accuracy.

9.4. Review Limitations

Several limitations should be considered when interpreting the synthesized evidence. First, the literature search was limited to Scopus, Web of Science Core Collection, Engineering Village–Compendex, and IEEE Xplore on 29–30 July 2026; relevant studies outside these sources or published subsequently may have been missed.

Second, the initial study screening and data extraction were conducted by one reviewer following predefined eligibility criteria. During manuscript revision, a second author reviewed all 55 studies included in the formal performance synthesis for eligibility, methodological quality, and evidence confidence, while a third author independently rechecked the exclusion decisions for the 226 studies excluded during full-text assessment and the 19 studies excluded during the final evidence assessment. Nevertheless, the assessment of methodological quality and confidence in the evidence was conducted using a structured qualitative approach, without applying a standardized risk-of-bias tool or a formal certainty-of-evidence grading framework.

Finally, the 55 performance synthesis studies represented a subset of the 198 primary studies. Differences in datasets, prediction tasks, output formulations, and evaluation metrics constrained direct numerical comparisons and limited the feasibility of pooled statistical synthesis.

9.5. Concluding Remarks

Regarding RQ1, the PPUO framework organizes aircraft trajectory prediction into preparation, prediction, update, and output stages, providing a system-level view of the relationship between observations, operational context, prediction mechanisms, and downstream use.

Regarding RQ2, physics-based, state estimation, data-driven, and hybrid methods have different input requirements, modeling assumptions, and operating conditions. The evidence supports selecting a method according to the prediction task and operating conditions, but does not establish a universal ordering of these approaches.

Regarding RQ3, prediction horizon, target cardinality, and interaction context jointly affect task formulation, input information, output structure, and method selection. Short-horizon prediction relies more strongly on recent motion, while longer-horizon prediction often places greater importance on intent, procedures, meteorology, and aircraft performance information. Interaction awareness describes the use of surrounding aircraft information, while joint multi-aircraft prediction specifies a coupled multi-target output.

Regarding RQ4, the reviewed studies employ trajectory error, probabilistic, calibration, and engineering metrics, although experimental conditions and reporting practices remain heterogeneous. In the 55-study formal performance subset, 51 studies use single-target prediction and four use joint multi-aircraft prediction; 38 report deterministic outputs and 17 use probabilistic or generative formulations. Within-study comparisons reveal horizon-dependent error patterns, while uncertainty calibration and engineering performance remain inconsistently reported. These findings emphasize the importance of task- and condition-specific performance evaluation.

Overall, this review establishes an evidence framework spanning system-level organization, modeling methodology, task dimensions, and performance evaluation. Future work should improve experimental reproducibility and cross-scenario validation while advancing uncertainty calibration, inter-aircraft dependency modeling, computational reporting, and operational assurance. These priorities provide a basis for trajectory prediction systems

that are accurate, transferable, uncertainty-aware, computationally feasible, and suitable for civil air traffic management applications.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/aerospace13111016/s1: Supplementary File S1: Database Specific Search Strategies and Search Log; PRISMA 2020 Checklist, based on Page et al. [14].

Author Contributions: Conceptualization, C.Z. and N.W.; methodology, C.Z., B.Z. and D.L.; investigation, C.Z., B.Z., C.Y. and H.B.; data curation, C.Z., C.Y., H.B. and Y.C.; formal analysis, C.Z., Y.C. and L.W.; validation, B.Z., D.L. and N.W.; visualization, C.Z. and Y.C.; writing—original draft preparation, C.Z.; writing—review and editing, B.Z., C.Y., H.B., Y.C., L.W., D.L. and N.W.; supervision, N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Aerospace Information Research Institute, Chinese Academy of Sciences, under the Special Research Assistant (Talent Program), Division II (Execution), Project No. E23307010D.

Data Availability Statement: No new experimental or participant data were generated in this systematic review. The evidence synthesized in this article is drawn from the published literature cited in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Civil Aviation Organization. *Global Air Navigation Plan*, 5th ed.; Technical Report Doc 9750; International Civil Aviation Organization: Montreal, QC, Canada, 2016.
2. Enea, G.; Porretta, M. A comparison of 4D-trajectory operations envisioned for NextGen and SESAR, some preliminary findings. In Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences (ICAS 2012), Brisbane, Australia, 23–28 September 2012; Volume 5, pp. 4152–4165, Paper ICAS 2012-10.3.1.
3. Post, J. The next generation air transportation system of the United States: Vision, accomplishments, and future directions. *Engineering* **2021**, *7*, 427–430. https://doi.org/10.1016/j.eng.2020.05.026.
4. Bolić, T.; Ravenhill, P. SESAR: The past, present, and future of European air traffic management research. *Engineering* **2021**, *7*, 448–451. https://doi.org/10.1016/j.eng.2020.08.023.
5. SESAR Joint Undertaking. *SESAR 2020 Concept of Operations*; Technical Report, SESAR Joint Undertaking, 2017; PJ19 Content Integration; SESAR: Brussels, Belgium, 2020.
6. Federal Aviation Administration, Air Traffic Organization, NextGen and Operations Planning. *NextGen Mid-Term Concept of Operations for the National Airspace System*; Initial Coordination Draft, Version 1.0; 30 June 2009; Federal Aviation Administration: Washington, DC, USA, 2009.
7. Mondoloni, S.; Rozen, N. Aircraft trajectory prediction and synchronization for air traffic management applications. *Prog. Aerosp. Sci.* **2020**, *119*, 100640. https://doi.org/10.1016/j.paerosci.2020.100640.
8. Zeng, W.; Chu, X.; Xu, Z.; Liu, Y.; Quan, Z. Aircraft 4D trajectory prediction in civil aviation: A review. *Aerospace* **2022**, *9*, 91. https://doi.org/10.3390/aerospace9020091.
9. Sankararaman, S.; Daigle, M. Uncertainty quantification in trajectory prediction for aircraft operations. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Grapevine, TX, USA, 9–13 January 2017; AIAA 2017-1724. https://doi.org/10.2514/6.2017-1724.
10. Guo, D.; Zhang, Z.; Yang, B.; Zhang, J.; Yang, H.; Lin, Y. Integrating spoken instructions into flight trajectory prediction to optimize automation in air traffic control. *Nat. Commun.* **2024**, *15*, 9662. https://doi.org/10.1038/s41467-024-54069-5.
11. Yin, Y.; Zhang, S.; Zhang, Y.; Zhang, Y.; Xiang, S. Context-aware aircraft trajectory prediction with diffusion models. In *Proceedings of the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 24–28 September 2023*; IEEE: New York, NY, USA, 2023; pp. 5312–5317. https://doi.org/10.1109/itsc57777.2023.10422124.
12. Xie, X.; Tian, Y.; Li, J.; Wan, L. Shifting from physical models to learning methods for trajectory prediction: A review. *Expert Syst. Appl.* **2026**, *297*, 129303. https://doi.org/10.1016/j.eswa.2025.129303.
13. Kwak, N.; Lee, B. The Evolution and Taxonomy of Deep Learning Models for Aircraft Trajectory Prediction: A Review of Performance and Future Directions. *Appl. Sci.* **2025**, *15*, 10739. https://doi.org/10.3390/app151910739.

14. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. https://doi.org/10.1136/bmj.n71.
15. Mondoloni, S.; Paglione, M.; Green, S. Trajectory modeling accuracy for air traffic management decision support tools. In Proceedings of the 23rd Congress of the International Council of the Aeronautical Sciences (ICAS 2002), Toronto, ON, Canada, 8–13 September 2002; Paper ICAS 2002-R45.
16. Alligier, R.; Gianazza, D.; Durand, N. Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights. *Transp. Res. Part C Emerg. Technol.* **2013**, *36*, 45–60. https://doi.org/10.1016/j.trc.2013.08.006.
17. Rohani, A.S.; Puranik, T.G.; Kalyanam, K.M. Machine learning approach for aircraft performance model parameter estimation for trajectory prediction applications. In *Proceedings of the 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), Barcelona, Spain, 1–5 October 2023*; IEEE: New York, NY, USA, 2023; pp. 1–9. https://doi.org/10.1109/dasc58513.2023.10311271.
18. Mondoloni, S.; Swierstra, S. Commonality in disparate trajectory predictors for air traffic management applications. In *Proceedings of the 24th Digital Avionics Systems Conference, Washington, DC, USA, 30 October–3 November 2005*; IEEE: New York, NY, USA, 2005; Volume 1, pp. 3.C.2-1–3.C.2-10. https://doi.org/10.1109/dasc.2005.1563348.
19. Carreño, V.; Muñoz, C. Implicit intent information for conflict detection and alerting. In *Proceedings of the 23rd Digital Avionics Systems Conference, Salt Lake City, UT, USA, 24–28 October 2004*; (IEEE Cat. No. 04CH37576); IEEE: New York, NY, USA, 2004; Volume 2. https://doi.org/10.1109/dasc.2004.1390765.
20. Yepes, J.L.; Hwang, I.; Rotea, M. New algorithms for aircraft intent inference and trajectory prediction. *J. Guid. Control. Dyn.* **2007**, *30*, 370–382. https://doi.org/10.2514/1.26750.
21. Tran, P.N.; Nguyen, H.Q.; Pham, D.T.; Alam, S. Aircraft trajectory prediction with enriched intent using encoder-decoder architecture. *IEEE Access* **2022**, *10*, 17881–17896. https://doi.org/10.1109/access.2022.3149231.
22. Mondoloni, S.; Bayraktutar, I. Impact of factors, conditions and metrics on trajectory prediction accuracy. In Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD, USA, 27–30 June 2005.
23. Rudnyk, J.; Ellerbroek, J.; Hoekstra, J.M. Trajectory prediction sensitivity analysis using Monte Carlo simulations based on inputs' distributions. *J. Air Transp.* **2019**, *27*, 181–198. https://doi.org/10.2514/1.d0156.
24. Zhang, J.; Liu, J.; Hu, R.; Zhu, H. Online four dimensional trajectory prediction method based on aircraft intent updating. *Aerosp. Sci. Technol.* **2018**, *77*, 774–787. https://doi.org/10.1016/j.ast.2018.03.037.
25. Sotiriou, D.; Kopsaftopoulos, F.; Fassois, S. An adaptive time-series probabilistic framework for 4-D trajectory conformance monitoring. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1606–1616. https://doi.org/10.1109/tits.2015.2511024.
26. Reynolds, T.G.; Hansman, R.J. Investigating conformance monitoring issues in air traffic control using fault detection techniques. *J. Aircr.* **2005**, *42*, 1307–1317. https://doi.org/10.2514/1.10055.
27. Zheng, Q.M.; Zhao, Y.J. Probabilistic approach to trajectory conformance monitoring. *J. Guid. Control. Dyn.* **2012**, *35*, 1888–1898. https://doi.org/10.2514/1.53951.
28. Fukuda, Y.; Shirakawa, M.; Senoguchi, A. Development and evaluation of trajectory prediction model. In Proceedings of the 27th International Congress of the Aeronautical Sciences, Nice, France, 19–24 September 2010; pp. 1–8.
29. Schuster, W.; Porretta, M. High-performance trajectory prediction for civil aircraft. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Salt Lake City, UT, USA, 3–7 October 2010*; IEEE: New York, NY, USA, 2010; IEEE Document 5655515. https://doi.org/10.1109/dasc.2010.5655515.
30. Gallo, E.; Navarro, F.A.; Nuic, A.; Iagaru, M. Advanced aircraft performance modeling for ATM: BADA 4.0 results. In *Proceedings of the 2006 IEEE/AIAA 25th Digital Avionics Systems Conference, Portland, OR, USA, 15–18 October 2006*; IEEE: New York, NY, USA, 2006; pp. 1–12. https://doi.org/10.1109/dasc.2006.313660.
31. Gallo, E.; López-Leonés, J.; Vilaplana, M.; Navarro, F.; Nuic, A. Trajectory computation infrastructure based on BADA aircraft performance model. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Dallas, TX, USA, 21–25 October 2007*; IEEE: New York, NY, USA, 2007; IEEE Document 4391830; https://doi.org/10.1109/dasc.2007.4391830.
32. Thipphavong, D.P.; Schultz, C.A.; Lee, A.G.; Chan, S.H. Adaptive algorithm to improve trajectory prediction accuracy of climbing aircraft. *J. Guid. Control. Dyn.* **2013**, *36*, 15–24. https://doi.org/10.2514/1.58508.
33. Bronsvoort, J.; McDonald, G.; Paglione, M.; Garcia-Avello, C.; Bayraktutar, I.; Young, C. Impact of missing longitudinal aircraft intent on descent trajectory prediction. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Seattle, WA, USA, 16–20 October 2011*; IEEE: New York, NY, USA, 2011; pp. 3E2-1–3E2-14; IEEE Document 6096062. https://doi.org/10.1109/dasc.2011.6096062.
34. Legrand, K.; Rabut, C.; Delahaye, D. Wind networking applied to aircraft trajectory prediction. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Prague, Czech Republic, 13–17 September 2015*; IEEE: New York, NY, USA, 2015; IEEE Document 7311330. https://doi.org/10.1109/dasc.2015.7311330.

35. Pan, Y.; Nascimento, M.; Sander, J. Online stochastic prediction of mid-flight aircraft trajectories. In Proceedings of the 12th International Workshop on Computational Transportation Science, Chicago, IL, USA, 5 November 2019. https://doi.org/10.1145/3357000.3366144.
36. Lin, Y.; Yang, B.; Zhang, J.; Liu, H. Approach for 4-D trajectory management based on HMM and trajectory similarity. *J. Mar. Sci. Technol.* **2019**, *27*, 246–256. https://doi.org/10.6119/jmst.201906_27(3).0007.
37. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*; John Wiley & Sons: Hoboken, NJ, USA, 2001. https://doi.org/10.1002/0471221279.
38. Chatterji, G. Short-term trajectory prediction methods. In Proceedings of the Guidance, Navigation, and Control Conference and Exhibit, Portland, OR, USA, 9–11 August 1999; AIAA 1999-4233. https://doi.org/10.2514/6.1999-4233.
39. Gustafsson, F.; Gunnarsson, F.; Bergman, N.; Forssell, U.; Jansson, J.; Karlsson, R.; Nordlund, P.J. Particle filters for positioning, navigation, and tracking. *IEEE Trans. Signal Process.* **2002**, *50*, 425–437. https://doi.org/10.1109/78.978396.
40. Lymperopoulos, I.; Lygeros, J. Adaptive aircraft trajectory prediction using particle filters. In Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit, Honolulu, HI, USA, 18–21 August 2008; AIAA 2008-7387. https://doi.org/10.2514/6.2008-7387.
41. Lymperopoulos, I.; Lygeros, J. Sequential Monte Carlo methods for multi-aircraft trajectory prediction in air traffic management. *Int. J. Adapt. Control Signal Process.* **2010**, *24*, 830–849. https://doi.org/10.1002/acs.1174.
42. Liu, Y.; Li, X.R. Intent based trajectory prediction by multiple model prediction and smoothing. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, USA, 5–9 January 2015; AIAA 2015-1324. https://doi.org/10.2514/6.2015-1324.
43. Baumgartner, T.; Maeder, U. Trajectory prediction for low-cost collision avoidance systems. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Orlando, FL, USA, 25–29 October 2009*; IEEE: New York, NY, USA, 2009; IEEE Document 5347567. https://doi.org/10.1109/dasc.2009.5347567.
44. Liu, W.; Hwang, I. Probabilistic 4D trajectory prediction and conflict detection for air traffic control. In *Proceedings of the IEEE Conference on Decision and Control, Atlanta, GA, USA, 15–17 December 2010*; IEEE: New York, NY, USA, 2010; IEEE Document 5717681. https://doi.org/10.1109/cdc.2010.5717681.
45. Ayhan, S.; Samet, H. Aircraft trajectory prediction made easy with predictive analytics. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 21–30. https://doi.org/10.1145/2939672.2939694.
46. Tastambekov, K.; Puechmorel, S.; Delahaye, D.; Rabut, C. Aircraft trajectory forecasting using local functional regression in Sobolev space. *Transp. Res. Part C Emerg. Technol.* **2014**, *39*, 1–22. https://doi.org/10.1016/j.trc.2013.11.013.
47. Gong, C.; Sadovsky, A. A final approach trajectory model for current operations. In Proceedings of the 10th AIAA Aviation Technology, Integration and Operations Conference 2010, Fort Worth, TX, USA, 13–15 September 2010; AIAA 2010-9117. https://doi.org/10.2514/6.2010-9117.
48. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
49. Song, Y.; Cheng, P.; Mu, C. An improved trajectory prediction algorithm based on trajectory data mining for air traffic management. In *Proceedings of the 2012 IEEE International Conference on Information and Automation, Shenyang, China, 6–8 June 2012*; IEEE: New York, NY, USA, 2012; pp. 981–986. https://doi.org/10.1109/icinfa.2012.6246959.
50. Hernández, A.M.; Magaña, E.J.C.; Berna, A.G. Data-driven aircraft trajectory predictions using ensemble meta-estimators. In *Proceedings of the 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK, 23–27 September 2018*; IEEE: New York, NY, USA, 2018; pp. 1–10. https://doi.org/10.1109/dasc.2018.8569535.
51. Shi, Z.; Xu, M.; Pan, Q.; Yan, B.; Zhang, H. LSTM-based flight trajectory prediction. In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018*; IEEE: New York, NY, USA, 2018; pp. 1–8. https://doi.org/10.1109/ijcnn.2018.8489734.
52. Han, P.; Wang, W.; Shi, Q.; Yang, J. Real-time short-term trajectory prediction based on GRU neural network. In *Proceedings of the 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC), San Diego, CA, USA, 8–12 September 2019*; IEEE: New York, NY, USA, 2019; pp. 1–8. https://doi.org/10.1109/dasc43569.2019.9081618.
53. Pang, Y.; Xu, N.; Liu, Y. Aircraft trajectory prediction using LSTM neural network with embedded convolutional layer. In Proceedings of the Annual Conference of the PHM Society 2019, Scottsdale, AZ, USA, 21–26 September 2019; Volume 11, pp. 1–8. https://doi.org/10.36001/phmconf.2019.v11i1.849.
54. Zeng, W.; Quan, Z.; Zhao, Z.; Xie, C.; Lu, X. A deep learning approach for aircraft trajectory prediction in terminal airspace. *IEEE Access* **2020**, *8*, 151250–151266. https://doi.org/10.1109/access.2020.3016289.
55. Guo, D.; Wu, E.Q.; Wu, Y.; Zhang, J.; Law, R.; Lin, Y. FlightBERT: Binary encoding representation for flight trajectory prediction. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1828–1842. https://doi.org/10.1109/tits.2022.3219923.
56. Guo, D.; Zhang, Z.; Yan, Z.; Zhang, J.; Lin, Y. FlightBERT++: A non-autoregressive multi-horizon flight trajectory prediction framework. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 127–134. https://doi.org/10.1609/aaai.v38i1.27763.

57. Luo, A.; Luo, Y.; Liu, H.; Du, W.; Wu, X.; Chen, H.; Yang, H. An improved transformer-based model for long-term 4D trajectory prediction in civil aviation. *IET Intell. Transp. Syst.* **2024**, *18*, 1588–1598. https://doi.org/10.1049/itr2.12530.
58. Yoon, S.; Lee, K. Aircraft trajectory prediction with inverted transformer. *IEEE Access* **2025**, *13*, 26318–26330. https://doi.org/10.1109/access.2025.3539742.
59. Fan, Y.; Tan, Y.; Wu, L.; Ye, H.; Lyu, Z. Global and local interattribute relationships-based graph convolutional network for flight trajectory prediction. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 2642–2657. https://doi.org/10.1109/taes.2024.3357668.
60. Kuang, Y.; Yuan, S.; Zou, X.; Zhang, J.; Shi, Z.; Zhang, Y.; Meng, F.; Wang, Z. MTDA-STGCN: Modern Temporal and Dual-Attention-Based Spatiotemporal Graph Convolutional Network for 4D Trajectory Prediction. *IEEE Trans. Intell. Transp. Syst.* **2025**, 26, 21809–21823. https://doi.org/10.1109/tits.2025.3605670.
61. Cao, H.; Li, Y.; Mi, X.; Gao, Q. GAT-BiGRU-TPA City Pair 4D Trajectory Prediction Model Based on Spatio-Temporal Graph Neural Network. *Aerospace* **2025**, *12*, 999. https://doi.org/10.3390/aerospace12110999.
62. Xu, Z.; Zeng, W.; Chu, X.; Cao, P. Multi-aircraft trajectory collaborative prediction based on social long short-term memory network. *Aerospace* **2021**, *8*, 115. https://doi.org/10.3390/aerospace8040115.
63. Chu, X.; Zeng, W.; Wu, L. Joint prediction of multi-aircraft trajectories in terminal airspace: A Flight Pattern-Guided Social Long-Short Term Memory network. *Eng. Appl. Artif. Intell.* **2025**, *147*, 110325. https://doi.org/10.1016/j.engappai.2025.110325.
64. Pang, Y.; Liu, Y. Probabilistic aircraft trajectory prediction considering weather uncertainties using dropout as Bayesian approximate variational inference. In Proceedings of the AIAA SciTech 2020 Forum, Orlando, FL, USA, 6–10 January 2020; AIAA 2020-1413. https://doi.org/10.2514/6.2020-1413.
65. Chen, Z.; Guo, D.; Lin, Y. A deep gaussian process-based flight trajectory prediction approach and its application on conflict detection. *Algorithms* **2020**, *13*, 293. https://doi.org/10.3390/a13110293.
66. Pang, Y.; Liu, Y. Conditional generative adversarial networks (CGAN) for aircraft trajectory prediction considering weather effects. In Proceedings of the AIAA SciTech 2020 Forum, Orlando, FL, USA, 6–10 January 2020; AIAA 2020-1853. https://doi.org/10.2514/6.2020-1853.
67. Wu, X.; Yang, H.; Chen, H.; Hu, Q.; Hu, H. Long-term 4D trajectory prediction using generative adversarial networks. *Transp. Res. Part C Emerg. Technol.* **2022**, *136*, 103554. https://doi.org/10.1016/j.trc.2022.103554.
68. Ho, J.; Jain, A.N.; Abbeel, P. Denoising diffusion probabilistic models. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 6840–6851.
69. Yang, S.; Liu, L.; Chen, B.; Cheng, S.; Shi, Z.; Zou, Z. GooDFlight: Goal-Oriented Diffusion Model for Flight Trajectory Prediction. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 7447–7465. https://doi.org/10.1109/taes.2025.3536436.
70. Ma, L.; Tian, S. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction. *IEEE Access* **2020**, *8*, 134668–134680. https://doi.org/10.1109/access.2020.3010963.
71. Huang, J.; Ding, W. Aircraft trajectory prediction based on Bayesian optimised temporal convolutional network–bidirectional gated recurrent unit hybrid neural network. *Int. J. Aerosp. Eng.* **2022**, *2022*, 2086904. https://doi.org/10.1155/2022/2086904.
72. Schimpf, N.; Wang, Z.; Li, S.; Knoblock, E.J.; Li, H.; Apaza, R.D. A generalized approach to aircraft trajectory prediction via supervised deep learning. *IEEE Access* **2023**, *11*, 116183–116195. https://doi.org/10.1109/access.2023.3325053.
73. Jarry, G.; Dalmau, R.; Olive, X.; Very, P. A Neural ODE Approach to Aircraft Flight Dynamics Modelling. In Proceedings of the 15th SESAR Innovation Days, Bled, Slovenia, 1–4 December 2025; http://arxiv.org/abs/2509.23307.
74. Choi, H.C.; Deng, C.; Hwang, I. Hybrid machine learning and estimation-based flight trajectory prediction in terminal airspace. *IEEE Access* **2021**, *9*, 151186–151197. https://doi.org/10.1109/access.2021.3126117.
75. Liu, B.; Shi, Q.; Han, P. Short-Term 4D Trajectory Prediction Method Based on LSTM-IMM. In *Proceedings of the 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC), Portsmouth, VA, USA, 18–22 September 2022*; IEEE: New York, NY, USA, 2022; IEEE Document 9925832. https://doi.org/10.1109/dasc55683.2022.9925832.
76. Shi, Z.; Xu, M.; Pan, Q. 4-D flight trajectory prediction with constrained LSTM network. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7242–7255. https://doi.org/10.1109/tits.2020.3004807.
77. Feng, S.; Wang, G.; Zhao, P.; Chao, X.; Cai, K. Short-term aircraft trajectory prediction considering weather effect. In *Proceedings of the 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), Barcelona, Spain, 1–5 October 2023*; IEEE: New York, NY, USA, 2023; pp. 1–6. https://doi.org/10.1109/dasc58513.2023.10311245.
78. Hu, Q.; Huang, G.; Shi, H.; Lin, Y.; Guo, D. A short-term aircraft trajectory prediction framework using conditional generative adversarial network. In *Proceedings of the 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 12–14 October 2022*; IEEE: New York, NY, USA, 2022; pp. 433–439. https://doi.org/10.1109/iccasit55263.2022.9986799.
79. Hwang, I.; Seah, C.E. Intent-based probabilistic conflict detection for the next generation air transportation system. *Proc. IEEE* **2008**, *96*, 2040–2059. https://doi.org/10.1109/jproc.2008.2006138.

80. Zhang, J.; Qin, M.; Yuan, Y.; Yang, Y.; Meng, X. Flight Trajectory Prediction Based on TCN-Transformer Method. In *Proceedings of the International Conference on Intelligent Computing, Ningbo, China, 26–29 July 2025*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 393–404. https://doi.org/10.1007/978-981-96-9986-5_33.
81. Tong, Q.; Hu, J.; Chen, Y.; Guo, D.; Liu, X. Long-term trajectory prediction model based on transformer. *IEEE Access* **2023**, *11*, 143695–143703. https://doi.org/10.1109/access.2023.3343800.
82. Dong, X.; Tian, Y.; Dai, L.; Li, J.; Wan, L. A New Accurate Aircraft Trajectory Prediction in Terminal Airspace Based on Spatio-Temporal Attention Mechanism. *Aerospace* **2024**, *11*, 718. https://doi.org/10.3390/aerospace11090718.
83. Patrikar, J.; Moon, B.; Oh, J.; Scherer, S. Predicting like a pilot: Dataset and method to predict socially-aware aircraft trajectories in non-towered terminal airspace. In *Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022*; IEEE: New York, NY, USA, 2022; pp. 2525–2531. https://doi.org/10.1109/icra46639.2022.9811972.
84. Zhong, X.; Cai, K.; Wang, G.; Zhao, P. Multiple Aircraft Trajectory Prediction Based on Bayesian Social LSTM. In Proceedings of the AIAA SCITECH 2025 Forum, Orlando, FL, USA, 6–10 January 2025; AIAA 2025-1732. https://doi.org/10.2514/6.2025-1732.
85. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. https://doi.org/10.3354/cr030079.
86. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Fei-Fei, L.; Savarese, S. Social LSTM: Human Trajectory Prediction in Crowded Spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016*; IEEE: New York, NY, USA, 2016; pp. 961–971. https://doi.org/10.1109/cvpr.2016.110.
87. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting: Methods and Applications*, 3rd ed.; Wiley India Pvt. Limited: New Delhi, India, 2008.
88. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech, Signal Process.* **1978**, *26*, 43–49. https://doi.org/10.1109/tassp.1978.1163055.
89. Zhao, H.; Gao, J.; Lan, T.; Sun, C.; Sapp, B.; Varadarajan, B.; Shen, Y.; Shen, Y.; Chai, Y.; Schmid, C.; et al. TNT: Target-Driven Trajectory Prediction. In *Proceedings of the 2020 Conference on Robot Learning, Virtual, 16–18 November 2020*; PMLR: Cambridge, MA, USA, 2021; Volume 155, pp. 895–904; Proceedings of Machine Learning Research.
90. Mercat, J.; Zoghby, N.E.; Sandou, G.; Beauvois, D.; Gil, G.P. Kinematic single vehicle trajectory prediction baselines and applications with the NGSIM dataset. *arXiv* **2019**, arXiv:1908.11472. http://arxiv.org/abs/1908.11472.
91. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. https://doi.org/10.1198/016214506000001437.
92. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Trans. Neural Networks* **2011**, *22*, 1341–1356. https://doi.org/10.1109/tnn.2011.2162110.
93. Cheng, Y.; Wang, D.; Zhou, P.; Zhang, T. Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Process. Mag.* **2018**, *35*, 126–136. https://doi.org/10.1109/msp.2017.2765695.
94. Schäfer, M.; Strohmeier, M.; Lenders, V.; Martinovic, I.; Wilhelm, M. Bringing up OpenSky: A large-scale ADS-B sensor network for research. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN 2014), Berlin, Germany, 15–17 April 2014*; IEEE: New York, NY, USA, 2014; pp. 83–94. https://doi.org/10.1109/ipsn.2014.6846743.
95. Olive, X.; Krummen, J.; Figuet, B.; Alligier, R. Filtering techniques for ADS-B trajectory preprocessing. *J. Open Aviat. Sci.* **2025**, *2*. https://doi.org/10.59490/joas.2024.7882.
96. Kim, E.; Sivits, K. Blended secondary surveillance radar solutions to improve air traffic surveillance. *Aerosp. Sci. Technol.* **2015**, *45*, 203–208. https://doi.org/10.1016/j.ast.2015.05.018.
97. Deshannavar, U.B. High dimensional weather data used in a deep generative model to predict trajectories of aircraft. *J. Airl. Oper. Aviat. Manag.* **2022**, *1*, 80–88. https://doi.org/10.56801/jaoam.v1i1.10.
98. Hodgkin, A.; Pepper, N.; Thomas, M. Conditioning Aircraft Trajectory Prediction on Meteorological Data with a Physics-Informed Machine Learning Approach. In Proceedings of the AIAA Science and Technology Forum and Exposition (AIAA SciTech Forum 2026), Orlando, FL, USA, 12–16 January 2026; AIAA 2026-1792. https://doi.org/10.2514/6.2026-1792.
99. Georgiou, H.; Pelekis, N.; Sideridis, S.; Scarlatti, D.; Theodoridis, Y. Semantic-aware aircraft trajectory prediction using flight plans. *Int. J. Data Sci. Anal.* **2020**, *9*, 215–228. https://doi.org/10.1007/s41060-019-00182-4.
100. Hoekstra, J.M.; Ellerbroek, J. BlueSky ATC simulator project: An open data and open source approach. In Proceedings of the 7th International Conference on Research in Air Transportation (ICRAT 2016), Philadelphia, PA, USA, 20–24 June 2016.
101. Sekine, K.; Bouchaudon, P.; Tatsukawa, T.; Fujii, K.; Itoh, E. Validating flow-based en route AMAN coupling escape light with AirTOp software. In Proceedings of the 34th Congress of the International Council of the Aeronautical Sciences (ICAS 2024), Florence, Italy, 9–13 September 2024; Paper ICAS2024_0968.
102. Warren, A.W.; Ebrahimi, Y.S. Vertical path trajectory prediction for next generation ATM. In *Proceedings of the 17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference, Bellevue, WA, USA, 31 October–7 November 1998*; Proceedings (Cat. No.98CH36267); IEEE: New York, NY, USA, 1998; IEEE Document 739809. https://doi.org/10.1109/dasc.1998.739809.

103. Zammit, B.; Zammit-Mangion, D. Accuracy Considerations of a Simple Aircraft Trajectory Prediction Model for Idle Thrust Descents. In *Proceedings of the 2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC), East Syracuse, NY, USA, 5–10 October 2013*; IEEE: New York, NY, USA, 2013; IEEE Document 6712526. https://doi.org/10.1109/dasc.2013.6712526.
104. Cásek, P.; Weber, R.; Kubalcík, J. Analysis of FMS-generated trajectory prediction accuracy and sensitivity. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Dallas, TX, USA, 21–25 October 2007*; IEEE: New York, NY, USA, 2007; IEEE Document 4391827. https://doi.org/10.1109/dasc.2007.4391827.
105. Paglione, M.; Musialek, B.; Pankok, C.; Young, C. Trajectory predictor performance experiment using Required Time of Arrival during descent. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Seattle, WA, USA, 16–20 October 2011*; IEEE: New York, NY, USA, 2011; IEEE Document 6095999. https://doi.org/10.1109/dasc.2011.6095999.
106. Ramchandra, A.R.; Ranganathan, P.; Semke, W. Trajectory Prediction Model for Accurate Navigation and Surveillance Using kNN Regressor. In *Proceedings of the IEEE Aerospace Conference Proceedings, Big Sky, MT, USA, 2–9 March 2024*; IEEE: New York, NY, USA, 2024; IEEE Document 10521217. https://doi.org/10.1109/aero58975.2024.10521217.
107. Christien, R.; Zeghal, K.; Hoffman, E. Predicting Conflict Free Trajectories Using Supervised Machine Learning, Initial Investigations. In *Proceedings of the 2020 Integrated Communications Navigation and Surveillance Conference (ICNS), Virtual, 8–10 September 2020*; IEEE: New York, NY, USA, 2020; IEEE Document 9222959. https://doi.org/10.1109/icns50378.2020.9222959.
108. de Leege, A.; van Paassen, M.; Mulder, M. A machine learning approach to trajectory prediction. In Proceedings of the AIAA Guidance, Navigation, and Control (GNC) Conference, Boston, MA, USA, 19–22 August 2013; AIAA 2013-4782. https://doi.org/10.2514/6.2013-4782.
109. Petrou, P.; Tampakis, P.; Georgiou, H.; Pelekis, N.; Theodoridis, Y. Online Long-Term Trajectory Prediction Based on Mined Route Patterns. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer, Cham, Switzerland, 2020. https://doi.org/10.1007/978-3-030-38081-6_4.
110. Rocha, M.M.; De, O.M. A data-driven probabilistic trajectory model for predicting and simulating terminal airspace operations. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, San Antonio, TX, USA, 11–15 October 2020*; IEEE: New York, NY, USA, 2020; IEEE Document 9256644. https://doi.org/10.1109/dasc50938.2020.9256644.
111. Arnaldo, C.; Jurado, R.A.; Suarez, M.; Castan, J.; Moreno, F. 4-D Aircraft Trajectory Prediction for ATM Capacity and Demand Optimization in a Strategic Time Horizon Using Artificial Intelligence Models. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Montreal, QC, Canada, 14–18 September 2025*; IEEE: New York, NY, USA, 2025; IEEE Document 11257313. https://doi.org/10.1109/dasc66011.2025.11257313.
112. Guo, D.; Zhang, Z.; Liu, J.; Zhang, J.; Lin, Y. Multi-horizon flight trajectory prediction enabled by time-frequency wavelet transform. *Nat. Commun.* **2026**, *17*, 633. https://doi.org/10.1038/s41467-025-67399-9.
113. Ma, X.; Zheng, L.; Lu, X. 4D trajectory prediction and conflict detection in terminal areas based on an improved convolutional network. *PLoS ONE* **2025**, *20*, e0317549. https://doi.org/10.1371/journal.pone.0317549.
114. Lu, T.T.; Liu, B.; Li, C.Z. Aircraft Trajectory Prediction within Terminal Area Based on Seq2Seq-Attention Model. *Sci. Technol. Eng.* **2024**, *24*, 3882–3895. https://doi.org/10.12404/j.issn.1671-1815.2304239.
115. Yoon, S.; Lee, K. Multi-Agent Inverted Transformer for Flight Trajectory Prediction. *IEEE Trans. Intell. Transp. Syst.* **2026**, *27*, 11516–11526. https://doi.org/10.1109/tits.2026.3689290.
116. Zhang, J.; Chao, P.; Fang, J.; Peng, H.; Xu, J. SWIFT: Scene-Aware Dual Cross-Attention Flight Trajectory Prediction. In *Proceedings of the IEEE International Conference on Big Data, BigData, Macau, China, 8–11 December 2025*; IEEE: New York, NY, USA, 2025; IEEE Document 11401796. https://doi.org/10.1109/bigdata66926.2025.11401796.
117. Wang, L.; Tuo, H.; Li, M.; Jing, Z. IDEA-FastTEN: A Real-Time ADS-B Flight Trajectory Prediction Method With Independent Distance Extrapolation. *IEEE Trans. Intell. Transp. Syst.* **2026**, *27*, 9492–9503. https://doi.org/10.1109/tits.2026.3683520.
118. Zhang, H.; Liu, Z. 4D aircraft trajectory prediction considering severe weather effects. *Aeronaut. J.* **2025**, *129*, 3619–3643. https://doi.org/10.1017/aer.2025.10053.
119. Zhu, X.; Zhang, K.; Zhang, Z.; Tan, L. Predicting Flight Trajectory in Convective Weather through Boosted Spatiotemporal Deep Learning Ensemble. *J. Adv. Transp.* **2024**, *2024*, 6400839. https://doi.org/10.1155/2024/6400839.
120. Acharya, A.; Bulusu, V.; Idris, H. VFR Trajectory Forecasting using Deep Generative Model for Autonomous Airspace Operations. In *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, San Diego, CA, USA, 29 September–3 October 2024*; IEEE: New York, NY, USA, 2024; IEEE Document 10748987. https://doi.org/10.1109/dasc62030.2024.10748987.
121. Zhang, X.; Mahadevan, S. Bayesian neural networks for flight trajectory prediction and safety assessment. *Decis. Support Syst.* **2020**, *131*, 113246. https://doi.org/10.1016/j.dss.2020.113246.
122. He, P.; Zhang, Z.; Yu, Y.; Jiang, G.; Hong, F.; Wang, B. FlightDiff: A dual-constraint guided two-phase diffusion framework for accurate flight prediction. *GeoInformatica* **2026**, *30*, 1. https://doi.org/10.1007/s10707-025-00559-7.
123. Ye, S.; Gombolay, M.C. Efficient Trajectory Forecasting and Generation with Conditional Flow Matching. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Abu Dhabi, United Arab Emirates, 14–18 October 2024*; IEEE: New York, NY, USA, 2024; IEEE Document 10802208. https://doi.org/10.1109/iros58592.2024.10802208.

124. Zhu, X.; Liu, Z.; Huang, G.; An, Y.; Wang, S. DST-GAT: Predicting Multi-Flight Trajectories with Spatio-Temporal Flight Correlation Dynamics Mining. *Expert Syst.* **2025**, *42*, e70156. https://doi.org/10.1111/exsy.70156.
125. Mueller, K.; Sorensen, J.; Couluris, G. Strategic aircraft trajectory prediction uncertainty and statistical sector traffic load modeling. In Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Monterey, CA, USA, 5–8 August 2002; AIAA 2002-4765. https://doi.org/10.2514/6.2002-4765.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.