Center for Western Weather and Water Extremes

SCRIPPS INSTITUTION OF OCEANOGRAPHY AT UC SAN DIEGO

Technological advances, the power of big data

Western States Water Council

May 19, 2017

Scott L. Sellars[1], Phu Nguyen[2], John Graham[3], Joulien Tatar[3], Tom DeFanti[3] ,Larry Smarr[3], F. Martin Ralph[1] and Soroosh Sorooshian[2]

[1]Center for Western Weather and Water Extremes, UCSD, La Jolla, CA

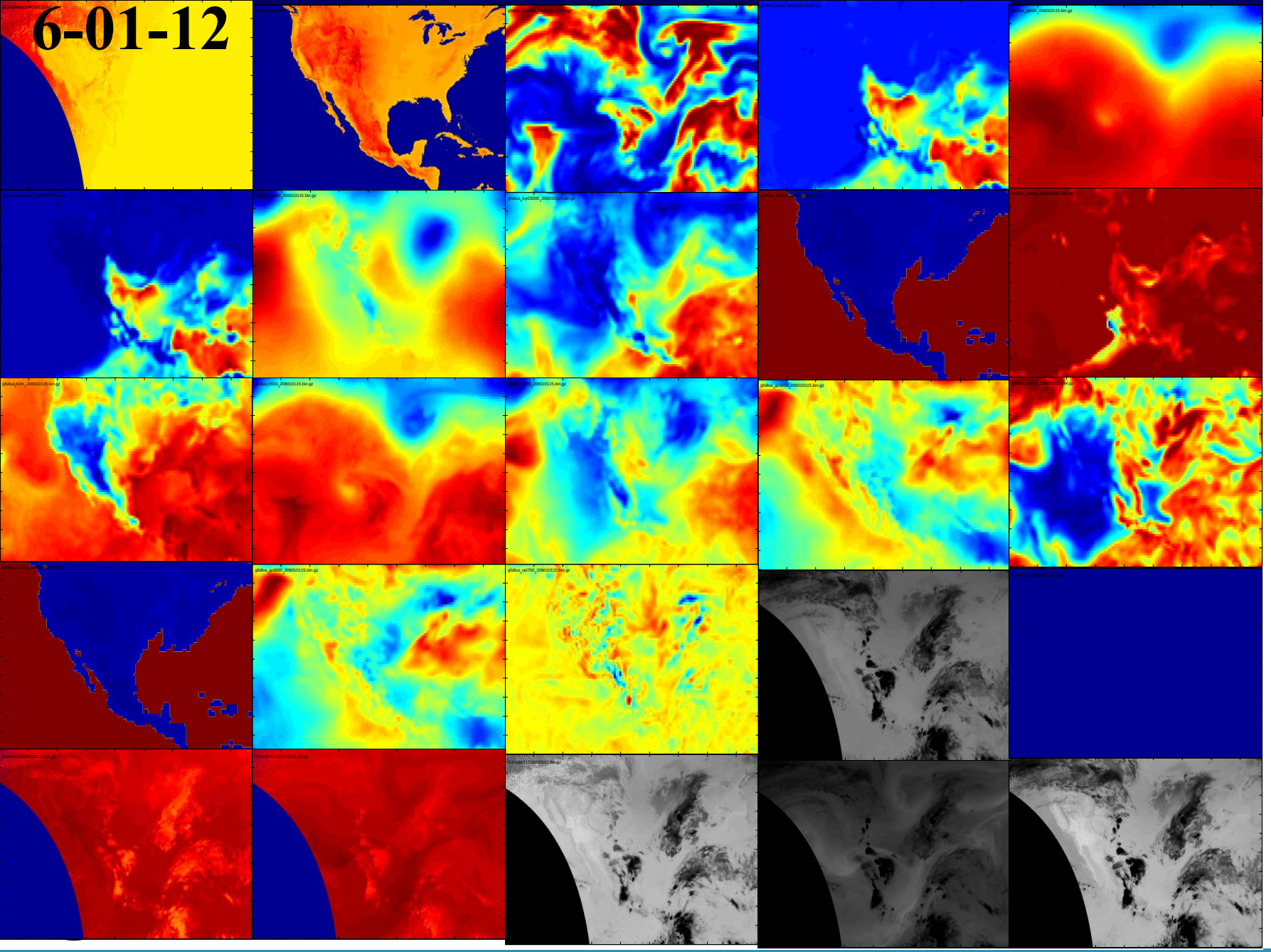
[2]Center for Hydrometeorology and Remote Sensing, UCI, Irvine, CA

[3]California Institute for Telecommunications and Information Technology (Calit2), UCSD, La Jolla, CA

Outline

- Complexity in Predictability of Earth Sciences
- Big Data and Technology
  - CONNECT/PRP (Pilot Project) – How we are using the PRP
- Earth Science Knowledge Discovery
- “Integrated Understanding” in Earth Sciences?

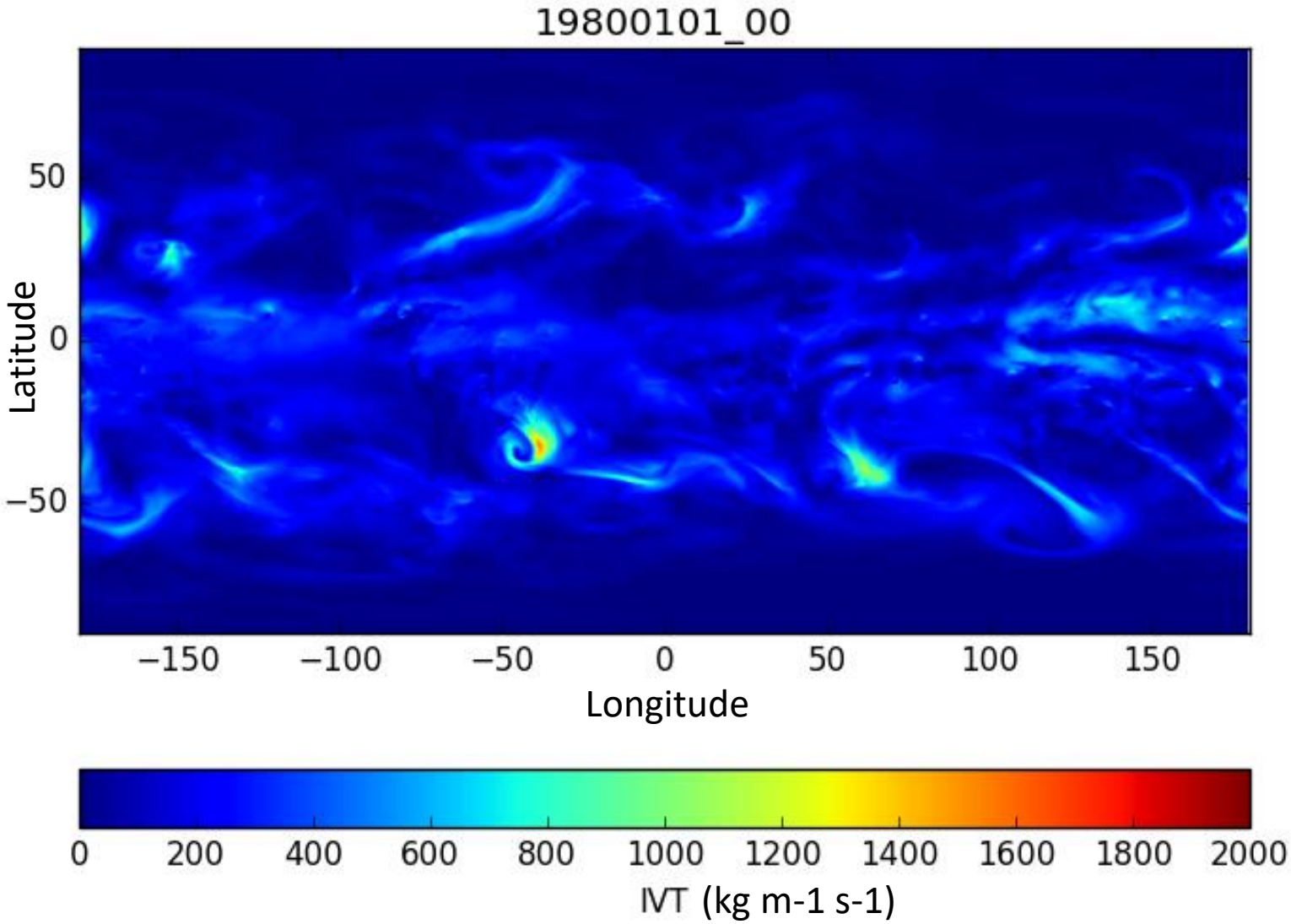
*NASA Apollo 17 crew traveling toward the Moon



Hyper-dimensional data from model reanalysis and satellite based sources for June 1, 2012 over the continental United States. Each image represents a specific atmospheric variable or feature.

CW3E

NASA MERRA2 IVT (kg m-1 s-1)



CW3E

Name	Source	Domain	Period of Record	available timestep(s)	available resolution	available format(s)	Model Resolution	scheme & model vintage
Arctic System Reanalysis (ASR)	Center, The Ohio State University/ David Bromwich,NCAR,CIRES,U Illinois	Arctic	2000/01 to 2012/12	Sub-daily, Monthly	30 km; 71 levels; 10hPA top, 10 km	netCDF	30 km and 10 km	WRF-VAR
Climate Forecast System Reanalysis (CFSR)	NCEP	Global	1979/01 to 2011/01	Sub-daily, Monthly	.5°x.5° & 2.5°x2.5°, 0.266 hPA top	GRIB	T382 x 64 levels	3DVAR | 2009
ERA-15	ECMWF	Global	1979/01 to 1993/12	Sub-daily, Monthly	T106, 2.5 x 2.5	GRIB	T106 (1.125)	
ERA-20C: ECMWF's atmospheric reanalysis of the 20th century (and comparisons with NOAA's 20CR)	ECMWF	Global	1900/01 to 2011/01	Sub-daily, Daily, Monthly	~ 125km; 160 x 320; 91 model levels/ 37 pressure levels / 16 potential temperature levels, and the 2 PVU potential vorticity level	netCDF, GRIB		4DVAR
ERA-Interim	ECMWF	Global	1979/01 to 2016/12	Sub-daily, Daily, Monthly	0.75°x0.75°x60 lev 0.1 hPA top	netCDF, GRIB	T255, 60 levels	4DVAR | 2006
ERA40	ECMWF	Global	1957/09 to 2002/08	Sub-daily, Monthly	2.5°x2.5° / 1.125°x1.125°; 60 levels 0.1 hPA top	netCDF, GRIB	T159, 60 levels	3DVAR | 2004
JRA-25	Japanese Meteorological Agency	Global	1979/01 to 2004/12	Sub-daily, Monthly	1.125x1.125/2.5x2.5; 0.4 hPA top	GRIB	T106, 40 levels	3DVAR | 2004
JRA-55	Japanese Meteorological Agency	Global	1957/12 to 2016/07	Sub-daily, Monthly	T319 x 60 levels, 0.1 hPA top	GRIB	T319 x 60 levels	4DVAR | 2009
NASA MERRA	NASA	Global	1979/01 to 2016/11	Sub-daily, Monthly	0.5° x 0.667° x 72 , 0.01 hPA top	netCDF, HDF	0.5° x 0.667° x 72	GEOS IAU | 2009
NCEP NARR	NCEP	North America	1979/01 to 2017/02	Climatology, Sub-daily, Monthly	32km	GRIB	32km x 45 eta	3DVAR | 2003
NCEP Reanalysis (R2)	NCEP,DOE	Global	1979/01 to 2017/02	Sub-daily, Daily, Monthly	2.5°x2.5° 28 levels 3 hPA top	netCDF, GRIB	T62 28 levels	3DVAR | 2001
NCEP-NCAR (R1): An Overview	NCEP,NCAR	Global	1948/01 to 2017/02	Sub-daily, Daily, Monthly	2.5°x2.5°; 3 hPA top	netCDF, GRIB	T62 - 28 levels	3DVAR | 1995
NOAA 20th-Century Reanalysis, Version 2 and 2c	NOAA ESRL,CIRES CDC / Gil Compo	Global	1850/12 to 2014/12	Sub-daily, Daily, Monthly	2°x2°, 28 levels 10 hPA top	netCDF, GRIB	T62 28 levels	Ensemble Kalman Filter | 2009

CW3E

Why can’t we forecast seasonal climate?

***Each are well known climate phenomena**

Enfield et al. 2001
Gray et al. 2001

Atlantic Multidecadal Oscillation

Zhang et al. 1997

Pacific Decadal Oscillation

Barlow et al. 1997
Cane et al, 2004

El Nino Southern Oscillation

Madden Julian Oscillation

Gaun et al. 2013
Madden and Julian 1993

Arctic Oscillation

Higgins et al. 2001
Larson et al. 2000

*NASA

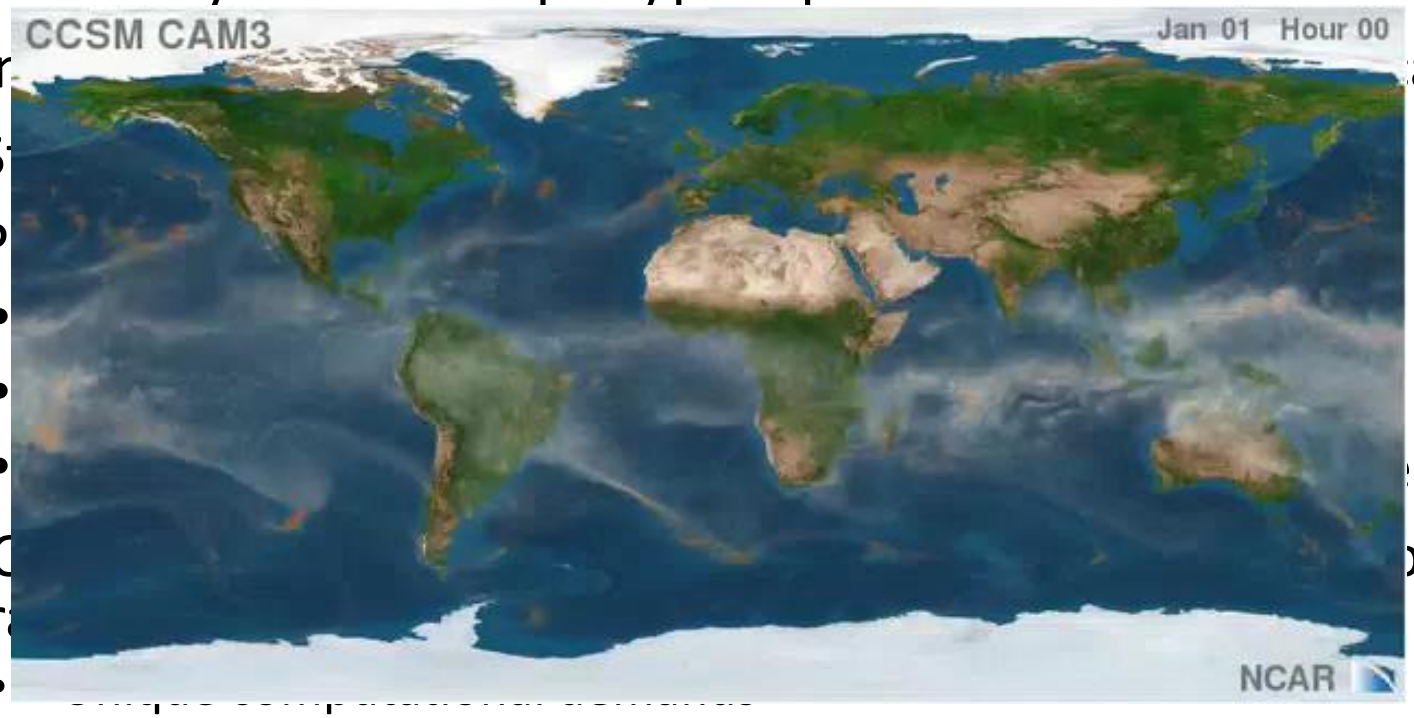
?

What causes extreme floods and droughts on the **West coast**?

CW3E

What should we do?

1. How can we untangle the fundamental causes in the variability of water vaport/precipitation?
2. Ir…asts?
3. S…
4. P…
   - …
   - …
   - …ery
5. CC…on and char…
   - …
   - Unique data storage demands
   - Unique analysis approaches

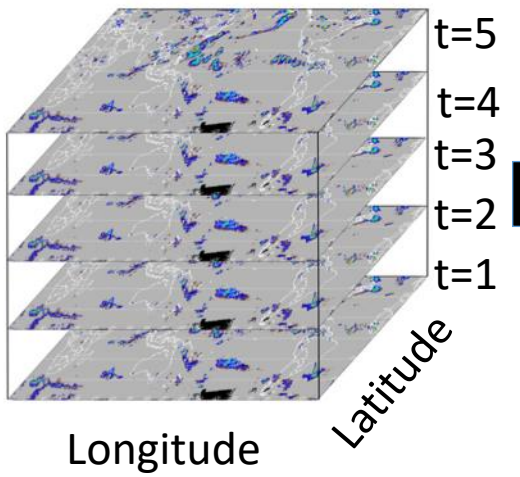


Atmospheric Moisture

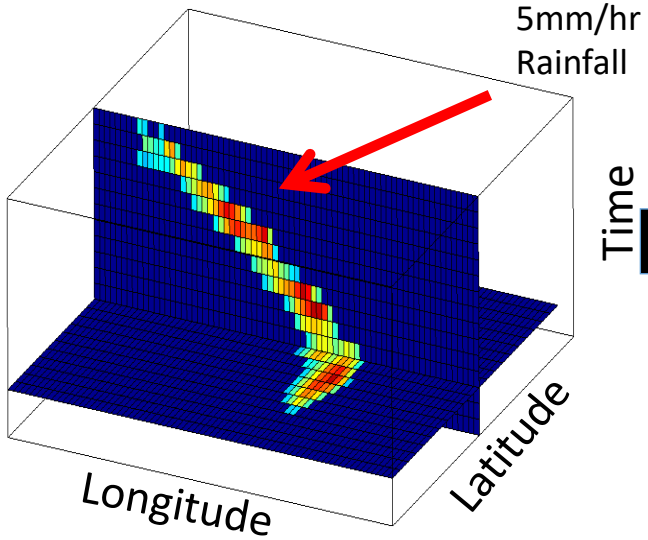
Multi-Institutional Big Data Transformed Into Insight

- For Computational Earth Sciences (Sellars et al. 2013, 2015)
  - **CONNected objECT (CONNECT) Algorithm, developed at UCI-CHRS**
    - **Team: Wei Chu, Scott Sellars, Phu Nguyen, Xiaogang Gao, Kuo-lin Hsu, and Soroosh Sorooshian**
  - Most algorithms do not track the events over it's life cycle

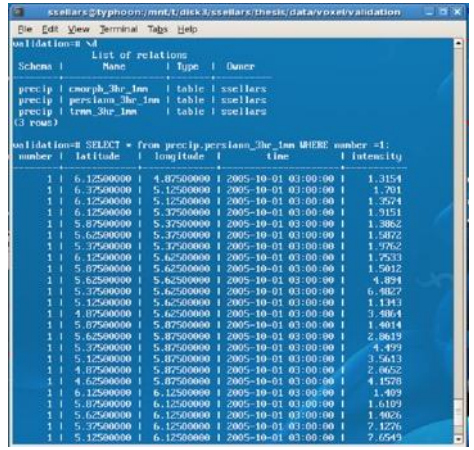
Data Hypercube:



CONNECT: Object Segmentation



Object Storage (PostgreSQL)



Data

1. 60N-60S, 0-360 lat and long
2. Hourly time step
3. March 1st, 2000 to January 1st, 2011

Set Object Criteria:

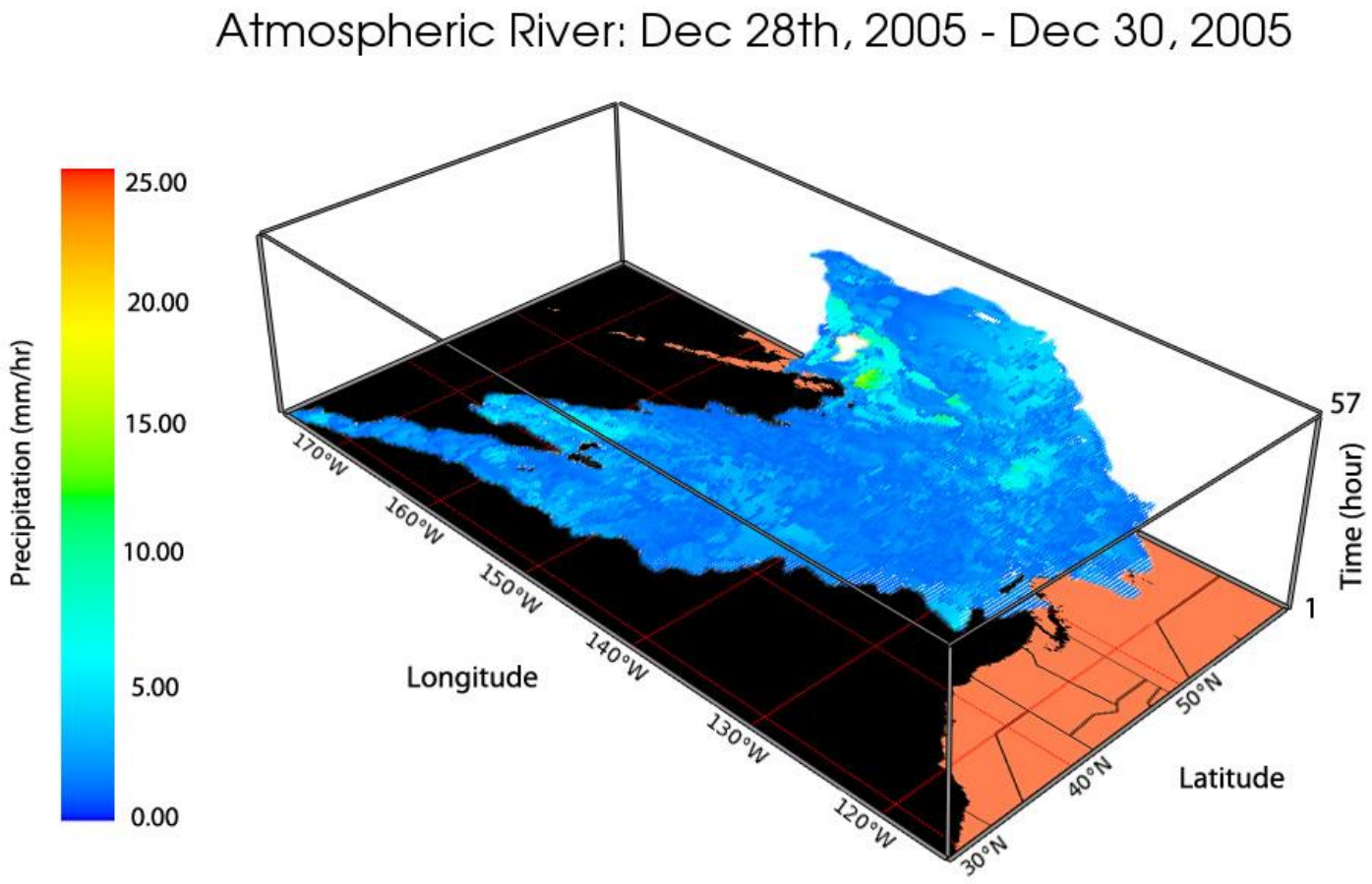
1. Each voxel must have 1mm/hr
2. Each object must exist for 24 hours
3. 6 voxel connections

Database Indexes:

1. Object ID Number
2. Latitude (of each voxel in objects)
3. Longitude (of each voxel in objects)
4. Time (hour)

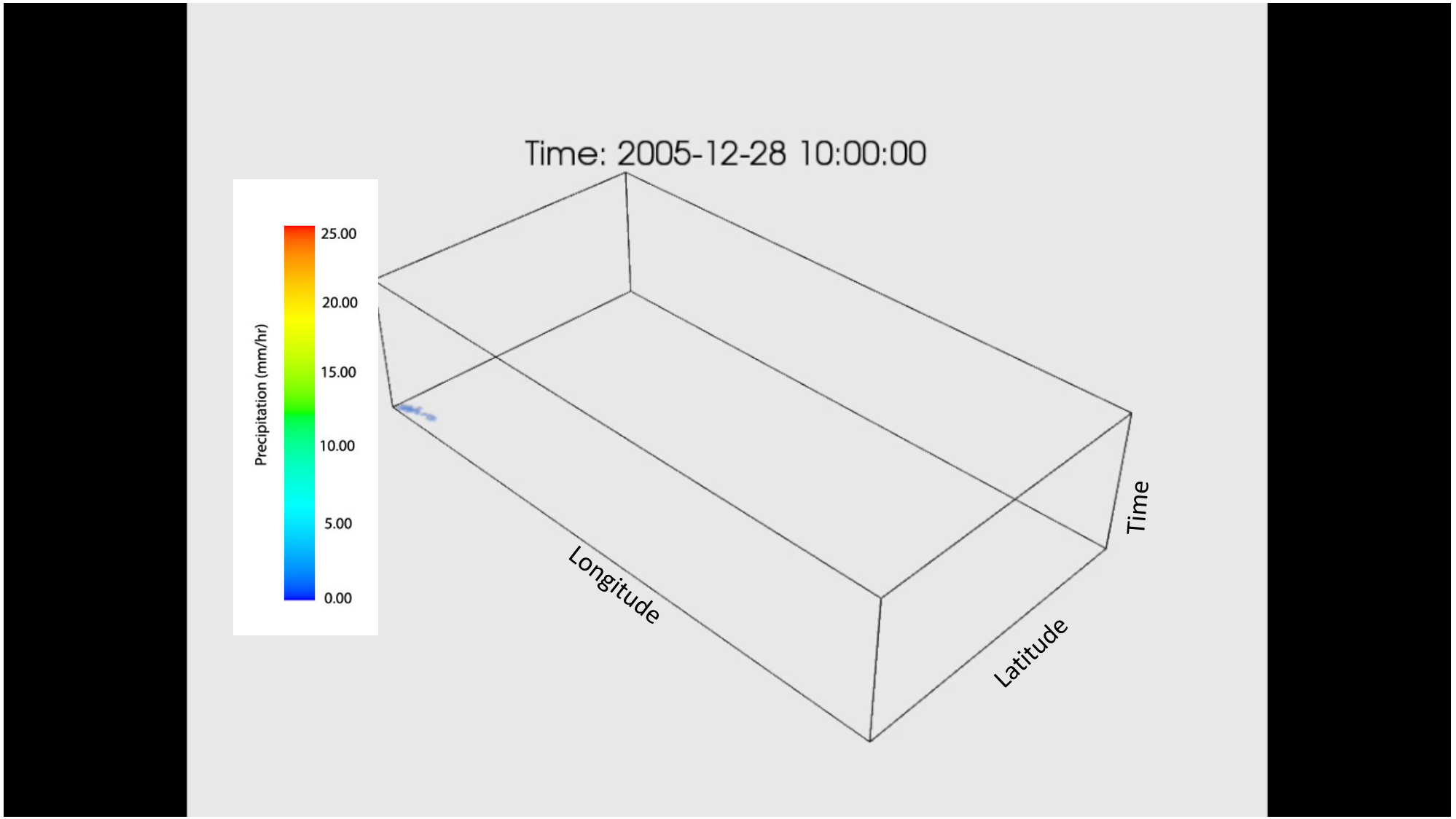
CW3E

4-D Precipitation object example



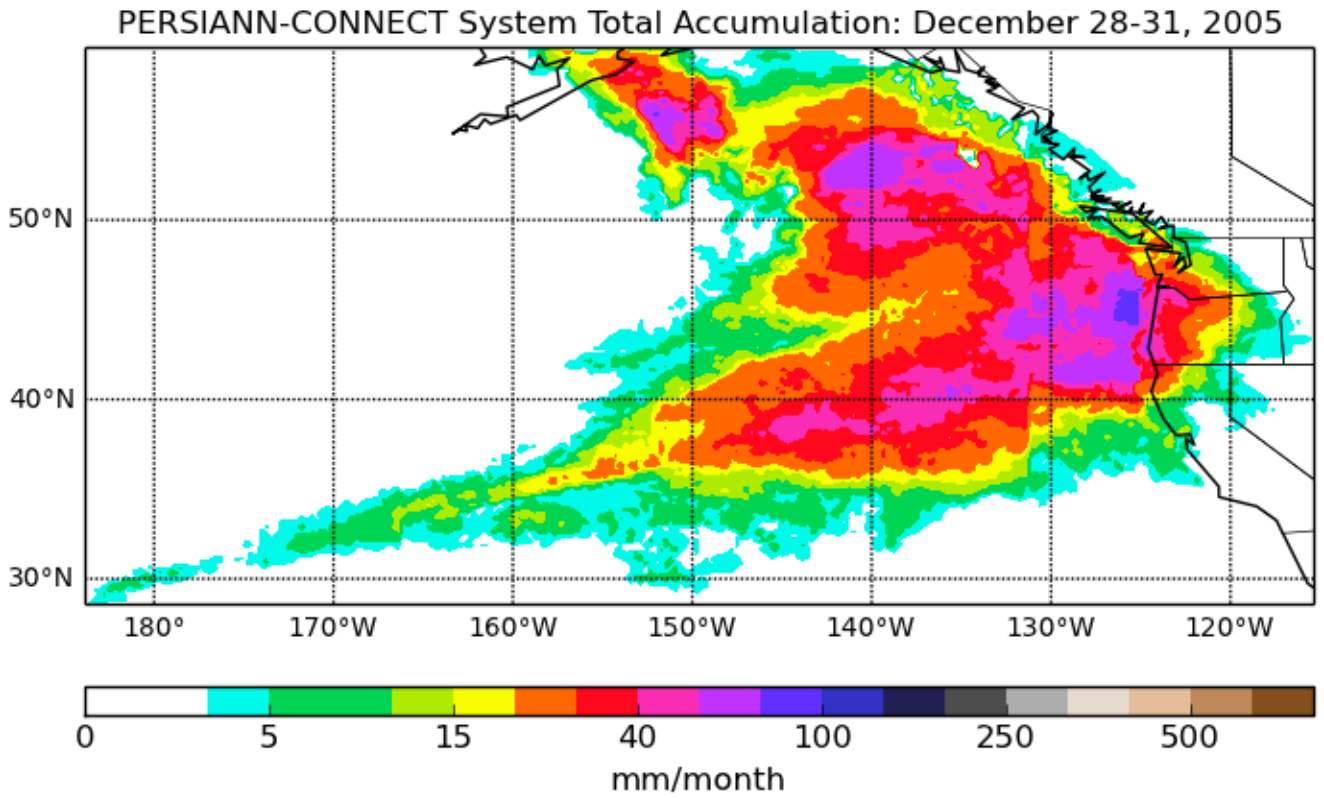
CW3E

4-D Precipitation object example

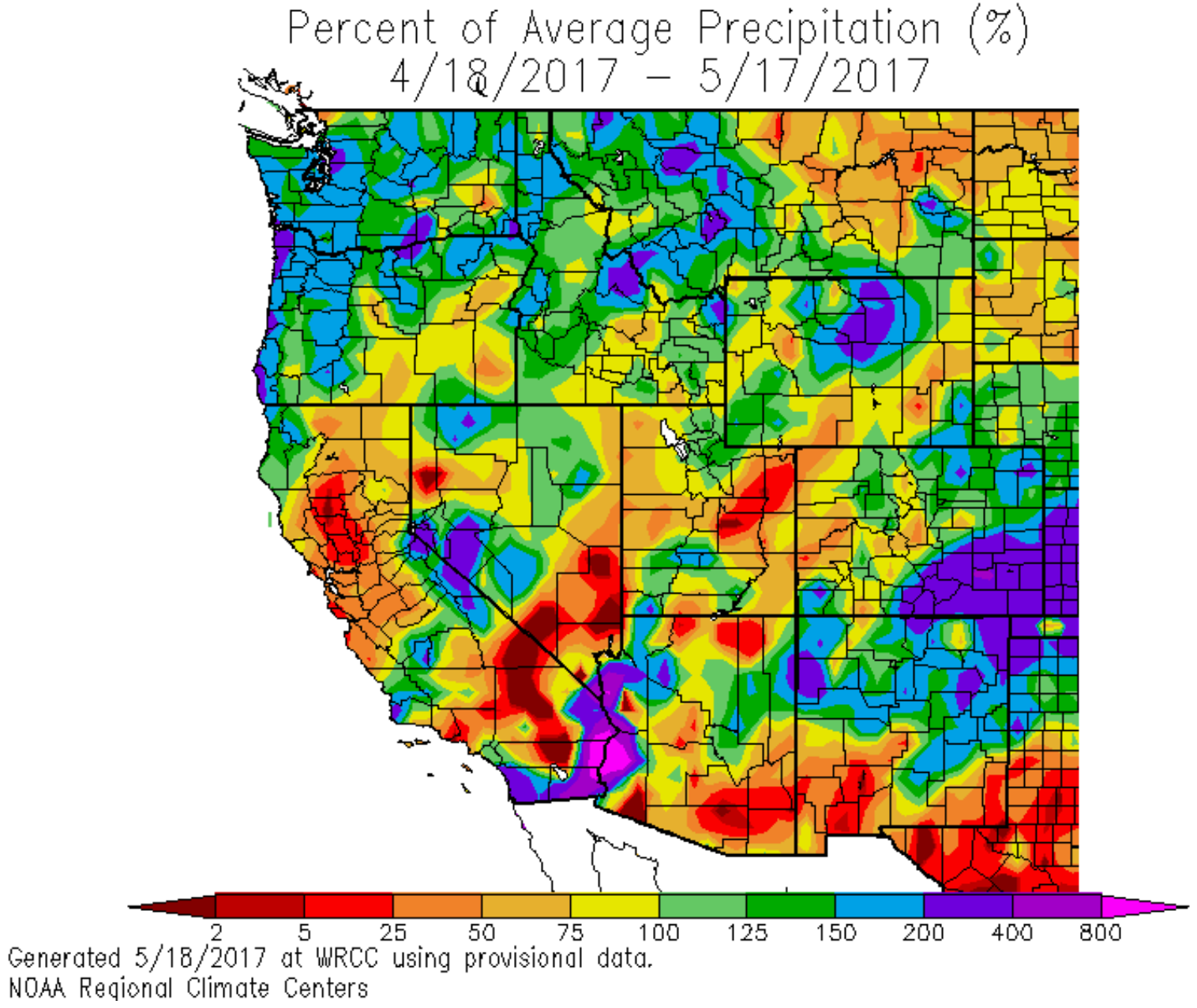


CW3E

4-D Precipitation Object Precipitation Accumulation Map



CW3E


Percent of Average Precipitation (%)
4/18/2017 – 5/17/2017
2
5
25
50
75
100
125
150
200
400
800
Generated 5/18/2017 at WRCC using provisional data.
NOAA Regional Climate Centers

CW3E

TABLE A.1: 4D Precipitation Object Features

Object Features		
Climate Features	Physical Features	Reanalysis Features
ANOM12	Avg Intensity	d_-102.5x15.0_130.0x12.5
ANOM3	Cenroid Latitude	d_-102.5x15.0_150.0x15.0
ANOM34	Centroid Longitude	d_-12.5x45.0_-10.0x67.5
ANOM4	Duration	d_-15.0x77.5_-12.5x45.0
CAR	Ending Latitude - Centroid	d_-155.0x40.0_60.0x12.5
NTA	Ending Longitude Centriod	d_-157.5x67.5_170.0x35.0
AAO	Id	d_-160.0x75.0_130.0x42.5
AMMSST	Max Intensity	d_-160.0x75.0_170.0x35.0
AMON	Median Intensity	d_-162.5x22.5_-157.5x67.5
AO	Speed	d_-167.5x55.0_170.0x35.0
EA	Starting Latitude Centriod	d_-167.5x55.0_175.0x20.0
EPO	Starting Longitude Centriod	d_-180.0x67.5_-162.5x22.5
GLAAM	Standard deviation Intensity	d_-180.0x67.5_130.0x42.5
GMSST	Time of Year	d_-180.0x67.5_170.0x35.0
MEI	Volume	d_-180.0x67.5_175.0x20.0
MJO RMM1		d_-27.5x35.0_-10.0x67.5
MJO RMM2		d_-45.0x37.5_-10.0x67.5
MJO Amplitude		d_-50.0x17.5_-35.0x52.5
MJO Phase		d_-90.0x20.0_130.0x12.5
NAO		d_-90.0x20.0_150.0x15.0
NP		d_110.0x25.0_132.5x80.0
ONI		d_112.5x32.5_132.5x80.0
PDO		d_130.0x42.5_132.5x80.0
PNA		d_55.0x57.5_95.0x35.0
QBO		d_65.0x72.5_110.0x25.0
SOI		d_65.0x72.5_112.5x32.5
Solar		d_65.0x72.5_95.0x35.0
SW MONSOON		
TNA		
TNI		
TSA		
WHWP		
WP		

Object characteristics

Gather characteristics from each object:

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & & & \\ x_1^N & x_2^N & \cdots & x_d^N \end{bmatrix}$$

N = # of objects

d = # of dimensions (72 in this case)

***Sellars et al. 2015 JHM**

CW3E

Department of Water Resources

FLOOD EMERGENCY RESPONSE INFORMATION EXCHANGE (FERIX) - PERSIANN-CONNECT Storms

HOME | REAL TIME HYDROLOGY | FLOOD CONTROL SYSTEM | CLIMATE | FLOOD H&H | FMO

PERSIANN-CONNECT STORMS

Search for enter address here... Search

PERSIANN-CONNECT FROM CHRS, UC IRVINE

SEARCH STORMS

- Start Date: to
- Storm Volume (km3): to
- Storm Duration (hours): to
- Max Intensity (mm/hr): 30 to 100
- IDs (comma separated):

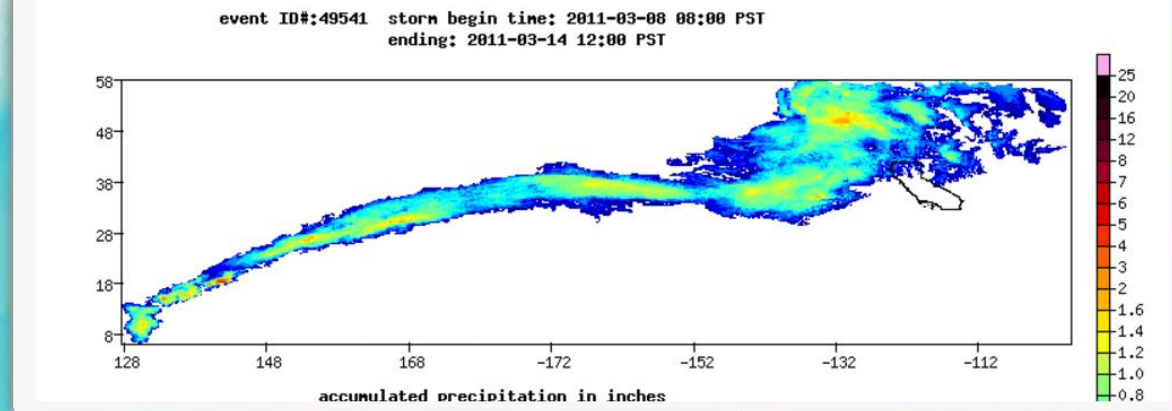
Show Clear

Click on a row to view storm on map. Export

Storm ID	Start Date	End Date ▲	Latitude	Longitude	Start Lat	Sta Lon
73373	2016-01-12 12:00	2016-01-19 20:00	30.516	-198.639	21.069	-237.
73340	2016-01-09 01:00	2016-01-14 19:00	35.719	-183.943	20.518	-249.
73289	2016-01-04 06:00	2016-01-09 01:00	42.522	-157.992	42.247	-192.
73236	2015-12-29 10:00	2016-01-05 05:00	42.212	-142.951	38.375	-185.
73028	2015-12-11 07:00	2015-12-19 13:00	31.385	-182.045	8.638	-222.
72237	2015-10-11 14:00	2015-10-28 14:00	28.104	-156.532	8.925	-108.

Info | Select | Graphics | Tools | Bookmark | Base Map | Layer | Legend

MEI	MJO RMM1	MJO RMM2	MJO Amplitude	MJO Phase	NAO	NOI
-1.511	-1.1169273	0.16821386	1.129523	8	0.24	0
NP Pattern	**Average Intensity**	**Duration**	**Max Intensity**	**Median Intensity**	**Speed**	**STDDEV Intensity**
1,009.34	2.013 mm/hr	149 hours	34.812 mm/hr	1.671 mm/hr	87.216 km/hr	1.183 mm/hr
Time of Year	**Volume**	**ONI**	**PAC Warm**	**PDO**	**PNA**	**QBO**
3	137.065 km3	-0.74	0	-0.69	0.38	10.44
SOI	**Solar Flux**	**SW Monsoon**	**TNA**	**TNI**	**TSA**	**WHWP**
4.2	1,158	0	0.41	1.546	0.72	-0.85
WP						
-0.1						

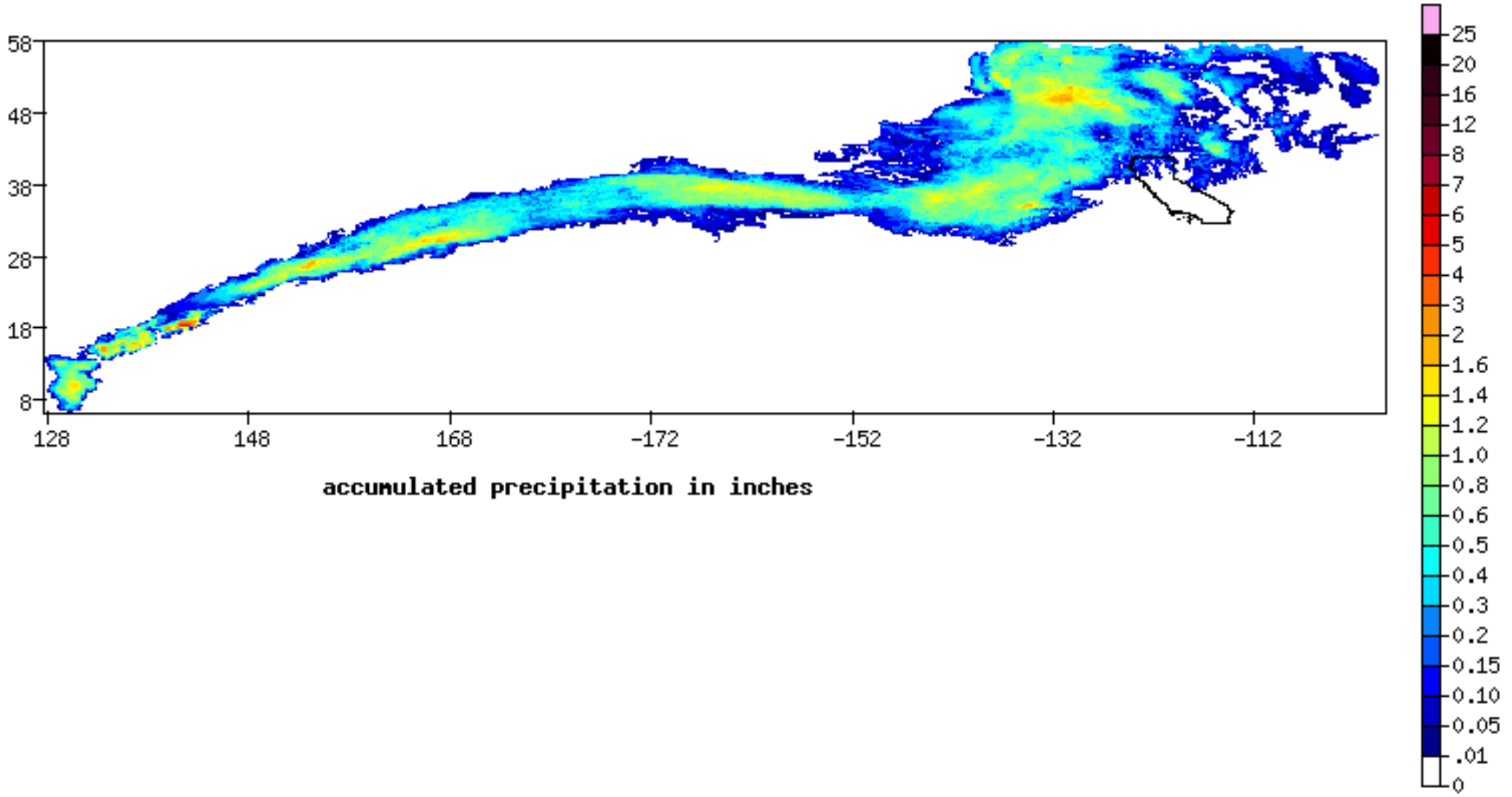


44966 34880 48459 24881 54861 73340 48057 11784 35278 16536 44053 63622 34573 25649 53778 44288 3601

0 200 400mi

Current Tool: Identify Active Layer: Lat/Long:47.983900, -145.906339

event ID#:49541 storm begin time: 2011-03-08 08:00 PST
ending: 2011-03-14 12:00 PST



CW3E

Western U.S. "Objects" – How does ENSO impact them?

	Avg intensity ($mm\ h^{-1}$)	Max intensity ($mm\ h^{-1}$)	Speed ($km\ h^{-1}$)	Hit California	Volume (km^3)	Duration (h)	Lat centroid (°N)	Lon centroid (°W)
ENSO								
El Niño (185)	2.5	16.9	44.6	73.5%	95.05	62.7	39.8	206.5
Neutral (248)	2.5	**15.3**	43.7	**65.3%**	**60.75**	54.4	40.3	211.3
La Niña (193)	2.6	**19.3**	46.9	70.5%	90.97	62.1	39.3	207.3

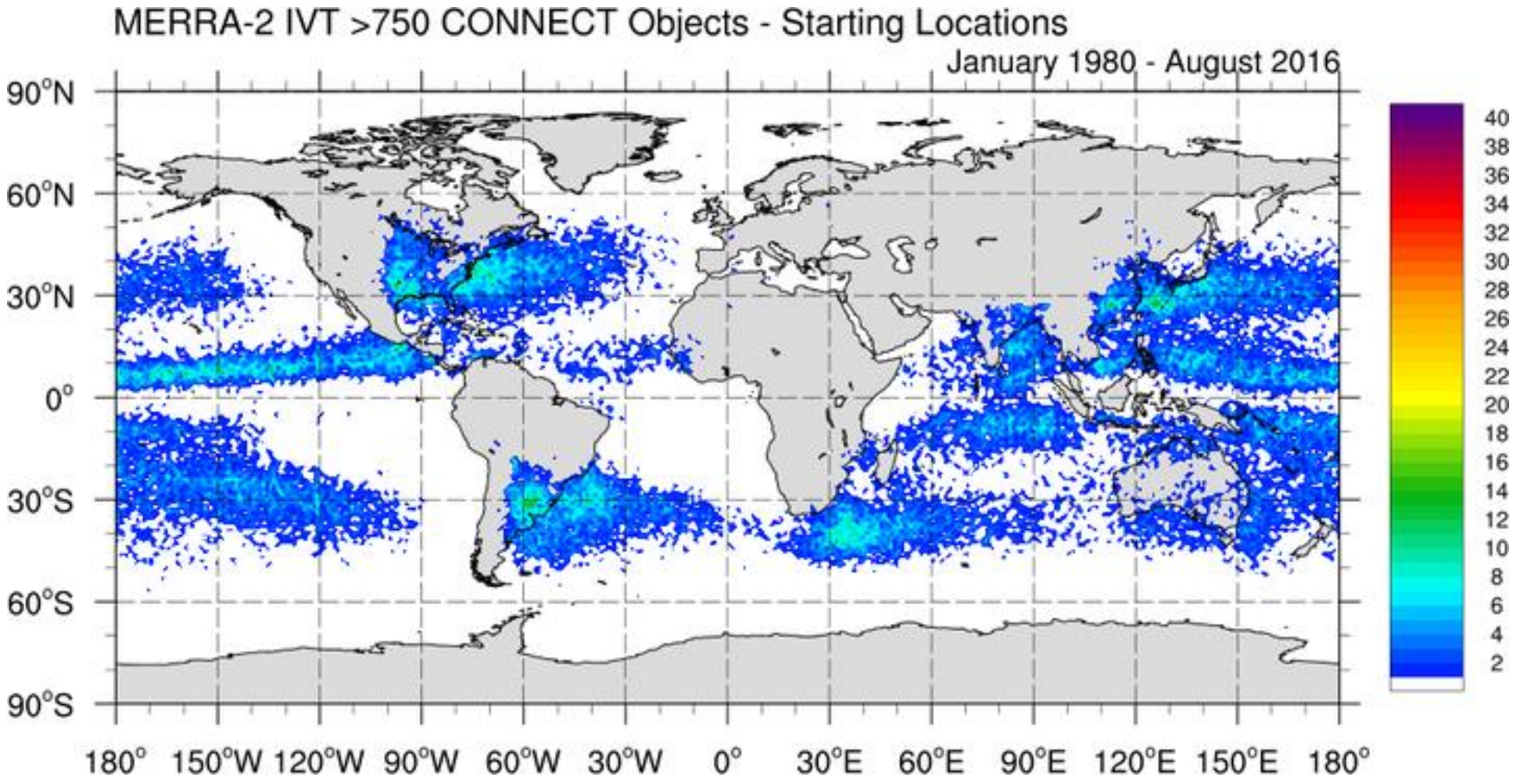
Bold letters indicate statistical significance at the 95% level.

El Nino/La Nina phases of ENSO have objects that are **larger**, **faster**, **longer lasting**, and begin slightly further east in the Pacific Ocean.

Sellars et al. 2015 JHM

CW3E

IVT Objects > 750kg m-1 s-1



*Sellars et al., 2017 (in prep)

CW3E

PAST: Big Data Analysis Pipeline: One Variable

2.4T
Time:
7d 10h 49min

NASA

Download

2.4T to 100GB
Time:
10d 5h 05min

CW3E

Data organization
Variable format

100GB to 50GB
Time:
~1d 14h 00m

CHRS

CONNECT
Segmentation

50GB to 100MB
Time:
~1d 5h 00m

C3WE

CONNECT
Characteristic Calc.

Total time: ~20d 11h 0m

- **Not including data visualization**
- **Not including data mining/machine learning jobs**
- **Assumes we know what we are doing**

CHRS

Data Visualization
And Search

CW3E

Data mining and
Discovery,
Machine Learning

CW3E

Then I Talked to PRP and Research IT Engineers!

CENIC and Calit2's PRP

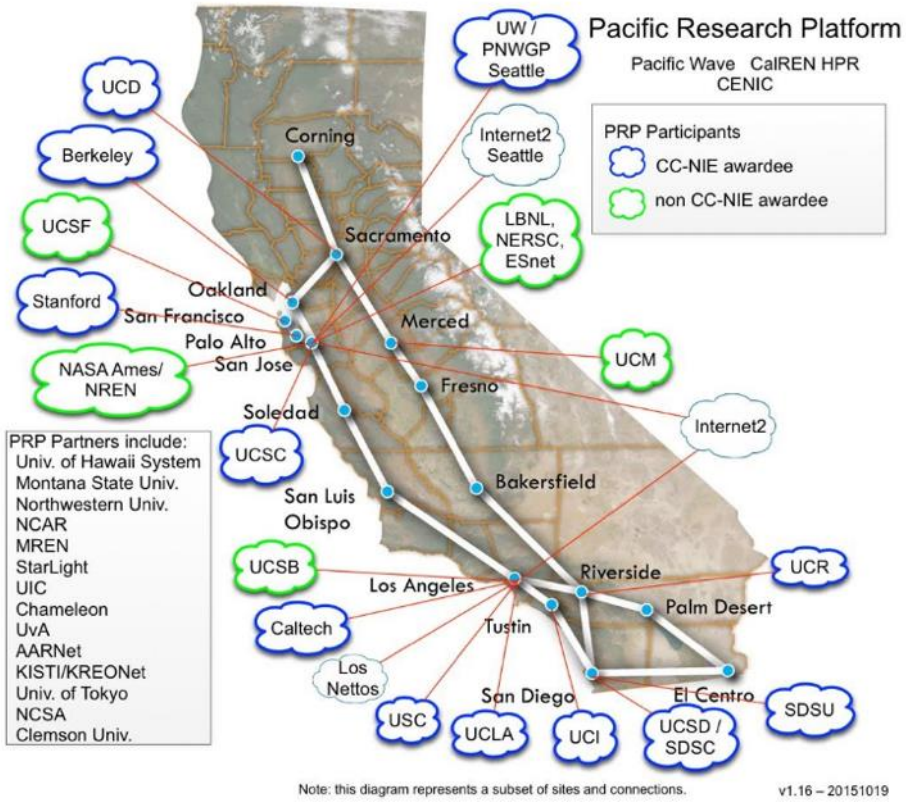


*image courtesy of Larry Smarr

Tools

FIONA – Flash I/O Network Appliance: Linux PCs Optimized for Big Data on DMZs

*SC 2015

Working with:
John Graham and Joseph Keefe

CW3E

Results, in Megabytes, to Date

- Before PRP (**7/16**) NASA **---->** local connection **---->** SIO
  - **10MB/s** download critical NASA data (2.4TB took over 7 days)

- Initial PRP test (**8/16**) NASA **---->** PRP connection **---->** SIO FIONA
  - **40MB/s (4x increase simply using the PRP and a FIONA DTN)**

- PRP/CONNECT (**4/17**) SIO FIONA **---->** via THREDDS **---->** UCI FIONA
  - **559MB/s (56x increase)**
  - **Could be as much as 1896MB/s between FIONAs using Globus (almost 200x better) based on local testing**

FUTURE:

Pacific Research Platform (10-100 Gb/s)

jupyter

GPUs

Calit2's FIONA

Complete workflow time: 20 Minutes!

SDSC's COMET

jupyter

GPUs

Calit2's FIONA

SCRIPPS INSTITUTION OF OCEANOGRAPHY UCSD

CENTER FOR HYDROMETEOROLOGY & REMOTE SENSING CHRS UCIrvine

CW3E

Conclusions

- Big Data Approaches: CONNECT/PRP (Pilot Project) - Multi-investigator, multi-institutional, and need more rapid access to ultra-large heterogeneous distributed datasets
- Using PRP network via FIONA
  - **Download speed from NASA increased 4x** (40MB/s which is 4x faster than the 10MB/s standard connection previously being used by researchers at SIO).
  - **Removing data transfer as a limiting constraint changes everything!**
  - Researchers and scientists need to rethink how we are approaching data transfer and analysis
- PRP/CW3E/SIO "Big Data and the Earth Sciences: Grand Challenges Workshop"
  - May 31st to June 2nd, Four keynote lectures
  - 75+ registered participants

CW3E

Thank you!

scottsellars@ucsd.edu

Acknowledgements:
NSF The Pacific Research Platform Project #ACI-1541349
California Department of Water Resources
NASA Earth and Space Science Fellowship 2014 (NESSF)

CW3E