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
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

• 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
  – 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

<table>
<thead>
<tr>
<th>NCEP</th>
<th>ECMWF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation (precip)</td>
<td>TP (total precipitation)</td>
</tr>
<tr>
<td>T 2 meters (t2m)</td>
<td>2T</td>
</tr>
</tbody>
</table>
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

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

* 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

<table>
<thead>
<tr>
<th>Status on 2018-10-25</th>
<th>Time range</th>
<th>Resolution</th>
<th>Ens. Size</th>
<th>Frequency</th>
<th>Re-forecasts</th>
<th>Rfc length</th>
<th>Rfc frequency</th>
<th>Rfc size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoM (ammc)</td>
<td>d 0-62</td>
<td>T47L17</td>
<td>3*11</td>
<td>2/week</td>
<td>fix</td>
<td>1981-2013</td>
<td>6/month</td>
<td>3*11</td>
</tr>
<tr>
<td>CMA (babj)</td>
<td>d 0-60</td>
<td>T106L40</td>
<td>4</td>
<td>daily</td>
<td>fix</td>
<td>1994-2014</td>
<td>daily</td>
<td>4</td>
</tr>
<tr>
<td>CNR-ISAC (isac)</td>
<td>d 0-32</td>
<td>0.75x0.56 L54</td>
<td>41</td>
<td>weekly</td>
<td>fix</td>
<td>1981-2010</td>
<td>every 5 days</td>
<td>5</td>
</tr>
<tr>
<td>CNRM (lfpw)</td>
<td>d 0-32</td>
<td>T255L91</td>
<td>51</td>
<td>weekly</td>
<td>fix</td>
<td>1993-2014</td>
<td>4/month</td>
<td>15</td>
</tr>
<tr>
<td>ECCC (cwao)</td>
<td>d 0-32</td>
<td>0.45x0.45 L40</td>
<td>21</td>
<td>weekly</td>
<td>on the fly</td>
<td>1998-2017</td>
<td>weekly</td>
<td>4</td>
</tr>
<tr>
<td>ECMWF (ecmf)</td>
<td>d 0-46</td>
<td>Tco639/319 L91</td>
<td>51</td>
<td>2/week</td>
<td>on the fly</td>
<td>past 20 years</td>
<td>2/week</td>
<td>11</td>
</tr>
<tr>
<td>HMCR (rums)</td>
<td>d 0-61</td>
<td>1.1x1.4 L28</td>
<td>20</td>
<td>weekly</td>
<td>on the fly</td>
<td>1985-2010</td>
<td>weekly</td>
<td>10</td>
</tr>
<tr>
<td>JMA (rjtd)</td>
<td>d 0-33</td>
<td>T1479/T1319L100</td>
<td>50</td>
<td>weekly</td>
<td>fix</td>
<td>1981-2010</td>
<td>3/month</td>
<td>5</td>
</tr>
<tr>
<td>KMA (rksl)</td>
<td>d 0-60</td>
<td>N216L85</td>
<td>4</td>
<td>daily</td>
<td>on the fly</td>
<td>1991-2010</td>
<td>4/month</td>
<td>3</td>
</tr>
<tr>
<td>NCEP (kwbc)</td>
<td>d 0-44</td>
<td>T126L64</td>
<td>16</td>
<td>daily</td>
<td>fix</td>
<td>1999-2010</td>
<td>daily</td>
<td>4</td>
</tr>
<tr>
<td>UKMO (egrr)</td>
<td>d 0-60</td>
<td>N216L85</td>
<td>4</td>
<td>daily</td>
<td>on the fly</td>
<td>1993-2016</td>
<td>4/month</td>
<td>7</td>
</tr>
</tbody>
</table>
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
TIGGE Data Flow in 2018

TIGGE: The International Grand Global Ensemble
## Enhanced web infrastructure

### TIGGE

<table>
<thead>
<tr>
<th>Category</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Website</td>
<td><a href="http://tigge.ecmwf.int/">http://tigge.ecmwf.int/</a></td>
</tr>
<tr>
<td>Technical Website</td>
<td><a href="http://confluence.ecmwf.int/display/TIGGE">http://confluence.ecmwf.int/display/TIGGE</a></td>
</tr>
<tr>
<td>Data Portal</td>
<td><a href="http://apps.ecmwf.int/datasets/data/tigge/">http://apps.ecmwf.int/datasets/data/tigge/</a></td>
</tr>
<tr>
<td>History</td>
<td><a href="http://apps.ecmwf.int/datasets/history/tigge-prod/">http://apps.ecmwf.int/datasets/history/tigge-prod/</a></td>
</tr>
<tr>
<td>Usage statistics</td>
<td><a href="http://confluence.ecmwf.int/display/TIGGE/Usage+statistics">http://confluence.ecmwf.int/display/TIGGE/Usage+statistics</a></td>
</tr>
</tbody>
</table>

### S2S

<table>
<thead>
<tr>
<th>Category</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Website</td>
<td><a href="http://s2s.ecmwf.int/">http://s2s.ecmwf.int/</a></td>
</tr>
<tr>
<td>Technical Website</td>
<td><a href="http://confluence.ecmwf.int/display/S2S/">http://confluence.ecmwf.int/display/S2S/</a></td>
</tr>
<tr>
<td>Data Portal</td>
<td><a href="http://apps.ecmwf.int/datasets/data/s2s/">http://apps.ecmwf.int/datasets/data/s2s/</a></td>
</tr>
<tr>
<td>History</td>
<td><a href="http://apps.ecmwf.int/datasets/history/s2s-prod/">http://apps.ecmwf.int/datasets/history/s2s-prod/</a></td>
</tr>
<tr>
<td>Usage statistics</td>
<td><a href="http://confluence.ecmwf.int/display/S2S/Usage+statistics">http://confluence.ecmwf.int/display/S2S/Usage+statistics</a></td>
</tr>
</tbody>
</table>
TIGGE & S2S databases: status & usage

<table>
<thead>
<tr>
<th>Contents</th>
<th>TIGGE</th>
<th>S2S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive size</td>
<td>3.3 PiB</td>
<td>98 TiB</td>
</tr>
<tr>
<td>Number of fields</td>
<td>7 billion</td>
<td>2 billion</td>
</tr>
<tr>
<td>Time span</td>
<td>2006-present (12 years)</td>
<td>2015-present (4 years)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity</th>
<th>TIGGE (since 2014)</th>
<th>S2S (since 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr active users</td>
<td>3122</td>
<td>1094</td>
</tr>
<tr>
<td>Delivered data</td>
<td>437 TiB</td>
<td>496 TiB</td>
</tr>
<tr>
<td>Retrieved fields</td>
<td>8,682,889,139</td>
<td>12,332,520,642</td>
</tr>
<tr>
<td>Nr requests</td>
<td>6,305,895</td>
<td>9,709,987</td>
</tr>
<tr>
<td>Volume data on disk</td>
<td>170 TiB (5%)</td>
<td>77 TiB (80%)</td>
</tr>
</tbody>
</table>
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
TIGGE Database @ ECMWF: Number of active users
TIGGE Database @ ECMWF: User distribution

- China (951 - 32%)  
- United States (108 - 12%)  
- United Kingdom (269 - 9%)  
- Japan (139 - 5%)  
- India (117 - 4%)  
- Germany (116 - 4%)  
- France (105 - 4%)  
- Netherlands (71 - 2%)  
- Korea, Republic of (66 - 2%)  
- Canada (64 - 2%)  
- Brazil (51 - 2%)  
- Spain (51 - 2%)  
- Iran, Islamic Republic of (46 - 2%)  
- Afghanistan (44 - 1%)  
- Taiwan, Province of (40 - 1%)  
- Italy (39 - 1%)  
- Australia (30 - 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 (8 - 0%)  
- Israel (8 - 0%)  
- Uruguay (8 - 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 (5 - 0%)  
- Peru (5 - 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

- China (148 - 15%)
- United States (139 - 14%)
- United Kingdom (106 - 11%)
- Korea, Republic of (70 - 7%)
- India (53 - 5.5%)
- Brazil (47 - 5%)
- Indonesia (42 - 4%)
- Germany (31 - 4%)  Japan (6 - 0.5%)
- Spain (27 - 3%)
- France (25 - 3%)
- Italy (22 - 2%)
- Iran, Islamic Republic of (22 - 2%)
- Australia (17 - 2%)
- Taiwan, Province of China (16 - 2%)
- Canada (10 - 2%)
- Netherlands (14 - 1%)
- Russian Federation (12 - 1.5%)
- Switzerland (12 - 1.5%)
- Singapore (12 - 1.5%)
- Norway (11 - 1%)
- Viet Nam (10 - 1%)
- Israel (8 - 1%)
- Finland (8 - 1%)
- South Africa (8 - 1%)
- Saudi Arabia (7 - 1%)
- Mexico (6 - 1%)
- Denmark (6 - 1%)
- Greece (6 - 1%)
- Senegal (6 - 1%)
- Thailand (6 - 1%)
- Portugal (6 - 1%)
- Mongolia (5 - 1%)
- Cameroon (6 - 1%)
- Turkey (5 - 1%)
- Argentina (5 - 1%)
- Egypt (5 - 1%)
- Serbia (5 - 1%)

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
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
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
TIGGE & S2S databases: Log analytics

• Using Big Data Analytics tools (such as Splunk) we are able to get a better insight of the utilisation of the data:
  – S2S: reforecast vs realtime forecast
TIGGE & S2S Log analytics: Parameters

Most popular parameters for requests issued from the ECMWF Public Datasets

TIGGE

S2S
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

<table>
<thead>
<tr>
<th>Category</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Website</td>
<td><a href="http://tigge.ecmwf.int/">http://tigge.ecmwf.int/</a></td>
</tr>
<tr>
<td>Technical Website</td>
<td><a href="http://confluence.ecmwf.int/display/TIGGE">http://confluence.ecmwf.int/display/TIGGE</a></td>
</tr>
<tr>
<td>Data Portal</td>
<td><a href="http://apps.ecmwf.int/datasets/data/tigge/">http://apps.ecmwf.int/datasets/data/tigge/</a></td>
</tr>
<tr>
<td>History</td>
<td><a href="http://apps.ecmwf.int/datasets/history/tigge-prod/">http://apps.ecmwf.int/datasets/history/tigge-prod/</a></td>
</tr>
<tr>
<td>Usage statistics</td>
<td><a href="http://confluence.ecmwf.int/display/TIGGE/Usage+statistics">http://confluence.ecmwf.int/display/TIGGE/Usage+statistics</a></td>
</tr>
</tbody>
</table>

### S2S

<table>
<thead>
<tr>
<th>Category</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Website</td>
<td><a href="http://s2s.ecmwf.int/">http://s2s.ecmwf.int/</a></td>
</tr>
<tr>
<td>Technical Website</td>
<td><a href="http://confluence.ecmwf.int/display/S2S/">http://confluence.ecmwf.int/display/S2S/</a></td>
</tr>
<tr>
<td>Data Portal</td>
<td><a href="http://apps.ecmwf.int/datasets/data/s2s/">http://apps.ecmwf.int/datasets/data/s2s/</a></td>
</tr>
<tr>
<td>History</td>
<td><a href="http://apps.ecmwf.int/datasets/history/s2s-prod/">http://apps.ecmwf.int/datasets/history/s2s-prod/</a></td>
</tr>
<tr>
<td>Usage statistics</td>
<td><a href="http://confluence.ecmwf.int/display/S2S/Usage+statistics">http://confluence.ecmwf.int/display/S2S/Usage+statistics</a></td>
</tr>
</tbody>
</table>
TIGGE & S2S Timelines

- Definition phase
- Implementation phase


S2S Phase I
TIGGE Phase I
TIGGE Phase II
TIGGE Phase III
S2S Phase II