The technical development of the TIGGE and S2S databases

Manuel Fuentes
Products Team, ECMWF
Chair TIGGE Panel
Chair S2S Technical Committee

Thanks to many colleagues @ ECMWF & TIGGE S2S contributors

ECMWF

Outline

- Creation of TIGGE & S2S Databases

- Status and usage statistics

- User surveys

- Future plans

TIGGE: THORPEX Interactive Grand-Global Ensemble

- A major component of THORPEX: a WMO World Weather Research Programme to accelerate the improvements in the accuracy of 1-day to 2-week high-impact weather forecasts

- Objectives:

    – **Enhance collaboration on ensemble prediction**, both internationally and between operational centres & universities.

    – **Facilitate research on ensemble prediction methods**, especially methods to combine ensembles and to correct systematic errors

    – Enable evolution towards a prototype operational system, the "**Global Interactive Forecast System**" GIFS

TIGGE: The beginning

- 1st TIGGE Workshop at ECMWF 1-4 March 2005:

    – Initially develop database of available ensembles, collected in near-real time

    – Users: Who? What for ? How ? When ?

    – Requirements: What data ? Format ? How to access ? Contributors ?

- Recommendations:

    – Create a TIGGE database of ensemble forecast data from interested production centres

    – Establish an Archive Working Group to decide on technical details

        • Meeting at ECMWF in 9-10 November 2005

TIGGE Database

- 3 Archive Centres

  – CMA, NCAR, ECMWF

- 8 Data providers (?)

  – NCEP, ECMWF, UKMO, JMA, BMRC, CPTEC, KMA, MSC

- Each Archive Centre will receive data from all the Data Providers

  – In near realtime

- Users will be able to get the same data from any of the Archive Centres

- No extra resources!

  – Use existing infrastructure

TIGGE: Key elements for success

- Homogeneity was paramount: the more consistent the archive, the easier to develop applications

  - Common terminology

NCEP	ECMWF
Precipitation (precip)	TP (total precipitation)
T 2 meters (t2m)	2T

  - Common data format

    - GRIB 1, GRIB 2, NetCDF, same units

  - Definition of a core dataset, a field uniquely identified by the following tuple:

    - base date, base time, time step, origin centre, ensemble number, level, parameter

  - Completeness: the objective was to have 100% complete Archives

- Organisation of the collaboration

  - Success directly linked to commitment of the partners

  - Tools, emails lists, web sites

- Experience from other projects: DEMETER, ENSEMBLES

# TIGGE Archive WG: Decisions & Recommendations

- Data transport:

  – LDM to provide a many-to-many data distribution

  – Test transfer rates, tune the network

- Use GRIB2 as common data format.

  – Agree on variables codes, units and common terminology

  – Strong governance

- Define a core dataset to ensure homogeneity of the TIGGE database

  – Sample model output provided, sample programs provided

- Establish Archive management communications:

  – Collaborative tools, list of contacts points

  – Archive Centres acting as coordinators

- Data Providers must commit to send missing data, even if it means rerunning a forecast cycle

Summary of TIGGE database configurations

Centre	Ensemble members	Output data resolution	Forecast length	Forecasts per day	Fields (out of 73)	Start date
BoM*	33	1.50° x 1.50°	10 day	2	55	3 Sep 07
CMA	15	0.56° x 0.56°	10 day	2	60	15 May 07
MSC	21	0.9° x 0.9°	16 day	2	56	3 Oct 07
CPTEC	15	0.94° x 0.94°	15 day	2	55	1 Feb 08
ECMWF	51	N200 (Reduced Gaussian) N128 after day 10	15 day	2	70	1 Oct 06
JMA	51	1.25° x 1.25°	9 day	1	61	1 Oct 06
KMA	24	0.56° x 0.38°	10 day	2	46	28 Dec 07
Météo-France	35	1.50° x 1.50°	4.5 day	2	62	25 Oct 07
NCEP	21	1.00° x 1.00°	16 day	4	69	5 Mar 07
UKMO	24	0.83° x 0.55°	15 day	2	72	1 Oct 06

* Delivery of BoM data currently suspended

TIGGE data flow in 2006
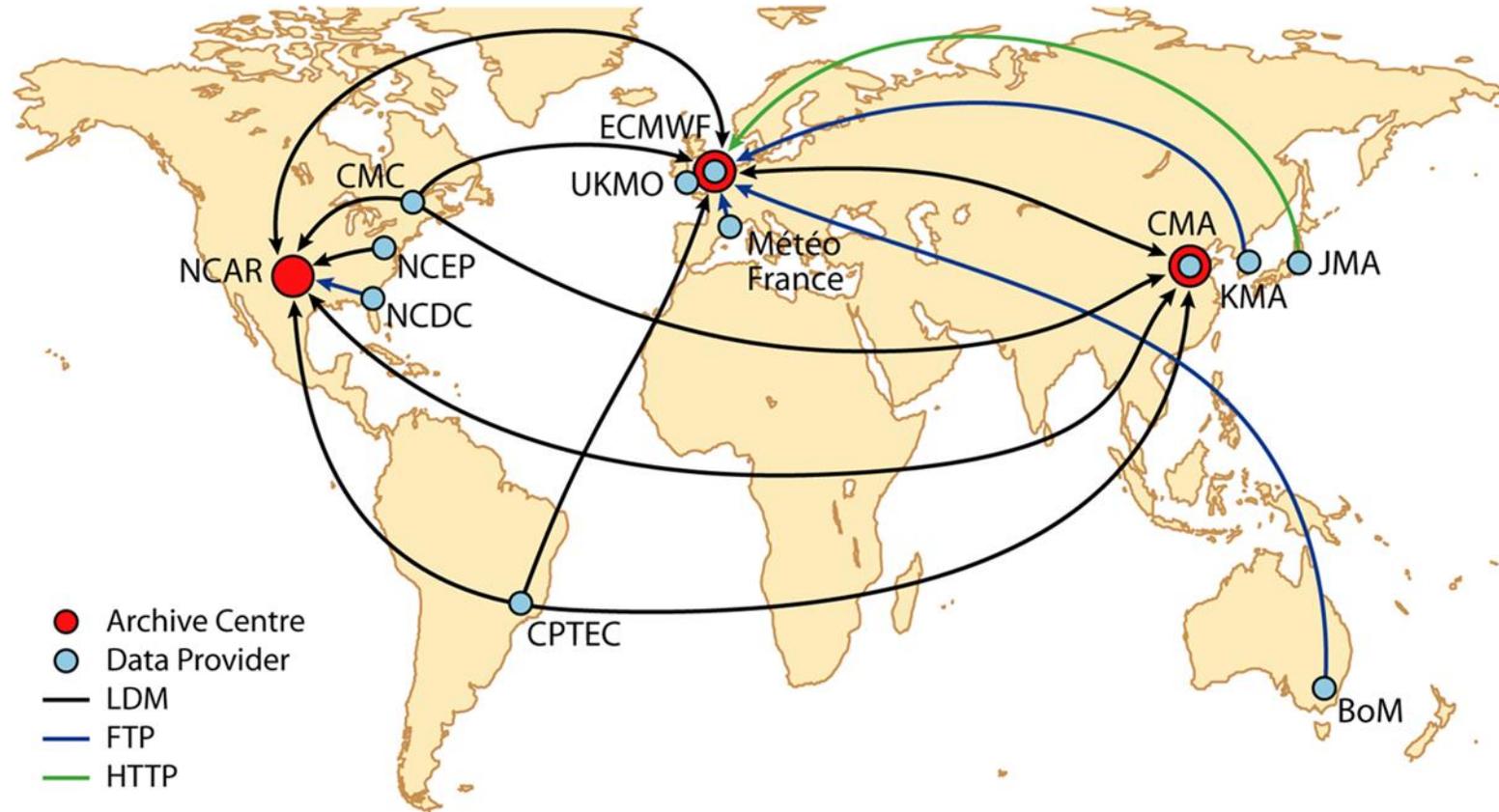
TIGGE Database: Phases

- Phase I: TIGGE-GIFS WG - 2006 until 2014

  – 3 Archive Centres, CMA, NCAR, ECMWF

  – BoM ceased in 2010

  – Extend TIGGE to Limited Area Models (TIGGE-LAM) from Jan 2013

- Phase II: PDEF WG - 2014 to 2019

  – At the end of THORPEX, commitment to continue for additional 5 years

  – 2 Archive Centres for gridded data: CMA, ECMWF

  – 1 Archive Centre for Tropical Cyclones in CXML: NCAR

  – NCMRWF (India) added in 2017

- Phase III ??

Sub-seasonal to Seasonal Project: S2S

In 2013 WWRP/THORPEX-WCRP established S2S project to improve forecast skill and understanding on the **sub-seasonal to seasonal** time scale, and promote its **uptake by operational centres** and **exploitation by the application community**.

- Phase I 2013 - 2018: Establish the S2S Database

- Followed the same key principles as TIGGE

  – Except many-to-many transfer was not needed, common 1.5/1.5 regular lat-lon grid

  – Heterogeneity: different reforecast methods, reforecast years, ensemble sizes

- Added difficulty of handling reforecast data in GRIB Edition 2

  – Submit proposals to WMO Expert Teams

- New variables:

  – Soil variables: soil moisture/temperature top 20cm/100cm, …

  – Daily averages

  – Ocean variables (on-going development)
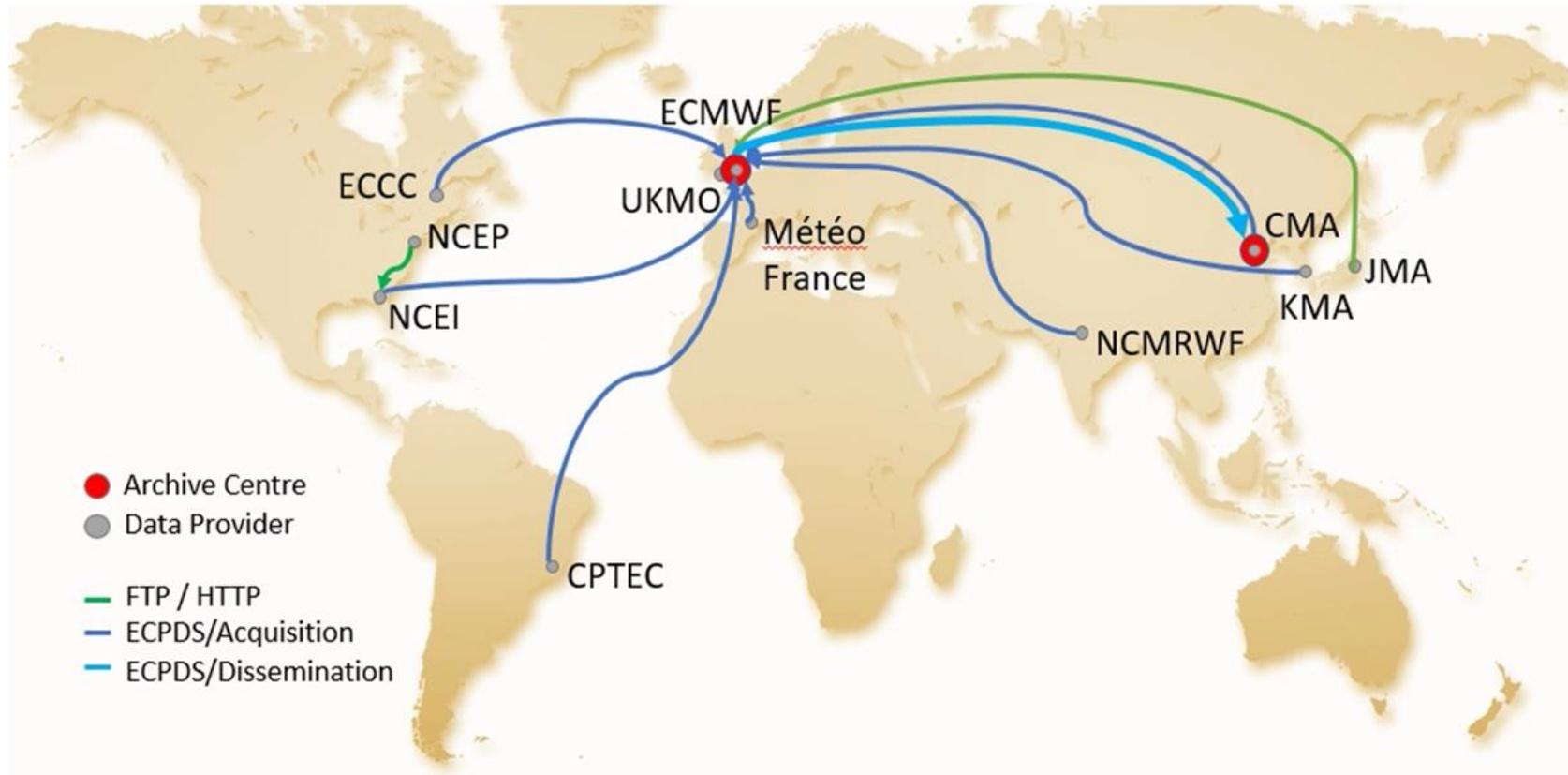
Summary of S2S Database configurations

Status on 2018-10-25	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size
BoM (ammc)	d 0-62	T47L17	3*11	2/week	fix	1981-2013	6/month	3*11
CMA (babj)	d 0-60	T106L40	4	daily	fix	1994-2014	daily	4
CNR-ISAC (isac)	d 0-32	0.75x0.56 L54	41	weekly	fix	1981-2010	every 5 days	5
CNRM (lfpw)	d 0-32	T255L91	51	weekly	fix	1993-2014	4/month	15
ECCC (cwao)	d 0-32	0.45x0.45 L40	21	weekly	on the fly	1998-2017	weekly	4
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10
JMA (rjtd)	d 0-33	Tl479/Tl319L100	50	weekly	fix	1981-2010	3/month	5
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2010	4/month	3
NCEP (kwbc)	d 0-44	T126L64	16	daily	fix	1999-2010	daily	4
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2016	4/month	7

Improvements of TIGGE & S2S databases

- TIGGE: many updates over 12 years

  – Migrate to use ECMWF's standard data transfer protocol ecPDS instead of LDM

- Delays downloading S2S datasets

  – Tapes were damaged due to over-use

  – Added more disk space to avoid reading many tapes

- Improved web infrastructure

  – Migrate Data Portal to new technologies (Django), providing powerful WebAPI

  – Provide visibility of better usage statistics

- User Surveys: engaging with the community

  – S2S (2017), TIGGE (2018)

TIGGE Data Flow in 2018

TIGGE: The International Grand Global Ensemble

Enhanced web infrastructure

TIGGE

Corporate Website	http://tigge.ecmwf.int/
Technical Website	http://confluence.ecmwf.int//display/TIGGE
Data Portal	http://apps.ecmwf.int/datasets/data/tigge/
History	http://apps.ecmwf.int/datasets/history/tigge-prod/
Usage statistics	http://confluence.ecmwf.int/display/TIGGE/Usage+statistics

S2S

| Corporate Website	http://s2s.ecmwf.int/
Technical Website	http://confluence.ecmwf.int//display/S2S/
Data Portal	http://apps.ecmwf.int/datasets/data/s2s/
History	http://apps.ecmwf.int/datasets/history/s2s-prod/
Usage statistics	http://confluence.ecmwf.int/display/S2S/Usage+statistics

TIGGE & S2S databases: status & usage

Contents	TIGGE	S2S
Archive size	3.3 PiB	98 TiB
Number of fields	7 billion	2 billion
Time span	2006-present (12 years)	2015-present (4 years)

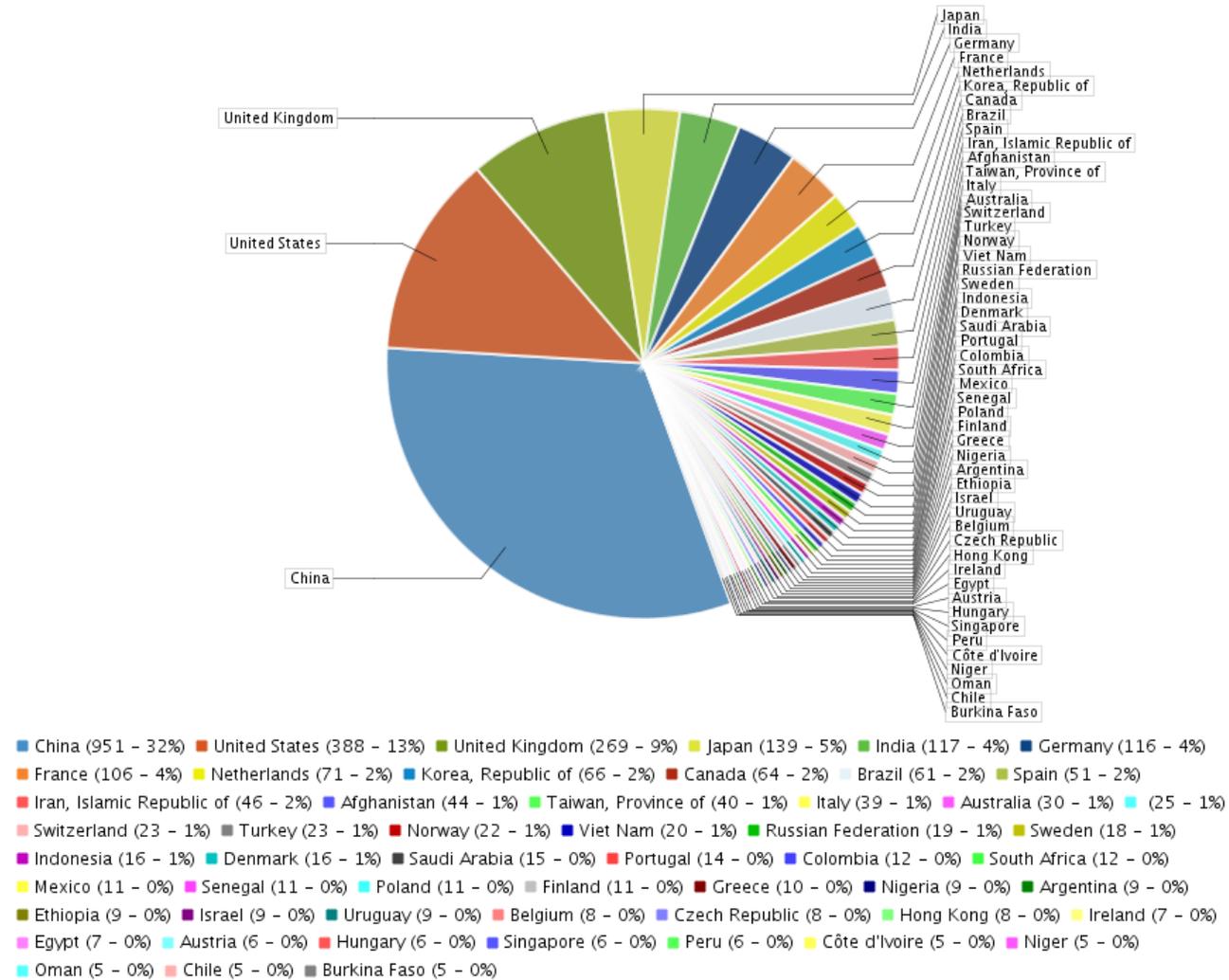| Activity	TIGGE (since 2014)	S2S (since 2015)
Nr active users	3122	1094
Delievered data	437 TiB	496 TiB
Retrieved fields	8,682,889,139	12,332,520,642
Nr requests	6,305,895	9,709,987
Volume data on disk	170 TiB (5%)	77 TiB (80%)

TIGGE Database @ ECMWF: Activity

- Retrieved data volumes per month
    - Volume of data retrieved from MARS before post-processing vs volume of data delivered to users after post-processing
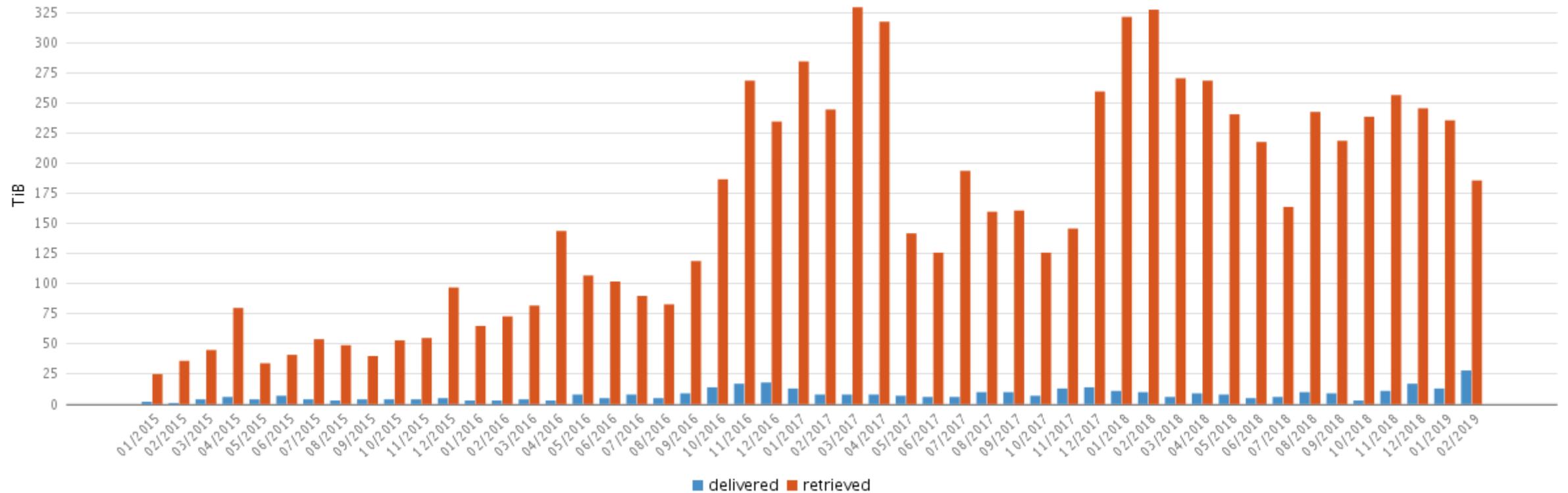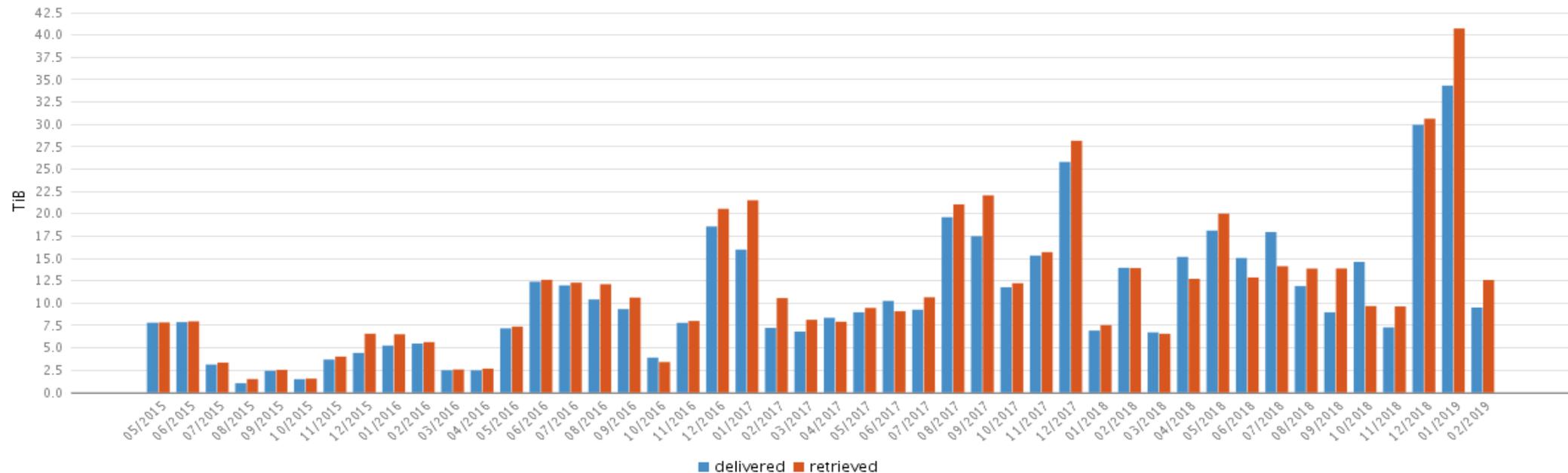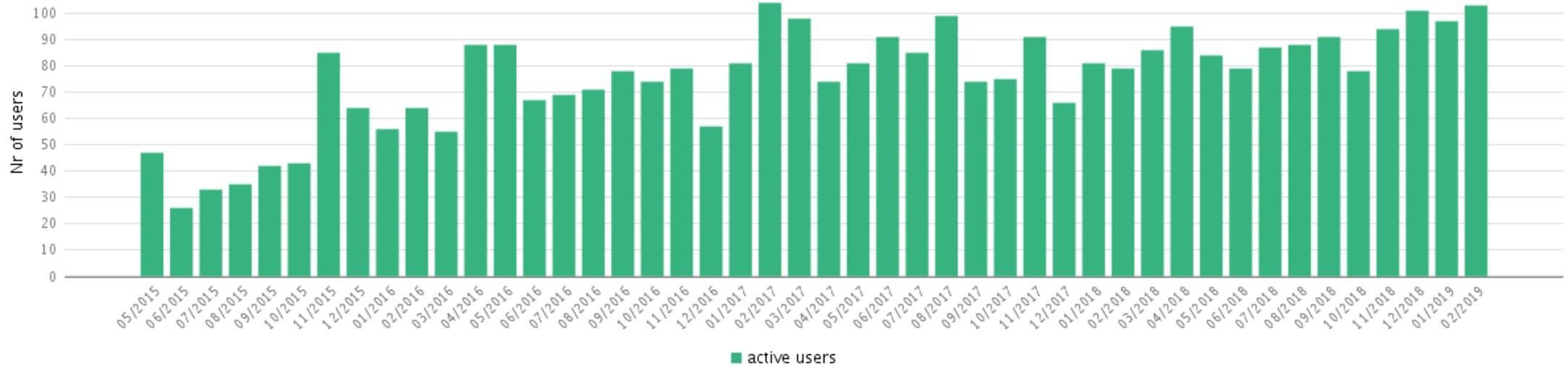

ECMWF

TIGGE Database @ ECMWF: Number of active users

TIGGE Database @ ECMWF: User distribution

China (951 – 32%) ■ United States (388 – 13%) ■ United Kingdom (269 – 9%) ■ Japan (139 – 5%) ■ India (117 – 4%) ■ Germany (116 – 4%)
France (106 – 4%) ■ Netherlands (71 – 2%) ■ Korea, Republic of (66 – 2%) ■ Canada (64 – 2%) ■ Brazil (61 – 2%) ■ Spain (51 – 2%)
Iran, Islamic Republic of (46 – 2%) ■ Afghanistan (44 – 1%) ■ Taiwan, Province of (40 – 1%) ■ Italy (39 – 1%) ■ Australia (30 – 1%) ■ (25 – 1%)
Switzerland (23 – 1%) ■ Turkey (23 – 1%) ■ Norway (22 – 1%) ■ Viet Nam (20 – 1%) ■ Russian Federation (19 – 1%) ■ Sweden (18 – 1%)
Indonesia (16 – 1%) ■ Denmark (16 – 1%) ■ Saudi Arabia (15 – 0%) ■ Portugal (14 – 0%) ■ Colombia (12 – 0%) ■ South Africa (12 – 0%)
Mexico (11 – 0%) ■ Senegal (11 – 0%) ■ Poland (11 – 0%) ■ Finland (11 – 0%) ■ Greece (10 – 0%) ■ Nigeria (9 – 0%) ■ Argentina (9 – 0%)
Ethiopia (9 – 0%) ■ Israel (9 – 0%) ■ Uruguay (9 – 0%) ■ Belgium (8 – 0%) ■ Czech Republic (8 – 0%) ■ Hong Kong (8 – 0%) ■ Ireland (7 – 0%)
Egypt (7 – 0%) ■ Austria (6 – 0%) ■ Hungary (6 – 0%) ■ Singapore (6 – 0%) ■ Peru (6 – 0%) ■ Côte d'Ivoire (5 – 0%) ■ Niger (5 – 0%)
Oman (5 – 0%) ■ Chile (5 – 0%) ■ Burkina Faso (5 – 0%)

S2S Database @ ECMWF: Activity

- Retrieved data volumes per month
  - Volume of data retrieved from MARS before post-processing vs volume of data delivered to users after post-processing

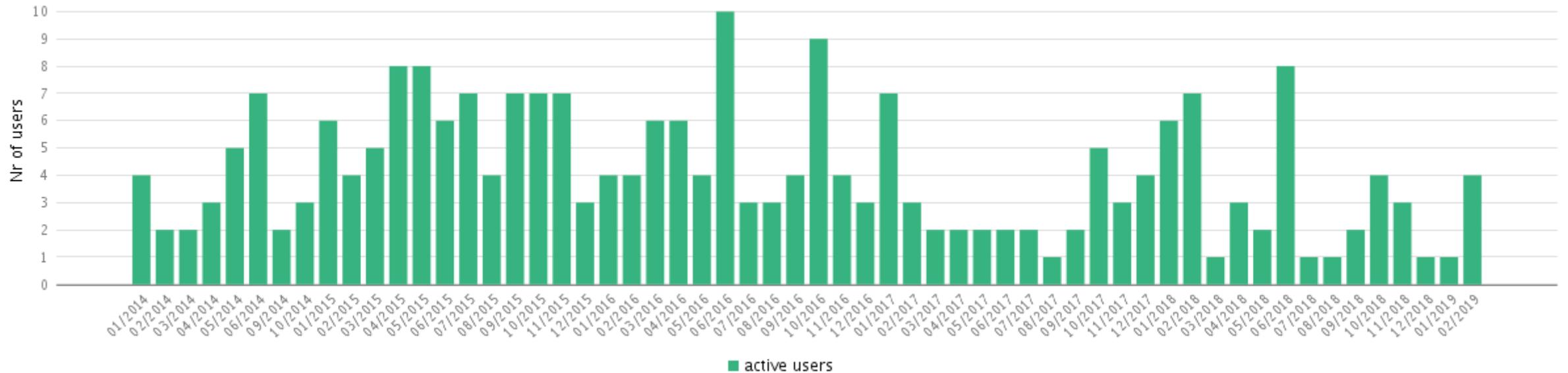# S2S Database @ ECMWF: Number of active users

S2S Database @ ECMWF: User distribution
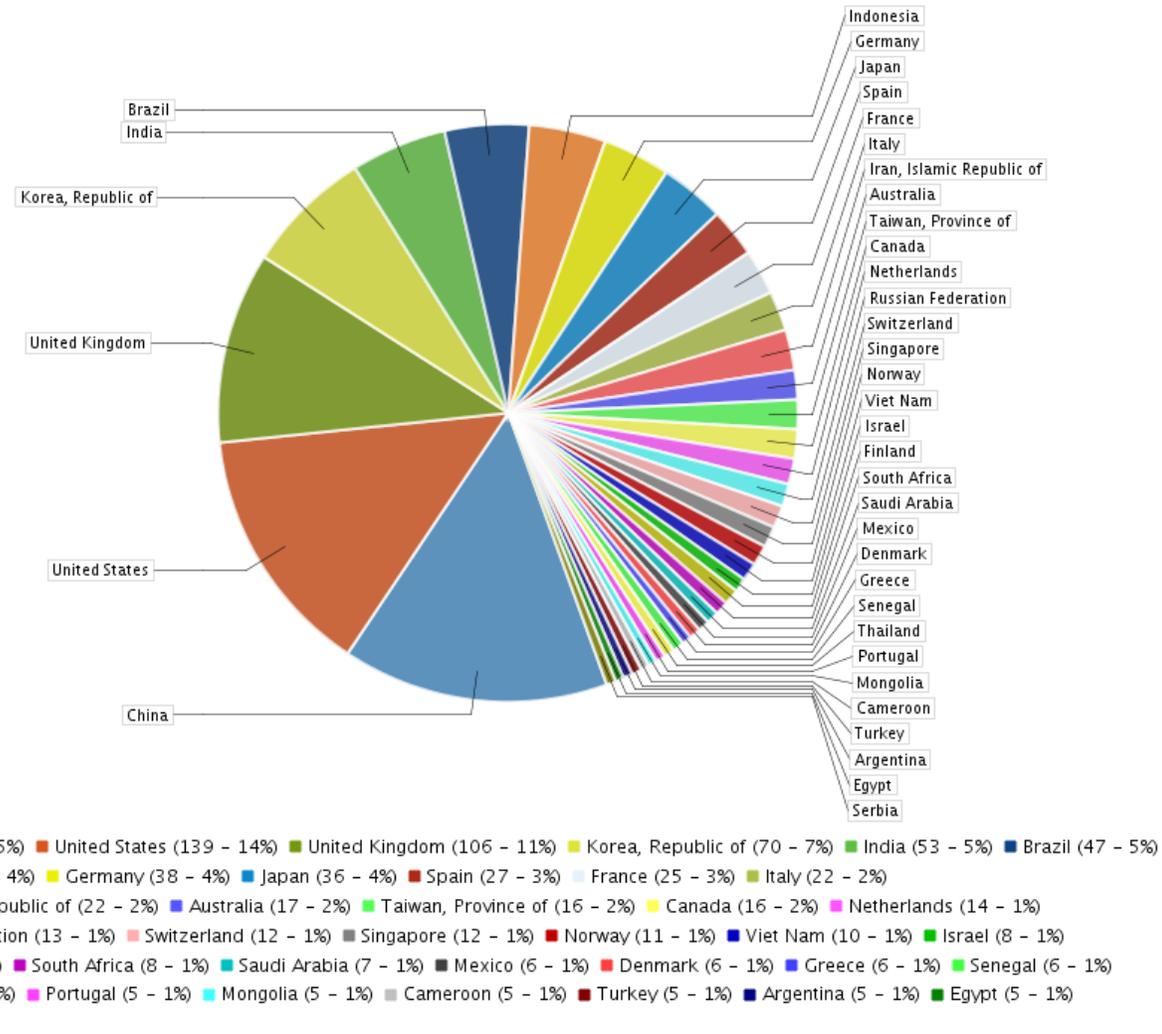
TIGGE-LAM: very low usage

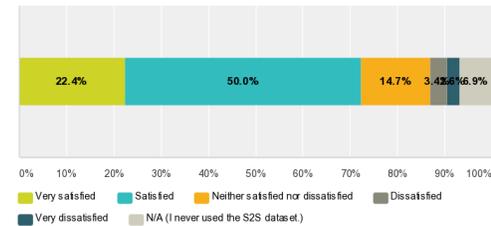- Due to the low uptake of TIGGE-LAM, we are in the process of discontinuing this dataset

S2S User Survey (2017)

- Sent to 700 registered users, received 116 replies (16%)

- Quick summary (percentage of replies):

  – overall, 72% satisfied or very satisfied with the S2S dataset

  – 76% at least satisfied with exploring the data via dedicated web Data Portals

  – 74% at least satisfied with the Web API

  – 80% at least satisfied with the documentation

- Full report available:

  – https://confluence.ecmwf.int/display/S2S/Related+research

Q1 Overall, how satisfied are you with the S2S dataset?

Answered: 116   Skipped: 0

22.4%	50.0%	14.7%	3.4% 2.6% 6.9%

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

■ Very satisfied   ■ Satisfied   ■ Neither satisfied nor dissatisfied   ■ Dissatisfied
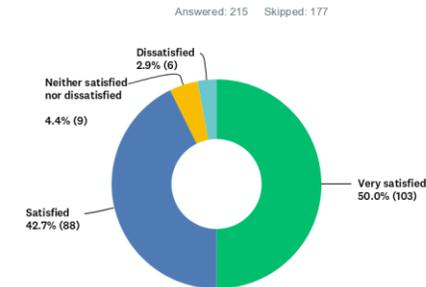■ Very dissatisfied   ■ N/A (I never used the S2S dataset.)

TIGGE User Survey (2018)

- Sent to 3,791 registered users, received 383 replies (10.3%)

- Percentage of replies:
  - 93% used TIGGE
  - 12% used TIGGE-LAM
  - 17% used TIGGE tropical cyclone tracks (in CXML format)

- Quick summary TIGGE (gridded data) related:
  - overall, 93% satisfied or very satisfied with TIGGE datasets as a research tool
  - 87% at least satisfied with exploring the data via the dedicated web Data Portals
  - 85% at least satisfied with the Web API interface to get the data programmatically
  - 86% at least satisfied with the TIGGE documentation
  - almost **50 additional research articles identified** (out of 270 in total)

- Full report available:
  - https://confluence.ecmwf.int/display/TIGGE/Related+research

Q4 Overall, how satisfied are you with the TIGGE dataset as a research tool?

Answered: 215   Skipped: 177

Dissatisfied 2.9% (6)
Neither satisfied nor dissatisfied 4.4% (9)
Very satisfied 50.0% (103)
Satisfied 42.7% (88)

TIGGE & S2S databases: Log analytics

• Using Big Data Analytics tools (such Splunk) we are able to get a better insight of the utilisation of the data:

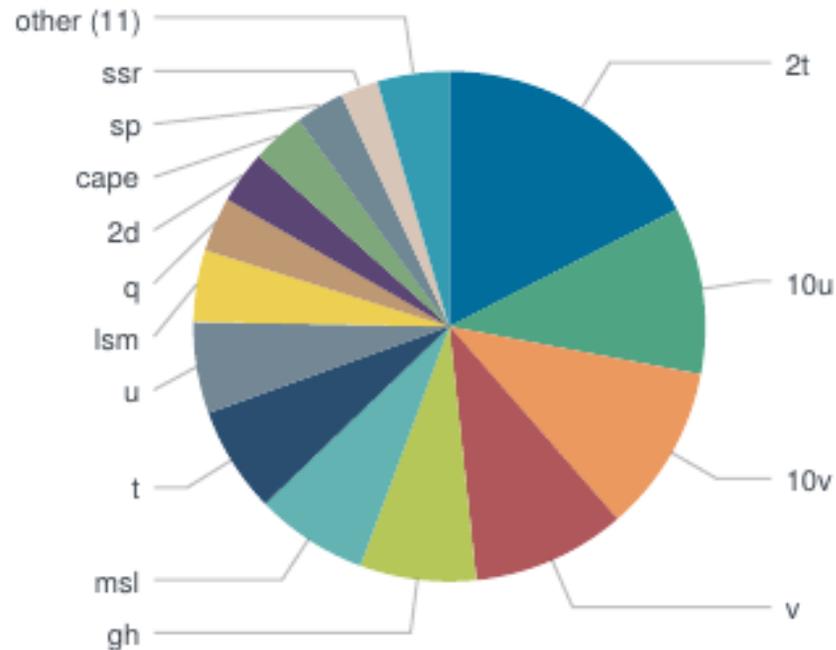– S2S: reforecast vs realtime forecast

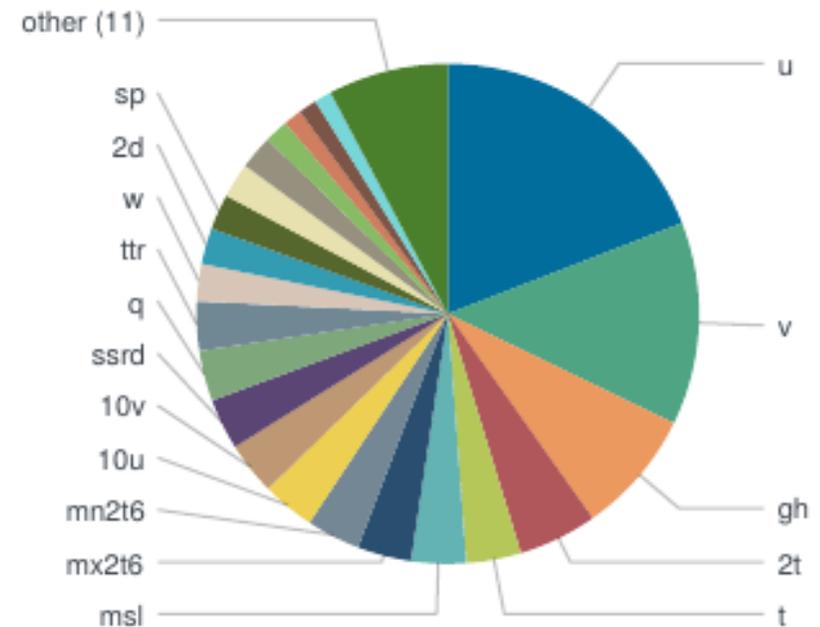Stream popularity for S2S dataset for requests coming from the ECMWF public datasets
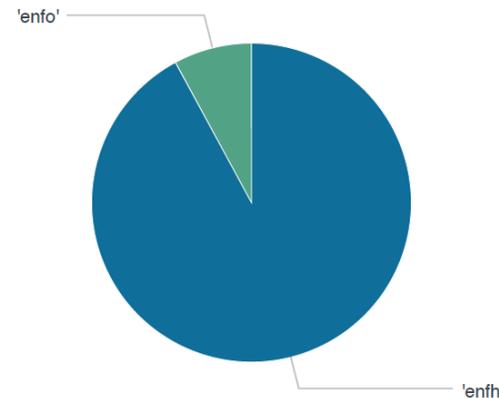
TIGGE & S2S Log analytics: Parameters

Most popular parameters for requests issued from the ECMWF Public Datasets
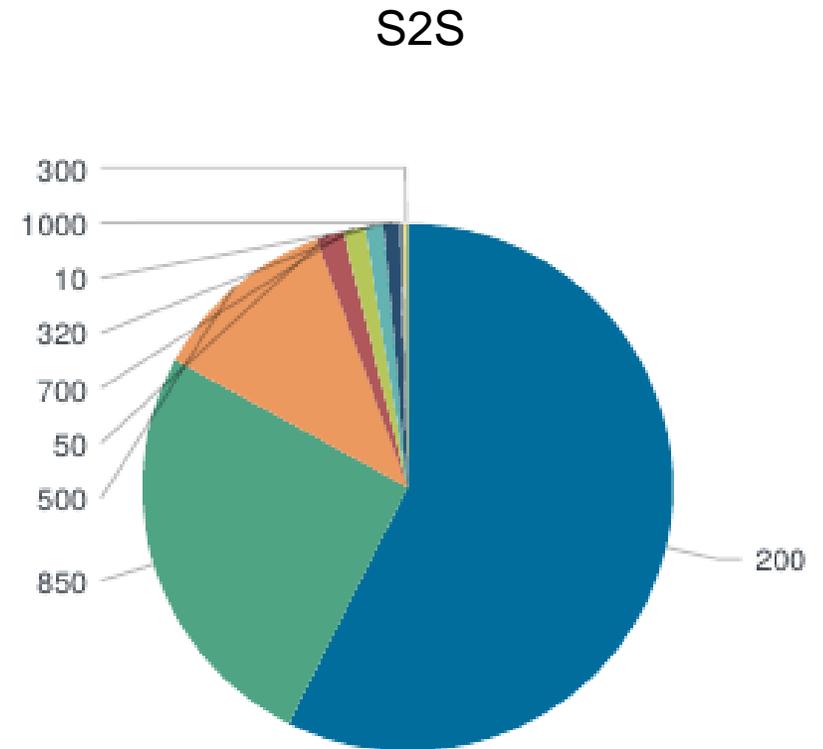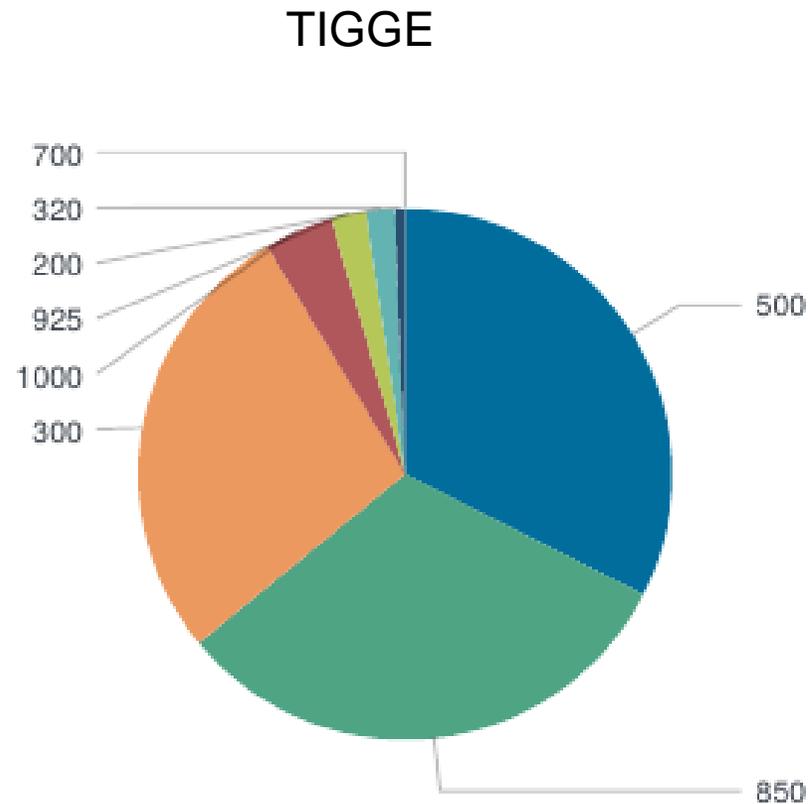
TIGGE

S2S

ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

TIGGE & S2S Log analytics: Levels

Most popular levels for requests issued from the ECMWF Public Datasets

TIGGE

S2S

Future plans

- S2S Phase II: 2019 - 2023
  - Add Ocean variables
  - Add more models to S2S: NASA, India
  - Pilot Realtime projects
- TIGGE Phase III: 2019 – 2023 (?)
  - ECMWF and CMA happy to continue with TIGGE Database
  - NCAR happy to continue archiving TIGGE Tropical cyclone tracks
  - Are all Data Providers committed to continue ?
- More additions to TIGGE:
  - DWD Icon-EPS
  - Deterministic forecasts at TIGGE resolution ?
- ECMWF Data Centre move to Bologna will cause disruption (Q2 2020)

Acknowledgments

Too many colleagues to name individually at NCAR, CMA, ECMWF and all technical and scientific contributors from data provider institutions as well as WMO WG that have supported TIGGE and S2S over the years… and THANK YOU to all the users!

Enhanced web infrastructure

TIGGE

Corporate Website	http://tigge.ecmwf.int/
Technical Website	http://confluence.ecmwf.int//display/TIGGE
Data Portal	http://apps.ecmwf.int/datasets/data/tigge/
History	http://apps.ecmwf.int/datasets/history/tigge-prod/
Usage statistics	http://confluence.ecmwf.int/display/TIGGE/Usage+statistics

S2S

Corporate Website	http://s2s.ecmwf.int/
Technical Website	http://confluence.ecmwf.int//display/S2S/
Data Portal	http://apps.ecmwf.int/datasets/data/s2s/
History	http://apps.ecmwf.int/datasets/history/s2s-prod/
Usage statistics	http://confluence.ecmwf.int/display/S2S/Usage+statistics

TIGGE & S2S Timelines

— Definition phase

— Implementation phase

S2S Phase I

S2S Phase II

TIGGE Phase III

TIGGE Phase II

TIGGE Phase I

2005 2006   2013 2014   2018 2019   2023