LA-UR-15-27738
Approved for public release; distribution is unlimited.

Title: Summer of `15: An Exit Talk

Author(s): Biswas, Ayan

Intended for: Web

Issued: 2015-10-05

Summer of `15: An Exit Talk

Ayan Biswas

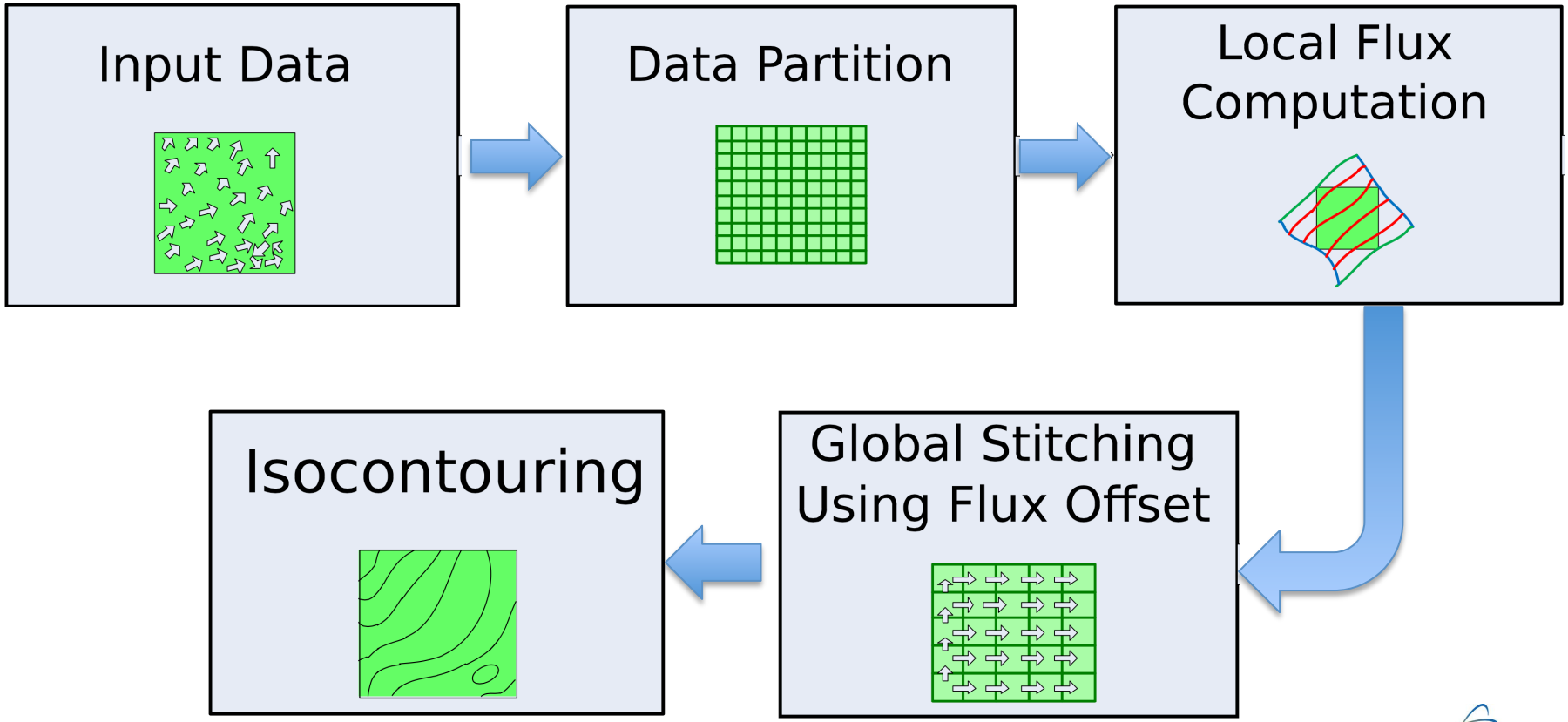
The Ohio State University
Los Alamos National Laboratory

Overview

- Parallel streamline work
  - Started last summer
  - Out-of-core to multi-core
  - Handle discontinuities
  - Add error based flux propagation
- Information theory based variable selection
  - R/C++ based implementation
  - Incorporate into ADR project

Parallel Streamline

- Conceptualized last year

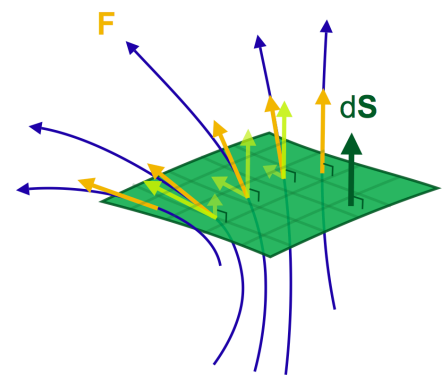


Parallel Stream Function Based Streamline Generation

- Stream functions compute total flux
- Convert a vector to scalar
- Two-dimensional incompressible flows
  - Divergence $\nabla \cdot v = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0$
- Key ideas
  - Flux is additive
  - Flux is path independent

Flux Properties

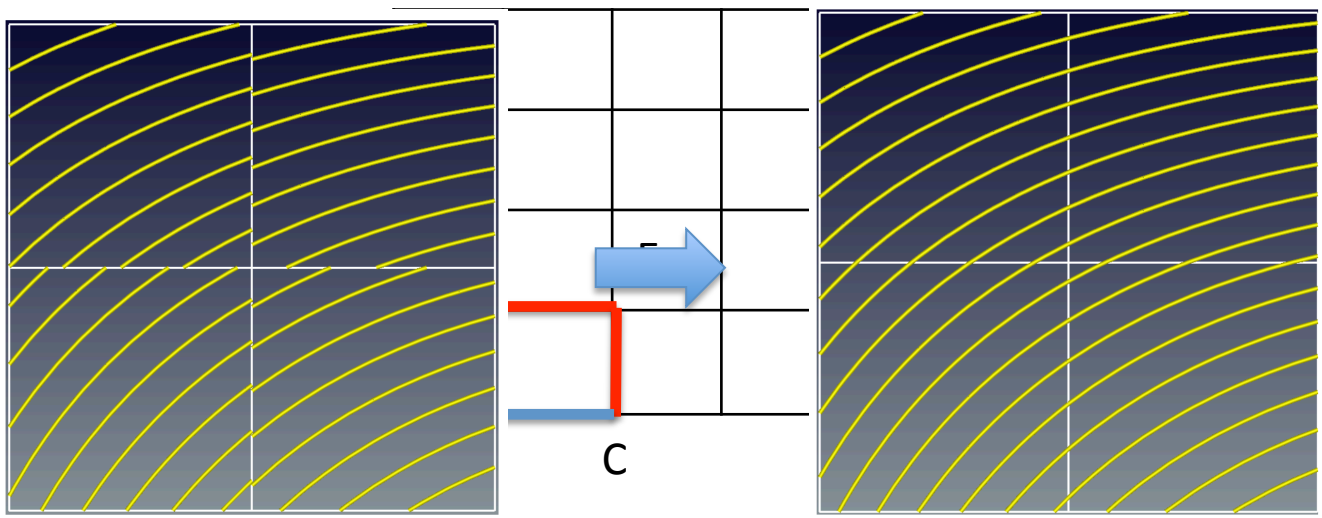
- Flux $\Phi = \int_C \mathsf{F} \cdot dc$



$\mathbf{F} \cdot d\mathbf{S}$
$= (F\cos\theta)dS$
$= F(dS\cos\theta)$

Flux Properties

- Flux $\Phi = \int_C \mathsf{F} \cdot dc$

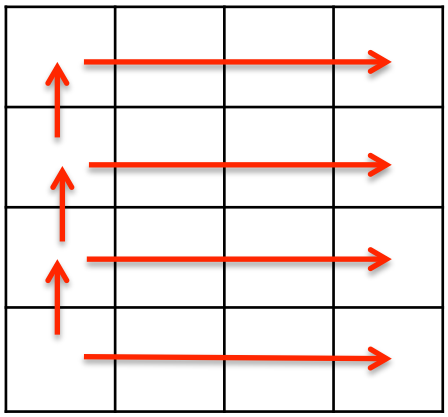


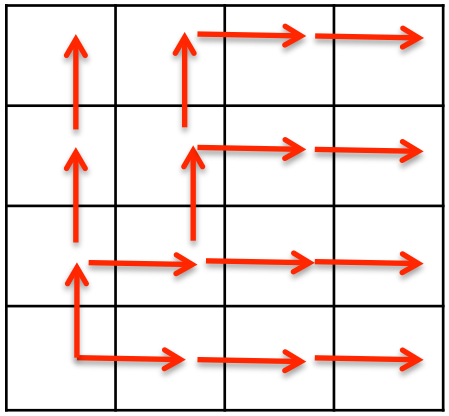
1. Flux(A,B) = Flux(A,C) + Flux(C,B)
2. Flux(A,C) = Flux(A,D) + Flux(D,E) + Flux(E,C)

Issues Solved (1)

- Out-of-Core to Multi-Core

Data

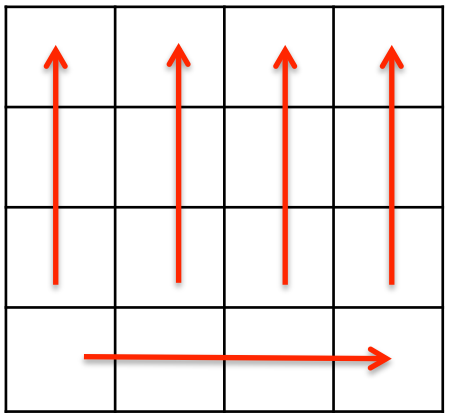


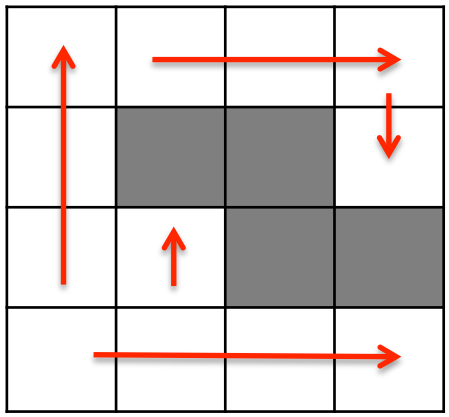


Issues Solved (2)

- Presence of land mass

Block



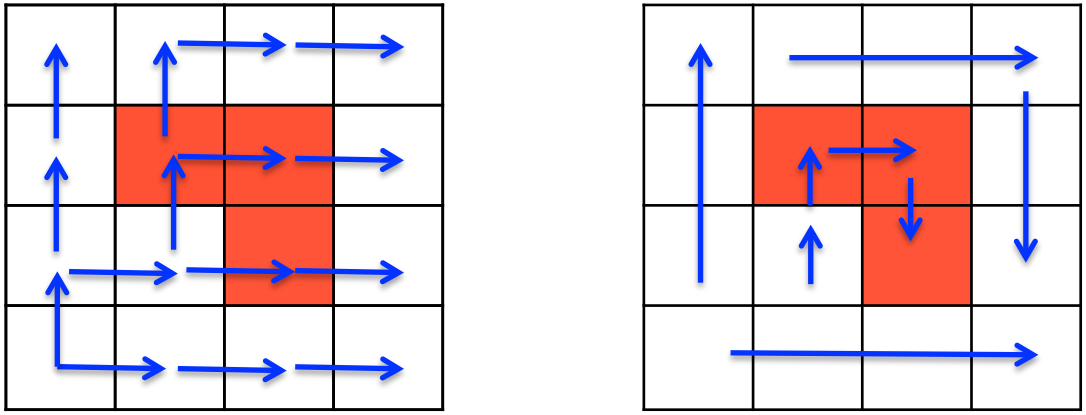


Issues Solved (3)

- Avoid erroneous regions in flux propagation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} > 0$$

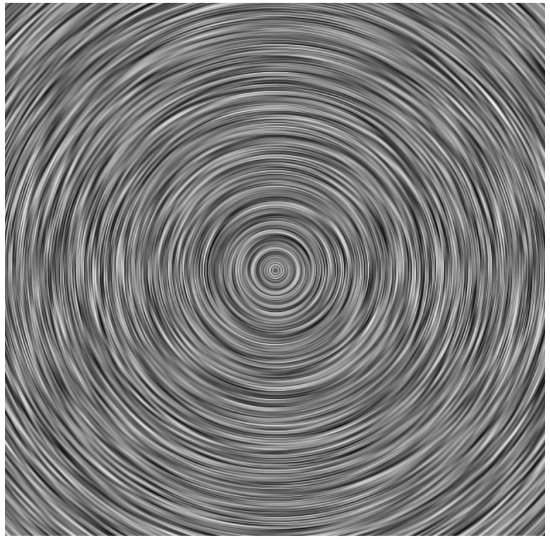
Block

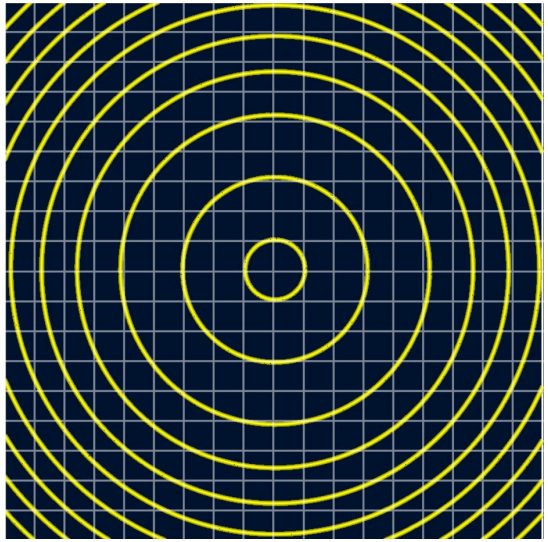


- Spanning tree minimizing the accumulated divergence values

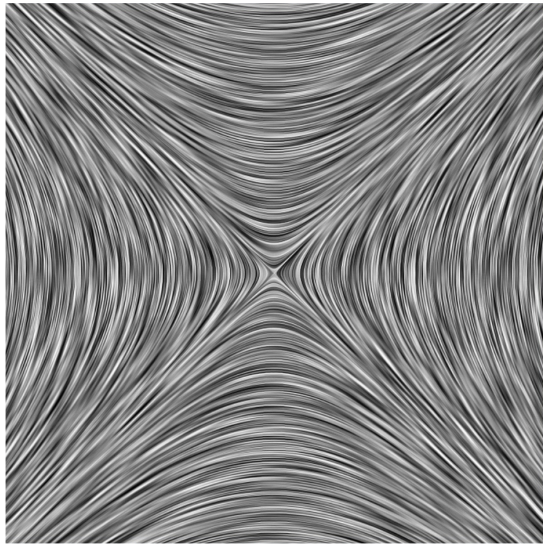
Outputs

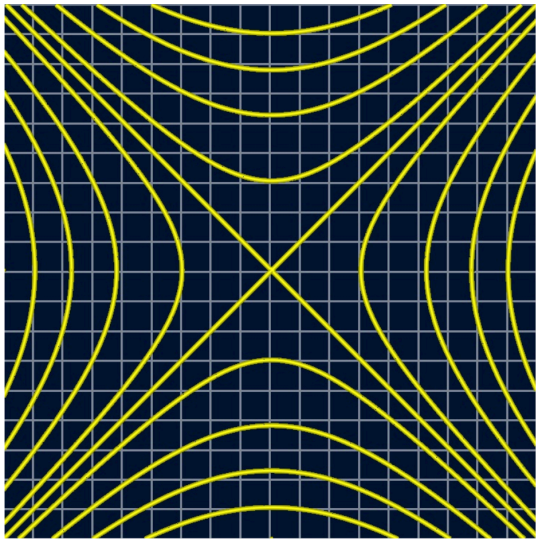
Circle





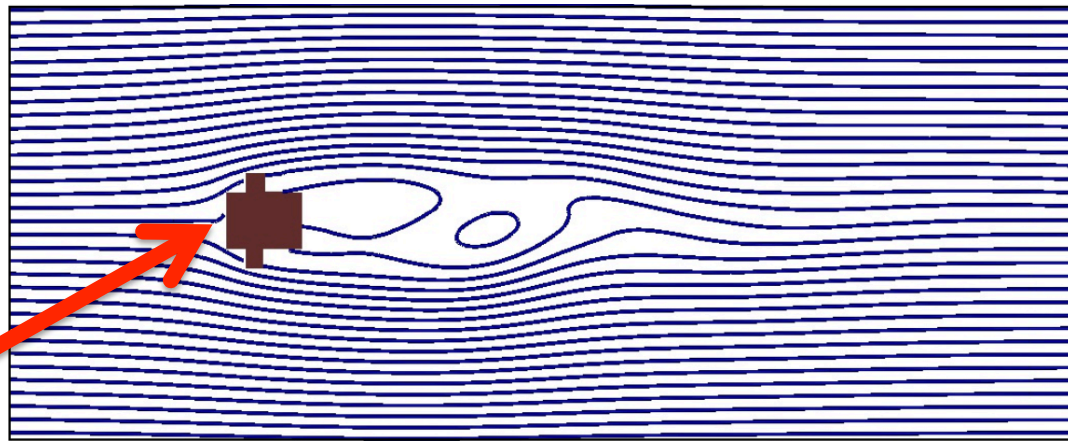
Saddle



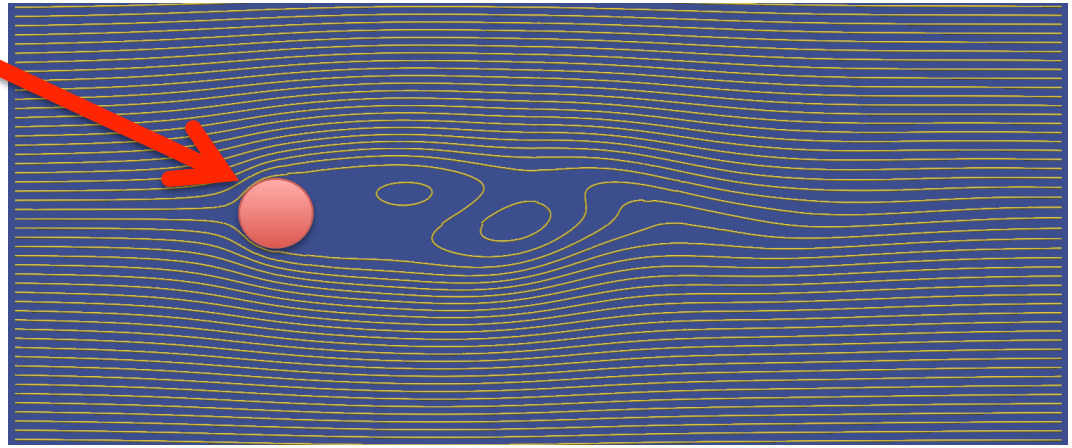


Output

Before

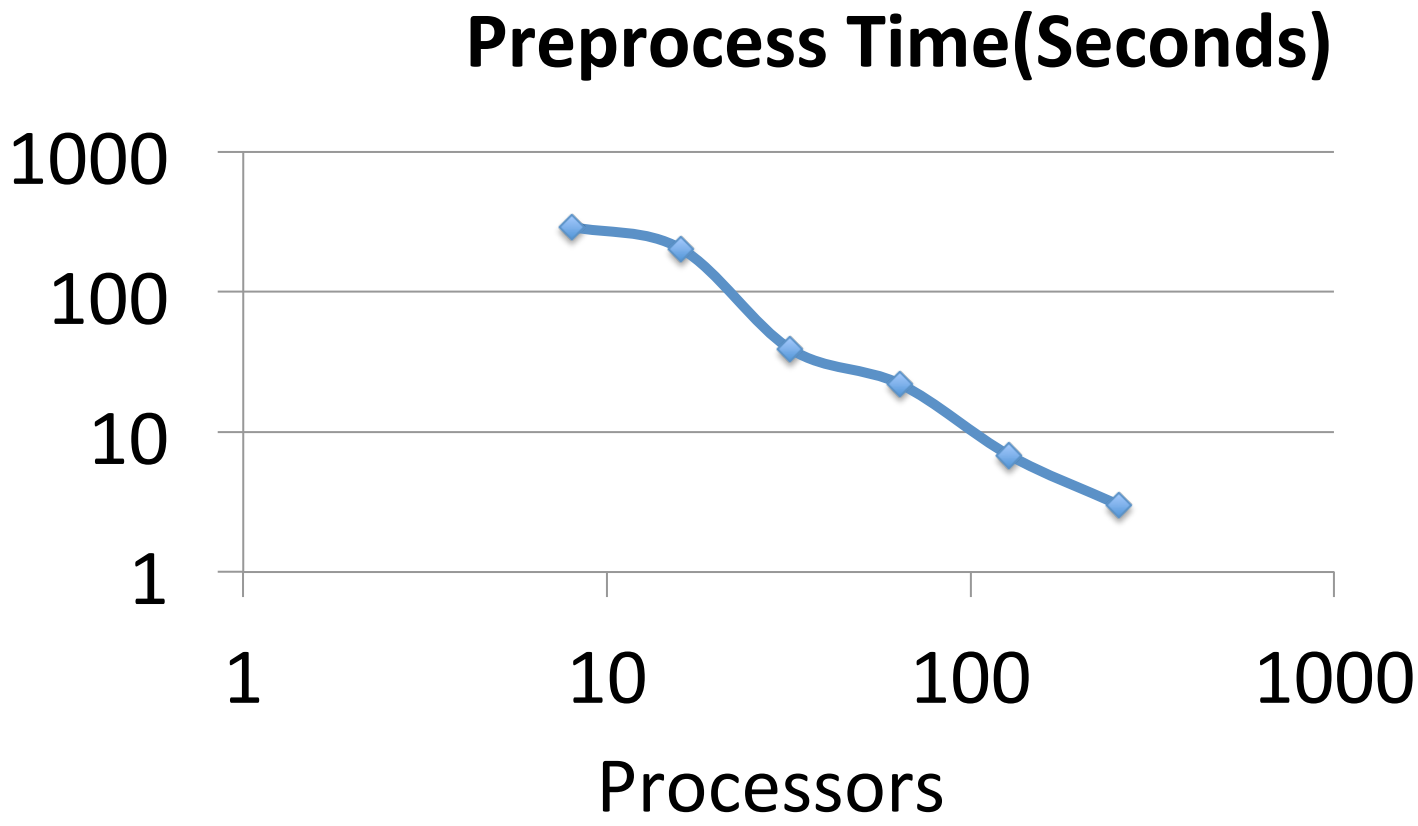


After



Scaling

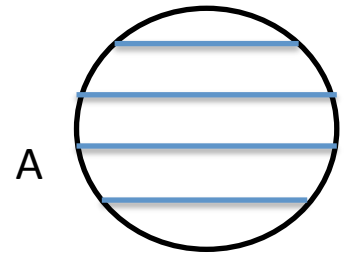
Dataset = 8000 X 4000



Information Theory Based Variable Selection

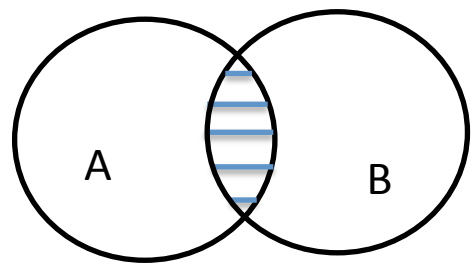
- Entropy

$$H(X) = -\sum_{x \in X} p(x) \log p(x)$$

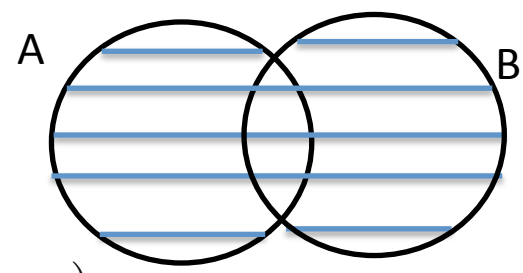


- Mutual Information

$$I(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$



- Joint Entropy



$$H(X_1, \ldots, X_n) = -\sum_{x_1 \in X_1} \ldots \sum_{x_n \in X_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$$

Variable Selection

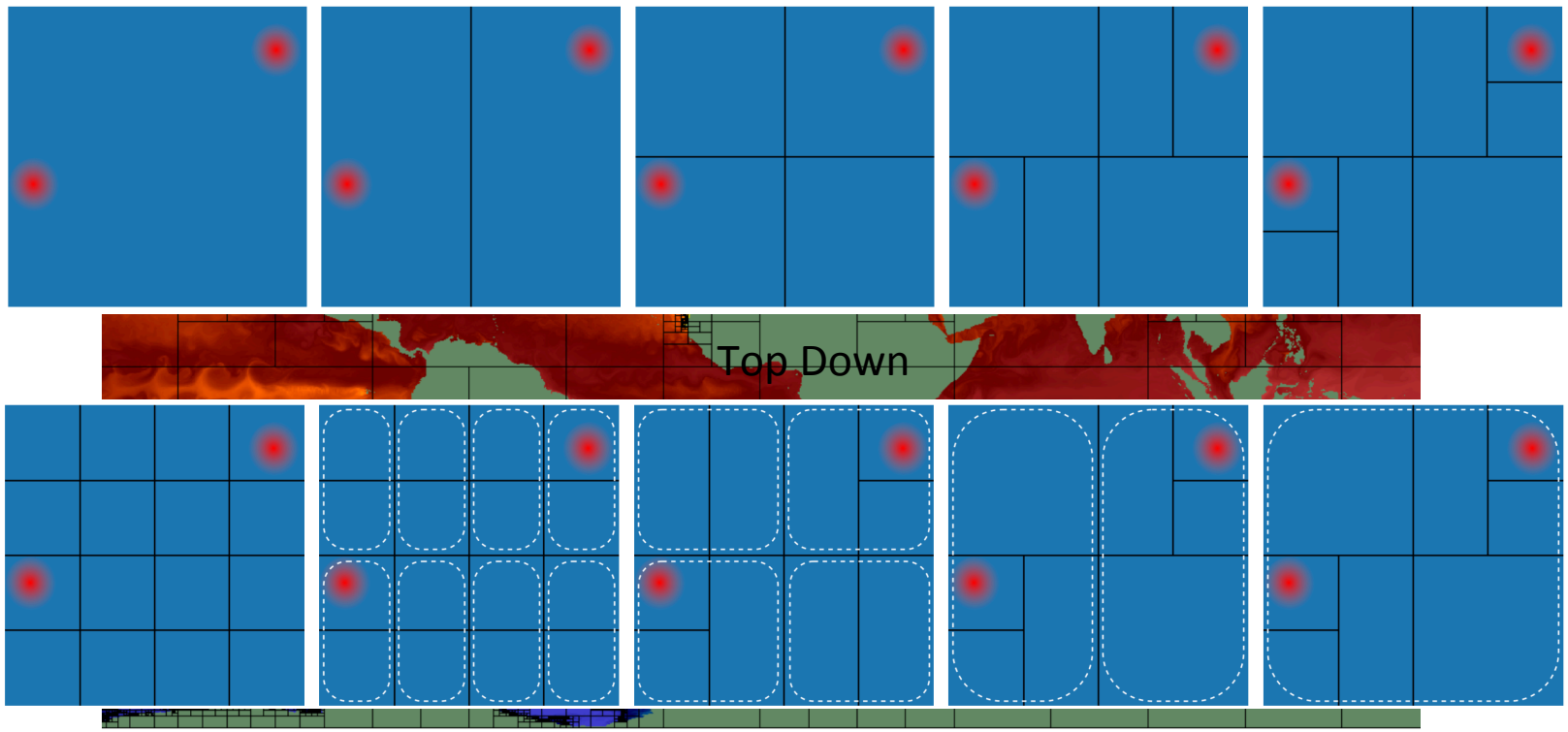
- Choose the most informative variable
  - A variable that reduces the uncertainty of the system the most
  - Variable with highest conditional entropy

$$H(X_1,\ldots,X_n|X_{k1},\ldots,X_{km}) = H(X_1,\ldots,X_n) - H(X_{k1},\ldots,X_{km})$$

- For N variables of the system, N-dimensional histogram is needed
  - C++ "map" can be used to create a sparse representation

ADR

- K-d creation for analysis driven refinement



R/C++ Connection

- ADR project
  - Compute a k-d partitioning tree (using Longitude, Latitude, Z-top) reducing the entropy of another variable (temperature)
  - R markdown and R code base
  - Added C++ code, used RCpp to create an R package, connected it to ADR project base
  - Data input assumed to be in sqlite3 database

Results

- Invoke a function get_top_variables(,,) with arguments

```
Console ~/adr_git/adr_r/
Importance 1 for variable lon
Importance 2 for variable displacedDensity
Importance 3 for variable lat
Importance 4 for variable velocityMeridional
Importance 5 for variable relativeVorticityCell
Importance 6 for variable layerThickness
Importance 7 for variable velocityZonal
Importance 8 for variable tracer1
Importance 9 for variable zTop
Importance 10 for variable potentialDensity
>
```

Results

	splitType	partitionRep	errorRep	variable	stopRange_Flag	rangeVal	stopCellCnt_Flag	minCntVal	imsToSplit	storage	raw_Size	pAIC	sse	estErr
0	medianSplit	mean	max	temperature	T	20	T	100	] "temperature" "temperature" temperature"\n	0.00048	7.8249	0.00125	23.38749	114.62790
1	medianSplit	mean	max	temperature	T	20	T	100	] "temperature" "temperature" "lon" n	0.00048	7.8249	0.00125	23.38749	114.62790
2	medianSplit	mean	max	temperature	T	20	T	100	] "temperature" "temperature" pot...	0.00048	7.8249	0.00125	23.38749	114.62790
3	medianSplit	mean	max	temperature	T	20	T	100	] "temperature" "lon" "lat" \n	0.00144	7.8249	0.00431	79.51687	400.58440
4	medianSplit	mean	max	temperature	T	20	T	100	] "temperature" "zTop" "pot...	0.00064	7.8249	0.00496	43.49770	521.73689
5	midPtSplit	mean	median	temperature	F	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
6	midPtSplit	mean	median	temperature	T	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
7	midPtSplit	mean	median	temperature	T	25	F	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
8	midPtSplit	median	median	temperature	F	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
9	midPtSplit	median	median	temperature	T	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
10	midPtSplit	median	median	temperature	T	25	F	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
11	midPtSplit	mean	mean	temperature	F	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
12	midPtSplit	mean	mean	temperature	T	25	T	100	] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
13	midPtSplit	mean	mean	temperature	T	25	F	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
14	midPtSplit	median	mean	temperature	F	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
15	midPtSplit	median	mean	temperature	T	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
16	midPtSplit	median	mean	temperature	T	25	F	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
17	midPtSplit	midpt	mean	temperature	F	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
18	midPtSplit	midpt	mean	temperature	T	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
19	midPtSplit	midpt	mean	temperature	T	25	F	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
20	midPtSplit	midpt	median	temperature	F	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
21	midPtSplit	midpt	median	temperature	T	25	T	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
22	midPtSplit	midpt	median	temperature	T	25	F	100	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
23	meanSplit	mean	median	temperature	F	25	T	100	[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01117	480.77522	786.26542

Results

'al	dimsToSplit	storage	raw_Size	pAIC	sse	estErr
	[1] "temperature" "temperature" "temperature"\n	0.00048	7.8249	0.00125	23.38749	114.62790
	[1] "temperature" "temperature" "lon" \n	0.00048	7.8249	0.00125	23.38749	114.62790
	[1] "temperature" "temperature" "pot...	0.00048	7.8249	0.00125	23.38749	114.62790
	[1] "temperature" "lon" "lat" \n	0.00144	7.8249	0.00431	79.51687	400.58440
	[1] "temperature" "zTop" "pot...	0.00064	7.8249	0.00496	43.49770	521.73689
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01022	423.89056	726.01771
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
	[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283

Results

dimsToSplit	storage	raw_Size	pAIC	sse	estErr
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01025	439.28235	716.09524
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01030	423.89056	735.02283
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01074	439.28235	773.27814
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01116	471.47647	782.10347
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.01117	471.47647	783.39042
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01117	480.77522	786.26542
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01117	480.77522	786.26542
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01117	480.77522	786.26542
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01121	500.67075	775.07022
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01121	500.67075	775.07022
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01121	500.67075	775.07022
[1] "lat" "lon" "zTop"\n	0.01056	7.8249	0.01153	480.77522	827.93370

Results

dimsToSplit	storage	raw_Size	pAIC	sse	estErr
[1] "lat" "displacedDensity" ...	0.01056	7.8249	0.06116	1230.65859	5999.32021
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.06134	2876.39039	4360.58344
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.06154	2876.39039	4385.52915
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.06687	1003.07475	7035.65393
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.06687	1003.07475	7035.65393
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.06687	1003.07475	7035.65393
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.07365	2876.39039	5842.65369
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.08861	1384.84726	9481.91599
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.08861	1384.84726	9481.91599
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.10525	2126.52240	10823.14845
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.12647	1594.04676	14248.55117
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.12647	1594.04676	14248.55117
[1] "lon" "displacedDensity" "lat...	0.01056	7.8249	0.14560	2754.96702	15786.21419
[1] "tracer1" "zTop" "pot...	0.03072	7.8249	0.31673	3575.96139	47064.96263
[1] "tracer1" "zTop" "pot...	0.03072	7.8249	0.31673	3575.96139	47064.96263

Questions?

- Thank you!!