TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Accuracy of Air Traffic Forecasts
Causes and Consequences

InterFAB Expert Talks: ATM performance data - can we do better?

02 March 2021

Prof. Dr.-Ing. habil. Hartmut Fricke
Institute of Logistics and Aviation, Chair of Air Transport Technology and Logistics

Agenda

1. The Need to Forecast
2. Traffic Forecasts Framing and Inputs
3. Traffic Forecasts Performance Assessment
4. Way Forward
5. Wrap Up

The Need to Forecast

Why and how to predict air traffic demand?

Background

- Air Navigation Service Providers (ANSPs) responsibility:
  - Efficient deployment of resources
  - To ensure service provision at minimum costs to Stakeholders.
- Resource planning relies on expected strategic demand. The optimum competes with
  - Sufficient resources for robust operations,
  - Minimum resources for cost effectiveness.
- Crucial resource: the ATCO
  A prediction horizon of 5 Years is required.
- Poor forecasting leads to either capacity shortage or extra production costs

Traffic Forecast Framing and Inputs

How are forecasts built?

STATFOR predictions

- Forecast horizons:
  - Short-Term (2 years)
  - **Medium Term (7-years)**
    published in spring and autumn every year
  - Long Term (20 years)
- STATFOR provides predictions for both **flights** and **service units**.
- Main (typical) Input parameters shown right

STATFOR Strategy and Claim:

- Generating large number of forecasts (low-, base-, high scenarios) to be both user-specific and robust
  - Over large series – started in 1990
  - better than linear extrapolation on average
  - Assumably better than FAA's figures in the U.S.

See also: EUROCONTROL: "Seven-Year Forecast February 2019 Main Report"

Traffic Forecasts Performance Assessment

How did STATFOR evaluate the perfomance? (2015-2019)

STATFOR's Accuracy (Relative Performance, RP), Risk (R) and Overall Performance (KPI) metrics

Accuracy (for base scenario):

Error E_{FC} = 3.2% - 3.9% = -0.7%

Absolute Error AE_{FC} = |-0.7%| = 0.7%

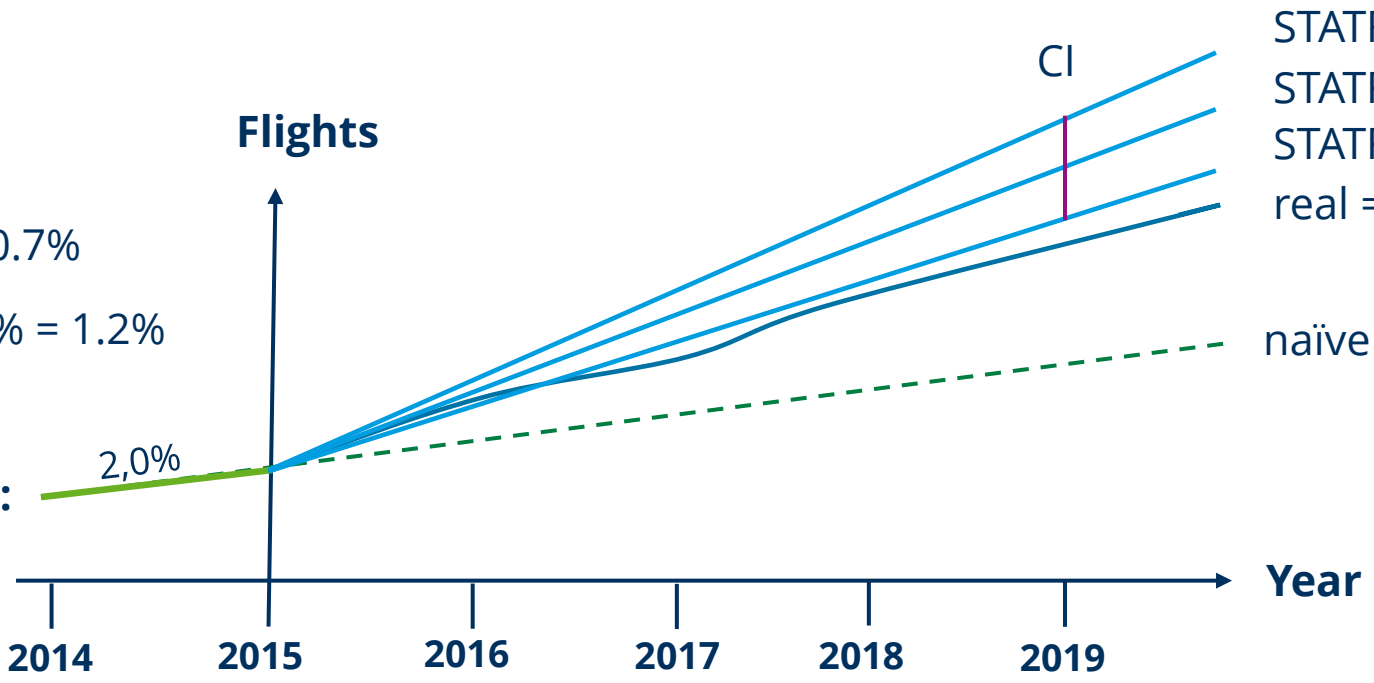
Absolute Error AE_{NAI} = 3.2% - 2.0% = 1.2%

Accuracy (RP) = AE_{NAI} / AE_{FC} = 1.7

Risk (Forecast-range accuracy):

With Confidence Interval (CI)

to catch CI = 50% forecasts



STATFOR high = 4.1%
STATFOR base = 3.9%
STATFOR low = 3.7%
real = 3.2%

naïve = 2.0%

Be 30% inside, then R = CI – 30% = 20% (more forecasts than we wanted outside) -0.5 < R < 0.5

Overall Performance KPI (combination of base and forecast-range accuracy):

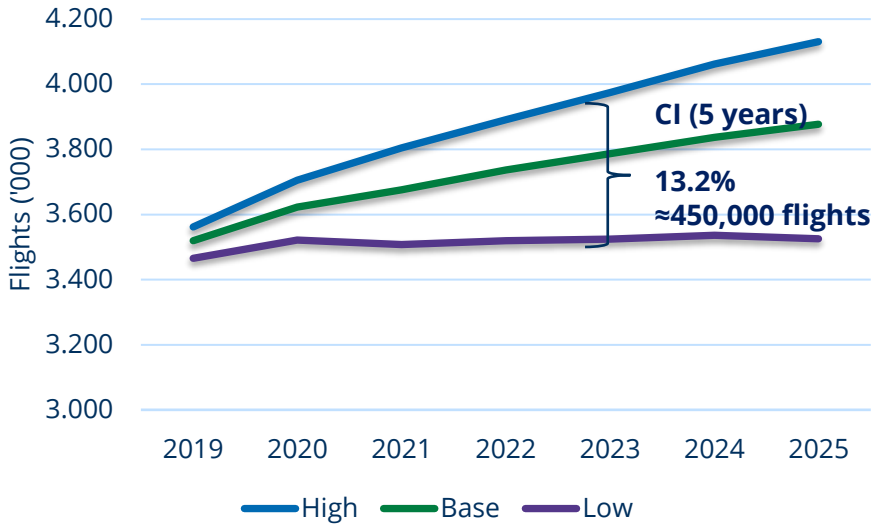
KPI = RP - | R | = 1.7 – 0.2 = 1.5 with KPI_{TARGET} > 2.0 and RP_{TARGET} > 1.5

Traffic Forecasts Performance Assessment

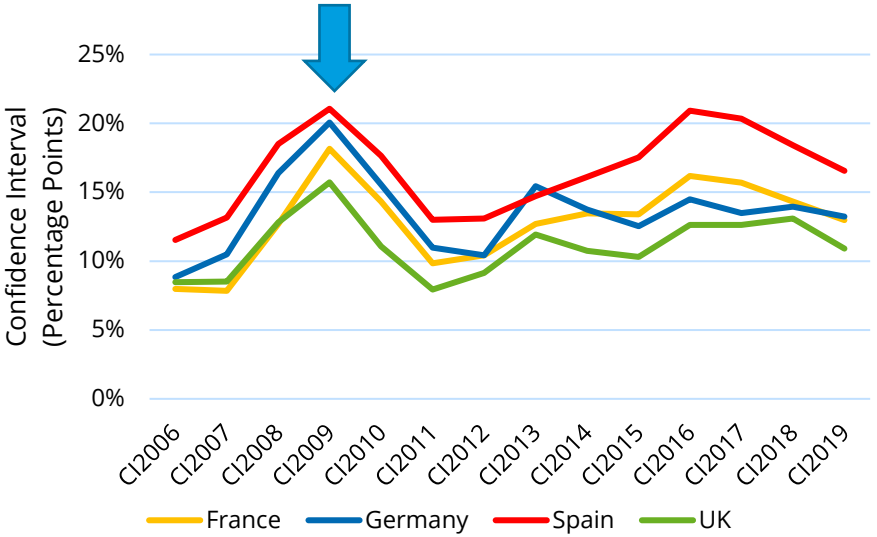
How to deal with uncertainties?

Consequences of the Used Boundaries

- STATFOR **high-** (blue line) to **low-scenario** difference (CI) revert directly to uncertainty in the predicted flight numbers
- Example: **Germany** - **spring report 2019**.
  - 5-year **CI** equals **13.2 %**.
  - The STATFOR prediction thus induce an uncertainty of **450,000 flights** for Germany in 2023 (2019 + 5 years).
- CIs change massively over time, impacting **performance metrics**.
- Reasons are
  - various adaptations to the scenario building technique and
  - major turning points in European traffic (e.g., 2009 economic crises - traffic decline & fuel price surge)
- CI however has **consequences** for ANSPs.



Source: STATFOR Seven-Year Forecast February 2019, Germany

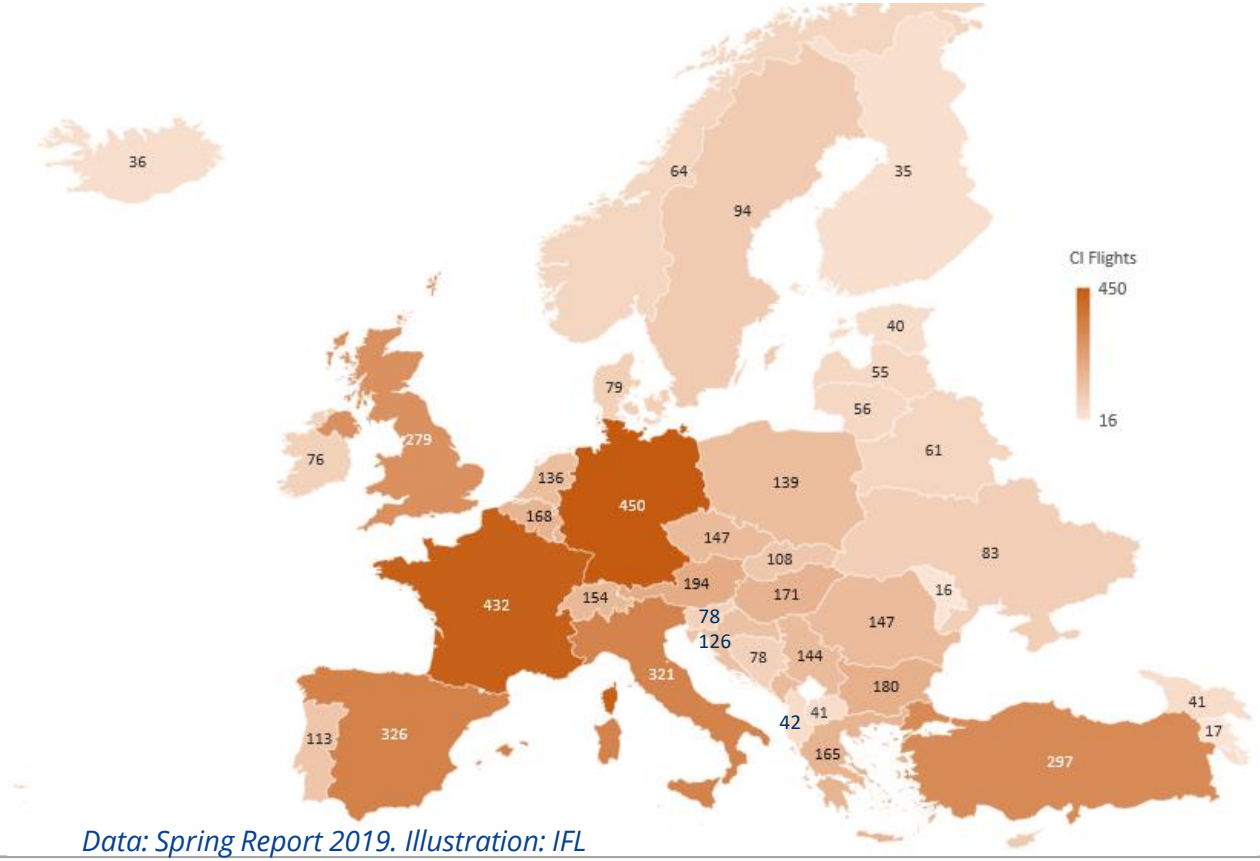


Data: STATFOR Seven-Year Forecasts Reports, Spring 2006-2019

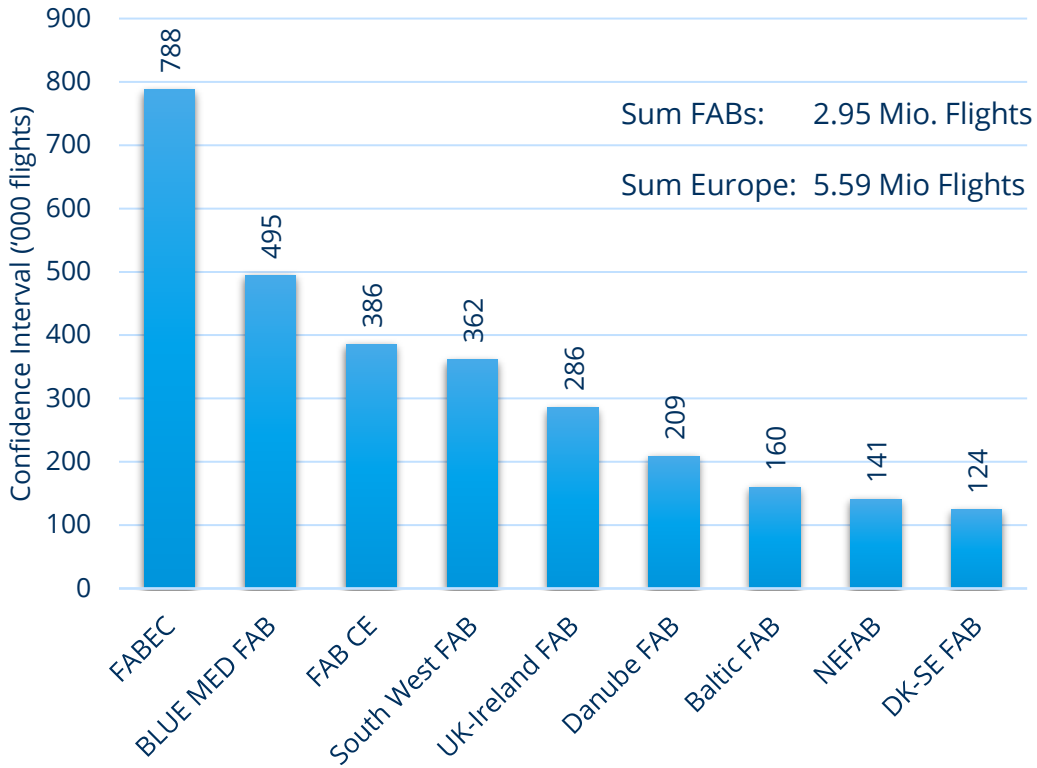
Traffic Forecasts and Implications

How does it affect the output (flights controlled)?

From Germany to whole Europe: **CI** to **flights** (in '000) conversion on country / FAB level (STATFOR spring report 02/2019, 5-year prediction, 2019-2023). **Fact**: Significant uncertainty with regards to flights for all countries:



Data: Spring Report 2019. Illustration: IFL

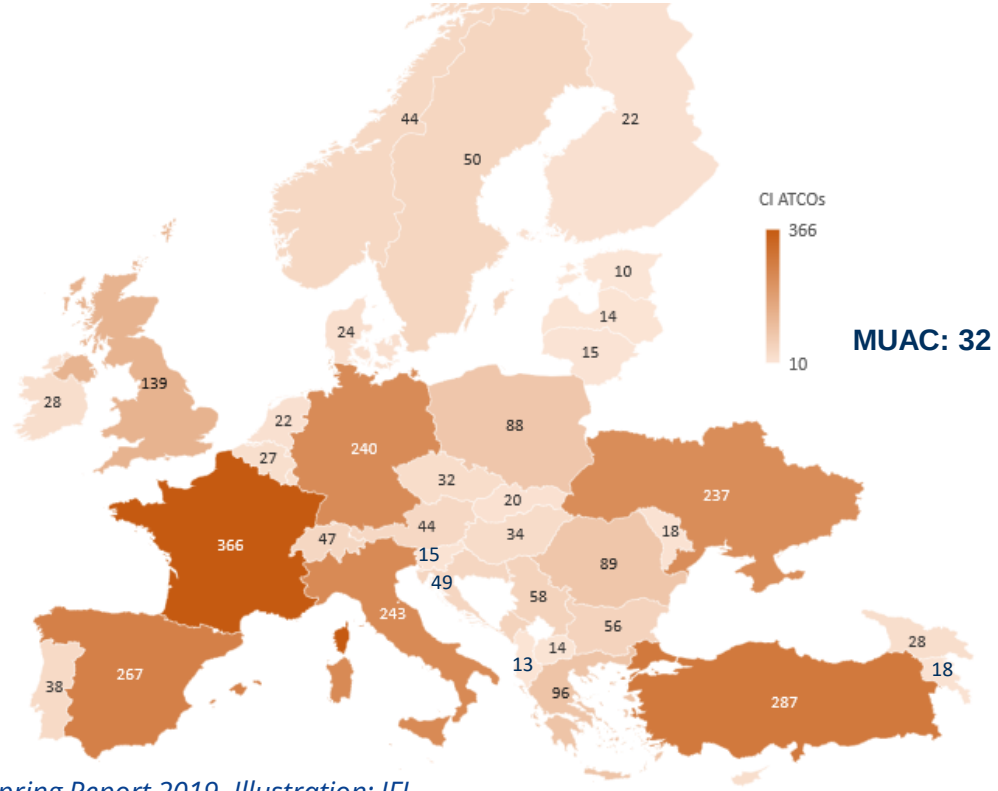


Data: Spring Report 2019. Illustration: IFL

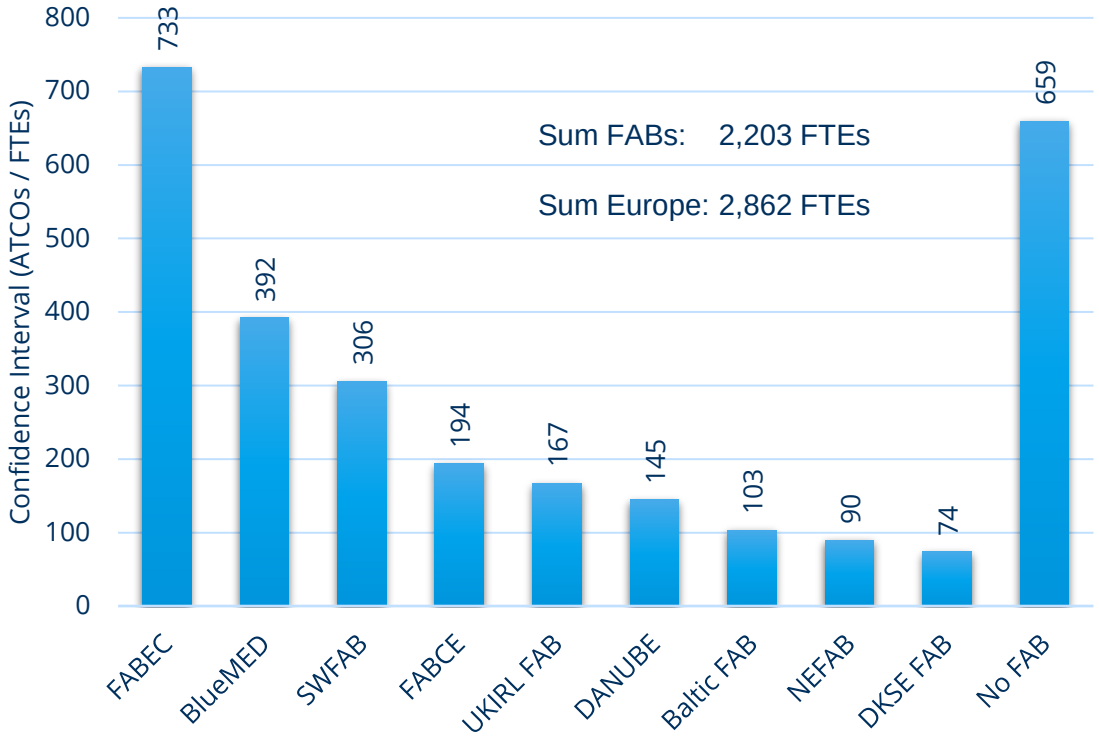
Traffic Forecasts and Implications

What does it mean for ANSP / FAB input (staffing)?

Demand triggers number of required human resources both on ANSP and FAB level. The figures illustrate **CI** conversion to **required ATCOs / FTEs** (Approximation based on ACE-data. 5-year prediction, 2019-2023).
Fact: **Linear interdependency** assumed (as such neglecting e.g., dis-/economies of scale) specific to each unit:



Data: Spring Report 2019. Illustration: IFL

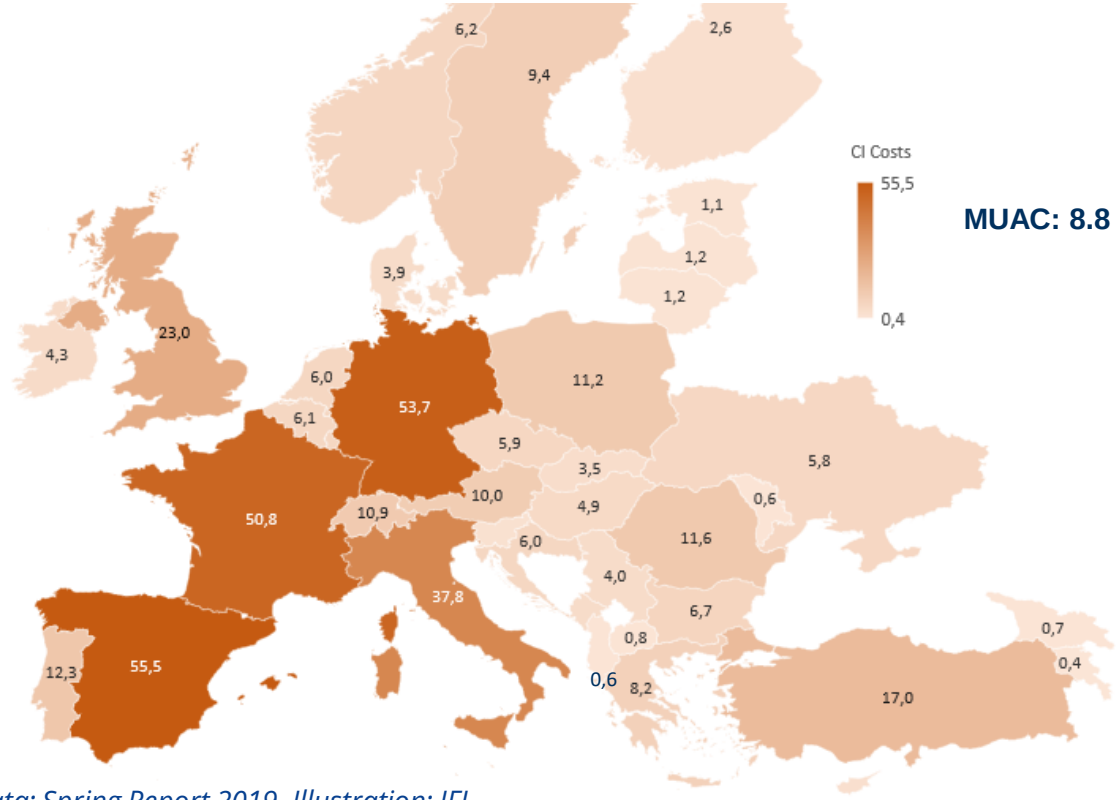


Data: Spring Report 2019. Illustration: IFL

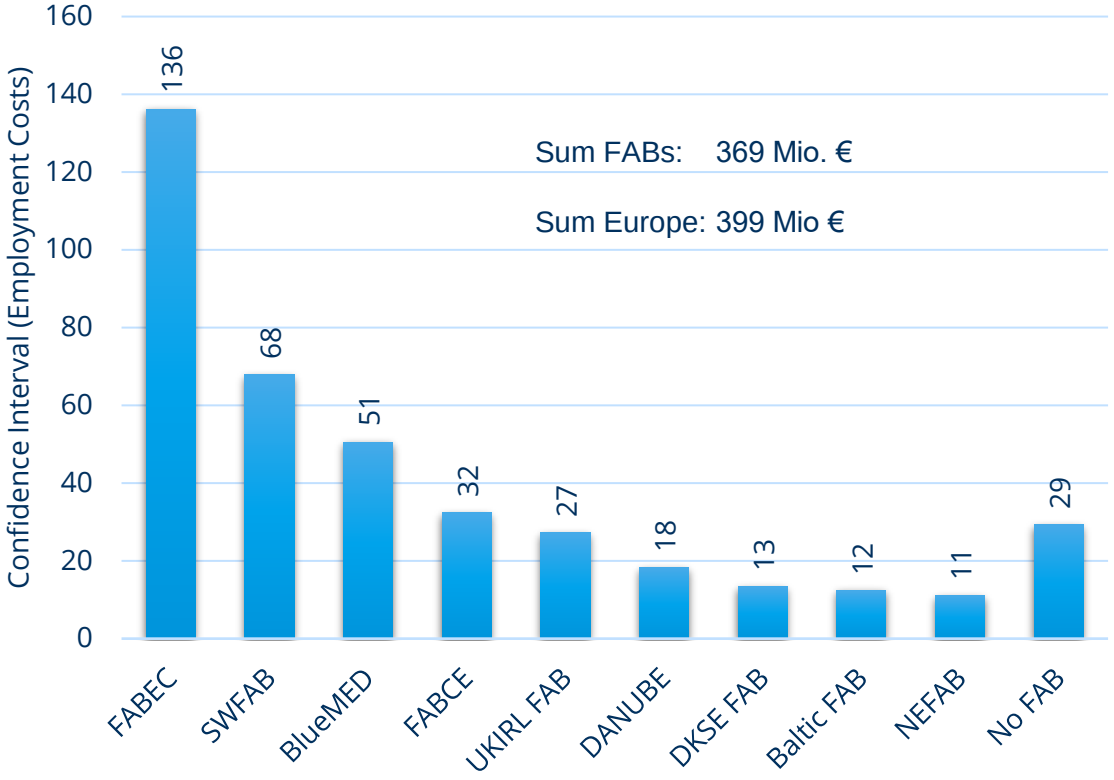
Traffic Forecasts and Implications

How does it affect costs of ANSPs?

CI Spreads lead to relevant uncertainty regarding costs for ANSPs (and FABs). The figures illustrate the **CI** conversion to **ATCO employment costs** in Mio. € (5-year prediction, 2019-2023).
Fact: High uncertainties in costs for some ANSPs, e.g., for DFS, DSNA, ENAIRE and ENAV.



Data: Spring Report 2019. Illustration: IFL



Data: Spring Report 2019. Illustration: IFL

Lessons learnt and and optional way forward

Throughput vs. confidence interval – the Risk

Recall: CI is a preset target (since 2015: 50%) leading to said “manageable” **uncertainties** / **risks** (R) for ANSP / FAB in **resource** and **cost planning**.

The right figure depicts how many actual flight cases did reside within CI in the past (**5-year horizon**, **eight observations** based on the spring reports 2006-2013). The share of ANSP related flight counts / year laying in the CI:

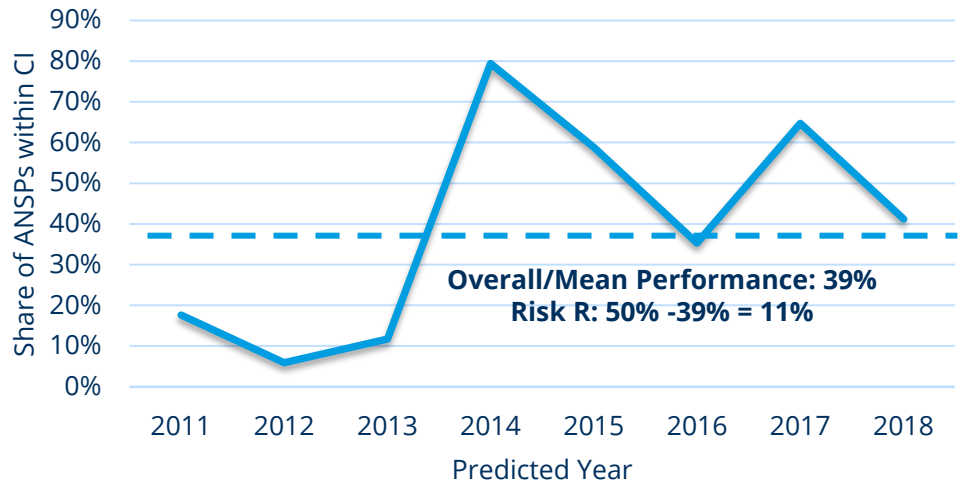


Illustration: IFL

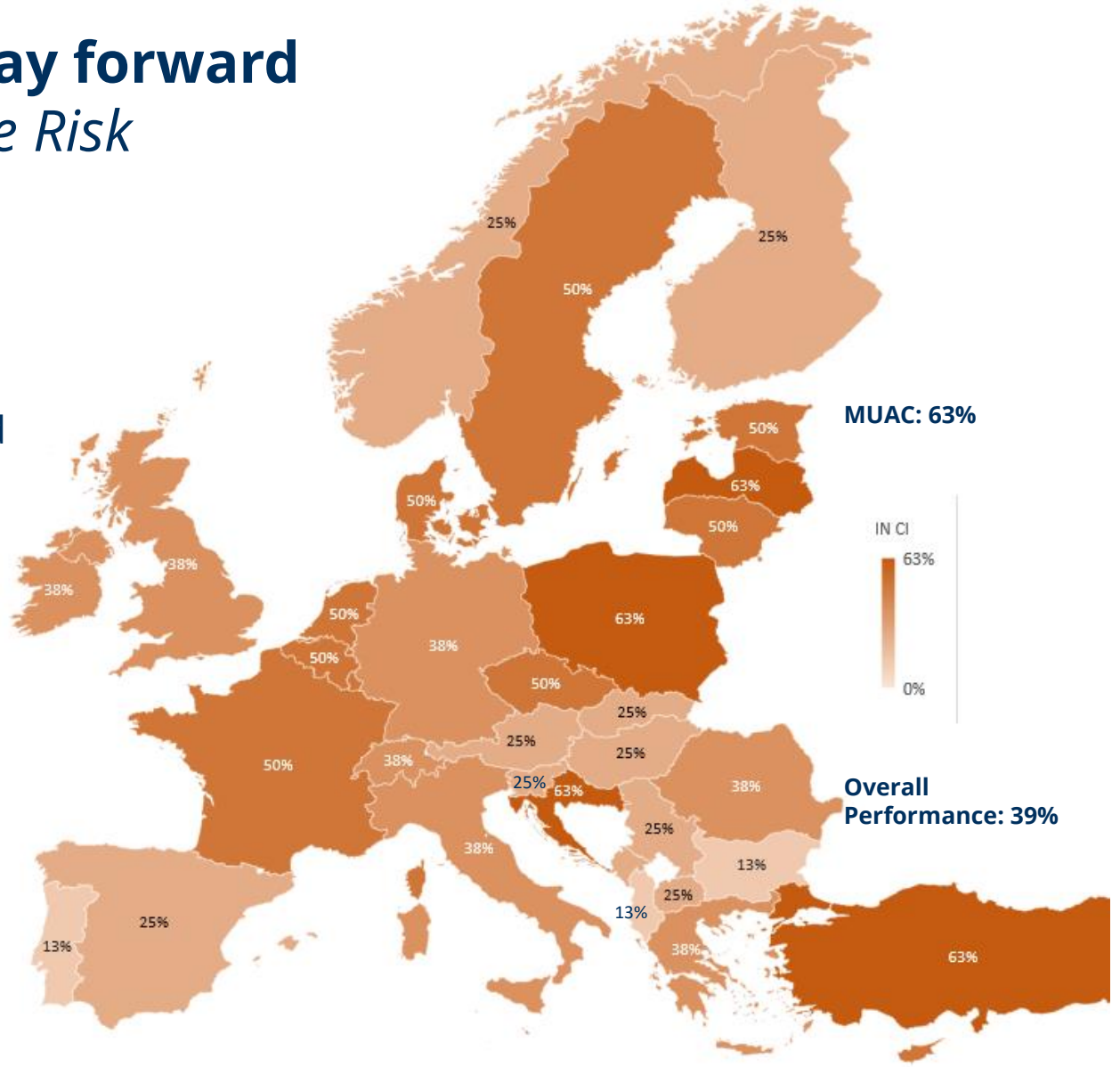


Illustration: IFL

Lessons learnt and an optional way forward

How to probably better measure „accuracy“?

Intermediate Conclusion

- For the pre-set **CI=50%,** only **39%** of actual flight numbers fall inside. Contradicts STATFOR intent to deliver reasonable forecasts in toto

Stakeholder Perspectives may additionally include

- The quality should measure the offset to actual traffic, not to a naïve forecast. It should strictly consider the 5-year time horizon. **Time- and unit-wise aggregability** is preferred.
- Candidate: ***Mean Average Percentage Error (MAPE):***
  - The score measures the **forecast bias. Accuracy = 1-MAPE.**
  - Application is recommended for **heterogenous industries** with **homogenous products** (given for ANSPs).
  - Thresholds exist → intuitive, easy to understand.
- Connecting forecast errors to cost- and capacity targets setting: Flow systems typically behave exponentially. However: Distinction between under-/overestimation distinction required → **0 > MAPE$_{directed}$ > 0**

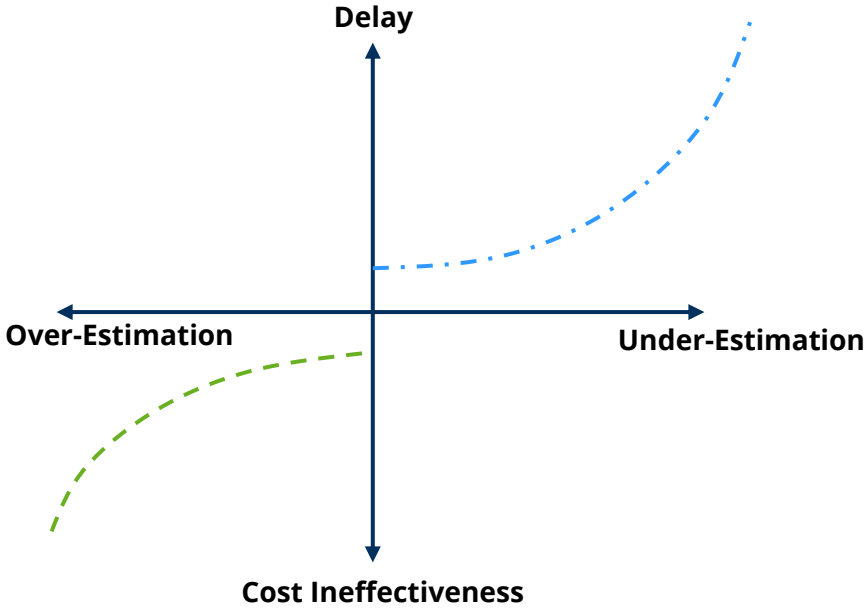


Illustration: IFL

$$MAPE_{dir} = \frac{Actual - Prediction}{Actual} = \frac{E(flights)_{FC}}{Throughput}$$

MAPE$_{directed}$	Meaning
= 0%	Actual = Prediction
≤ +5%	Very good under-estimation
≤ +10%	Good under-estimation
≤ - 5%	Very good over-estimation
≤ -10%	Good over-estimation

Lessons learnt and an optional way forward

How do STATFOR forecasts behave with MAPE?

Aggregated MAPE score

- The right figure shows the time-wise aggregated MAPE score per ANSP, the below figure the unit-wise aggregated score in a 5-year horizon for all considered reports (2006-2013).
- The MAPE scores comes to:
  - Very good prediction quality for one of 36 ANSP,
  - Good prediction quality for 13 of 36 ANSPs,
  - Overall: MAPE highlights well differences in accuracy

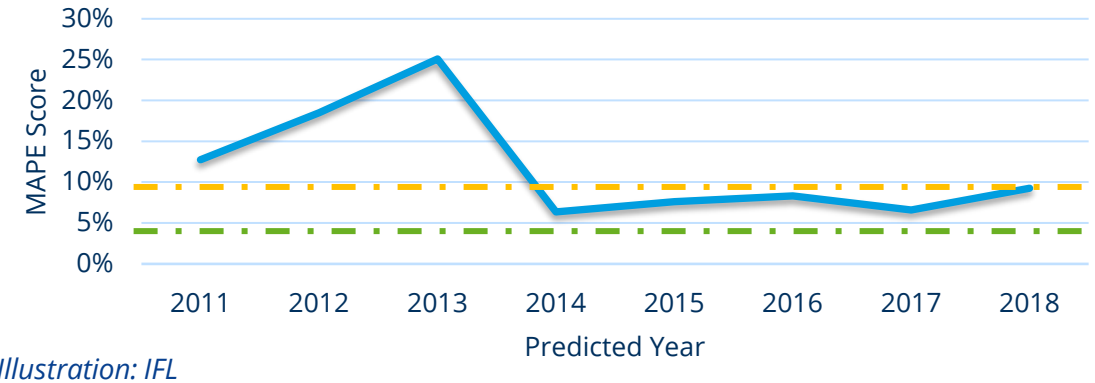


Illustration: IFL

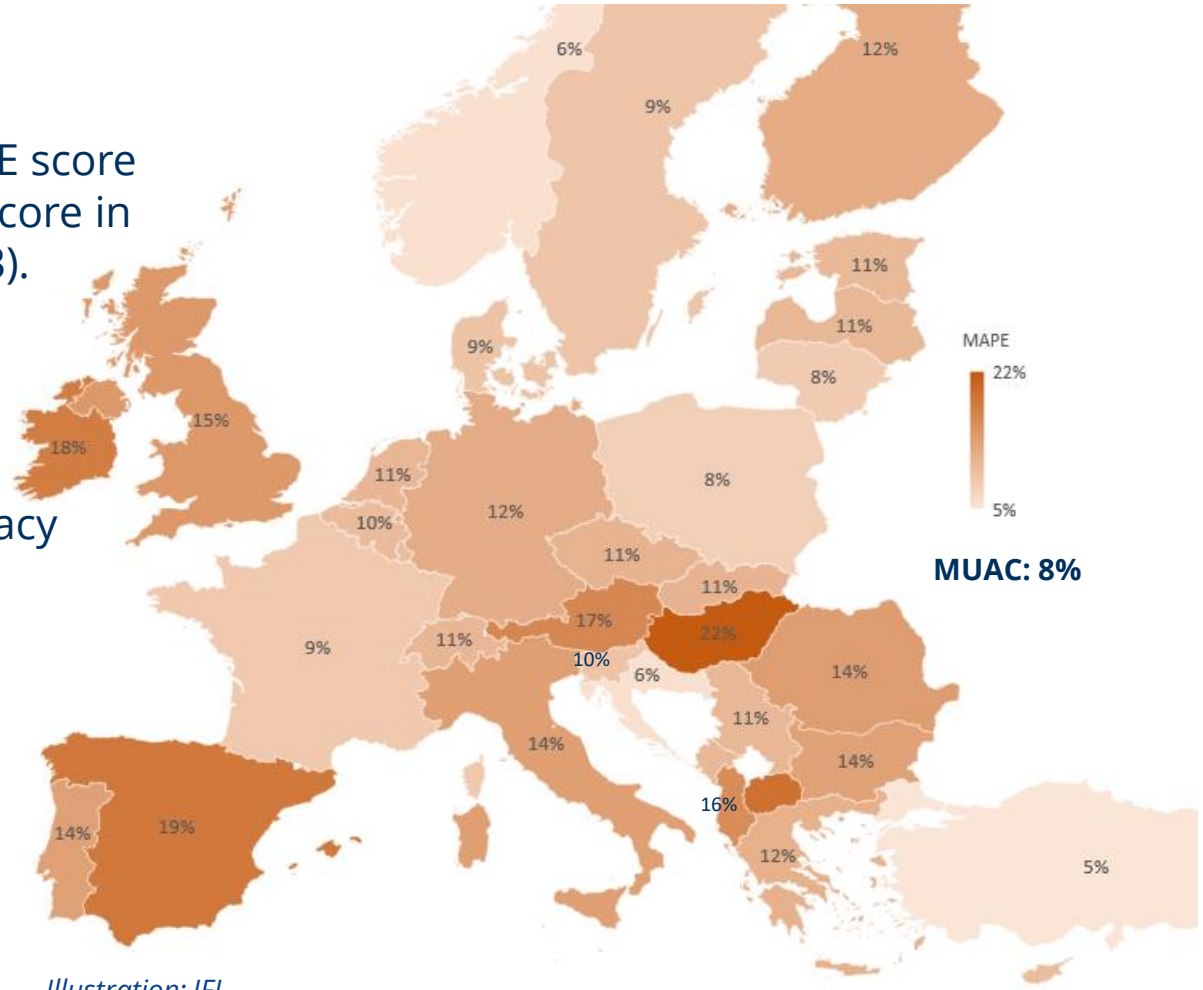


Illustration: IFL

Quality Assessment

How often occur over- and under-estimations?

Results:

- 272 observations (8 years x 34 ANSPs)
  - 167 over-estimations (61%),
  - 105 under-estimations (39%),
- 26.4 Mio. Flights*
  - Up to 20.4 Mio. over-estimated Flights,
  - 6.0 Mio. under-estimated flights.
- Is there a **systematic** over-estimation?

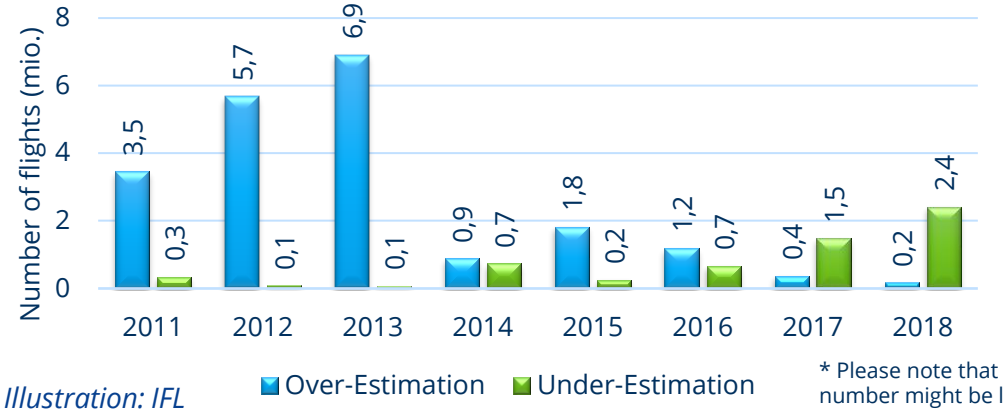


Illustration: IFL

* Please note that the actual number might be lower, since flights cannot be aggregated spatially.

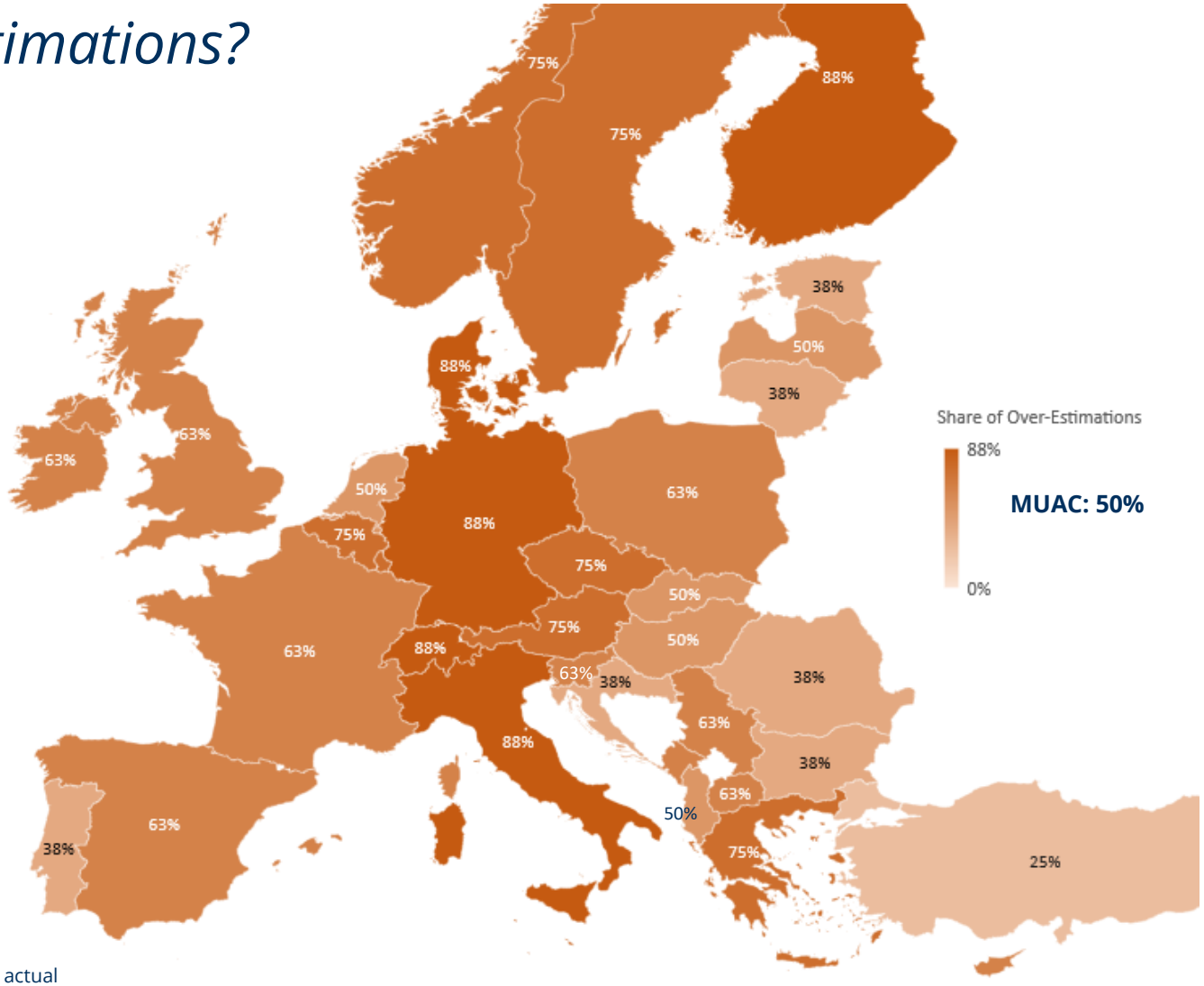
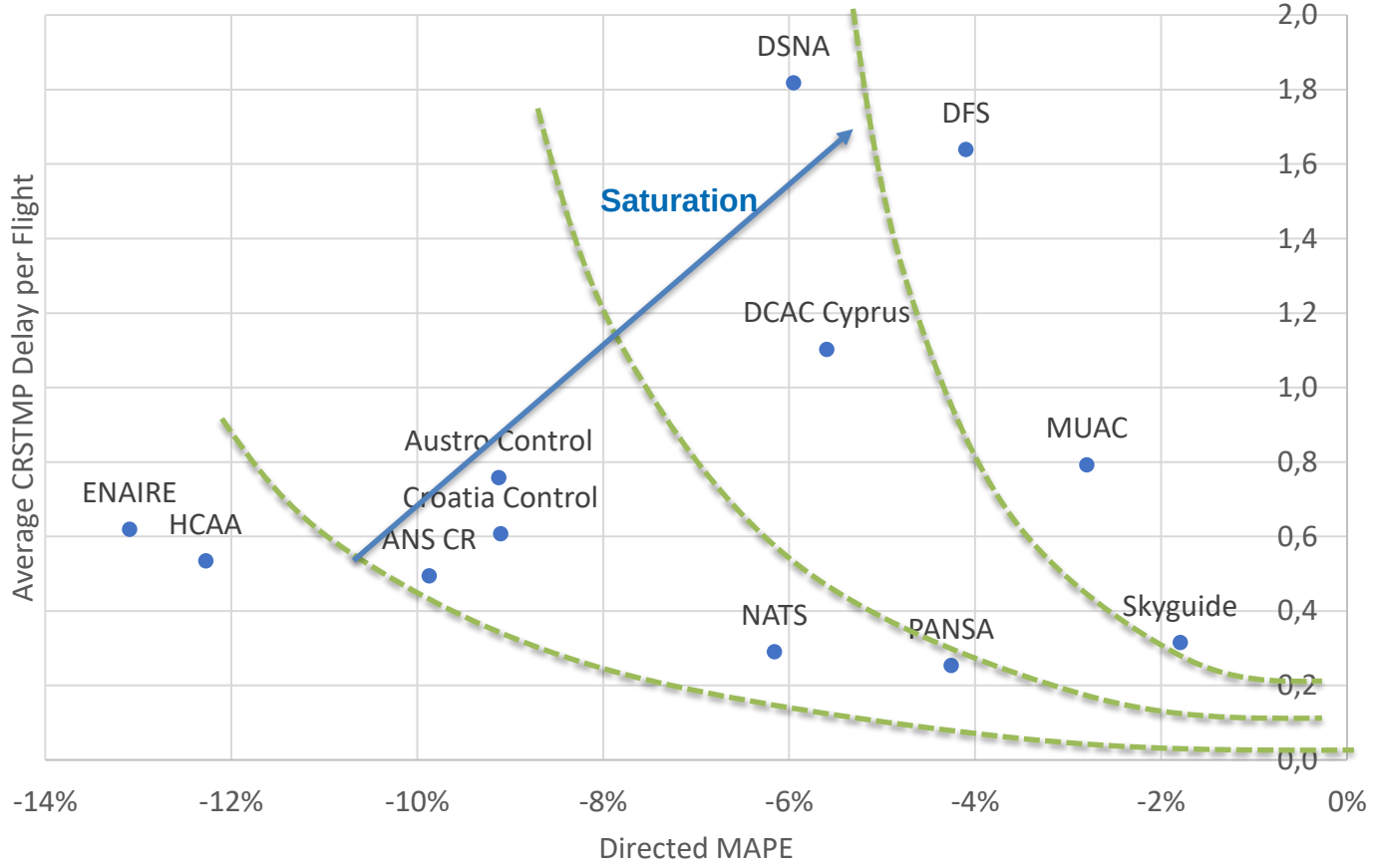


Illustration: IFL

Way forward

How to trigger forecast development with regard to ANSP performance?

- **Fact**: Delay increases exponentially with Demand.
- **Thesis:** Delay growth results in increasing under-estimation. The saturation may however follow different slopes.
- **Consequence:** Over-estimations may result from faulty slope modeling. Both effects do impact ANSPs performance.
- **Way forward**: Consideration in **economic benchmarking** and **target-setting**. The **interdependencies** will be checked by applying **regression analysis**.



Data: STATFOR Seven-Year Forecasts, 2006-2019, Illustration: IFL

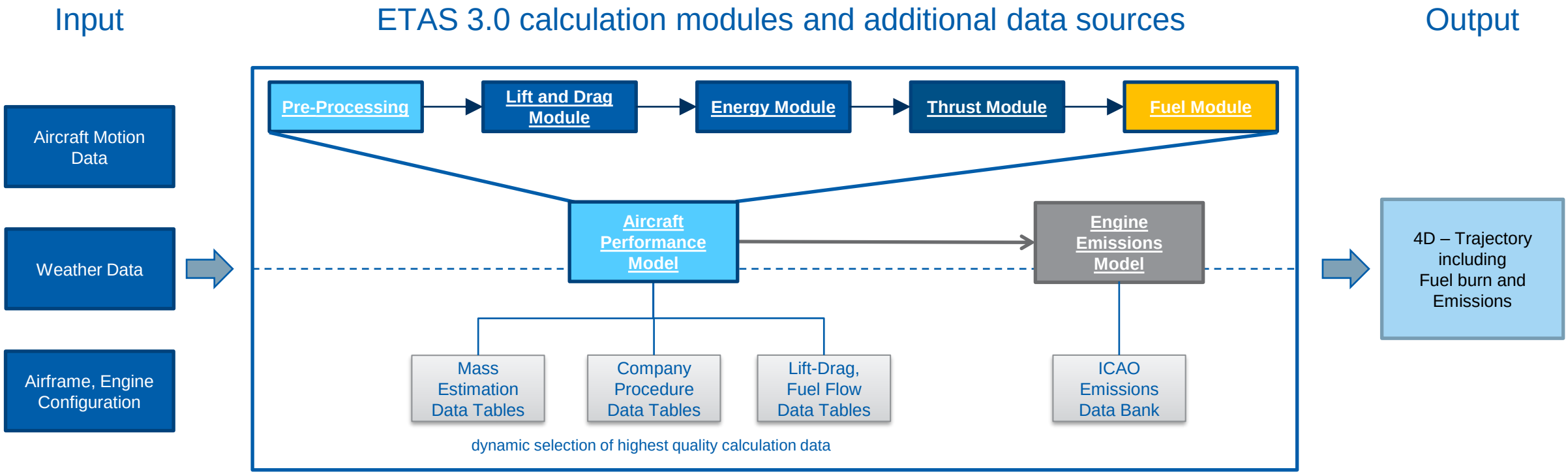
Wrap-up
What have we learned? How to go on?

1. The spread between „high“ and „low“ scenario has significant consequences for ANSPs with regards to resource- and cost-planning. However, 61% of observations did not match the 50% confidence interval.
2. The time horizon crucial for results (5 vs. 7 years?), but not considered in quality assessment by STATFOR.
3. Other quality indicator candidates exist. We tried (and adopted) MAPE, widely used in other industries. Considering all 34/36* ANSPs and eight time periods,
   - 148 observations (54%) showed a score below 10% (good quality).
   - For 82 observations (30%), MAPE score was below 5% (very good quality).
4. Injection to RP3 / RP 4 Performance Scheme
   - “ecology” as add-on to “capacity” and “cost eff.”
   - STATFOR forecasts should include predicted **fuel consumption** and **emissions,** the **overall performance** assess realised emissions (RP/MAPE).
   - Candidate: DFS runs **ETAS (Enhanced Trajectory Assessment System)**, already known to EUROCONTOL. Robust emission estimates for thousands of flights in German Airspace overnight since 2016.
   - Could be linked to any STATFOR forecast, can rely on EUROCONTROL BADA 4.
   - Accuracy (RP, MAPE), Risk and KPI can directly consider these additional variables.

* Depends on year

Wrap-up

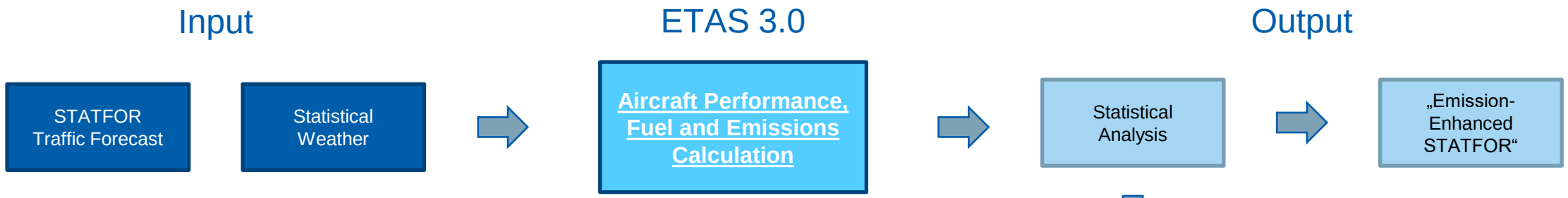
ETAS: How to link-in ecology in the STATFOR process (1)?



Wrap-up

ETAS: How to link-in ecology in the STATFOR process (2)?

Prediction / Planning Mode:



Performance Evaluation Mode:

Customer Contact: DFS OA/L – ATM Data & Services – frank.weber@dfs.de

Questions & Comments ?

The Accuracy of Forecasts

InterFAB Expert Talks: ATM performance data - can we do better?

02 March 2021

Prof. Dr.-Ing. habil. Hartmut Fricke / M.Sc. Thomas Standfuß
TU Dresden
Institute of Logistics and Aviation
Chair of Air Transport Technology and Logistics
hartmut.fricke@tu-dresden.de