

Understanding turbulent structures

Scott B. Baden

A Challenge

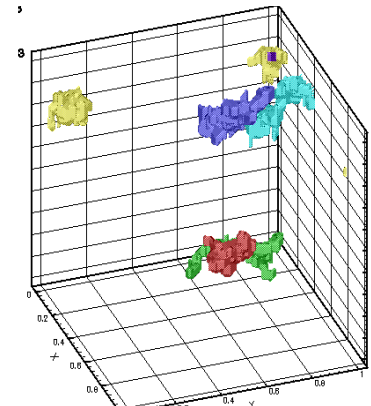
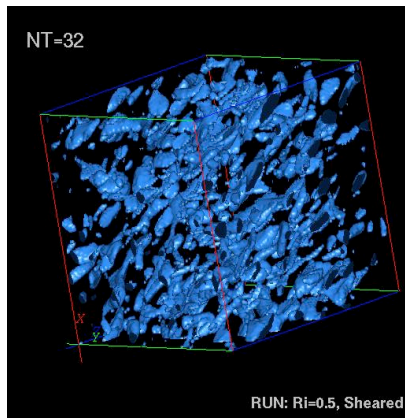
- Deriving knowledge from scientific data to understand technologically important physical phenomena
- How do we find what we are looking for?
- Do we know what we are looking for?
- How can we **remove barriers** limiting our ability to employ simulations of ever improving fidelity and accuracy?

A steady flood of data

- Legacy of exponential growth in computation and storage capacity
- Once a large dataset is committed to tertiary storage, it may never see the light of day

Observations

- The more we know about the how the data was generated ...
- ... the context in which the data will be used
- ... the more we can employ that knowledge to improve productivity



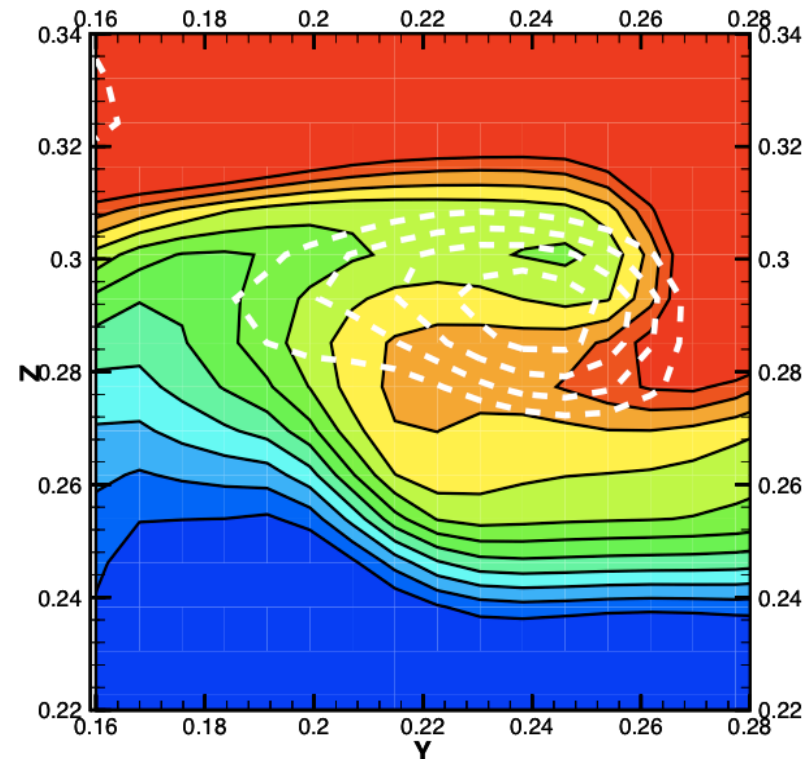
Diamessis, Kerney, Nomura, Baden [Para '02]

Roadmap

- Turbulence
- Data understanding
 - Feature tracking
 - Spatial data bases
 - The link between compression and data understanding

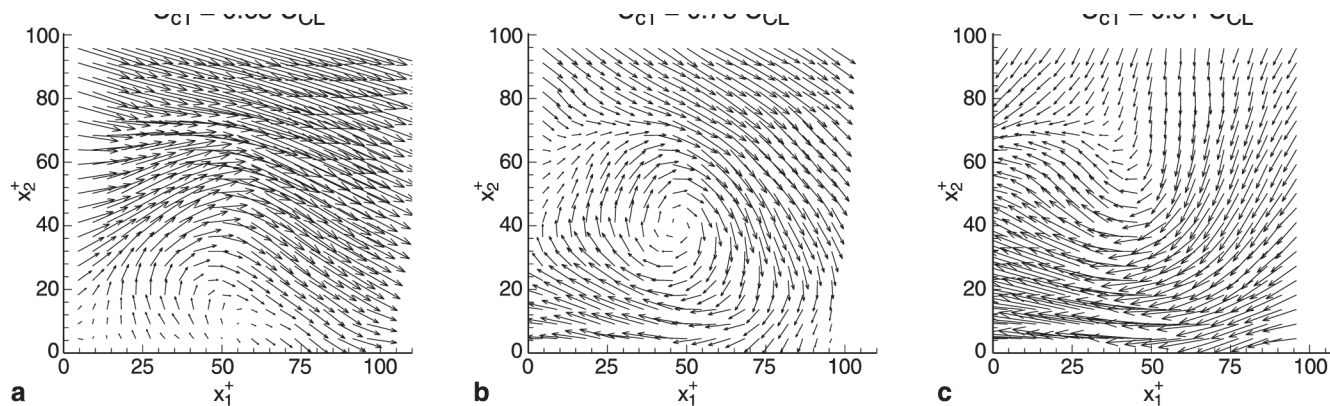
Motivating Application

- Turbulent vertical transport and mixing in the ocean and inland bodies of water (Diamessis and Nomura)
- Study *overturns*, regions of heavy fluid over lighter fluid
- Indicate the ability of turbulence to overcome the damping effect of mean stratification
- Do overturns constitute active mixing sites?



Turbulence

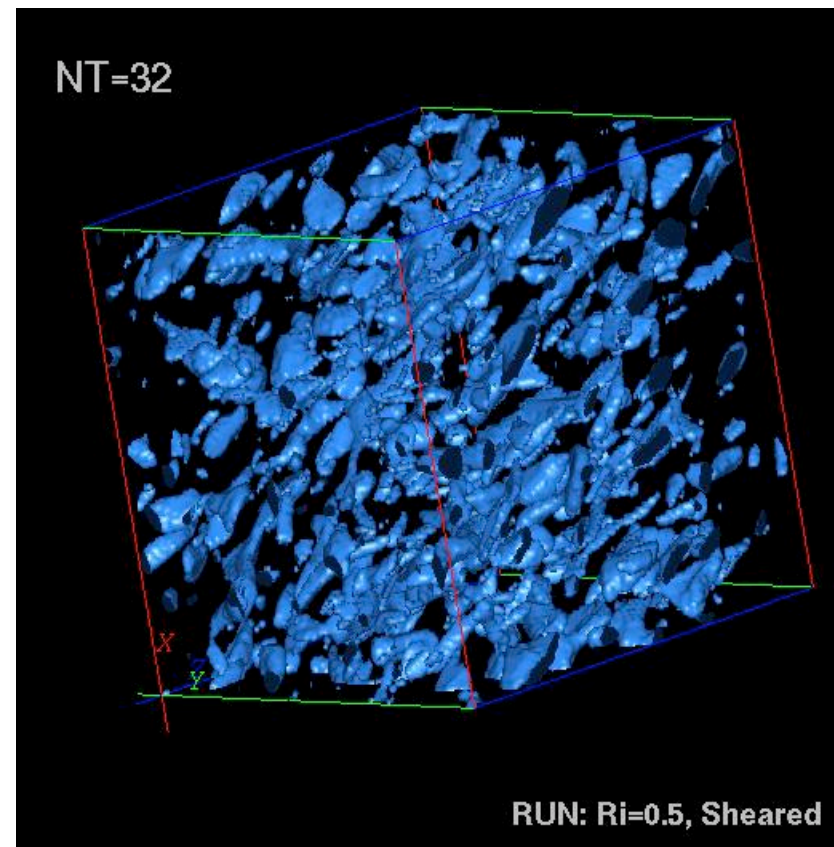
- Vorticity, curl of velocity: $\text{curl}(\mathbf{u})$
- Vortex structure (core) – no agreed upon identification criterion



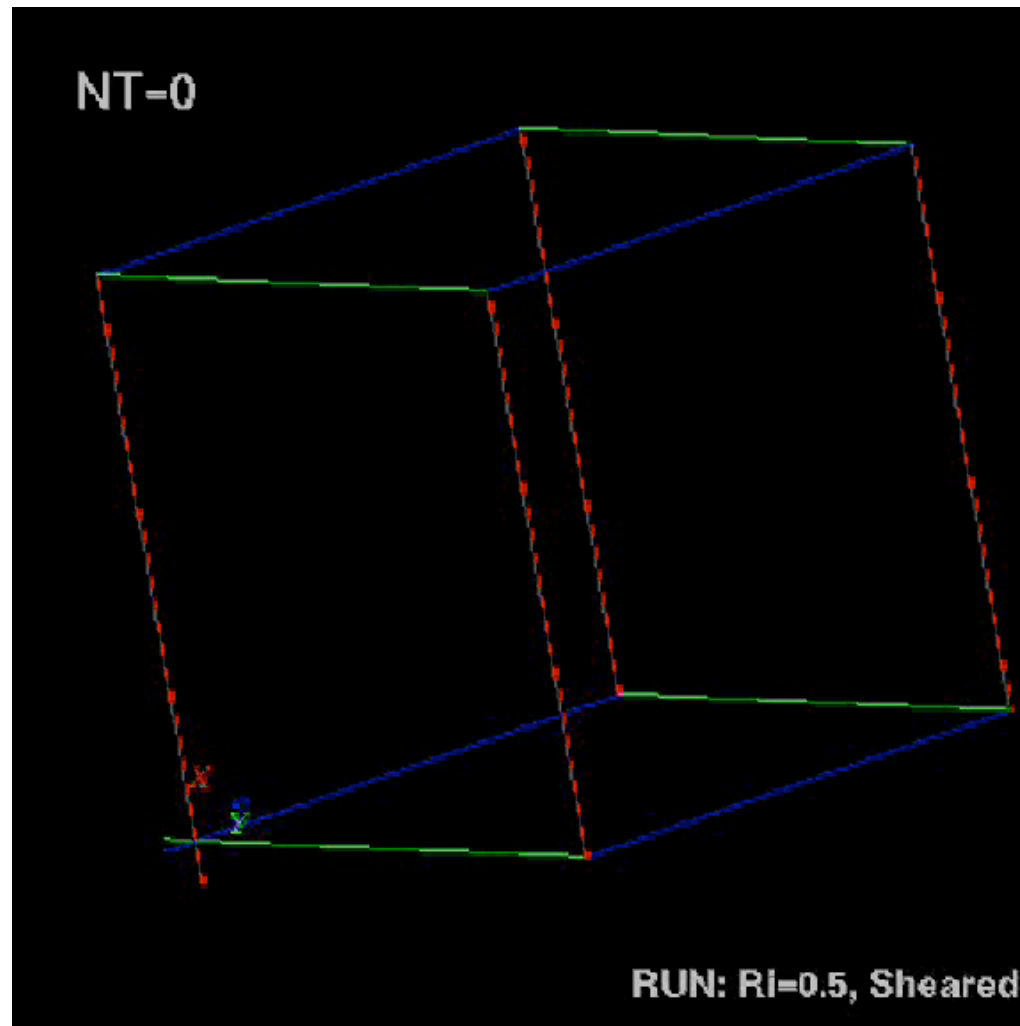
R. J. Adrian, K. T. Chistensen, Z.-C. Liu, *Exp. Fluids* **29**:275-290 (2000)

Numerical Simulaton

- Direct Numerical Simulation (DNS)
- $N=512$: 0.80 TB storage / run
- Parameter studies: 5-10 runs

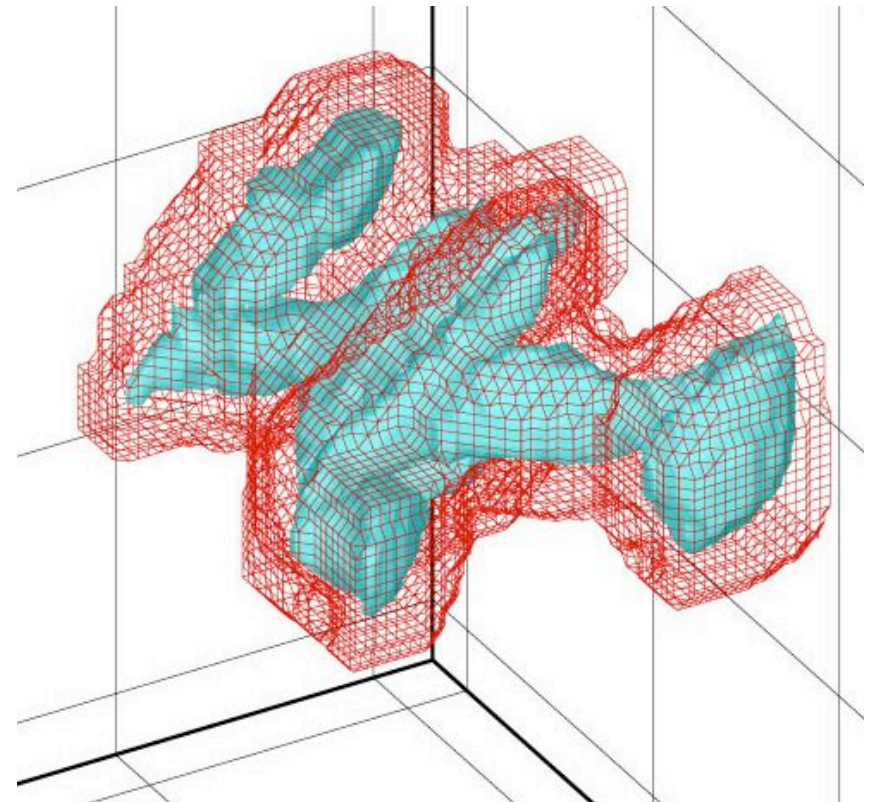


Animation



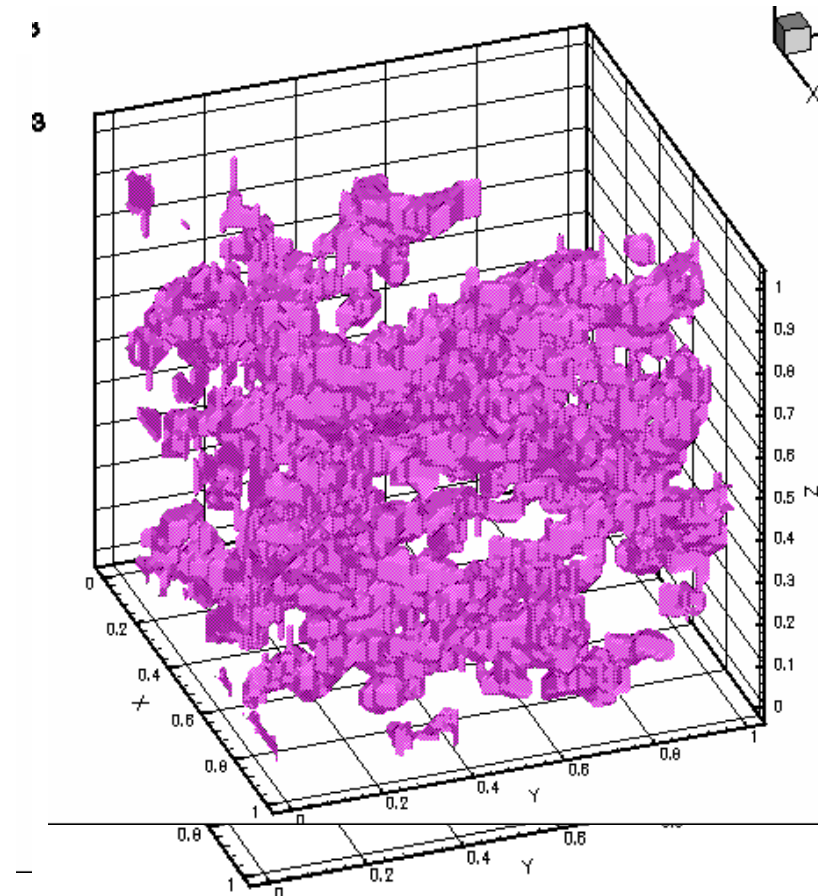
The traditional model of data discovery

- Analyze the dataset off-line
- Post process the data by computing derived quantities, e.g. vorticity
- Employ subjective visual analysis to guide the search
 - Find something that appears “interesting”
 - Apply computations to the interesting data



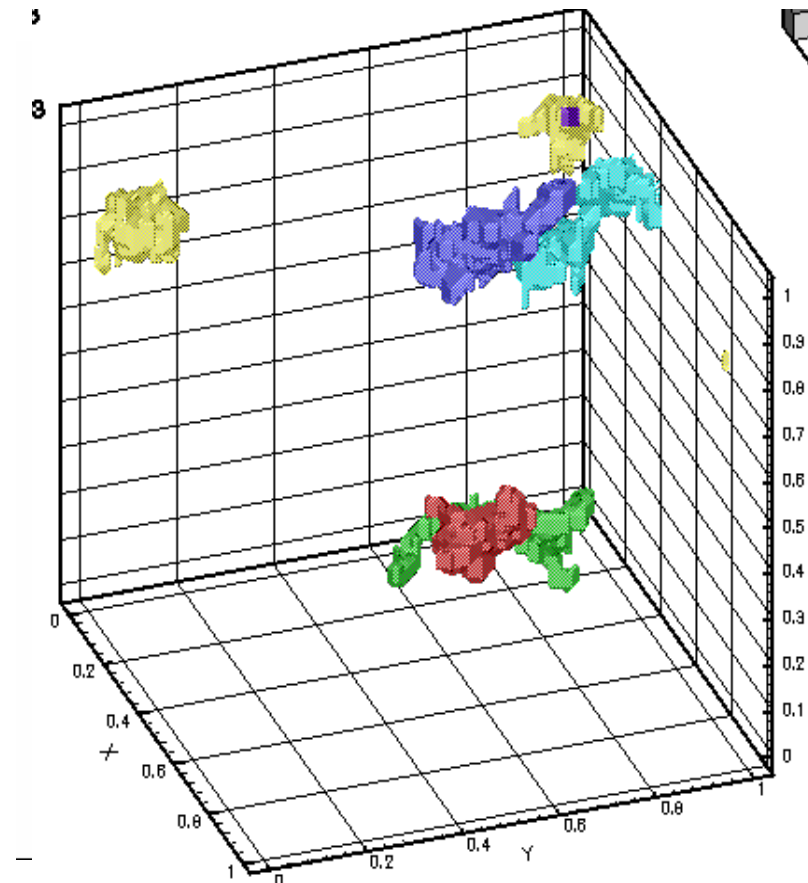
What is wrong with the model?

- This model is not scalable
 - Bulky data sets
 - Long transfer times
 - Analysis on a mainframe
- Without a precise definition of what is “interesting,” we can miss important information
- Systematic analysis of **feature populations** is impractical because it must be done by hand, one feature at a time



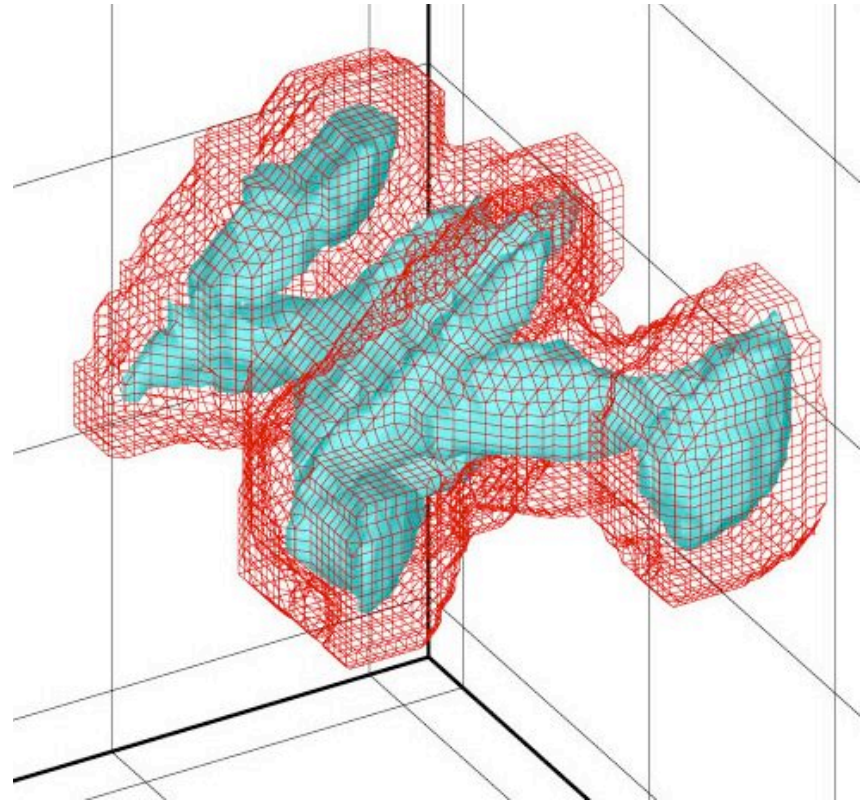
What is wrong with the model?

- This model is not scalable
 - Bulky data sets
 - Long transfer times
 - Analysis on a mainframe
- Without a precise definition of what is “interesting,” we can miss important information
- Systematic analysis of **feature populations** is impractical because it must be done by hand, one feature at a time



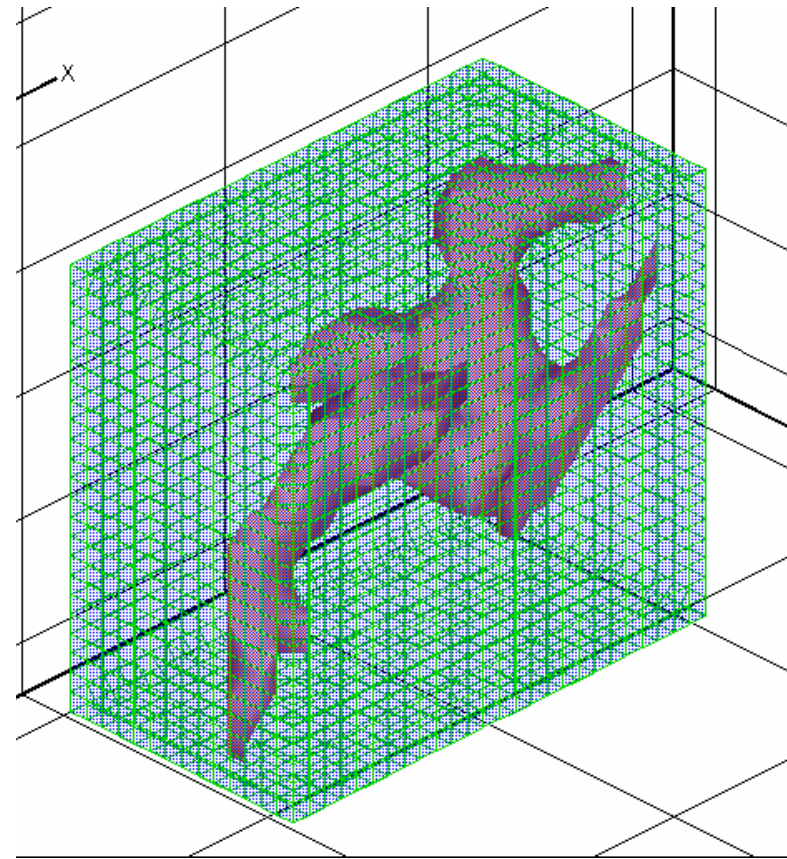
How do we extract the information?

- User function identifies features
- Run time support to apply identification systematically
- Supported by a system called MOLD [William Kerney '02]
- Diamessis, Kerney, Baden, Nomura '02]
- Volume Tracking [Silver & Wang '97] with added processing to avoid spurious tracking



Requirements

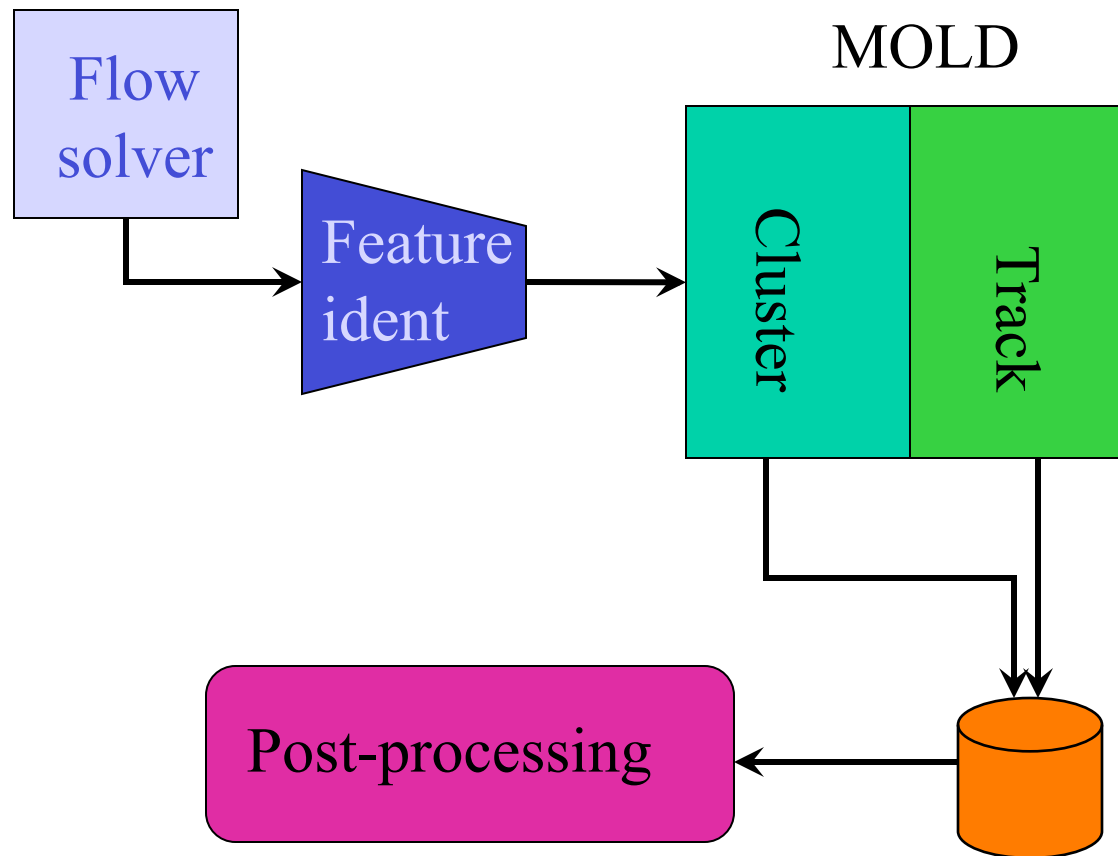
- Manage growth in storage, bandwidth, memory, etc.
- Support systematic analysis of feature populations



Feature Identification and Tracking

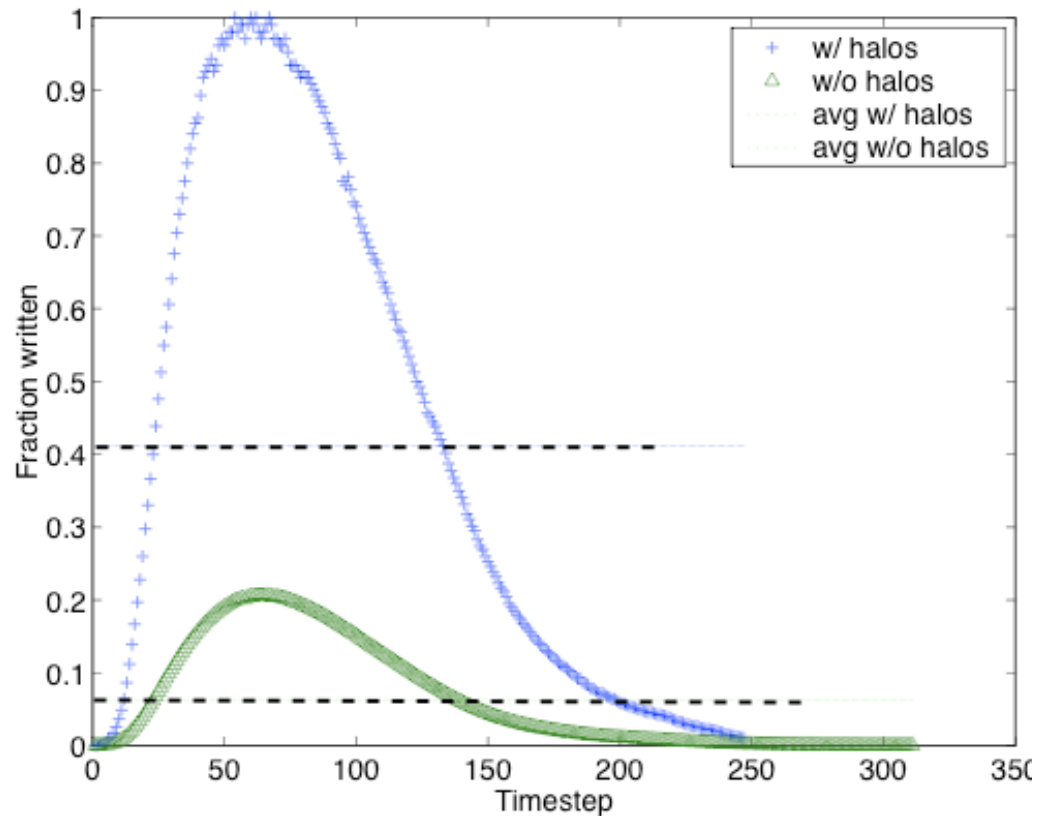
- Application identifies features (overtake regions)
- A library (MOLD) performs the analysis on-line
 - Generates a feature database
 - Stores irregular features, discarding other data
 - Compression is inversely proportional to volume fraction occupied by the features, i.e. the overtake regions

Analysis modules



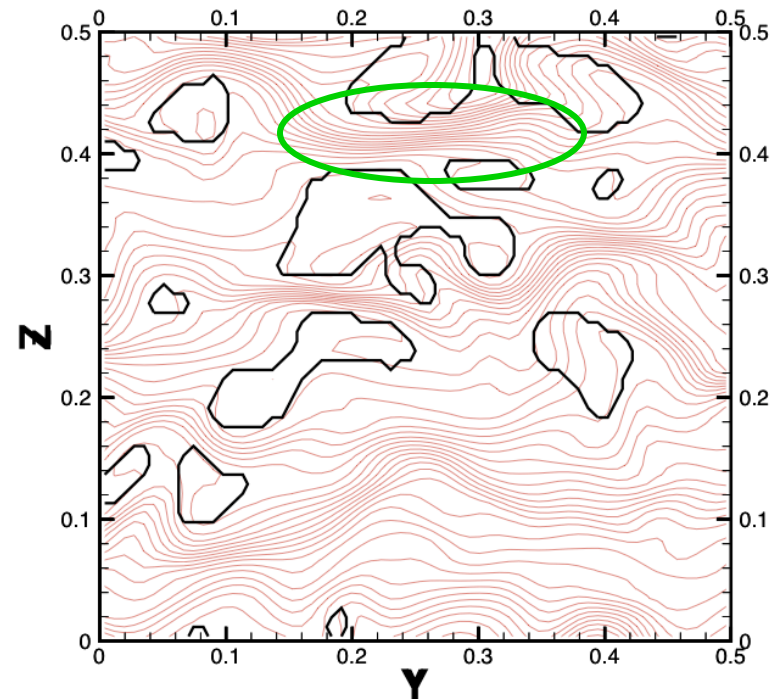
Compression via feature extraction

- Compression is inversely proportional to volume fraction occupied by the features
- Improvement over lossless compression (Lempel-Ziv, UNIX *gzip*)
 - With peripheral zones: 2.5:1
 - Without peripheral zones (animation): 15:1
- Compression ratio depends on the volume fraction of features



Results

- The identification and tracking techniques enabled us to extract new information on feature evolution in voluminous simulation output
- Overturns generate **peripheral zones** of mixing
 - Significant mixing in overturn periphery
 - Interior of overturn structures exhibit weaker mixing

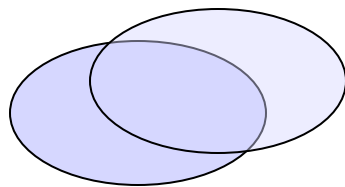


Feature Tracking

- Used to determine feature evolution and lifetime
- Using the feature database, draw correspondences between features in adjacent frames forward and backward in time
- Volume Tracking [Silver & Wang '97] with extra processing to avoid spurious continuations

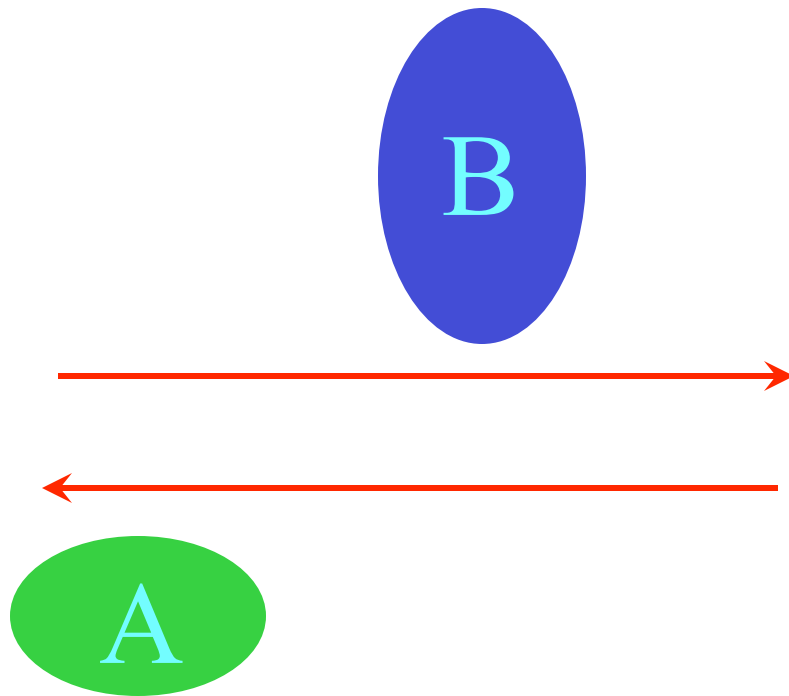
Continuations

- As features evolve, we record each continuation, which is the overlap between a feature and a descendent or an ancestor
- Creations and dissipations are not continuations



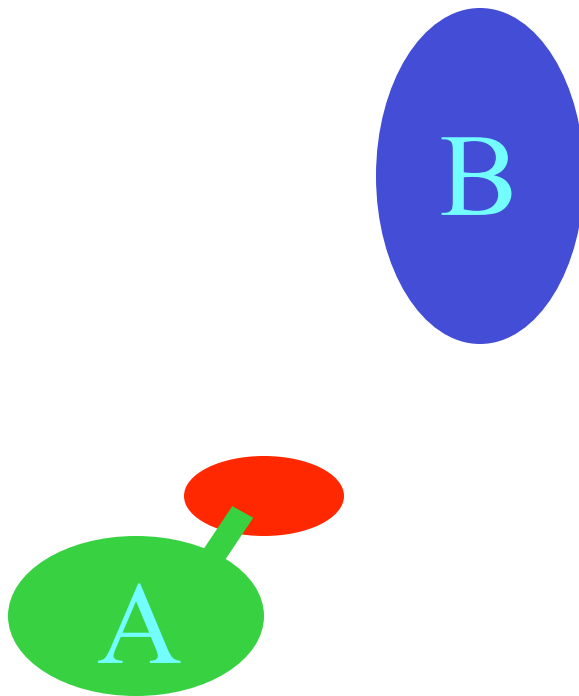
Removing spurious continuations

- Mean shear or local turbulent strain causes an overturn patch to pinch off and reconnect with another patch



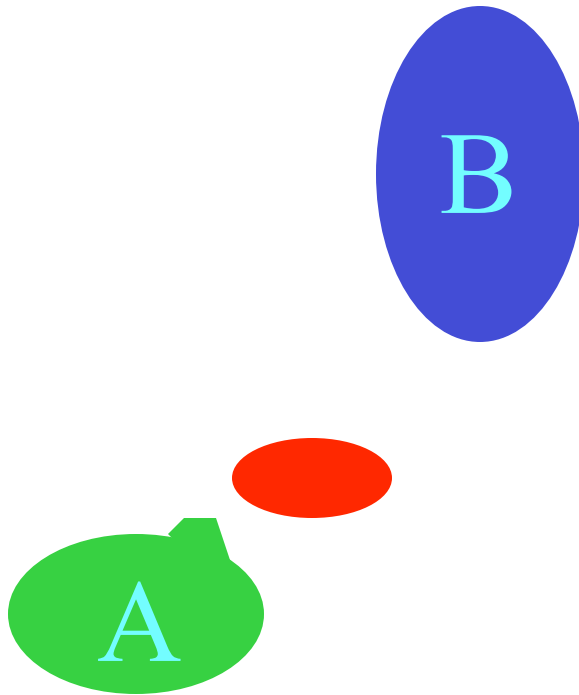
Removing spurious continuations

- Mean shear or local turbulent strain causes an overturn patch to **pinch off** and reconnect with another patch



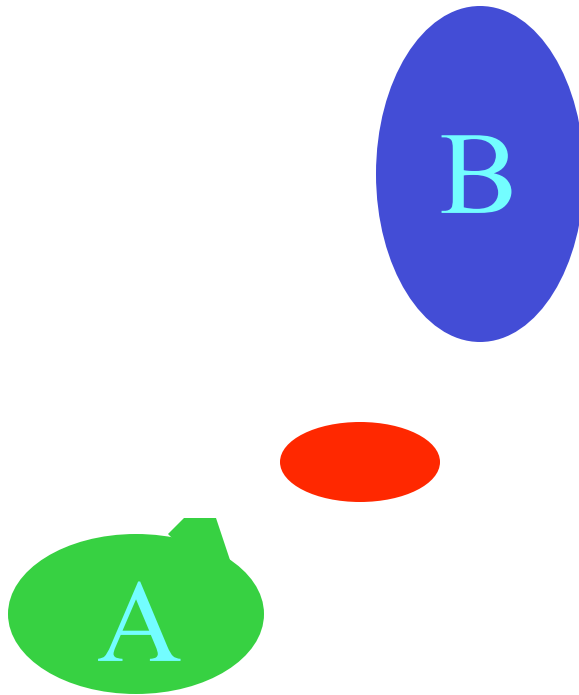
Removing spurious continuations

- Mean shear or local turbulent strain causes an overturn patch to **pinch off** and reconnect with another patch



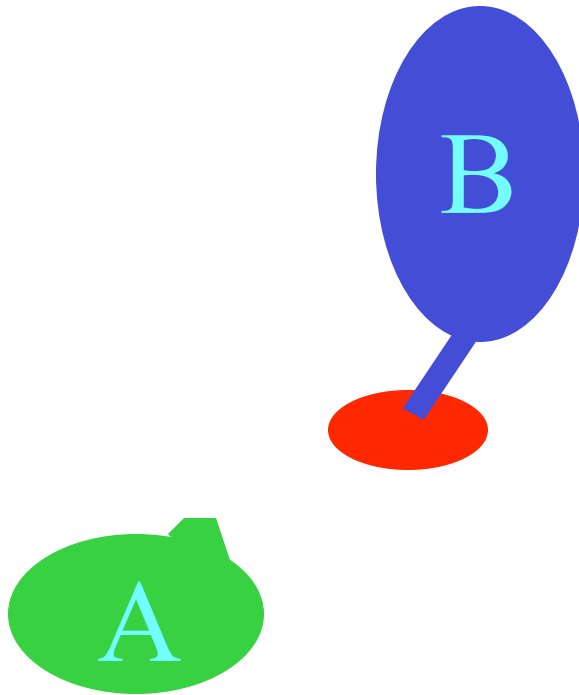
Removing spurious continuations

- Mean shear or local turbulent strain causes an overturn patch to pinch off and **reconnect** with another patch



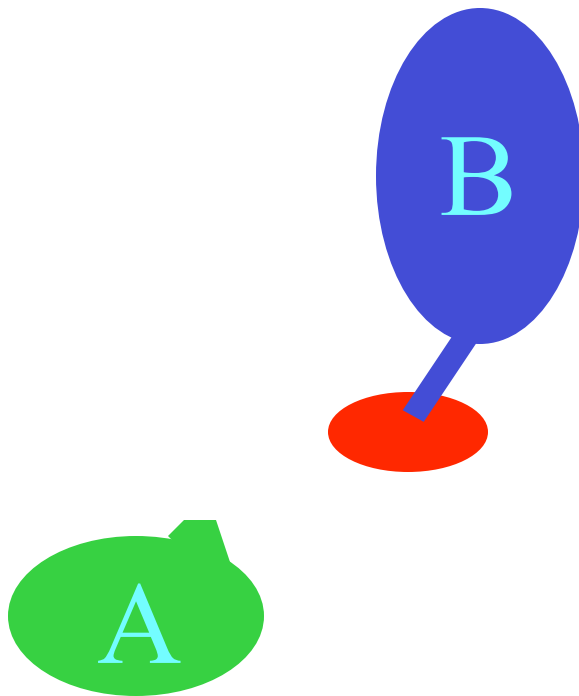
Removing spurious continuations

- Mean shear or local turbulent strain causes an overturn patch to pinch off and **reconnect** with another patch



Removing spurious continuations

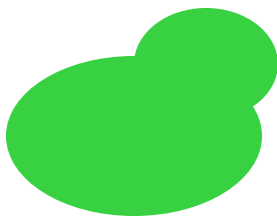
- Mean shear or local turbulent strain causes an overturn patch to pinch off and **reconnect** with another patch



B is misinterpreted as a descendent of A

Cutoff

- Ignore descendants that are above or below a certain threshold in size



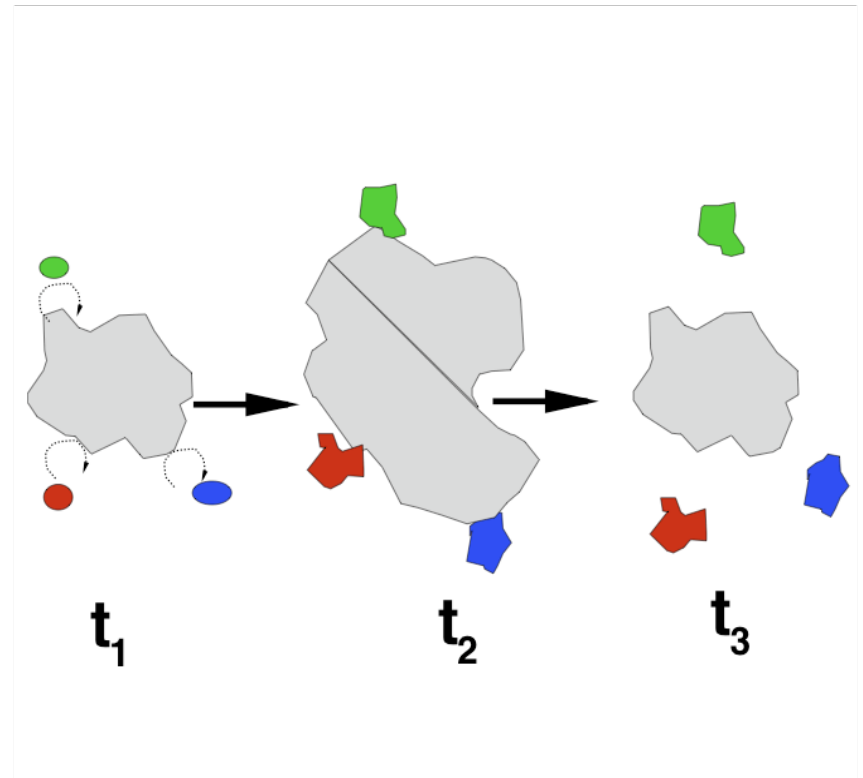
Cutoffs

- Ignore descendants that are above or below a certain threshold in size
- We require different cutoffs when tracking forward and backward in time
- Merging and splitting are different phenomenon



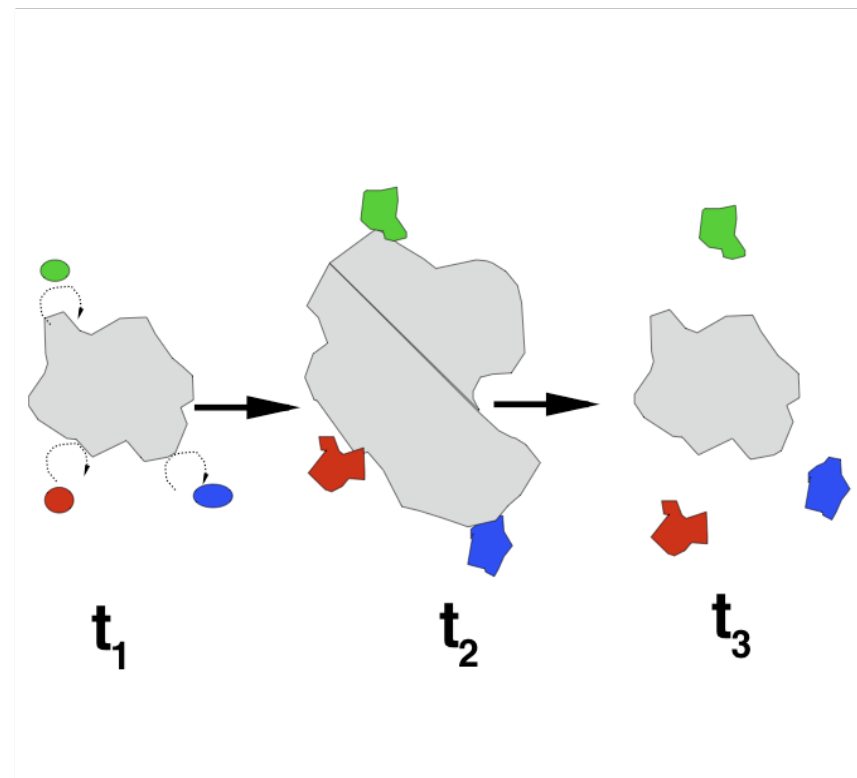
Patch age census

- Turbulent Froude number
 $Fr \equiv v / |N| = 0.5 :$
2.0 (inertial:buoyancy)
- As Fr increases, gravity is weaker, and large patches may appear:
~10% of domain volume
- Small features attach at t_2



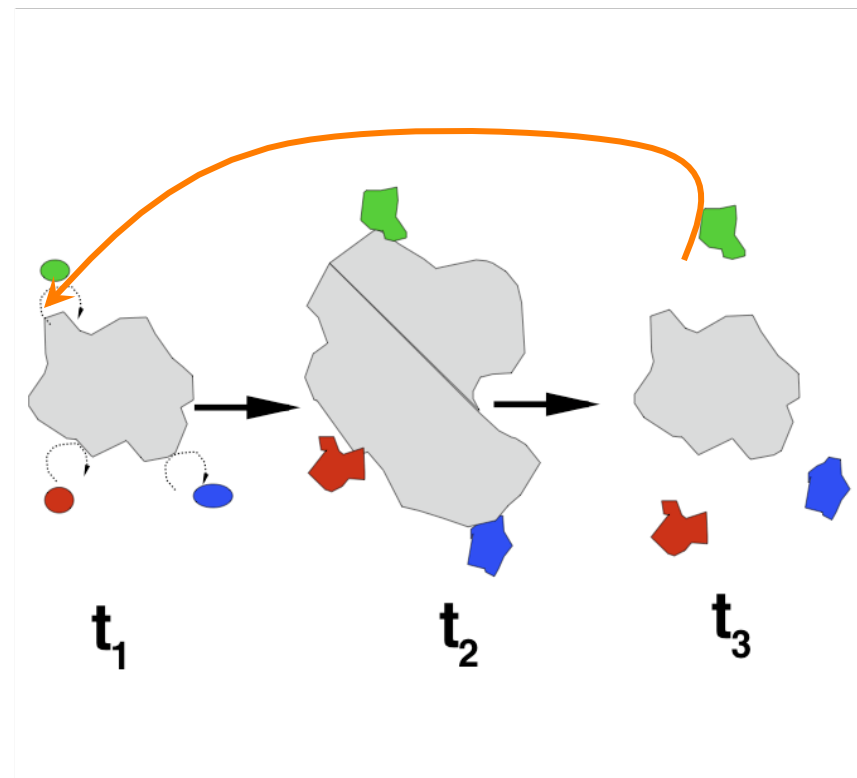
Ambiguity in determining age

- Large patches will grow, overcoming gravity
- Errors in identification process introduce spurious descendents
- Small features appear to attach at t_2 when a large overturn appears
- They appear to detach when gravitational collapse occurs at t_3



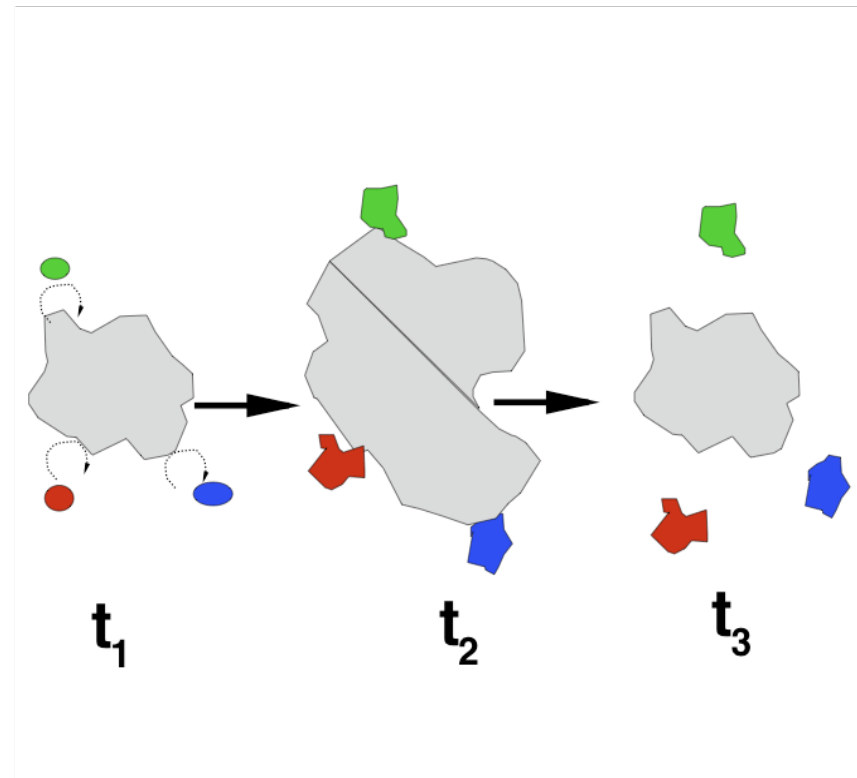
Resolving the ambiguity

- Smaller patches recorded as having same age as main patch
- Cannot correctly identify continuations prior to spurious attachment at t_2
- The ages of the separated patches are incorrect



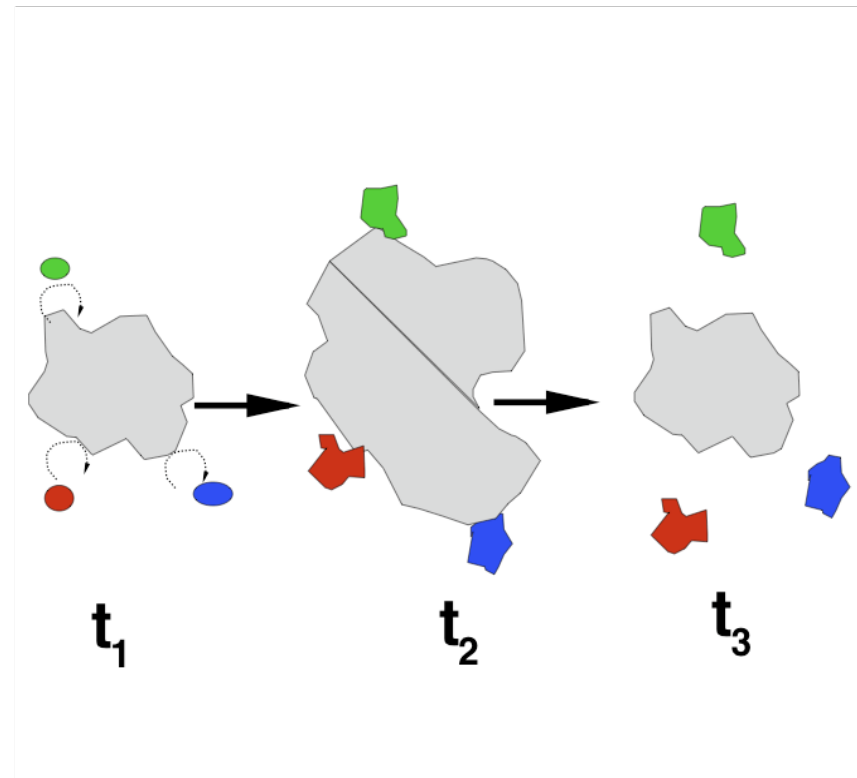
Filtering

- We start from time t_2 when the large feature reaches maximum size
- At time t_3 we retain the patch closest in size and location to the original large patch at t_2
- We need a better way to filter out noise

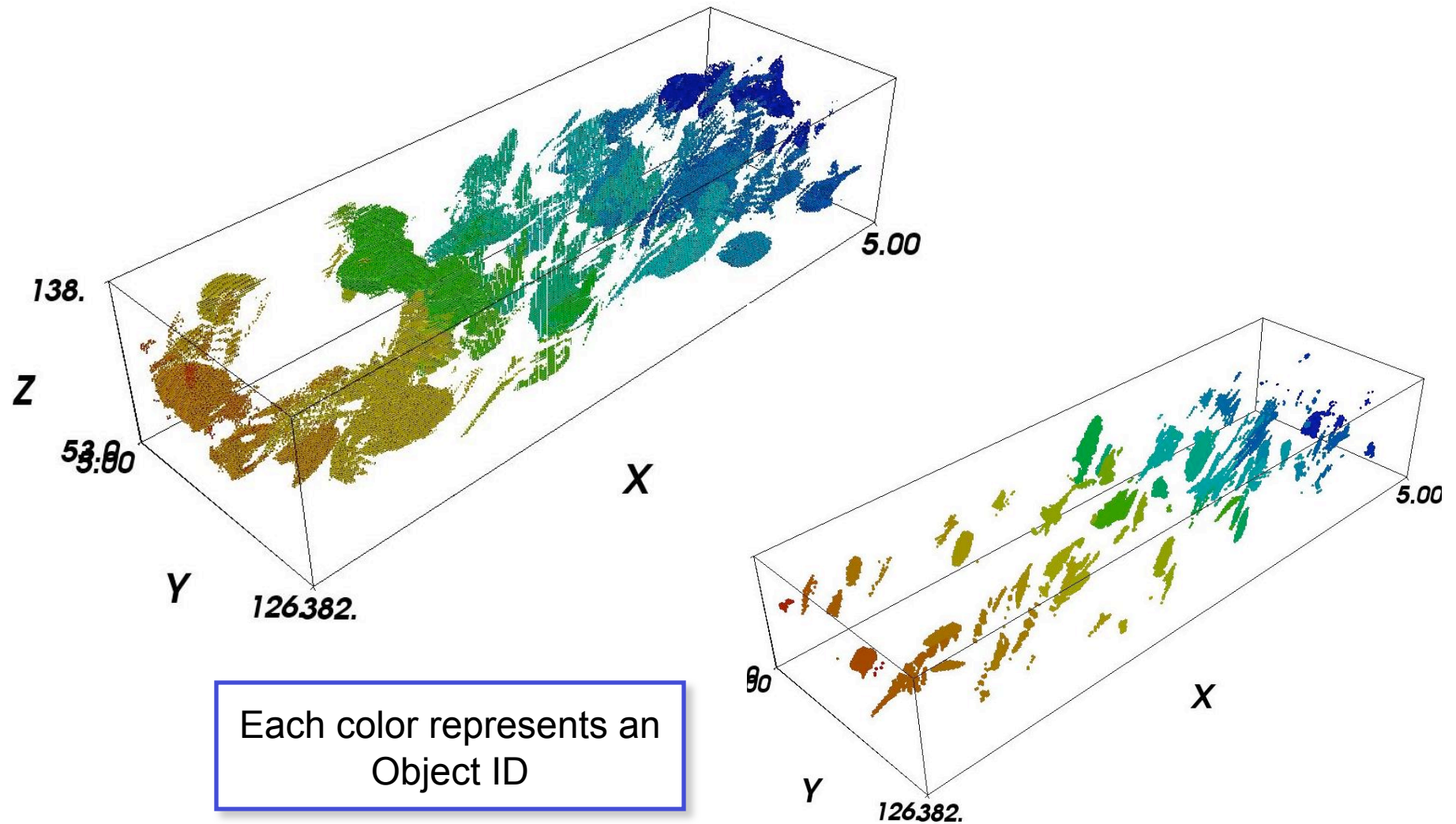


What is the appropriate analysis model?

- We have a feature database rather than a dataset
- Experiment with feature identification and tracking algorithms
- Heuristics
- Idea: let's link a spatial data base to abstractions for handling multiblock domains
- For that matter, multi-level



Component labeling (Han Kim)



Discussion

- Data sets →
Data bases of sparse feature domains
- Limitations of compression based on feature identification
 - Requires a feature identification mapping
 - What if we don't know what we are looking for?