

Detecting Topological Similarity and Dissimilarity in Multi-field Data

A PROJECT REPORT
SUBMITTED IN PARTIAL FULFIMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
Master of Engineering
IN
Computer Science and Engineering

BY
Suraj Kaushal



Computer Science and Automation
Indian Institute of Science
Bangalore – 560 012 (INDIA)

June, 2017

Declaration of Originality

I, **Suraj Kaushal**, with SR No. **04-04-00-10-41-15-1-12046** hereby declare that the material presented in the thesis titled

Detecting Topological Similarity and Dissimilarity in Multi-field Data

represents original work carried out by me in the **Department of Computer Science and Automation** at **Indian Institute of Science** during the years **2015-17**.

With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Date:

Student Signature

In my capacity as supervisor of the above-mentioned work, I certify that the above statements are true to the best of my knowledge, and I have carried out due diligence to ensure the originality of the report.

Advisor Name:

Advisor Signature

© Suraj Kaushal

June, 2017

All rights reserved

DEDICATED TO

The Student Community

Acknowledgements

I express my profound gratitude and sincere thanks to Prof. Vijay Natarajan for his valuable guidance, caring supervision and priceless feedback for the progress of this project. I am very grateful to Dr. Amit Chattopadhyay for his help in understanding the various concepts involved in this project and for his support throughout the project.

I am grateful to everyone who stood beside me during the course of my project. I would like to express my sincere gratitude to all my lab-mates for their help. I would like to thank all my classmates and colleagues for their unconditional support throughout my course.

Abstract

Here we address the problem to detect when two multi-fields are similar or dissimilar topologically. We propose a similarity measure between two multi-fields in terms of symmetric difference between their Reeb Spaces. For computation of Reeb Spaces, we use Joint Contour Net (JCN) which is a quantized approximation of the Reeb Space.

Contents

Acknowledgements	i
Abstract	ii
Contents	iii
List of Figures	iv
1 Introduction	1
2 Preliminaries	3
3 Related Work	5
4 Methods	7
4.1 Implementation	8
4.2 Dataset Description	9
5 Results	10
5.1 Height and Paraboloid Dataset	10
5.2 Plutonium Dataset	11
6 Conclusions and Future Work	21
Bibliography	22

List of Figures

2.1	Showing Level Set and Reeb Graph of the given Manifold	4
3.1	The method allows pattern matching in multi-fields such as the Hurricane Isabel data set. The user selects the eye of the hurricane at T=20 using a red box. This is the input pattern. The algorithm given in [7] uses the 3D SIFT features of 11 scalar fields simultaneously to find matching patterns in the following time steps. This amounts to a tracking of the eye of the hurricane. Image adapted from [7].	6
4.1	Showing bin count for a particular bin.	8
5.1	Histograms (count as no. of nodes) of unperturbed multi-field and height perturbed multi-field with perturbation amplitude 1.5	12
5.2	Plot of measure values against different perturbation amplitudes	12
5.3	Histograms (count as size) of unperturbed multi-field and height perturbed multi-field with perturbation amplitude 1.5	13
5.4	Plot of size measure values against different perturbation amplitudes	13
5.5	Histograms (count as no. of nodes) of unperturbed multi-field and paraboloid perturbed multi-field with amplitude 1.5	14
5.6	Plot of measure values against different perturbation amplitudes	14
5.7	Histograms of scission data-set for two consecutive time-steps	15
5.8	Plot of measure values against time-steps for resolution 16 x 2.	15
5.9	Histograms (count as size) of scission data-set for two consecutive time-steps	16
5.10	Plot of size measure values against time-steps for resolution 16 x 2.	16
5.11	Histograms of scission data-set for two consecutive time-steps	17
5.12	Plot of measure values against time-steps for resolution 8 x 2.	17
5.13	Histograms (count as size) of scission data-set for two consecutive time-steps	18

LIST OF FIGURES

5.14	Plot of size measure values against time-steps for resolution 8 x 2.	18
5.15	Joint Contour Nets for time steps 690 (top left), 692 (top right) and 693 (bottom)	19
5.16	Joint Contour Nets and their slab geometry in domain for time steps 690 (top left), 692 (top right and bottom left) and 693 (bottom right) for different selected portions (colored in blue) of corresponding JCN	20

Chapter 1

Introduction

Most of the data generated by scientific simulations and experiments include more than one scalar fields to study. The simulations of combustion, meteorology, molecular physics, fluid dynamics and other physical phenomena involve several scalar fields to study such as temperature, pressure, velocity, vorticity, kinetic energy, etc. One of the important physical phenomena of nuclear physics in which scientists are very interested is nuclear scission and to find the scission point at which a nucleus fragments into two parts is a challenging task.

Single field visualization techniques are not sufficient for showing and detecting interaction and correlations between multi-fields. The study of single scalar field representation, visualization and the relationship of one scalar field to another has been done extensively in [9] by V. Natarajan and M. Thomas and also in [10], [11] using different data structures for scalar fields and different similarity measures. The similarity between two scalar fields and self-similarity measure has been shown in [3] and [5] but relationship of one multi-field with another multi-field or similarity/dissimilarity of one multi-field with another is needed and there has not been much work done towards this.

So, we study the similarity/dissimilarity of two multi-fields and try to extract relevant features in the multi-fields and for this we need a measure which can capture the features between the multi-fields so that we can analyze the underlying phenomena between those multi-fields.

The study aims to find the similarity between two multi-fields by applying symmetric difference as a similarity measure on the JCN structure and we propose a method to find the similarity.

There has been some work on visualizing individual multi-fields and their abstractions like Multi-dimensional Reeb Graph [4] and JCN [1], [2] but because of very little work on the comparison of the multi-fields we are encouraged to work in this area.

By comparing multi-fields we will be able to find the interaction among several scalar fields and this will in turn help us visualize and understand the underlying phenomena and its relationships in a much better way.

Our main contributions in this work are:

1. We propose a similarity metric between two multi-fields in terms of symmetric difference between two JCNs.
2. We apply our method for automatic detection of nuclear scission point in a time-varying multi-field data of proton and neutron densities of plutonium atom.

Outline : In the next section we discuss some background needed for further discussion. In section 3, some of the previously done related work is discussed. Section 4 discusses our method of implementation and the data-sets used. In section 5, some of the experimental results and observations are shown and section 6 concludes the work with some future work.

Chapter 2

Preliminaries

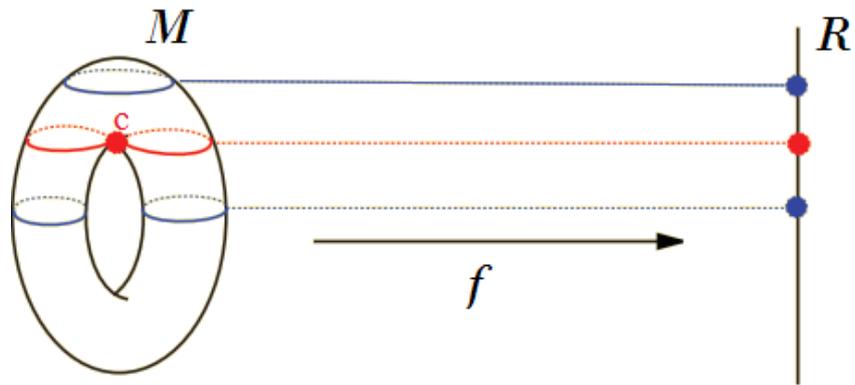
The topological properties of a scalar field refer to the topology of the level sets. Consider $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ to be a scalar field. For $c \in \mathbb{R}$, the set $f^{-1}(c) = \{x \in \mathbb{R}^3 | f(x) = c\}$ is called a *level set*. We can track the topological feature of the field by tracking the level sets, see Figure 2.1 (a).

Let f be a scalar field, the graph obtained by contracting each connected component of the level set to a point is called a *Reeb Graph*, see Figure 2.1 (b).

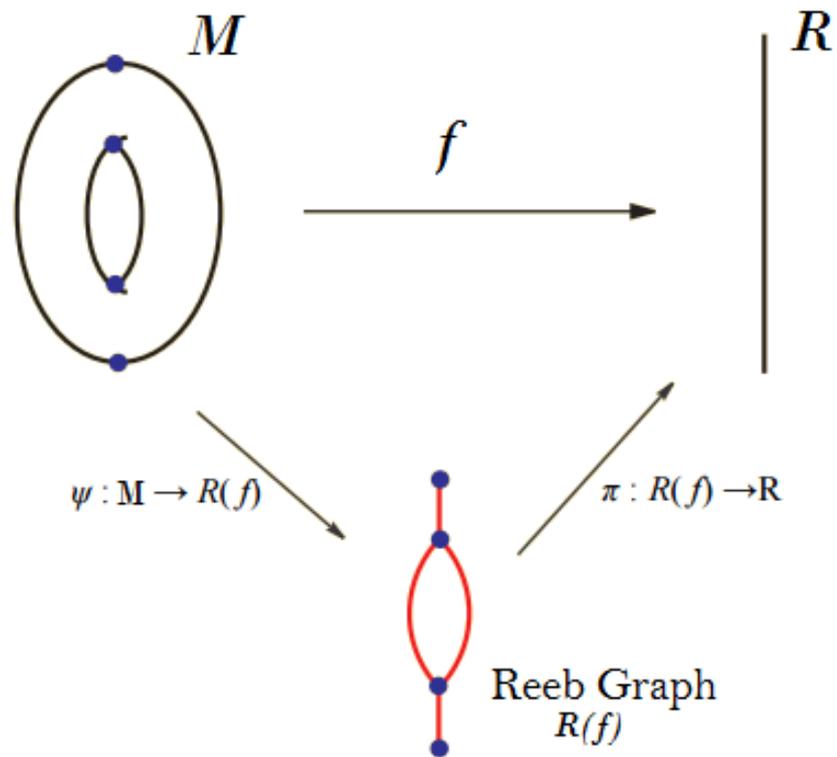
The differentiable map $f : \mathbb{R}^3 \rightarrow \mathbb{R}^m$ for $m \geq 1$ where $f(x) = (f_1(x), f_2(x), \dots, f_m(x))$ is called a *multi-field*. For $c \in \mathbb{R}^m$, the set $f^{-1}(c) = \{x \in \mathbb{R}^3 | f(x) = c\}$ is called a *fiber* (level set in case of scalar field). In a multivariate data fibers are the intersection of the *level sets* of each component of f .

Similar to Reeb graph of a scalar field, *Reeb Space* captures the fiber topology of a multi-field. Each connected component of a fiber is called a *joint contour* and corresponds to a point in the Reeb space [8].

Hamish Carr and David Duke in [1] present a quantized abstraction of the multi-field *Joint Contour Net*. For computing the JCN it takes multiple scalar fields as input, combines it to make a joint function, then in phase 1 of the JCN construction, it fragments each fields to form *Joint Contour Fragments*. In phase 2, it constructs a *Dual Graph* called *Joint Contour Graph* by keeping a node for each joint fragment called a *slab* and a node is connected to another node if they are adjacent to each other in the *Joint Contour Fragments*. In phase 3 and 4 it creates the JCN by removing the nodes having same function values in Joint Contour Graph.



(a) Critical level sets where topological changes occur.



(b) Obtaining Reeb Graph from the given Manifold

Figure 2.1: Showing Level Set and Reeb Graph of the given Manifold

Chapter 3

Related Work

There has been some extensive work done on finding similarity between scalar fields. Himangshu Saikia et al. [3], present a method to find repeating topological structures in the scalar data sets. It provides similarity between two scalar fields and self-similarity. It also introduces a new data structure Extended Branch Decomposition Graph (EBDG). It compares all the subtrees of two merge trees in an efficient manner exploiting the redundancy. The method to compute and compare EBDGs is based on Dynamic Programming. It compares two EBDGs by computing the cost level by level using bottom-up memoization. This method is stable to small perturbations and is fast and memory-efficient.

Another similarity measure provided in [5] is the Interleaving Distance between two merge trees which shows the measure of similarity between them. Interleaving Distance between two merge trees is the greatest lower bound on “epsilon” for which there are “epsilon-compatible maps” from one tree to other. Here Interleaving Distance has been defined and proof of stability against perturbations is given. Also, Bottleneck Distance is defined in [5] and its comparison to Interleaving Distance against noise is also shown. It is also shown that Interleaving Distance is never smaller than Bottleneck Distance between persistence diagrams. Bottleneck distance between two multi-sets of points X and Y is defined as:

$$d_B(X, Y) = \inf_{\gamma} \sup_x \|x - \gamma(x)\|_{\infty}$$

which is the length of the longest edge in the best matching, where γ goes over all possible bijections between X and Y and $\|x - \gamma(x)\|_{\infty} = \max\{|b_x - b_y|, |d_x - d_y|\}$, if $x = (b_x, b_y)$ and $y = (d_x, d_y)$. A significant property of bottleneck distance is that it is stable against the perturbations.

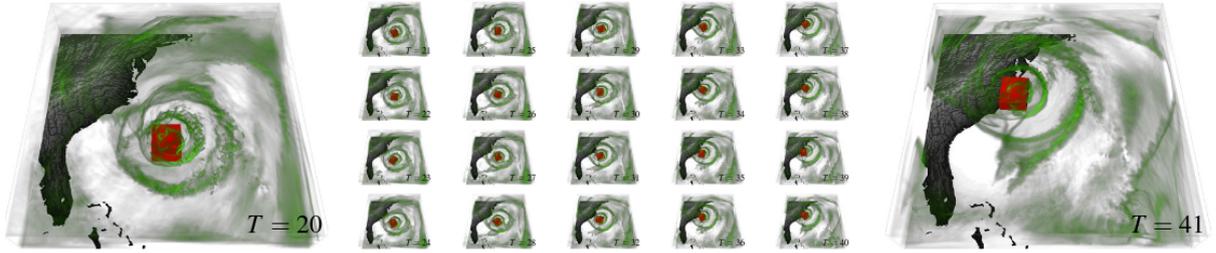


Figure 3.1: The method allows pattern matching in multi-fields such as the Hurricane Isabel data set. The user selects the eye of the hurricane at $T=20$ using a red box. This is the input pattern. The algorithm given in [7] uses the 3D SIFT features of 11 scalar fields simultaneously to find matching patterns in the following time steps. This amounts to a tracking of the eye of the hurricane. Image adapted from [7].

In Kenes Beketayev et al.[6], to find the topological similarity between two merge trees, they consider all branch decompositions of both trees and try to find a pair that minimizes the matching cost between them, where matching cost between two vertices is the maximum of absolute difference of their corresponding coordinates of the vertices. The branch decomposition B of a reduced merge tree (a merge tree with purged regular vertices) T is a pairing of all *minima* and *saddles* such that for each pair there exist at least one descending path from the saddle to the minimum. If T_f and T_g are two merge trees and $B_{T_f} = \{R_1^f, \dots, R_k^f\}$ and $B_{T_g} = \{R_1^g, \dots, R_k^g\}$ are all possible branch decompositions. Then the distance between merge trees T_f and T_g is defined as:

$$d_M(T_f, T_g) = \min(\varepsilon_{\min}(R_i^f, \dots, R_j^g))$$

where, $R_i^f \in B_{T_f}$, $R_j^g \in B_{T_g}$ and $\varepsilon_{\min}(R_i^f, \dots, R_j^g)$ is the smallest ε for which the overall cost does not exceed the value ε .

A closely related work as ours has been presented by Zhongjie Wang et al. in [7]. This is an approach to pattern matching in 3D multi-field data. This method does the joint analysis by bundling information from multiple fields into the “*description*” of a pattern. It extracts a sparse set of features for each 3D scalar fields using 3D-SIFT (Scale Invariant Feature transform) algorithm. Then the user defines a pattern as a set of SIFT features in multiple fields by “*brushing a region of interest*”. It locates and rank the matching patterns in the entire data set (see Figure 3.1).

Chapter 4

Methods

We define our measure in such a way that takes the symmetric difference of two JCNs and which can capture the changes or differences between two multi-fields. We take the JCN field values and combine those field values which have same range to form a histogram, where each bin in the histogram has a count of the no. of nodes in the JCN having field values in range same as a bin range (see Figure 4.1).

Consider multi-fields $f = (f_1, f_2)$ and $g = (g_1, g_2)$ and JCN corresponding to the multi-field f , which we name as JCN_1 .

The field f_1 ranges from $(f_{min_1}$ to $f_{max_1})$ and field f_2 ranges from $(f_{min_2}$ to $f_{max_2})$. We construct a 2D histogram with these 4 boundary points.

Each bin in the histogram has a count which is no. of nodes in the JCN_1 falling in that range of the bin. Similarly, we have another multi-field and its JCN, call it JCN_2 . We compute its 2D histogram as well. For example, in figure 4.1, we can see the histogram created with Paraboloid and Height fields, where Paraboloid and Height field values in the JCN ranges from -4.8615 to 54.1385 and from -6.49508 to 5.50492 respectively and the count range is shown in the colorbar.

In figure 4.1, we can see count value for the bin in the range 23 to 27 (x-axis) and -1 to 0 (y-axis) is 82.

We apply our similarity measure for each bin of two histograms which represents the no. of fiber components and we take difference of count values of all bins of a histogram with another histogram count values, those differ in no. of fiber components will survive and those which are equal in number will not contribute to the difference and then we sum up the overall difference to find how similar/dissimilar they are. The overall measure will be the sum of the absolute difference between the count values of two histograms divided by the total number of bins.

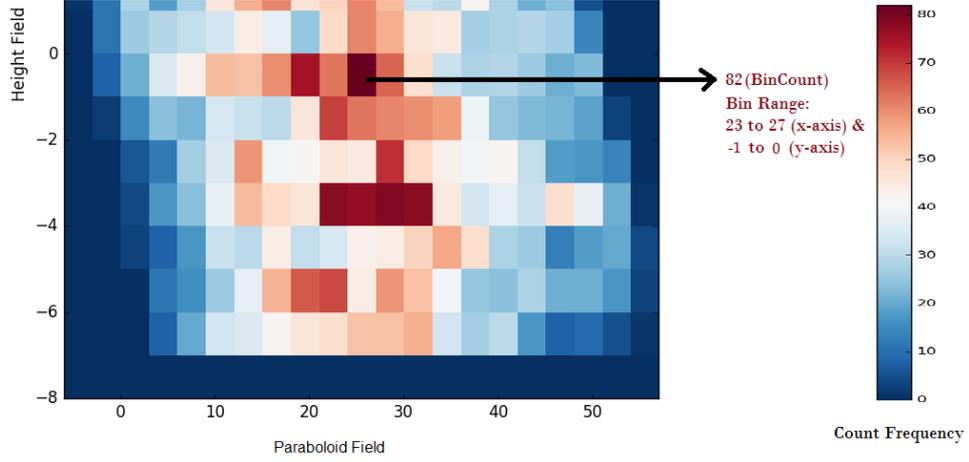


Figure 4.1: Showing bin count for a particular bin.

Mathematically, our measure can be defined as:

$$Measure = \frac{\sum_{i=1}^B |BinCount_i^1 - BinCount_i^2|}{B}$$

where, $BinCount^1$ and $BinCount^2$ are number of points falling in a bin of JCN_1 and JCN_2 respectively and B is the total number of bins in the histograms.

4.1 Implementation

1. First we take two multi-field data-sets and compute JCN for both multi-fields using the code given in [12].
2. We plot histograms for both JCNs by choosing the bin width by subdividing the range of the field values appropriately by experiment.
3. Then we take the *BinCounts* of each histograms and compute our measure as defined earlier.
4. We plot a graph of measures computed against different multi-fields to analyze the changes or relationship between them.

4.2 Dataset Description

We have used two multi-field data-sets to perform our experiments :

- First multi-field is a simulation data-set consists of Height and Paraboloid fields with domain $20 \times 20 \times 20$.
- Second multi-field is a real time-varying scission data-set consists of proton density and neutron density of plutonium atom as two fields having domain $40 \times 40 \times 66$. Scission data-set is available at 15 different time-steps i.e; [665, 670, 675, 680, 685, 686, 687, 688, 689, 690, 692, 693, 694, 695, 699].

Chapter 5

Results

5.1 Height and Paraboloid Dataset

First we take Height and Paraboloid data-set and perturb Height field at six different amplitudes i.e; at 0.25, 0.5, 0.75, 1.0, 1.25 and 1.5. For perturbing the Height field at amplitude a we add a uniform random value between $-a$ to $+a$ in the Height field values while keeping Paraboloid field unperturbed to generate another multi-field, then we compute JCN for this, plot histograms (Figure 5.1) and compute measure between both multi-fields. Colormap frequency in the histograms are set to be the maximum count among two histograms for the ease of comparing both visually. Similarly we compute the measure for several other perturbation amplitudes as well. We also plot a graph of measures against different perturbations (Figure 5.2).

We also plot histograms for the size of the JCN nodes which represents number of fragments (no. of triangles/tetrahedra) of a node (Figure 5.3). Here each bin count of a histogram contains the sum of sizes of all the points falling in that range. We plot the graph of measures against several perturbations and we found a somewhat similar behaviour compared to the previous plot (Figure 5.4).

Now we perturbed the other field i.e; Paraboloid field and keeping the Height field unperturbed and draw histograms (Figure 5.5) then plot the distance measures against different perturbations (Figure 5.6) and we found similar behaviour in this case as well.

5.2 Plutonium Dataset

Now we take the scission data-set for 15 time-steps and compute the distance measures between two histograms corresponding to two consecutive time steps (Figure 5.7). The measure is computed for (665, 670), (670, 675), (675, 680)...(695, 699) and plotted against the time steps (Figure 5.8). The plot is a line graph; for example, for the measure between time-steps 665 and 670 we have a horizontal line from time steps 665 to 670 (x-axis) at its corresponding measure value (y-axis). The highest measure value is marked as red horizontal line.

We have also computed the weighted (count as size) measure for consecutive time steps and similarly plotted histograms (Figure 5.9) and graph of size distance measures against time-steps (Figure 5.10). The measures are computed for slab width 16 for proton density and slab width 2 for neutron density with histogram's bin widths 16 and 2 for proton and neutron respectively.

We have computed the measure for different resolutions as well. When we computed the measures by taking slab width 8 and 2 for proton and neutron respectively, we found similar behaviour in the plot of measures vs time-steps (see Figures 5.11, 5.12, 5.13 and 5.14).

In our experiments, we notice several interesting points which was expected from our measure. First of all, we expect our measure to increase as we increase the perturbation value and we got the similar behaviour when we computed the measure for Paraboloid and Height multi-field. We can see in figures 5.2, 5.4 and 5.6 that the measure also gradually increases as we increase the perturbation amplitudes.

Secondly, we wanted our measure to capture an important topological phenomena happening in multi-field comparison. In figures 5.8 and 5.10, we notice that there is a huge jump (shown in red) in the measure value between time-step 690 and 692 and actually scission occurs at time-step 692. So, our measure is able to capture this important phenomena i.e; scission point in the given data-set.

We can also see the JCNs at time steps 690, 692 and 693 in the figure 5.15. We can easily observe the split (at tail) in the nodes of the JCN at time step 692 in comparison to JCN at time step 690.

The slab geometries in their domain corresponding to a selected portion (colored in blue) of a JCN is also shown in figures 5.16.

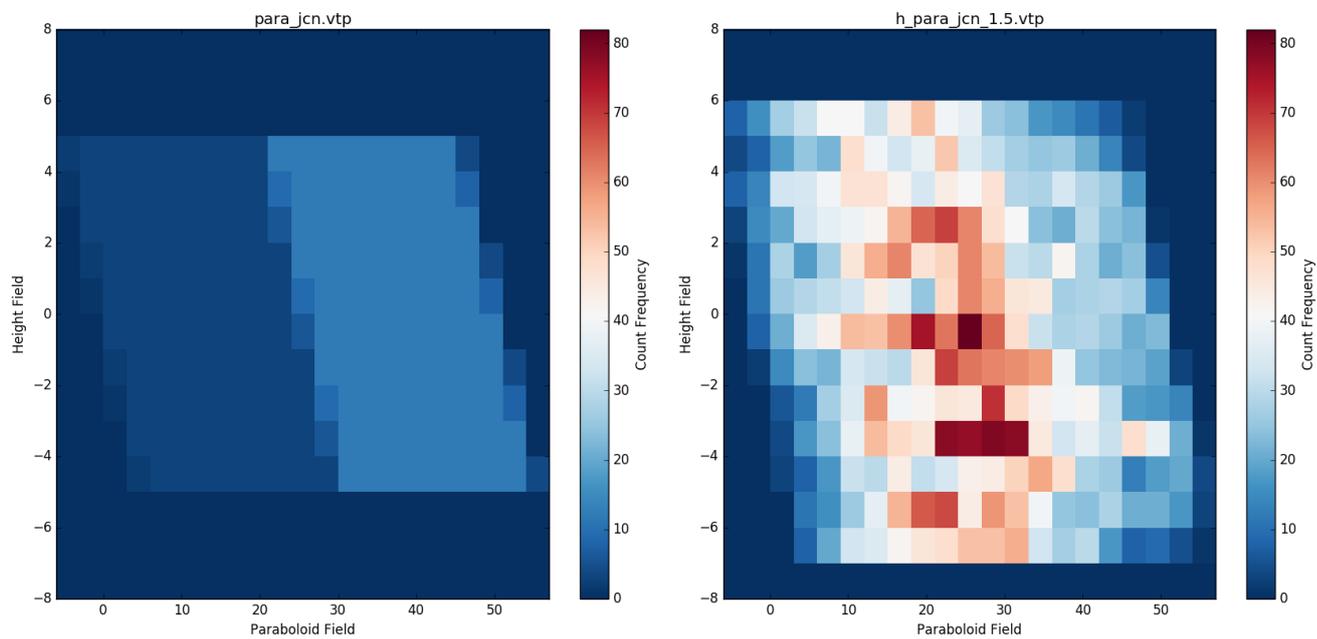


Figure 5.1: Histograms (count as no. of nodes) of unperturbed multi-field and height perturbed multi-field with perturbation amplitude 1.5

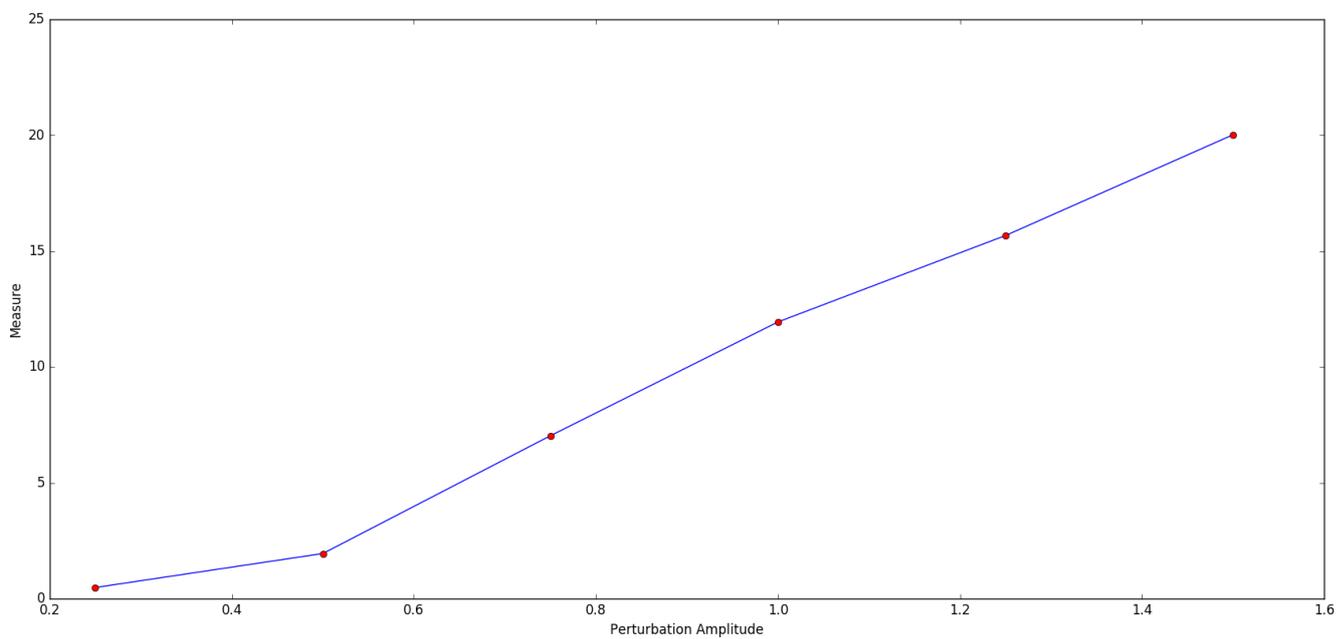


Figure 5.2: Plot of measure values against different perturbation amplitudes

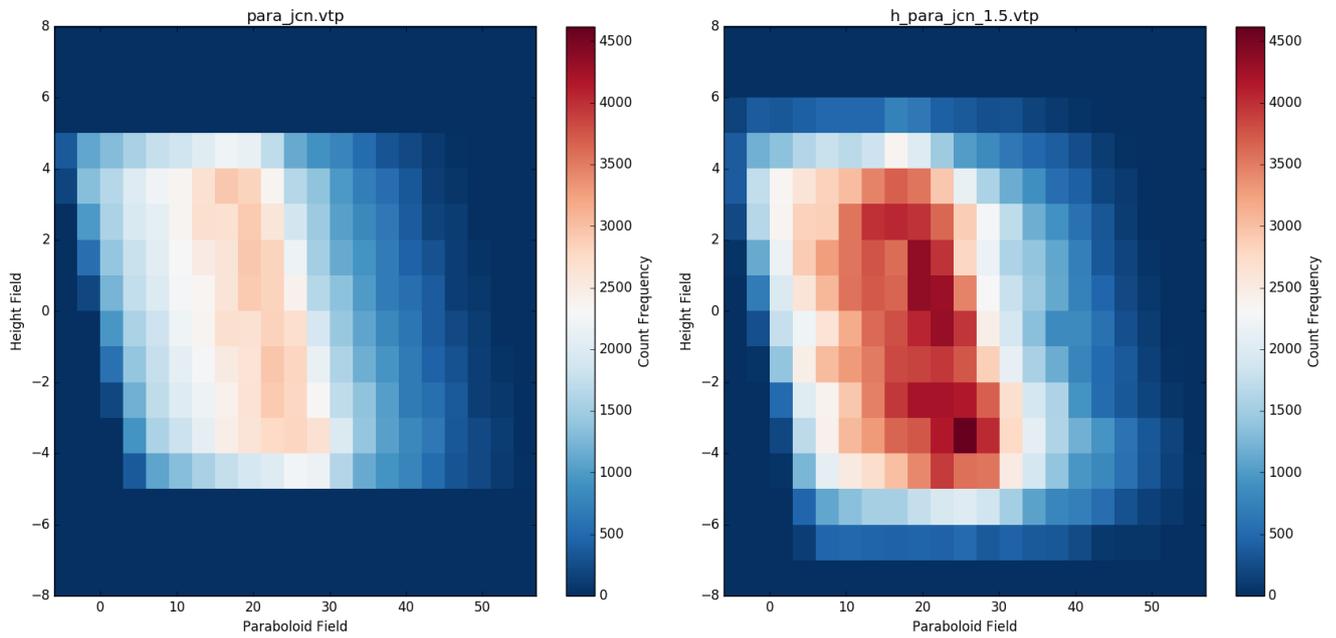


Figure 5.3: Histograms (count as size) of unperturbed multi-field and height perturbed multi-field with perturbation amplitude 1.5

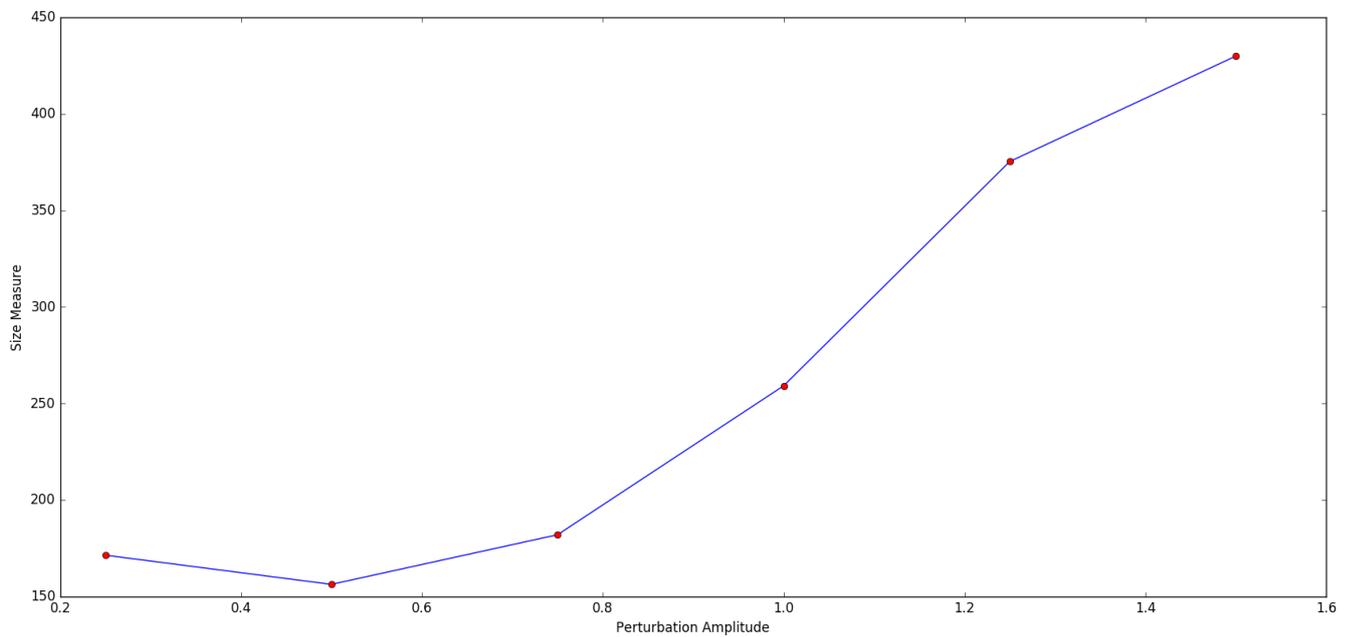


Figure 5.4: Plot of size measure values against different perturbation amplitudes

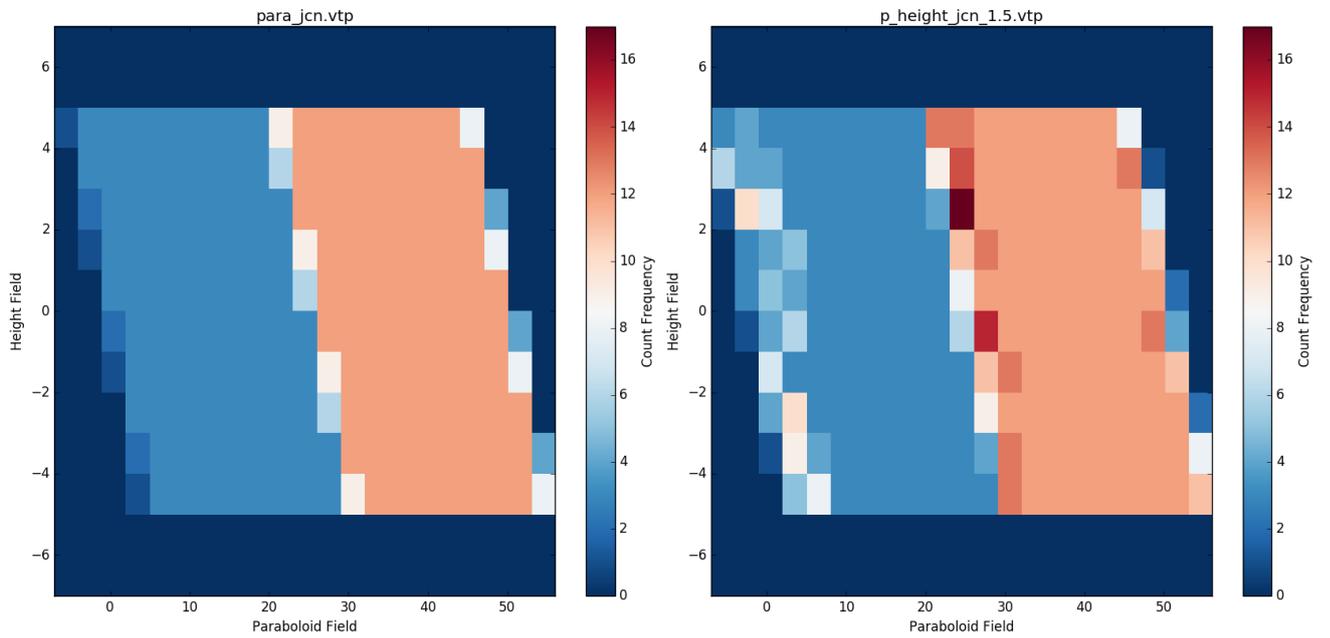


Figure 5.5: Histograms (count as no. of nodes) of unperturbed multi-field and paraboloid perturbed multi-field with amplitude 1.5

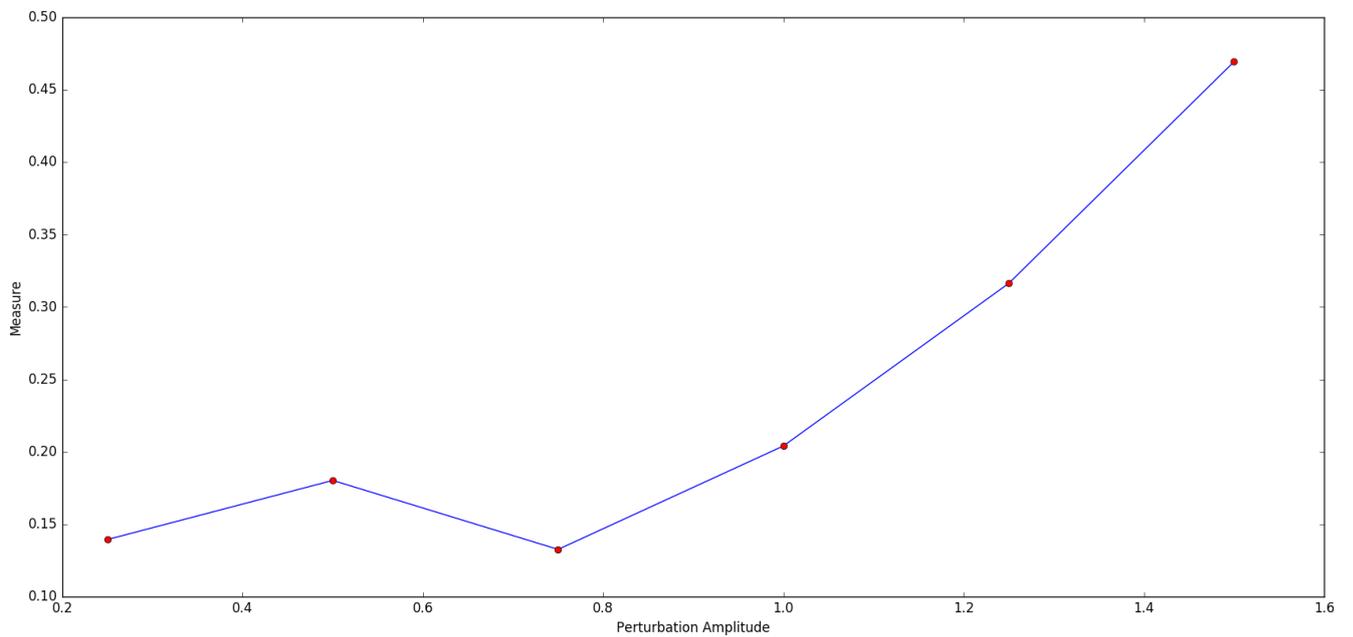


Figure 5.6: Plot of measure values against different perturbation amplitudes

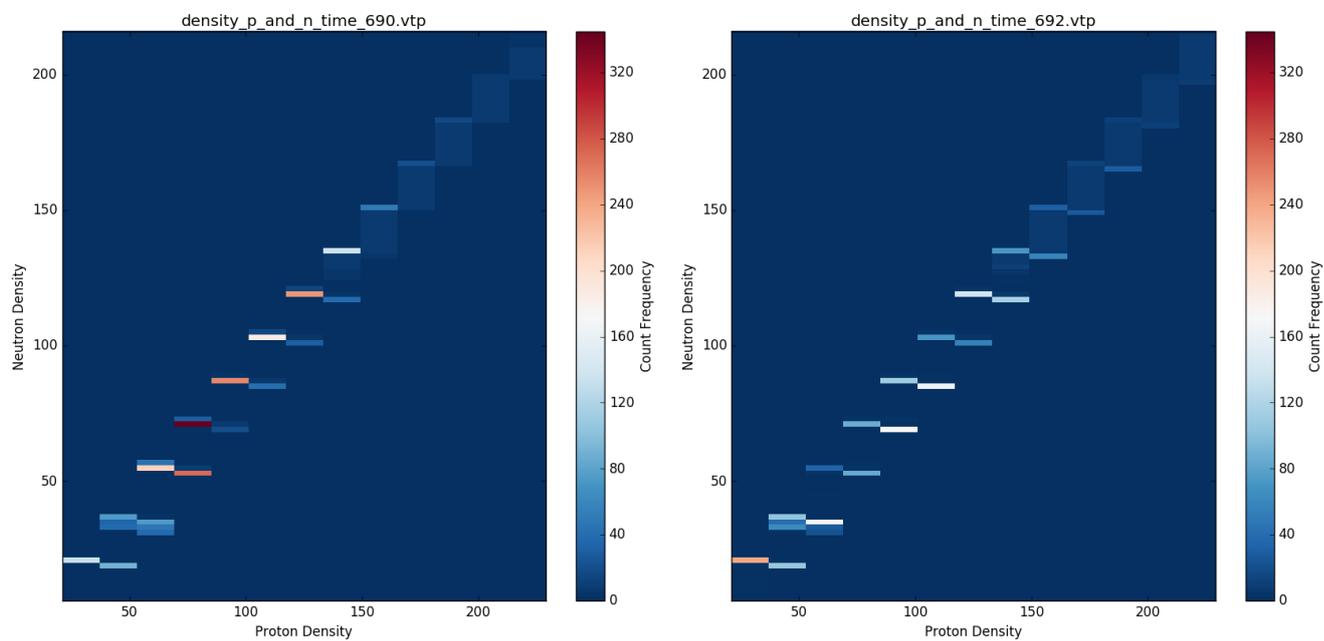


Figure 5.7: Histograms of scission data-set for two consecutive time-steps

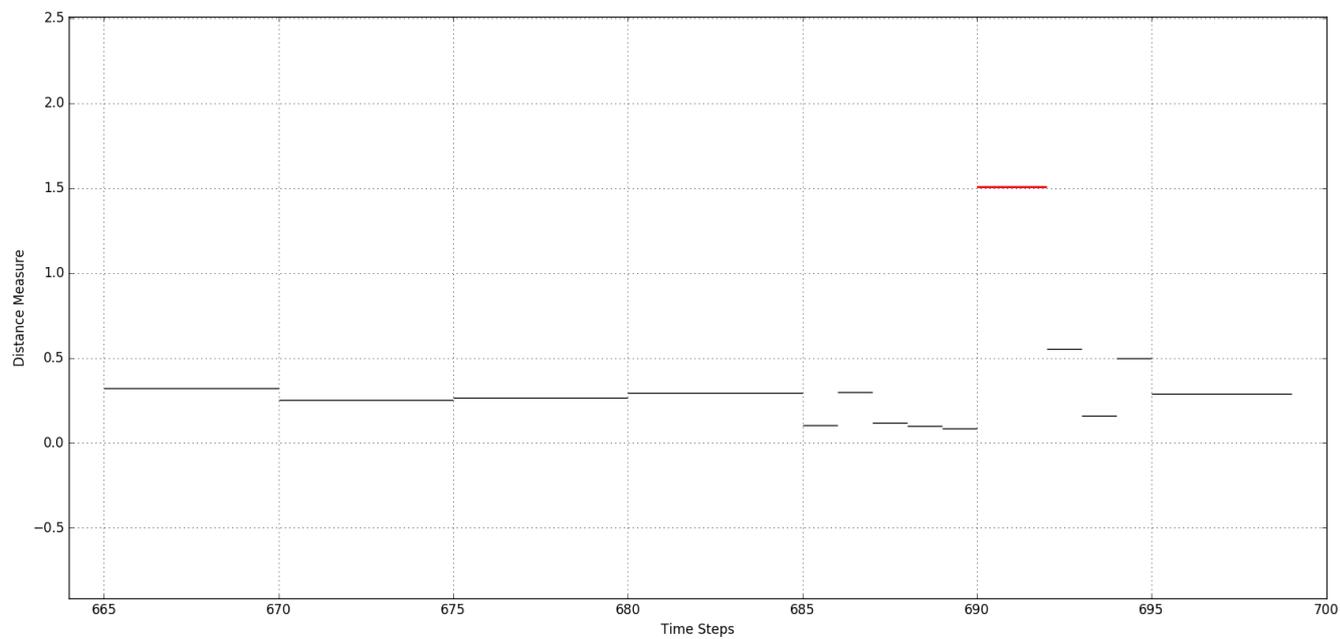


Figure 5.8: Plot of measure values against time-steps for resolution 16 x 2.

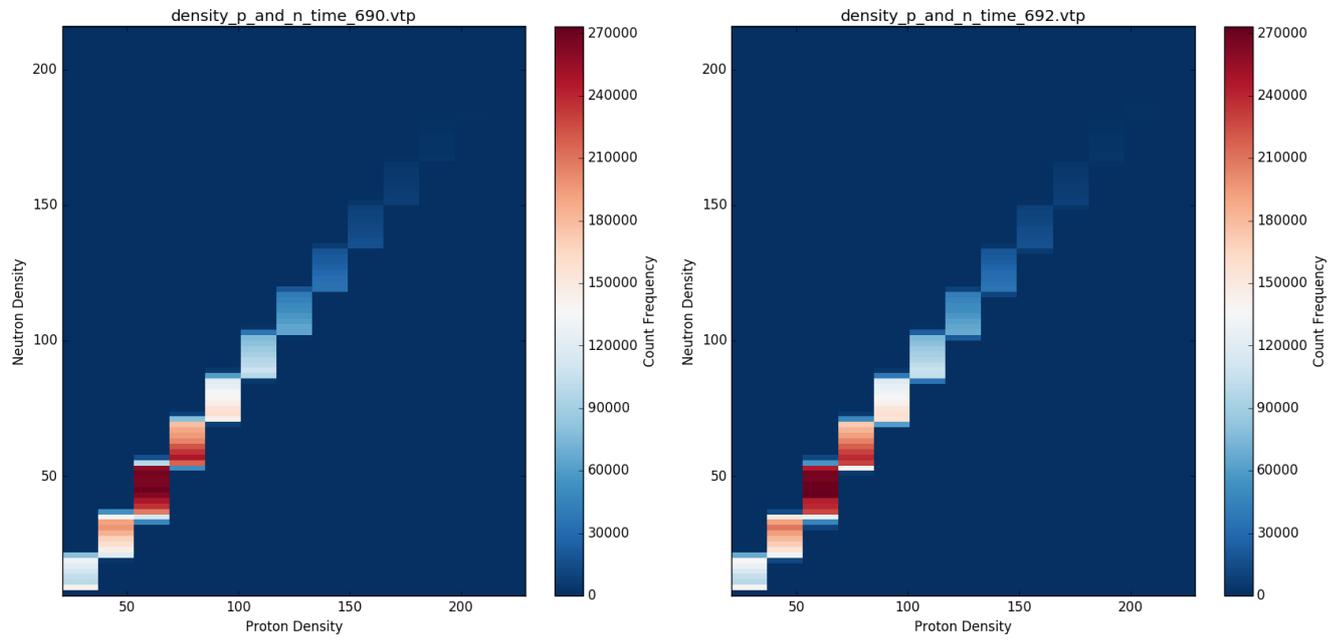


Figure 5.9: Histograms (count as size) of scission data-set for two consecutive time-steps

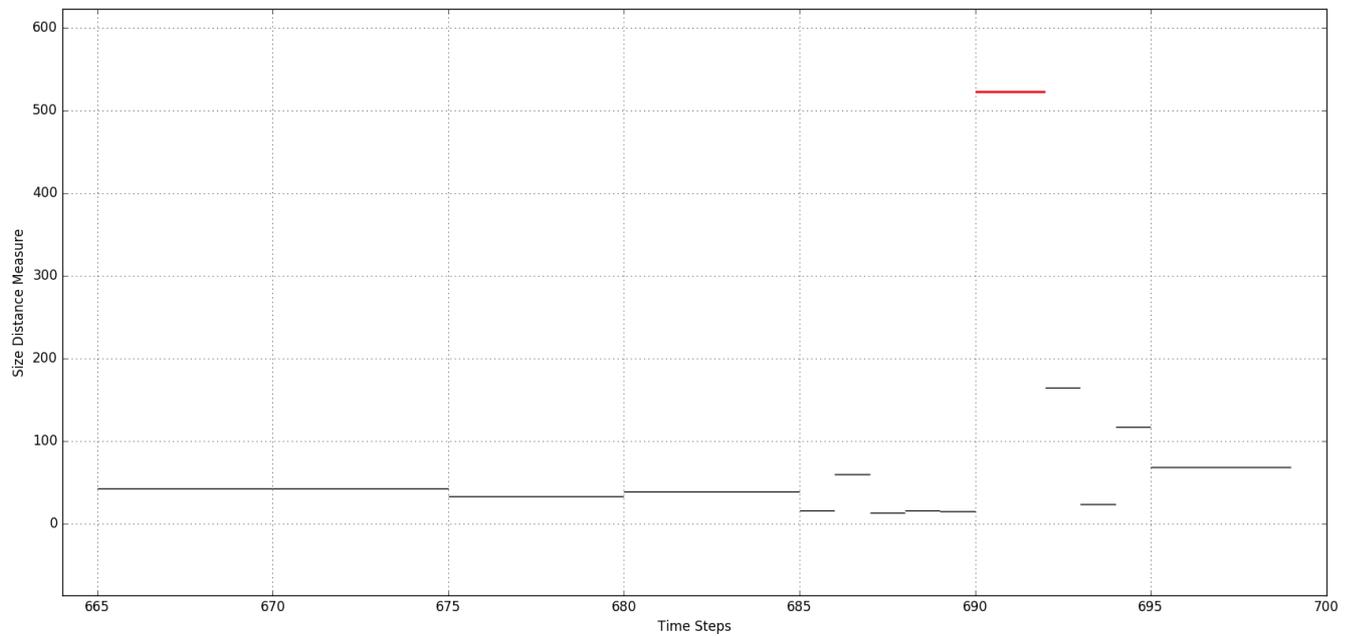


Figure 5.10: Plot of size measure values against time-steps for resolution 16 x 2.

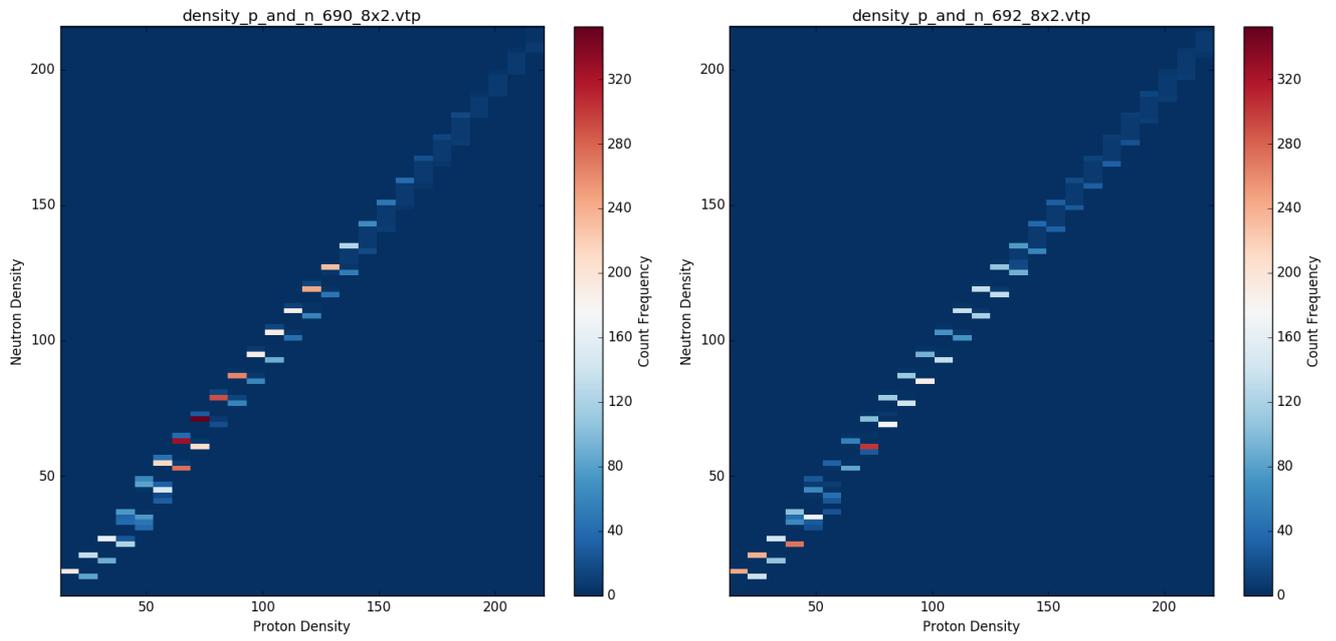


Figure 5.11: Histograms of scission data-set for two consecutive time-steps

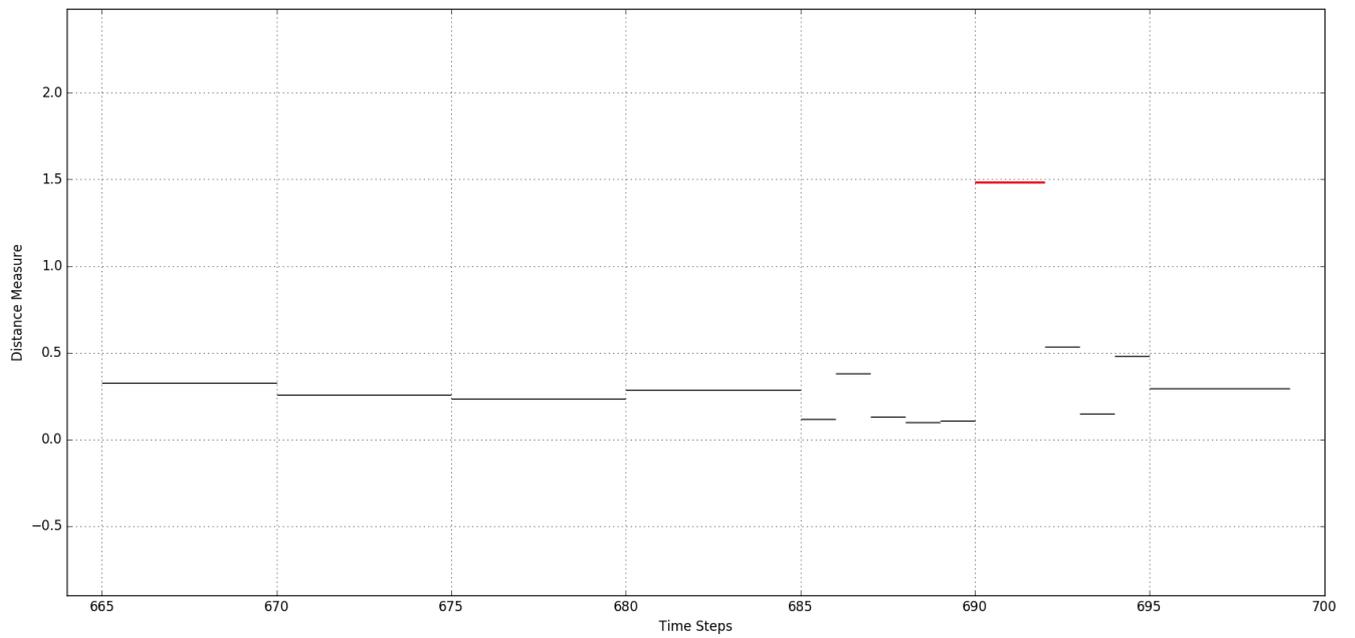


Figure 5.12: Plot of measure values against time-steps for resolution 8 x 2.

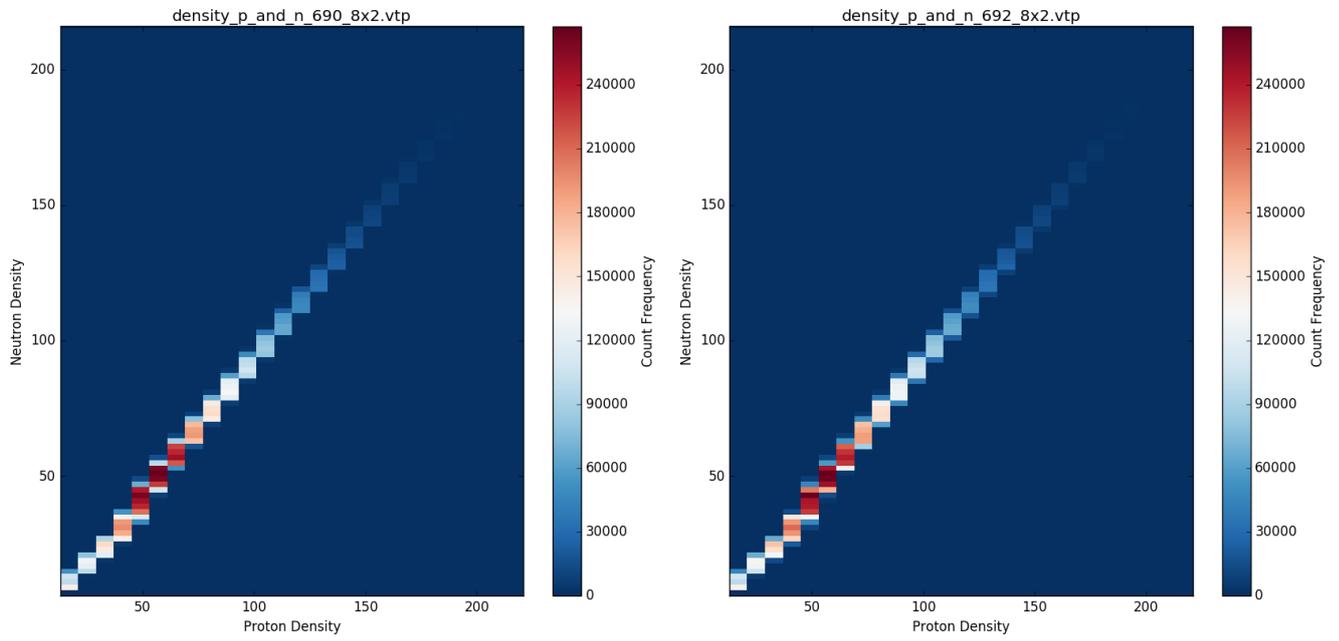


Figure 5.13: Histograms (count as size) of scission data-set for two consecutive time-steps

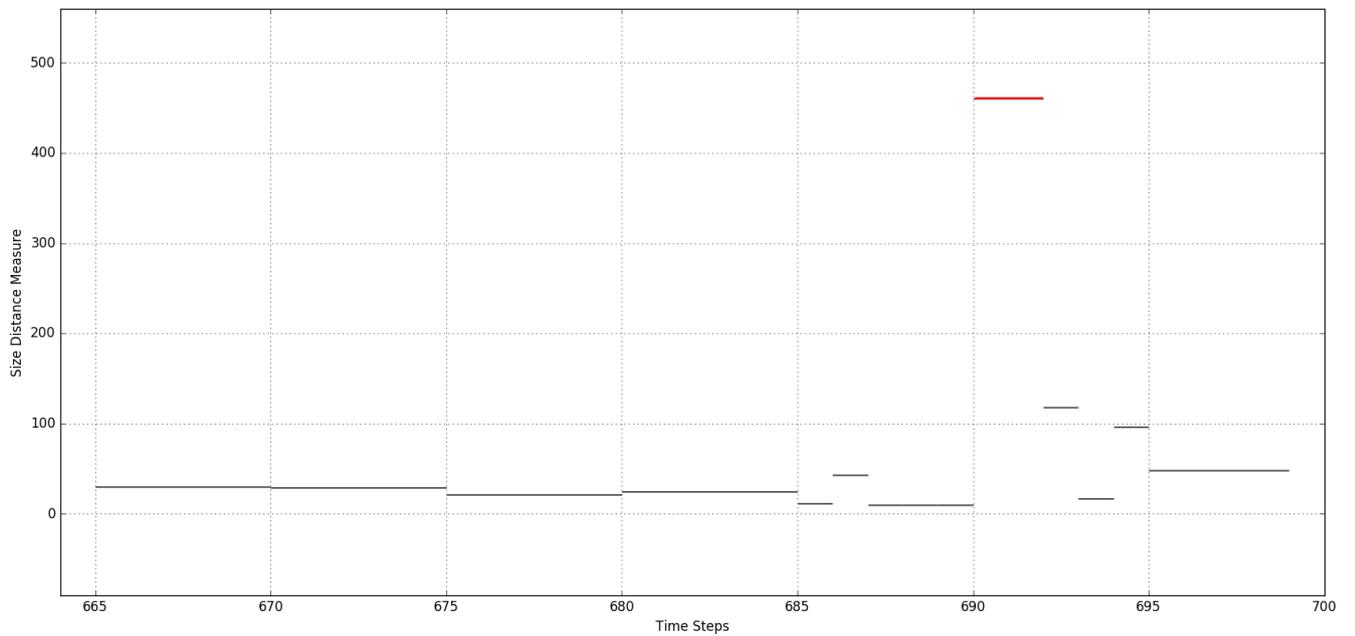


Figure 5.14: Plot of size measure values against time-steps for resolution 8 x 2.

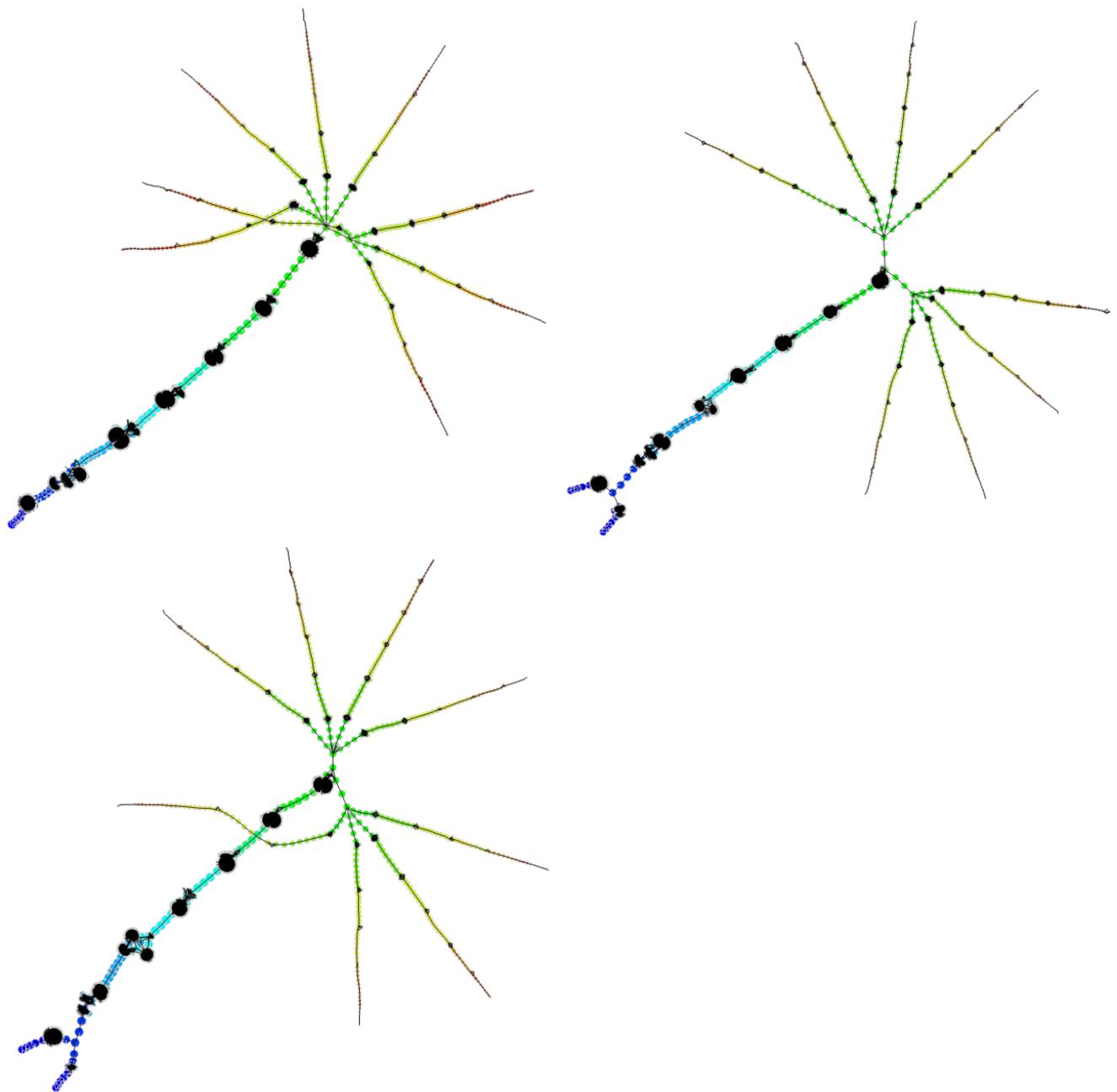


Figure 5.15: Joint Contour Nets for time steps 690 (top left), 692 (top right) and 693 (bottom)

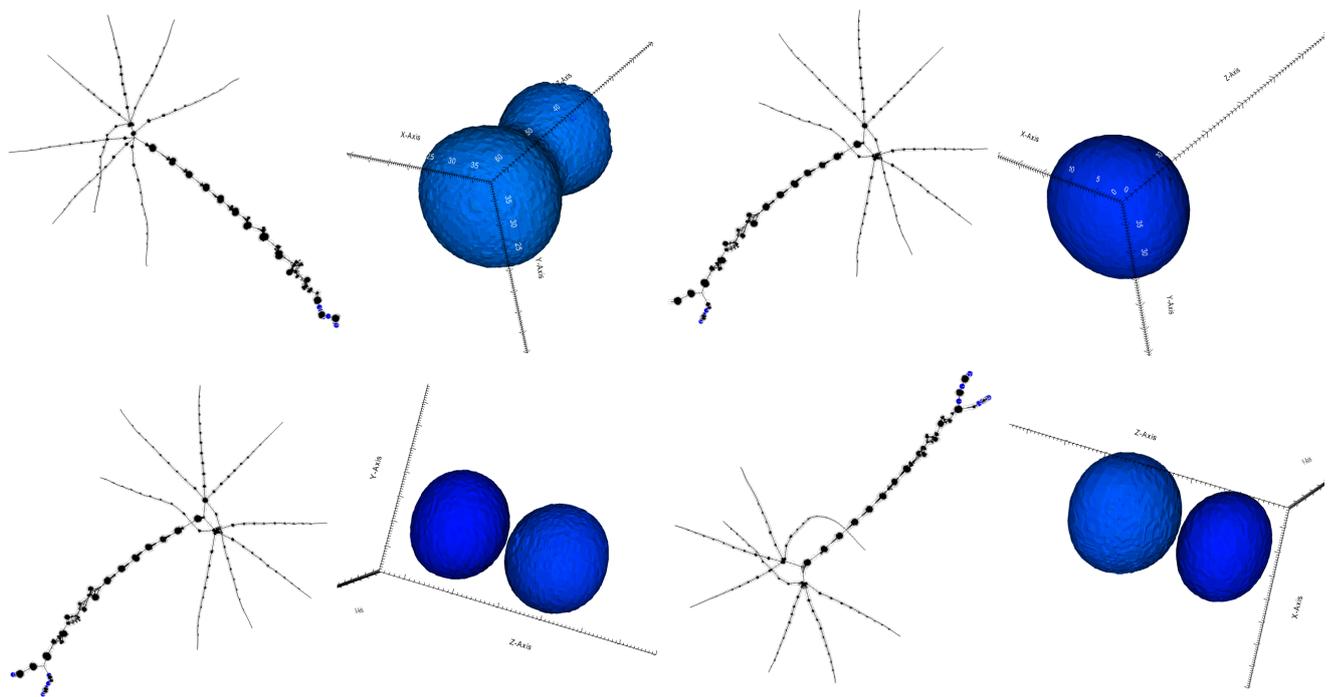


Figure 5.16: Joint Contour Nets and their slab geometry in domain for time steps 690 (top left), 692 (top right and bottom left) and 693 (bottom right) for different selected portions (colored in blue) of corresponding JCN

Chapter 6

Conclusions and Future Work

We proposed a measure to find the similarity and dissimilarity between two multi-fields which captures some interesting features of a multi-field. For further analysis one can try this measure on some other data-sets as well to see the behaviour of the measure. To know some of the properties of our measure further analysis is needed. The work can be extended to find a measure for more than two multi-fields and compare them.

Bibliography

- [1] H. Carr and D.J. Duke, Joint Contour Nets: Topological Analysis of Multivariate Data, IEEE Pacific Visualization, (2013) [1](#), [3](#)
- [2] D.J. Duke, H. Carr, A. Knoll, N. Schunck, H.A. Nam and A. Stszczak, Visualizing Nuclear Scission Through a Multifield Extension of Topological Analysis, Transactions on Visualization and Computer Graphics, 18(12), 2033-2040, IEEE Computer Society Press, (2012) [1](#)
- [3] Himangshu Saikia, Hans-Peter Seidel, Tino Weinkauff : Extended Branch Decomposition Graphs:Structural Comparison of Scalar Data In Eurographics Conference on Visualization (EuroVis 2014) [1](#), [5](#)
- [4] Amit Chattopadhyay, Hamish Carr, David Duke and Zhao Geng : Extracting Jacobi Structures in Reeb Spaces. In Eurographics Conference on Visualization, Short Paper, 9-13 June, Swansea, UK, (Eurovis 2014) [1](#)
- [5] Dmitriy Morozov, Kenes Beketayev, Gunther Weber : Interleaving Distance between Merge Trees, TopoInVis (2013) [1](#), [5](#)
- [6] Kenes Beketayev, Damir Yeliussizov, Dmitriy Morozov, Gunther H. Weber, and Bernd Hamann: Measuring Distance Between Merge Trees. In book: Topological Methods in Data Analysis and Visualization III, Publisher: Springer International Publishing, Editors: Bremer, Peer-Timo and Hotz, Ingrid and Pascucci, Valerio and Peikert, Ronald, pp.151-165 [6](#)
- [7] Zhongjie Wang, Hans-Peter Seidel, and Tino Weinkauff : Multi-field Pattern Matching based on Sparse Feature Sampling, IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No. 1, January (2016) [iv](#), [6](#)
- [8] H. Edelsbrunner, J. Harer, A. K. Patel: Reeb Spaces of Piecewise Linear Mappings. In SoCG (2008). [3](#)

BIBLIOGRAPHY

- [9] Dilip Mathew Thomas and Vijay Natarajan: Symmetry In Scalar Field Topology, In IEEE Transactions on Visualization and Computer Graphics, 17.12 (2011): 2035-2044. 1
- [10] Vidya Narayanan, Dilip Mathew Thomas, Vijay Natarajan: Distance between Extremum Graphs. In Visualization Symposium (PacificVis), 2015 IEEE Pacific 1
- [11] Kenes Beketayev, Damir Yeliussizov, Dmitriy Morozov, Gunther H. Weber, and Bernd Hamann: Measuring the Distance between Merge Trees. In LBNL-6629E (2014) 1
- [12] <https://institutes.engineering.leeds.ac.uk/computing/research/cse/meta/documentation.html> 8