Experiences Gained from Managing a Multi-Year, Multi-NWP Center Ensemble Archive

Doug Schuster – NCAR

Many Partners at NCAR, ECMWF, and CMA



Outline

- Background
- TIGGE Project Design
- Resulting Operational Implementation
- Access Options
- Reflection



Background: TIGGE Project

- THORPEX Interactive Grand Global Ensemble
- Accelerate improvements in the accuracy of 1-day to 2 week high impact weather forcasts
- Global ensemble forecasts to around 15 days generated routinely at different NWP centres around the world
- 850 TB, growing by > 3.5 TB/week







Background: TIGGE Project – Field Project Support



UCAR Africa Initiative

Probability of low humidity (10th percentile) West Sahel Initial:12UTC 14DEC2012 Day6-10

	Grand	ensemble	ECMWF	UKMO	NCEP	JMA	Grand ensemble	ECMWF	UKMO	NCEP	JMA
	Day6	mem:147	Day6 mem:51	Day6 mem:24	Day6 mem:21	Day6 mem:51	Day6 mem:147	Day6 mem:51	Day6 mem:24	Day6 mem:21	Day6 mem:51
			Carrow Manual Martin	Succession and the	and account of the second	a commence and		a stranger were and the	of second as	and the contraction of the contr	streeman we way
			And the second second second	and the second second		construction in a construction	an in	and the second second	and the second second second	which which we have the	and the manual of the
	- 4					A CONTRACTOR OF THE PARTY	03.241		CARENA CARENA CONTRACTOR	LVVEN Produces and the second	PRESE DON'S VELLED CLEAR AND
	Day7	mem:147	Day7 mem:51	Day7 mem:24	Day7 mem:21	Day7 mem:51	Day7 mem:147	Day7 mem:51	Day7 mem:24	Day7 mem:21	Day7 mem:51
		1	the way and a subvert of the	Survey of the same	Contraction of the second	Contract Proversity		The way and a service of the service	And a start of the	State Contraction	Control Control Control A
			pressent attender	and the future of the former	Construction of the	And and and a second	2	munition and the	and the second and th	construction of the second	and the second
	Qu				PANAL PARAL PROPERTY OF THE PANAL	A SALE AND A TANA TANA TANA TANA TANA TANA TAN	Con and	and the second sec	LALLEWER MAL TOTAL THE	ANT THE DOLL OF THE DECK LETTER	The same of the second of the second
	Dave	mem:147	Days mem 51	Dave mem:24	Dave mem:21	Dave mem:51	Dave mami147	Dave mami51	Dave mamile	Dave mam:21	Dave mami51
					Netto Anticipation	Duyo			Sad A STAALEY PARA	Nerrow Contraction of the second seco	
	- Carlos a		the town and the to the	A CONTRACTOR OF THE TAXA	the succession of the second	internation of the		Human energy carle	Marrie Milliant	the meret were to the	una una marine and
	50.			A TAST COLOR AND THE AND THE A	AND AND A CONTRACT OF A CONTRACT		Sol- Sol	CUT THE COMMUNICATION OF CALL	1 The Weeker ward better	L LA CONTRACTOR LA LA LA LA LA	A CONTRACT OF THE OWNER
				CONTRACTOR OF A	SECTOR AND ADDRESS OF	enered the terres	- Silver		MARCHALLEMANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	CCCCFFFF RAATTATTATAAA	encounterer Lives
	Day9	mem:147	Day9 mem:51	Day9 mem:24	Day9 mem:21	Day9 mem:51	Day9 mem:147	Day9 mem:51	Day9 mem:24	Day9 mem:21	Day9 mem:51
	4		The second constraints and the second s	The sufference we strand	en service service service	A CONTRACTOR OF CONTRACT	47 V	Manual and the state of the second	A STATE STATE STATE	and the second the second	Contraction of the states of t
	100	-	A Real Provide States and the second	A Statestate and the second	1 More Conserved Said	Martin Contraction of the second	and some	Automation and here	Muser and and and and	A Marine Marine State	And the second second
	- 04				A ATH A REAL PROPERTY AND	A MAR WALL POST AND VALUE	and a start of the		CARLEN AND THE PLANE AND THE P	ATT A STARLAGE AND A STARLAGE	FIT IN DOM WE CE DED VYLEER
	Day10	mem:96	Day10 mem:51	Day10 mem:24	Day10 mem:21	Day10	Day10 mem:96	Day10 mem:51	Day10 mem:24	Day10 mem:21	Day10
30N ·			A State of the second states	A A A A A A A A A A A A A A A A A A A	Harris Receiver Arths	1000		NY CALL MARKAN REPORT	Hill Contract After	Harris Charles VVVV APP	1000
20N -	2		A succession of the	Part and and and the	ununu ortu	Eno forecast	N	Marganeer mark	Marine Contractory	uccumment of the	Eno forecast
10N ·	- Qu		Res Providence State	1 147 T 4 DUNNE MEMORY BARRED	T I WE LE COMMUNE CONTROL MARCE	Col. The	N- Correction	PROCESSION AND PROCESSION	1 545- TO PROBLEMENT SAFE & C	ELVING STUDY LONG STUDY OF THE STUDY OF THE	Sol- Tall
EQ	ow 15w 10w	1 5W 0 5E 10E	NERS 1555555555555555555555555555555555555		SARASSA STATISTICS				***************************************	KKKKK1111111111111111111111111	
2					cont	rol surface wind	201 101 101 01 0 02 100			cont	rol surface wind
			10	30 50 70	90	20 [m/s]		10	30 50 70	90	20 [m/s]
			probability falli	ng below the 10th	percentile [%]			probability ex	ceeding the 90th p	percentile [%]	
									and the second		

Probability of high humidity (90th percentile) West Sahel Initial:12UTC 14DEC2012 Day6-10

RDA





TIGGE Project Design: Common Standards - Homogeneity

- Homogeneity is paramount for TIGGE to succeed
 - The more consistent the archive, the easier it is to manage and to develop applications
- There are many aspects to homogeneity:
 - Common terminology (parameter names, file names,...)
 - Common data structures (format –WMO GRIB-2, units, ...)
 - Definition of an agreed list of products (parameters, steps, levels, ...)
 - Data transmission protocol for providers Unidata IDD/LDM
 - Software tools to facilitate and enforce homogeneity



- Three archive centres: CMA, NCAR and ECMWF
- Ten data providers:
 - ECMWF, JMA (Japan), UK Met Office (UK), CMA (China), NCEP (USA), MSC (Canada), Météo-France (France), BOM (Australia), KMA (Korea), CPTEC (Brazil)
- Multiple data transmission methods
 - Problems implementing resend protocols



Inconsistent product definitions –instantaneous and 6-hourly



Time integrated surface net solar radiation (sfc), step 96, 20070608



- No NWP center produces all agreed upon parameters
- Models typically interpolated to lower resolution grids



• Incomplete time-series issues







NCAR

Access Options: User Data Access @ NCAR

	RPEX Intera	ctive Grand	d Globa	l Ensem	ble					5
itiona	Center for Atmos	oheric Research	oratory							-
lome	Get Forecast Data	Get Model Valida	ation Data	Tools Help	Documenta	tion Logo	ut	_		
_										
Dat	ta Subset Sele	ction								
Dat	ta Subset Sele	ction	TIC	GGE Po	rtal for S	Sub-set	ting			
Dat A B	ta Subset Sele Australia China I ureau of Administ Met.	ction Met. Met. Service ration of Canada	TIC ECMWF	GGE Po United Kingdom Met. Office	TTAL FOR S Select Centers USA NCEP	Sub-set ³ Met. France	ting Japan Met. Agency	. Korea Met. Administration	Brazil CPTEC	-

Output File Format

NetCDF

Retrieve Matching Model Validation Data:

Check the box above to retrieve observational data for model validation. This will be processed as a separate data request from the NCAR RDA data portal and will match the available parameters chosen in your TIGGE data request.

	Back	Submit	t Reque	st C	ancel		
		pii s					1
🗏 10v	10_meter_v_velocity		SL				
🗖 2d	surface_air_dew_point_temperature	К	SL				
🗖 2t	surface_air_temperature	к	SL				
Ism	land_sea_mask	0 or 1	SL				
mn2t6	surface_air_minimum_temperature	K	SL				
msl	mean_sea_level_pressure	Pa	SL				
mx2t6	surface_air_maximum_temperature	K	SL				
orog	orography		SL				
sd sd	snow_depth_water_equivalent	kg m ⁻²	SL				
skt	skin_temperature	K	SL				
sint	time_integrated_surface_net_solar_radiation	W m ⁻ s	SL				
■ sm	soil_moisture	kg m ^{-o}	SL				
■ sp	surface_pressure	Pa	SL				
ssnf	time_integrated_surface_sensible_heat_flux	W m ⁻² s	SL				
ssr	time_integrated_surface_net_solar_radiation	W m ⁻² s	SL				
u str	time_integrated_surface_net_thermal_radiation	W m ⁻² s	SL				
	total_column_water	kg m⁻²	SL				
⊠ tp	total_precipitation	kg m ⁻²	SL 25	55	6 12 18 24	hrs	
🔍 ttr	time_integrated_outgoing_long_wave_radiation	W m ⁻² s	SL				
		Next Cancel					





Access Options: User Data Access @ ECMWF

 ECMWF provides similar access options to the <u>complete</u> archive + batch access capability.

	Batch	access token for user Manuel Fuentes - Mozilla Firefox	×					
<u>F</u> ile	e <u>E</u> dit	<u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp	- 1					
		Home > TIGGE > Portal >						
		Batch access token for user Manuel Fuentes	1					
		This is a sample script to retrieve data in batch from this server. You can download a ver experimental version of ECMWF::DataServer from <u>here</u> . The module depends on LWP::UserAgent and HTTP::Request which are part of from <u>libwww-perl</u> . If you are b proxy, you can need to set the http_proxy environment variable:						
		http_proxy=http://proxy.my.place/ export http_proxy						
		Below is a sample script to retrieve data and save it in the file data.grib . The request is expressed in terms of the MARS language. Please see <u>here</u> for more details.						
		<pre>#!/usr/bin/perl -Ipath_to_where_ECMWF::DataServer_is_installed use ECMWF::DataServer; use strict;</pre>						
		<pre>my \$client = ECMWF::DataServer->new(portal => 'http://tigge-portal.ecmwf.int/d/dataserver/', token => ' email => 'MFuentes@ecmwf.int',);</pre>						
		<pre>\$client->retrieve(class => "tigge", number => "all", time => "00/12", date => "20071001/to/20071003".</pre>						
		origin => "all", step => "24/to/120/by/24", levtype => "sl", type => "pf", param => "tp"						
		area => "70/-130/30/-60", grid => "2/2", target => "data.grib",);						
•			Þ					
Do	ne		T					
	-							



Access Options: TIGGE Archive Usage







NCAR



Access Options: Data Downloaded from NCAR Tape Archive



Reflection: Scope

- What are the research objectives?
- What is **<u>REALLY</u>** needed to support those objective?
 - Is the project sustainable?
 - Can the project's support system scale as the project requires?
 - Is mission creep allowed?
 - Are all parties <u>completely</u> committed?
- Is the user community large enough, sufficiently aware, and well enough prepared to justify the storage and support costs?



Reflection: Science Partner Commitment

- Are all partners committed to implementing and supporting the agreed upon protocols, standards, etc..?
 - Data Transmission, Data Format, Format Conventions
 - Parameter definitions, Parameter Encoding
 - Science partners need to negotiate carefully the agreed set of parameter fields. There is a tendency to let this grow too large (many fields in TIGGE are infrequently or unused).
- Avoid 1-off solutions
 - <u>Very resource intensive!!!</u>





Reflection: Archive Structure

- What are the user/research requirements?
- What services will fulfill those requirements?
- What is the most efficient structure to support desired services?
 - TIGGE –Centralized Archive
 - Would distributed work better for this scale of archive (850 TB, 7 Million+ files? Use of common Service Oriented Architecture, and data brokering software?
 - TIGGE Archive @ NCAR organized in <u>file groups</u> by provider and forecast initialization time
 - Pressure Level, O Level, PV Level, Single/Surface Level
 - Contrast: TIGGE @ ECMWF is a one-off quasi-DB structure leveraging long-standing operational services



Conclusions

- Scope –Stay focused on achieving research objectives
 Are support systems scalable to meet future needs?
- Project Design –All partners must agree upon and support common standards and apply them
 - Ensures consistent parameter definitions/encoding
 - Supports completeness and interoperability across tools
- Archive Structure –Find an archive structure that best supports user research needs and is scalable
 - Data access options and services are impacted by choice of archive structure



Questions?

schuster@ucar.edu

International TIGGE Archive Centers						
ECMWF	http://tigge.ecmwf.int					
NCAR	http://tigge.ucar.edu					
CMA	http://wisportal.cma.gov.cn/tigge					

