



Using the OpenSky Network for Data-Driven Research

Engage TC2 Workshop, 25 January 2021

Dr. Martin Strohmeier, OpenSky Network (strohmeier@opensky-network.org)

Flight Schedule

- What is OpenSky?
- Interfaces for researchers
- Recent data-driven trajectory research
- Covid-19 & other use cases

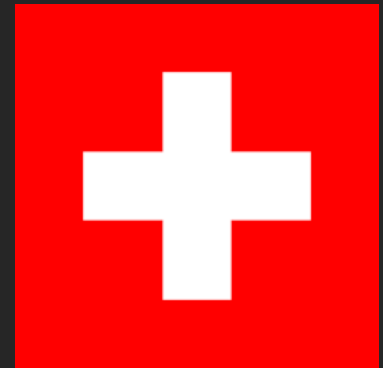
The OpenSky Network e.V.



Research-oriented



Non-Profit Association



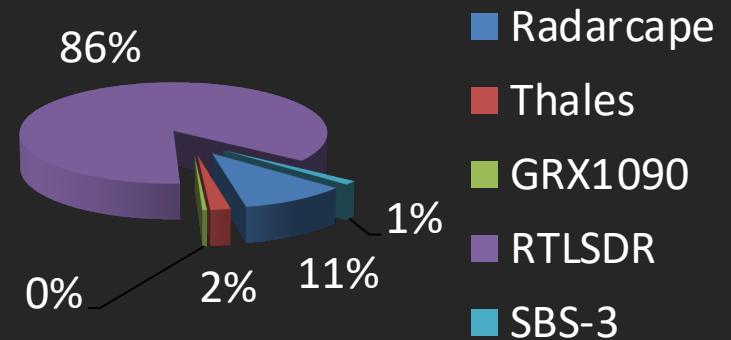
Based in Switzerland

Why collect large-scale ATC data?

- ANSPs operate massive ground infrastructure, but data not accessible due to numerous reasons:
 - Security concerns
 - Legal constraints
 - Privacy concerns
 - Liability issues
 - Commercial value
 - ...

Crowdsourcing ATC

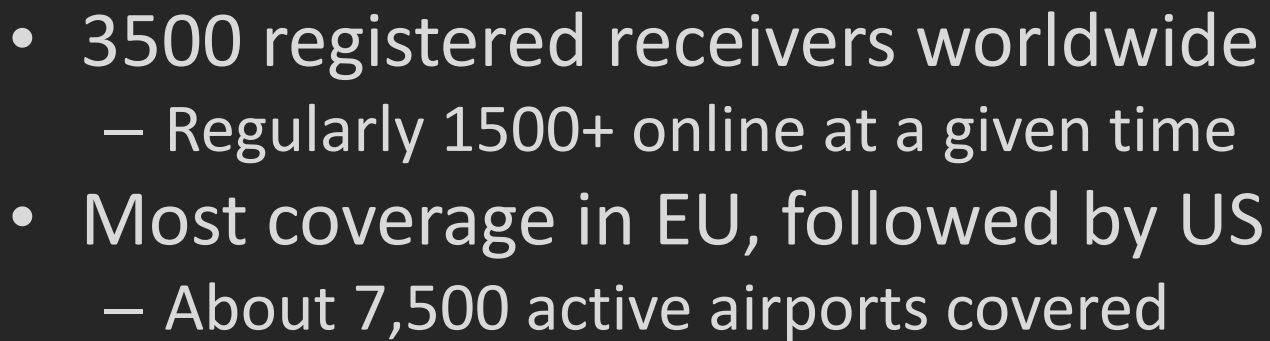
Use of cheap software-defined radio receivers proliferating around the world because of FlightAware, Flightradar24 and a lot of enthusiasts!



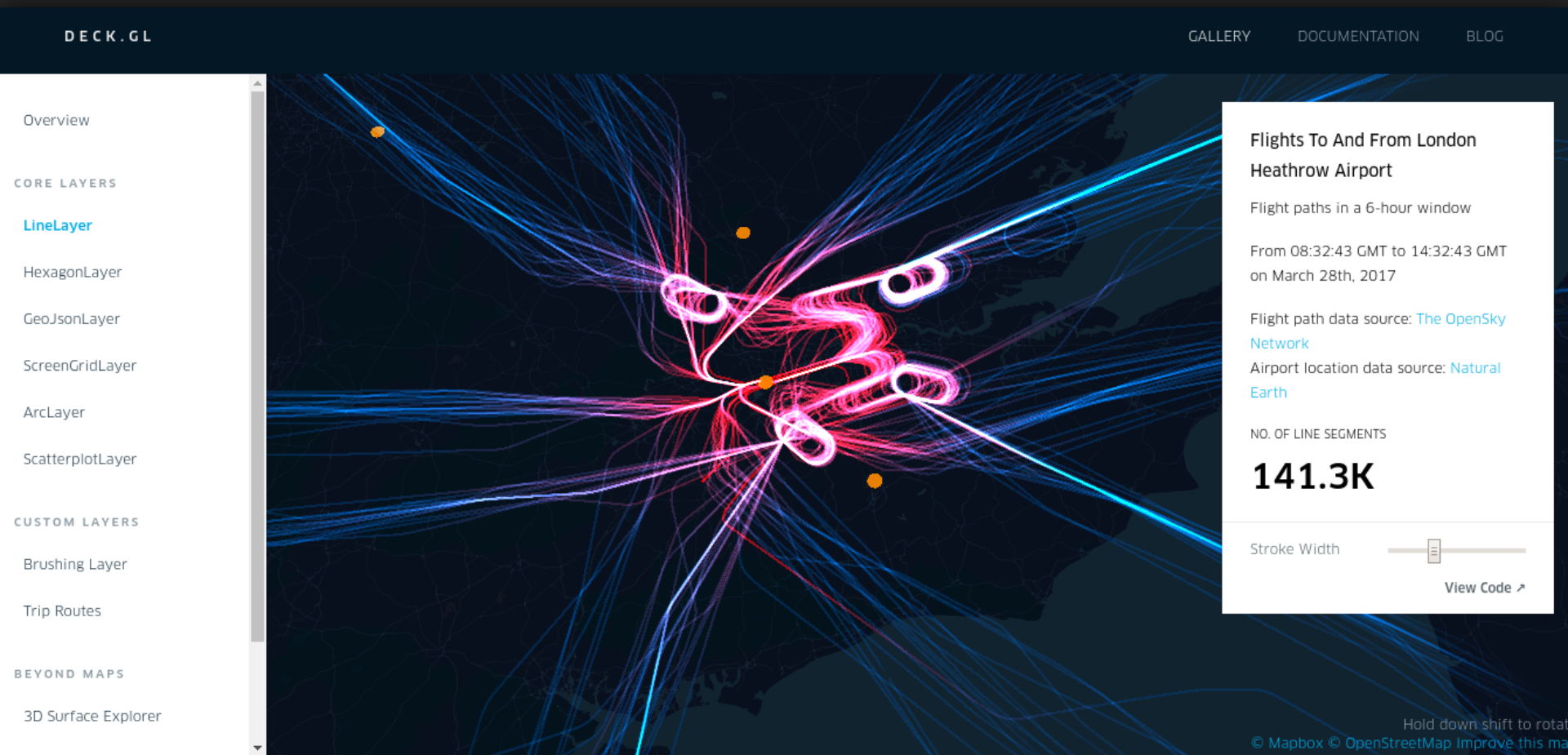
Mission & Achievements since 2012

“Our goal is to improve the **safety, security, efficiency,** and **reliability** of airspace usage by providing **open air traffic data**”

- Association provides free data to not-for-profit research
 - Contributed to over 150 published scientific studies
- Network collects ADS-B, Mode S, TCAS, FLARM
 - Data set contains almost 25 trillion signals recorded since 2013, equalling 2 Petabytes of data



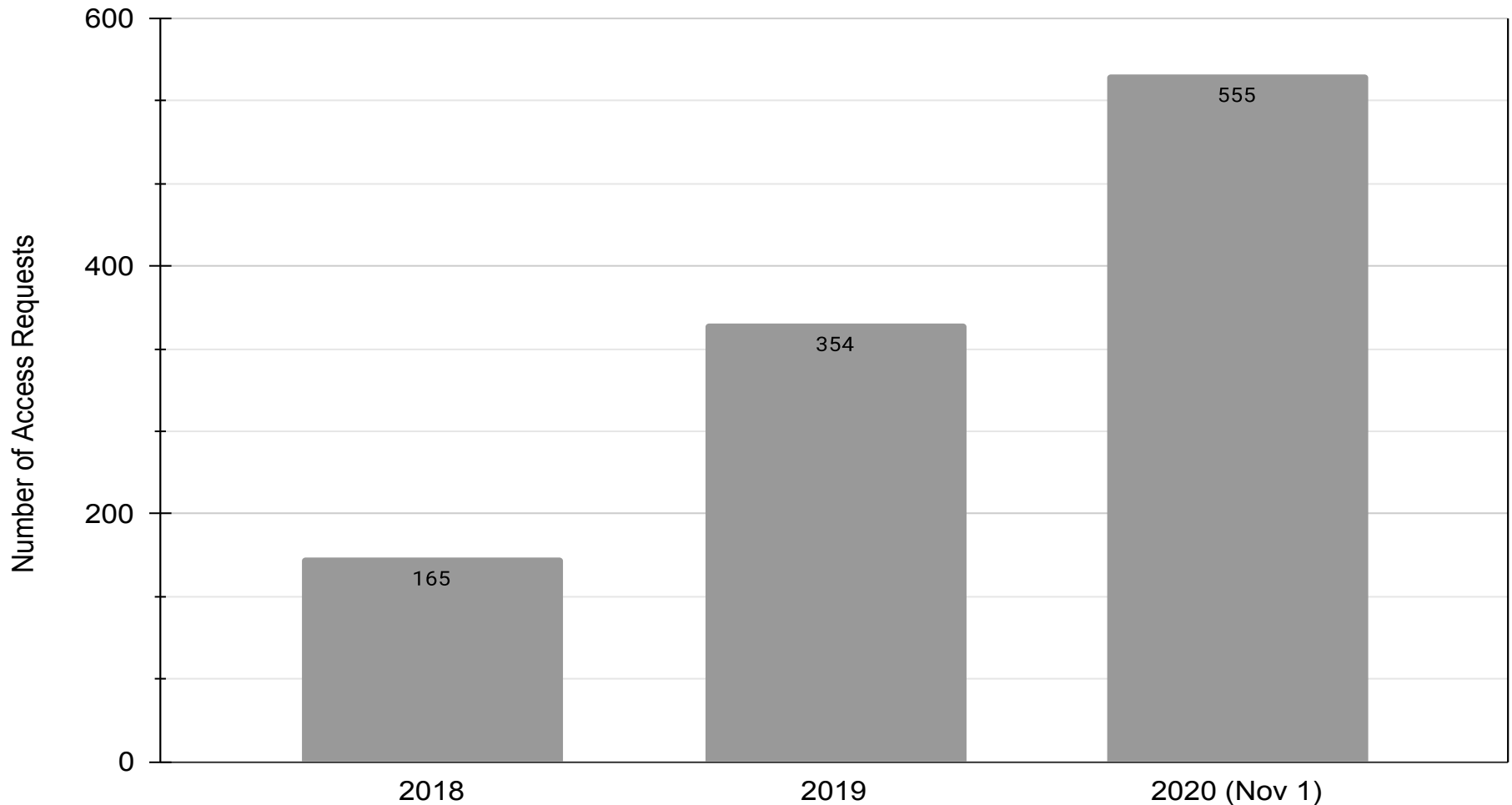
#dataviz



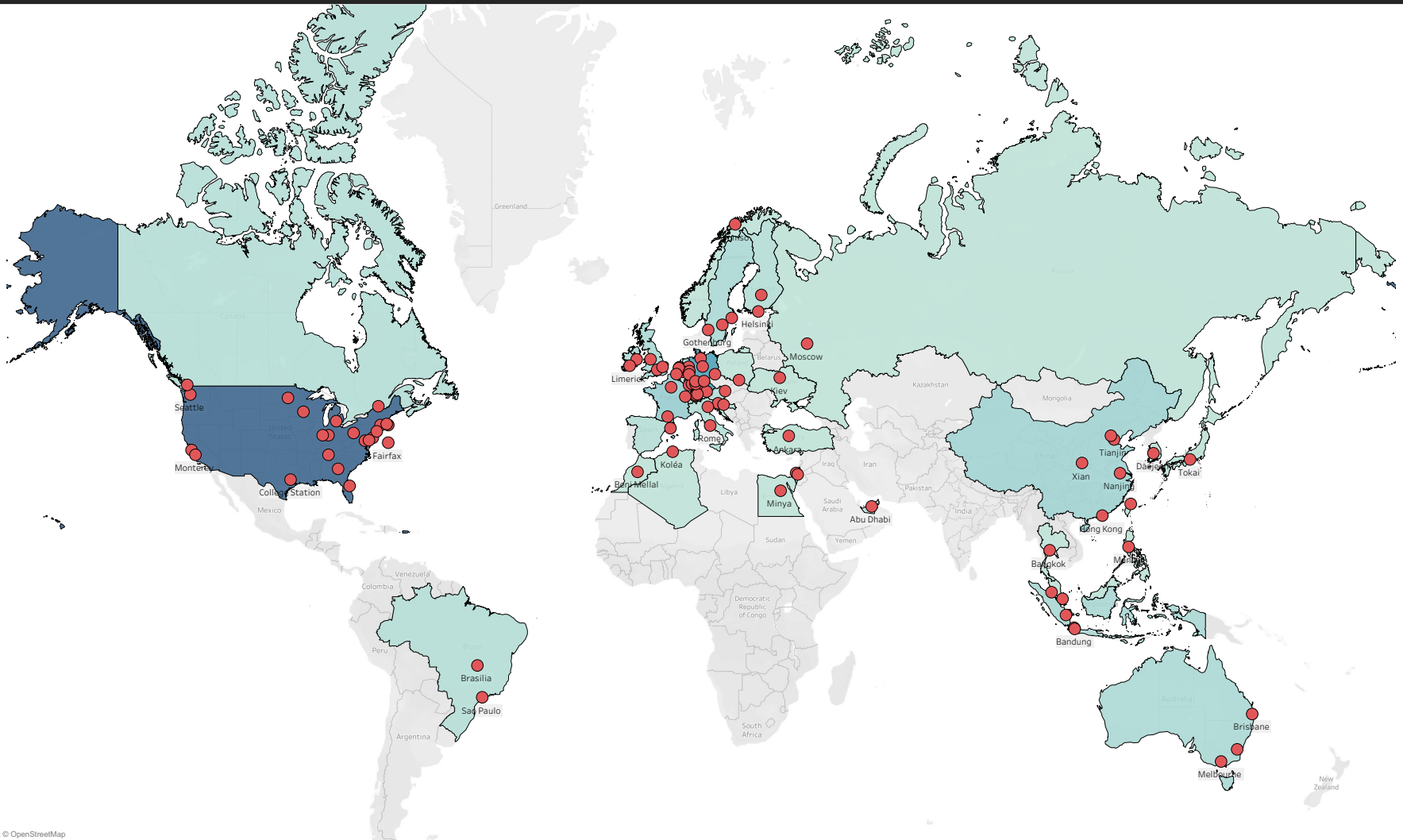
<https://deck.gl/examples/line-layer/>

OpenSky Data Requests

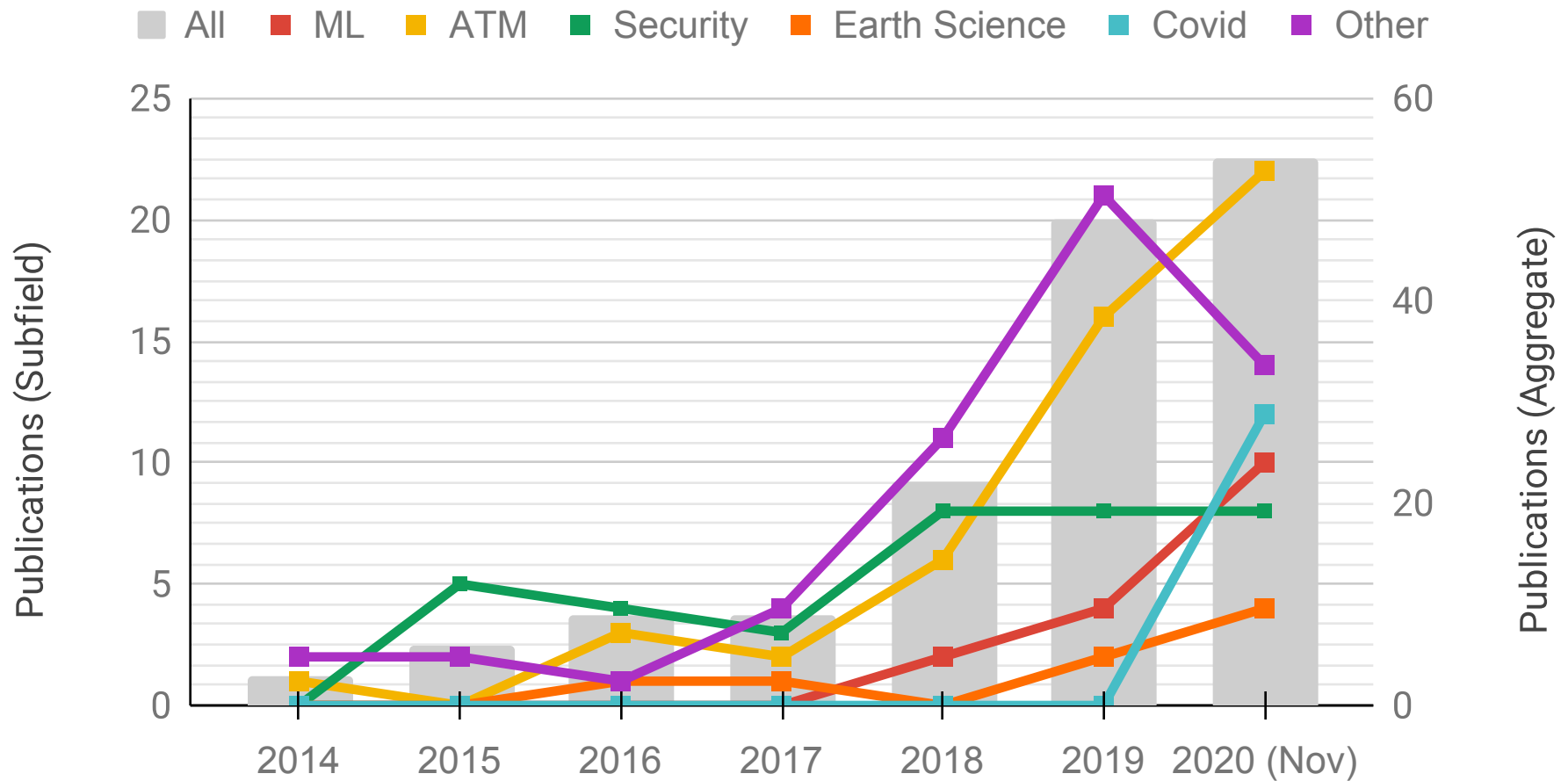
Breakdown of Data Access Requests between 2018 and 2020



Global Academic Requests 2019



Academic Publications using OpenSky



Source data: <https://opensky-network.org/community/publications>

OpenSky Symposium

The 8th OpenSky Symposium

12/13 November 2020 at EUROCONTROL in Brussels, Belgium

[Read more](#)

<https://symposium.opensky-network.org>

Interfaces for Researchers

Researcher's Requirements

- ✓ Record data over long periods
 - Typical requests ask for 1-12 months of data, but some also asked for 5+ years, e.g., to analyze progress
- ✓ Keep historical data
 - Data from specific points in time are needed, e.g., for incident investigation
- ✓ Keep low-level data
 - Every single transmission needed for frequency usage analyses or applications such as MLAT
- ✓ Calculate higher level abstractions
 - Most researchers are not very technical and do not want to deal with tons of low-level data
- ✓ Record data over large geographical areas
 - Requests for data from specific regions, countries, continents, or the whole world

Access to OpenSky Data

- Live API:
 - Free for non-commercial use
 - Data -> Live API (<https://opensky-network.org/apidoc/>)
 - Limitations for anonymous and normal users
- Access to historical data (Live and via API)
 - Explicitly for research; free for universities, research institutions and flight authorities
 - Data -> Apply for Data Access
 - Should get an answer within a few days (by me)
- Pre-prepared Scientific Datasets
 - 8 different types: <https://opensky-network.org/data/datasets>
 - From raw data at message level to abstract metadata

Impala Database

[Network](#)[Community](#)[Data](#)[About](#)[Contribute](#)[My OpenSky](#)[Home](#) > [Data](#) > [Historical Database](#)

A Quick Guide To OpenSky's Impala Shell

Besides our public API, we also grant free access to our full dataset over an SQL-like query interface to researchers and partners. If you don't have access to this free premium service yet, but good reasons to get it, you can apply for access by submitting this [application form](#) (account+login required). Note that this interface is mainly intended to serve university-affiliated researchers, governmental organisations, or aviation authorities as a data source for aviation-related research or incident investigation. If you are a private or commercial entity, please contact us for a license.

Since processing large queries is expensive in terms of resources and cluster energy consumption, we reserve the right

```
Last login: Tue Nov  3 18:02:51 2020 from 144.2.113.173
Starting Impala Shell without Kerberos authentication
*****
*****
OpenSky Network Impala Shell
*****
*****
[hadoop-1:21000] >
```

Access to OpenSky Data

- Fill in the forms:
 - Information about you
 - Institutional email address! (proof)
 - Information about your project and what you need the data for
 - Commercial use?

Historical Data Access Application

Please fill out the form below to request access to the historical data. All necessary information is marked with an asterisk (*).

Personal Information

Institution *	<input type="text" value="University or company name"/>
Full name *	<input type="text" value="Your full name"/>
Institutional Mail *	<input type="text" value="Your email address at your institution"/>
Role *	<input type="text" value="e.g. 'professor' or 'developer'"/>
Advisor	<input type="text" value="e.g. Prof. Dr. John Smith"/>

Project Information

Project URL	<input type="text"/>
Description *	<div><input type="text"/></div> <p>The description must contain at least 300 characters! 0</p>
Other users	<div><input type="text"/></div> <p>List all other persons who will work with the data. Make sure to include their roles and affiliations, e.g. "Tom Smith, project member, developer" or "Paul Lennart, company xy, data researcher".</p>
Commercial use *	<input type="text" value="Yes"/>

Impala Database Usage

- Overall, 415 distinct users since 2018
- 16,077,613 queries submitted

	2018	2019	2020 (31 Oct)	Total
Distinct Active Users	94	134	295	415
Submitted Queries	6,163,467	4,094,815	5,818,746	16,077,613

Trajectory-related Research

Implementation of Reference Trajectories for
Review

Predicting Airplane Go-Arounds Using Machine Learning and Open-Source Data†

by  **Benolt Figuet** 1,* 
enrico.spin...
rainer.koelle@e

Using Open Source Data with Machine Learning

by  **Gong Chen** 1,*   **Junzi Sun**

Flight clustering

Filippo Sestini

Detecting Events in Aircraft Trajectories: Rule-Based and Data-Driven Approaches†

by  **Xavier Olive** 1,*   **Junzi Sun** 2,*   **Adrien Lafage** 1,*   **Luis Basora** 1   **Go Wallenburg** and  **Ostap Okhrin** 1 

Time Prediction

Trajectory Clustering within the Terminal Airspace Utilizing a Weighted Distance Function†

by  **Samantha J. Corrado** *   **Tejas G. Puranik**   **Olivia J. Pinon**  and  **Dimitri N. Mavris** 

New Data Tools

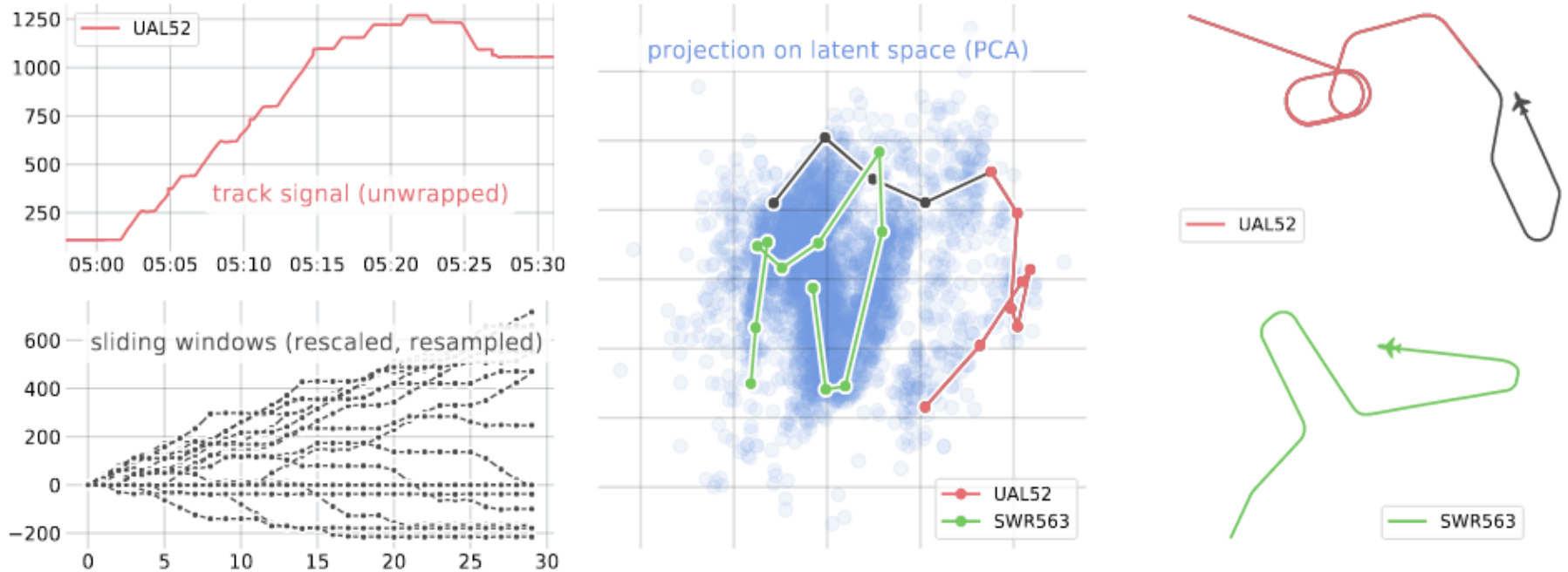


Figure 9. The 19,480 trajectories are split into sliding windows, rescaled and resampled (30 samples per window). The resulting 72,353 samples are then projected with a PCA. Holding patterns cluster in the latent space: the red part of the trajectory is well identified as a holding pattern, whereas the green trajectory, in spite of a pattern easily mistaken with rule-based models, stays in the regular cluster.

<https://opensky-network.org/data/data-tools>

<https://airtraffic.readthedocs.io>

Olive, Xavier. "Traffic, a toolbox for processing and analysing air traffic data." (2019).

Usage during Covid-19

Covid-19: Requests & Dataset

[Upload](#)[Communities](#)[Log in](#)[Sign up](#)

September 17, 2020

Dataset

Open Access

Crowdsourced air traffic data from The OpenSky Network 2020

Xavier Olive; Martin Strohmeier; Jannis Lübbe

Motivation

The data in this dataset is derived and cleaned from the full OpenSky dataset to illustrate the development of air traffic during the COVID-19 pandemic. It spans all flights seen by the network's more than 2500 members since 1 January 2019. More data will be periodically included in the dataset until the end of the COVID-19 pandemic.

License

See LICENSE.txt

Disclaimer

The data provided in the files is provided as is. Despite our best efforts at filtering out potential issues, some information could be erroneous.

- Origin and destination airports are computed online based on the ADS-B trajectories on approach/takeoff: no crosschecking with external sources of data has been conducted. Fields **origin** or **destination** are empty when no airport could be found.
- Aircraft information come from the OpenSky aircraft database. Fields **typecode** and **registration** are empty when the aircraft is not present in the database.

7,979

views

7,658

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

September 17, 2020

DOI:

DOI [10.5281/zenodo.4088202](https://doi.org/10.5281/zenodo.4088202)

Keyword(s):

aviation

flightlist

ads-b

covid19

<https://doi.org/10.5281/zenodo.3737101>

Covid-19: Pandemic Analysis

Science

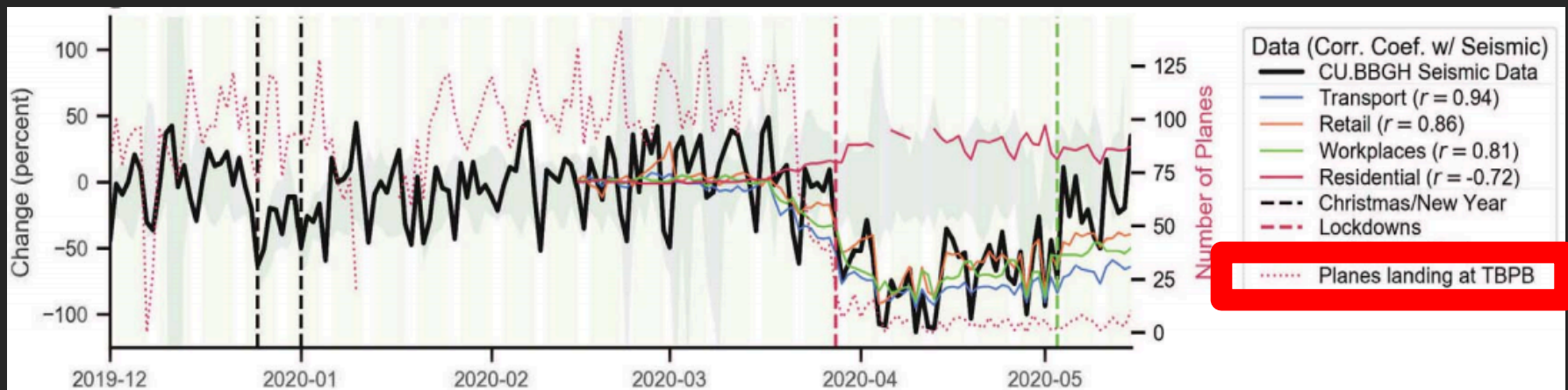
[Contents ▾](#)[News ▾](#)[Careers ▾](#)[Journals ▾](#)[Read our COVID-19 research and news.](#)**SHARE****REPORT**

Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures

Thomas Lecocq^{1,*}, Stephen P. Hicks², Koen Van Noten¹, Kasper van Wijk³, Paula Koelemeijer⁴, Raphael S...

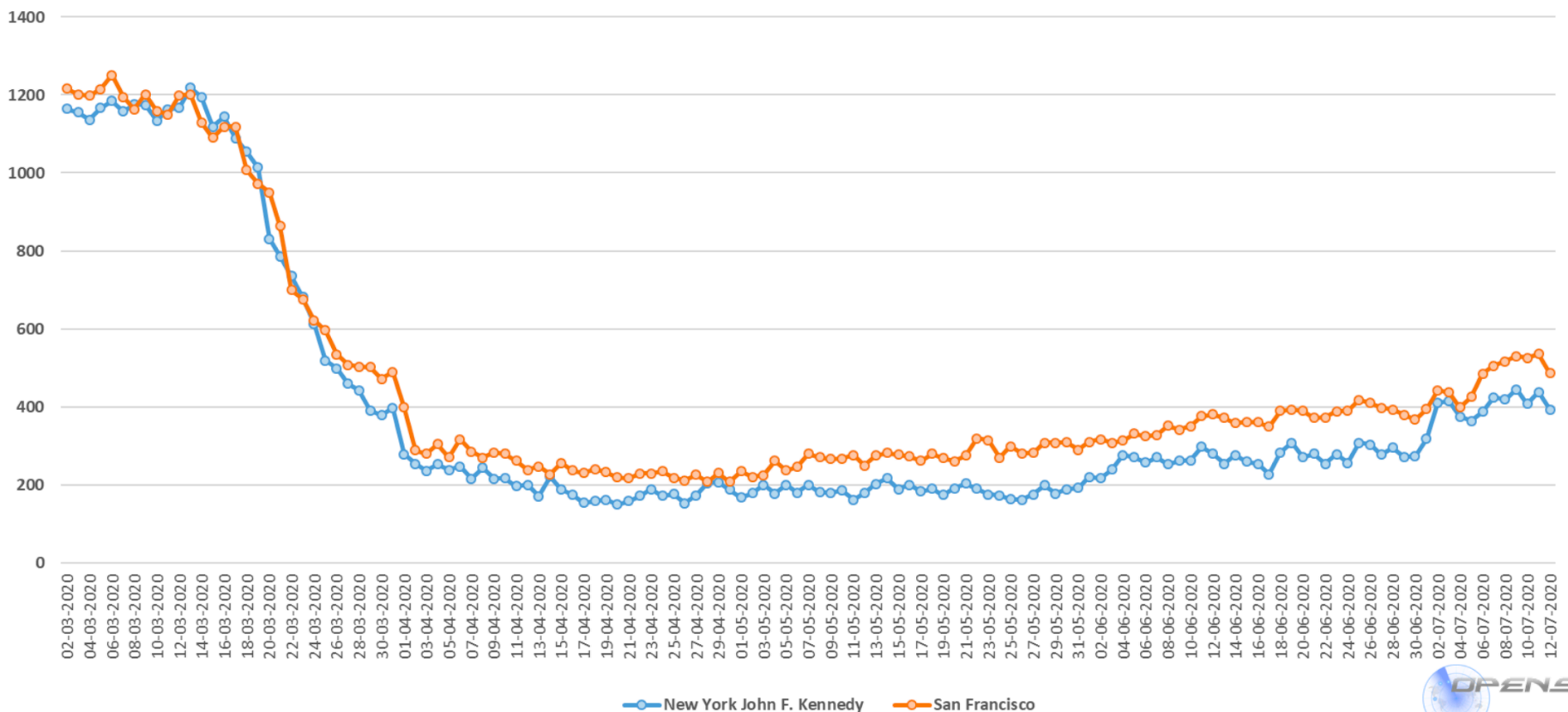
[+ See all authors and affiliations](#)

Science 11 Sep 2020:
Vol. 369, Issue 6509, pp. 1338-1343
DOI: 10.1126/science.abd2438



Covid-19: Pandemic Analysis

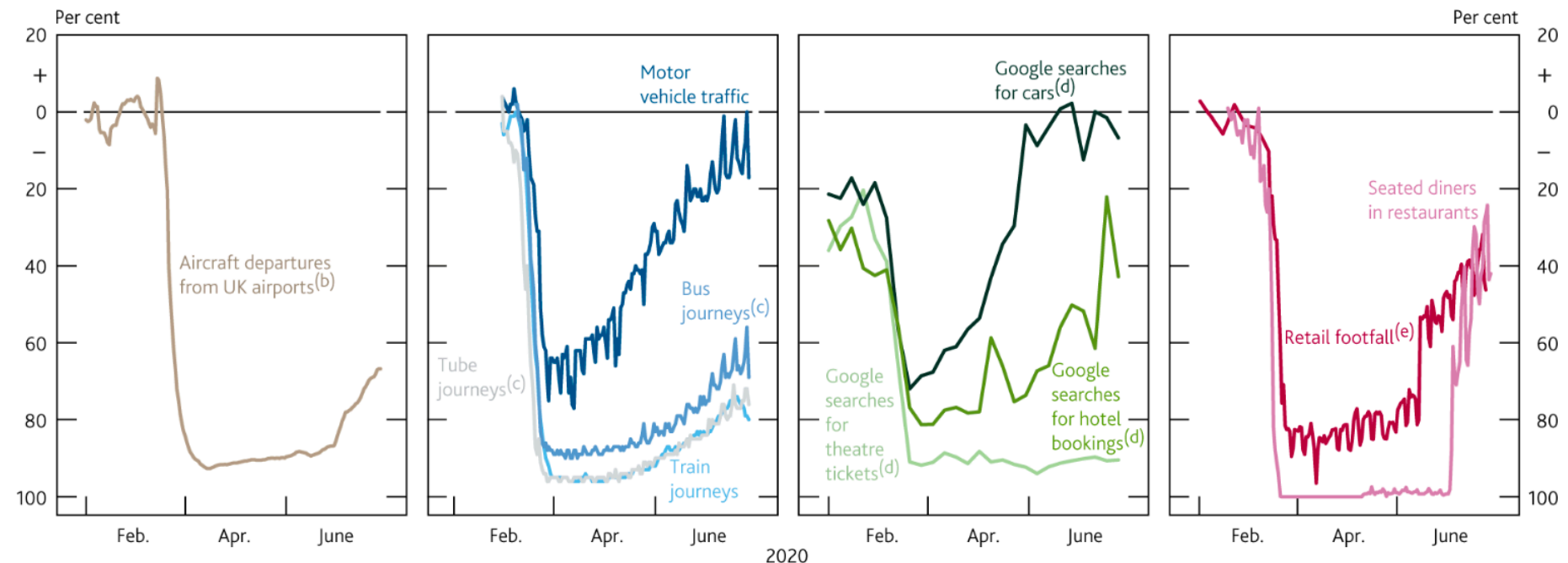
Domestic Traffic variation since mid February
at main airports in the US



Covid-19: Economic Nowcasting

Chart 2.18 High-frequency indicators show the change in behaviour that resulted from the pandemic

High-frequency indicators of economic activity^(a)



Sources: Department for Transport, Google Trends, OpenSky Network, OpenTable, ShopperTrak and Bank calculations.

(a) Data are not seasonally adjusted. Road and rail travel data are shown relative to normal levels. All other data are shown relative to a year earlier.

(b) Seven-day moving averages of flight departures tracked by the OpenSky Network from Birmingham, Gatwick, Heathrow, Luton, Manchester and Stansted airports.

(c) The number of tube journeys is based on Transport for London data and the number of bus journeys do not include London buses.

(d) Google searches data are weekly averages of the Google Trends index of UK search volumes. Searches for cars shows the average of changes in search volumes for the six highest-selling car brands in the UK.

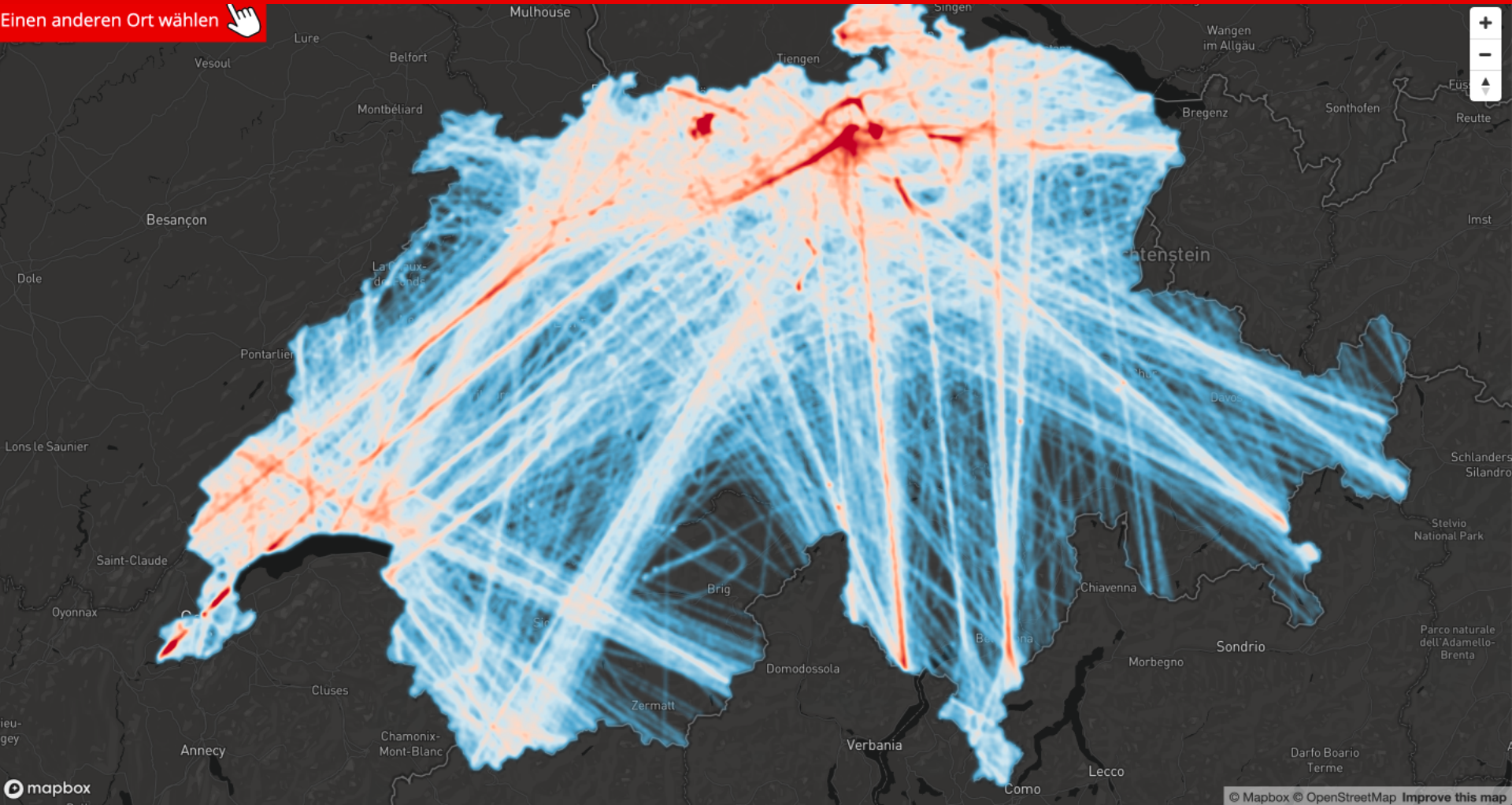
(e) Weekly changes to 14 March and daily thereafter. Data have been adjusted to remove distortions caused by bank holidays.

Non-Academic Use Cases

Data Journalism

Blick

Einen anderen Ort wählen 



Recreational

  LiveTraffic 1.22.190829

★★★★★ (131 reviews)



By TwinFan

[Find their other files](#)



Call Sign | Squawk UAE440 0577
Flight: Route

Simulated Time 2019-03-01 22:38:08
Last Data [s] | Chnl +57.4 OpenSky Live...
Position 25.2239 N / 55.2002 E
Bearing | Dist. [nm] 232 0.1

UAE440 Climb 8690 272			X
A/C key	896175	<input type="checkbox"/> AUTO	
Registr.	A6-EWC		
ICAO Type Class.	B77L	L2J	
Manufacturer	Boeing		
Model	Boeing 777 21HLR		
Operator	UAE Emirates Airline		
CSL Model	BB_Boeing B77L_UAE		
Call Sign Squawk	UAE440	0577	
Flight: Route			
Simulated Time	2019-03-01 22:38:08		
Last Data [s] Chnl	+57.4	OpenSky Live...	
Position	25.2239 N / 55.2002 E		
Bearing Dist. [nm]	232	0.1	
Flight Phase	Climb		
Gear Flaps	0.0	0.0	
Lights	nav/bcn/stbr/----/land		
Heading [°]	187		
Pitch Roll [°]	3	-9	
Altitude AGL [ft]	8690	8690	
Speed [kn] VSI [ft]	272	1501	
<input checked="" type="checkbox"/> Camera	<input checked="" type="checkbox"/> Visible		<input data-bbox="1796 1185 1816 1206" type="button" value="?"/>

How to contribute?

- Join the team
 - People can become a community manager and actively recruit new feeders
- Donations
 - We have created a “sponsoring program” that allows for supporting specific receiver locations
 - Infrastructure is expensive and requires institutional support
- Data
 - People and organizations can “donate” or exchange (meta) data
- Events
 - People can help organize or participate in events such as our symposium
- Marketing
 - Tweet, write, publish, support officially, and talk about OSN
- Code
 - Write and publish open-source tools using or improving OSN

Questions?

