FACULTY OF SCIENCE AND TECHNOLOGY

MASTER THESIS

Study programme / specialisation:

Computational Engineering

The spring semester, 2022

Open / Confidential

Author:
Aleksander Storhaug Strøm

.....
(signature author)

Course coordinator:
Aksel Hiorth
Supervisor(s):

Robert Schmidt
James Goldwater
Wiktor Waldemar Weibull
Aksel Hiorth

Thesis title:

Relative Geologic Time By Dynamic Time Warping

Credits (ECTS):
30
Keywords:

Geophysical

Computational

Signal Processing

Computer Science

Pages: 91

+ appendix: 17

Stavanger, 13.06.2022
date/year

University of Stavanger

CEGAL

Relative Geologic Time By Dynamic Time Warping

Aleksander Storhaug Strøm

The Thesis was formed in collaboration between Cegal Sysco and the University Of Stavanger

University of Stavanger

CEGAL

Acknowledgement

This thesis concludes my Master of Science in Computational Engineering. The thesis readily combines knowledge acquired in my Bachelor's in Petroleum Geoscience and my Master's in Computational Engineering. I am genuinely grateful for the opportunity to explore both the respective sciences. Petroleum Geoscience has successfully engaged my curiosity for learning methods and concepts that helped me understand the ways of the Earth. Computational Engineering has sparked the enthusiasm for powerful generic methods within mathematics and statistics, where I envision many exciting analyses to be explored in the future.

I want to provide a special thanks to Cegal, that believed in me during the summer internship and later proposed the exciting problem for the Master's thesis. Specifically, I want to thank Robert Schmidt, James Goldwater, and Espen Knudsen from Cegal, where Robert and James have continued by being my supervisors during the master thesis. I'm thankful for your collaboration, support, and many hours helping with practical, theoretical, and academic problems. I am grateful for all the help provided by my supervisors, Aksel Hiorth and Wiktor Waldemar Weibull; your guidance has been essential when working on the thesis. Thanks to the University of Stavanger for all its exciting lectures and for allowing me to undertake extra courses. Furthermore, I think my girlfriend, Hanne, is owed a colossal thanks for her patience and warmth during the time I've been preoccupied exploring my multiple projects. The last couple of years would not have been the same without you. Thanks to family and friends.

Abstract

This thesis considers an approach to tackle a core problem within seismic interpretation, which is bringing an autonomously generated interpretation of the seismic data, which is now known as a Relative Geologic Time. The proposed method readily utilizes the method of Dynamic Time Warping, which is an established method within signal processing. Using Dynamic Time Warping is thought to replicate similar interpretations an interpreter would conduct when fulfilling an interpretation of the subsurface. Utilizing Dynamic Time Warping to seismic data results in a fully autonomous interpretation of the subsurface, conducted in minutes and seconds. The method is simple and extendable, which can easily be further expanded. The workflow established during the thesis work results in a method that successfully produces an RGT volume. However, problems related to the method must be improved to enhance the outcome further and diminish errors present in the result. Furthermore, even with problems associated with the method, potential solutions are described in detail in the discussion and appendix. Discussion affiliated with previous attempts in solving Relative Geologic Time volumes is emphasized. The research conducted in Dynamic Time Warping is promising and emits potential for further research. LaTeX setup by Gunn and Patel (2017).

Contents

List of Figures

Nomenclature

δ	Binary label list
$\Im$	Imaginary Part of Signal
ϕ	Unwrapped Phase
$\Re$	Real Part of Signal
φ	Wrapped Phase
$A(f)$	Amplitude of signal in frequency-domain
BFS	Breadth First Search
CNN	Convolutional Neural Network
D	Cost Matrix
d	Similarity Matrix
DFS	Depth First Search
DTW	Dynamic Time Warping
FT	Fourier Transform
$G(f)$	Signal in frequency-domain
$g(t)$	Signal in time-domain
GAN	Generative Adversarial Network
$InSAR$	Interferometric Synthetic Aperture Radar
k	Integer list
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
$ReLU$	Rectified Linear Unit
RGB	Red Green Blue
RGT	Relative Geologic Time
RNN	Recurrent Neural Network
SAR	Synthetic Aperture Radar
TWT	Two Way Traveltime
UI	User Interface

Chapter 1

Introduction

This thesis seeks to deal with the problem of automatic interpretation. In this context, the objective is to automatically estimate Relative Geologic Time(RGT) from seismic reflection images. Seismic interpretation is typically performed manually, implying that the process is time-consuming. Seismic reflections are widely used within geophysics and hydrocarbon exploration. Many methods are associated with seismic inversion, and these methods often transform signals to increase insight regarding the content of fluid and lithologies to located areas of particular interest. Since seismic inversion is very mature, and many of the methods are very efficient and precise, there is more need to construct algorithms that provide fast structural knowledge of the subsurface. Estimating RGT would yield an automatic interpretation of the structural geology as a product of the time volume. Because of the relation between time and structures in the subsurface. Today's workflows can be tedious, entirely or sub-manually performed methods, where a geophysicist must evaluate the choice of seeds from the reflections in sub-manual methods.

The advantages of making seismic interpretation autonomous are multiple, primarily since autonomous procedure signifies the utilization of computing power of modern-day computers or cloud solutions. If an autonomous procedure provides a satisfactory outcome, it could rapidly increase the level of insight extracted from seismic data. The algorithm is directly coded as a plugin into Cegal's Avary, consisting of a user interface (UI) connected to the defined procedures. Implementing algorithms within this framework would ease the toll on the computer and provide a quicker insight gain.

An RGT volume will imply information on how the subsurface's age is distributed through the subsurface, and this can provide helpful information regarding structural properties within the subsurface. The product of tracking faults, folds, and unconformities is a volume describing the layers' ages in the subsurface. Contrary to RGT, faults, folds, and unconformities should be visible directly from the RGT volume. This reverses the interpretation workflow, where traditionally, the structures need to be determined before the relative age is interpreted. In contrast, the structures are identified by the generated RGT volume using the method described in this research. Moreover, since RGT is time-continuous, it implies that time volume has advanced its completeness over its traditional counterpart. Completeness can stimulate the use of the Wheeler Diagram, which readily can provide further information regarding sedimentary settings of the subsurface. This thesis will consider various methods that are, or could, be employed in estimating an RGT volume from seismic traces. Discussing the difference between the existing, potential methods, or variations is critical to emphasize the issues related to each of these methods.

The methodology used in the research is evaluating the RGT volume based on a solution derived from the principle of Dynamic Time Warping(DTW). Two other methods were evaluated before deciding on DTW: Phase Unwrapping and Deep Neural Networks. Although both methods are associated with promising findings in establishing RGT volumes in prior research, there are weaknesses affiliated with these methods. Hence, the principle of DTW appeared to be more simplistic, extensional, and an intermediate compromise between noise resistance and computational load among these methods(as discussed in chapter 3.1). The concept of DTW uses comparisons of the similarities amongst seismic traces as the fundamental feature. The comparison outcome yields a warped path, e.g., point-wise relations between signals where similarity is at its highest. Method satisfactorily finds non-linear relations between signals for which the method was originally developed. However, many extensions are associated with DTW, and many flexible options can be included in DTW for solving RGT. The outcome of this research appears to be promising. Nonetheless, some details indicate that further research must be a prerequisite before the method can be fully trusted to interpret the subsurface. Many potential additions are discussed to emphasize the way forward for the methodology.

1.1 Thesis Outline

- Chapter 2 Relative Geologic Time: Description of relevant conceptual background regarding seismic and how it relates to RGT, the description related to RGT additionally previous solutions on RGT or related topics within the field of seismic.
- Chapter 3 Theory and Method: Brief discussion considering choice of method.A description of the workflow established in this thesis is provided. Description of Phase Unwrapping. Further description of underlying theory related to Dynamic Time Warping, which is the central theorem in this paper. Description of the complete method and some description of attempted solution theory.
- Chapter 4 Result: A chapter that highlights the method's outcome and its currently implemented variation, both for synthetic data and real data sets. Comments associated with visible features observed in the outcome are provided.
- Chapter 5 Discussion: Description of performance of the method, additionally enlightenment of weaknesses are included. Evaluation of attempted solutions on the matter for solving issues encountered in the research. Discussion regarding plausible future works.
- Chapter 6 Conclusion: A final verdict of all content within this paper.
- Chapter 7 Bibliography: Literature and figure references.
- Chapter 8 Appendix: Describes non-implemented methods which yield potential solutions to RGT related to both Phase Unwrapping and Dynamic Time Warping extensions and a description of Deep Neural Network.

Chapter 2

Relative Geologic Time

Clarification of the concept of RGT requires elaborating fundamental background knowledge associated with the topic. This chapter will provide explanatory descriptions considering the advantages RGT volume yields. Emphasizes on future extensions onto the method are provided. Moreover, a discussion regarding RGT relevant solutions and explaining how the method functions behave. An informative description of how the RGT approach differs from the traditional approach, mainly explained by the DTW approach on an RGT volume.

2.1 Background knowledge

2.1.1 Seismic Signals

In this thesis, RGT is estimated from seismic reflection images. A whole range of factors influences the seismic response, for example, elastic properties of the lithology, the thickness of layers in the subsurface, and the amount and the type of fluid content. Usually, the procedure for gathering seismic reflection is initiated by firing a seismic source and then recording the response as a function of time at discrete receiver locations. This process is repeated by moving the source and receivers to different positions, thus increasing the coverage. Figure 2.1 shows an example of a marine seismic acquisition process. Here air -guns are used as a seismic source, and hydrophones are used to record the seismic wave fields.

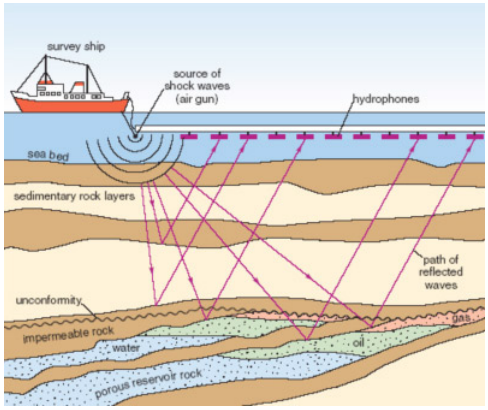


FIGURE 2.1: Seismic acquisition (Rigzone, 2022)

Any interface across which there are contrasts in elastic properties, such as density, P-wave velocity, or S-wave velocity, can produce reflections. Figure 2.2 shows an illustration of the partitioning of seismic waves across a reflector. In this case, an incident P-wave gets reflected as a P-wave and an S-wave. At the same time, a P-wave and an S-wave are also transmitted further into the subsurface. The reflected

waves propagate back to the surface and get recorded by the seismic receivers. In elastic media, there are two types of waves that are typically used for imaging the reflectivity of the subsurface. These are the P- and S-waves. Whereas the first is a longitudinal wave, the second is a transverse wave. Only P-waves can travel in non-rigid media, such as water, such that in a marine seismic acquisition such as the one shown in Figure 2.1, only P-waves are produced and recorded. S-waves can only be recorded in a marine setting when the receivers are planted on the seabed.

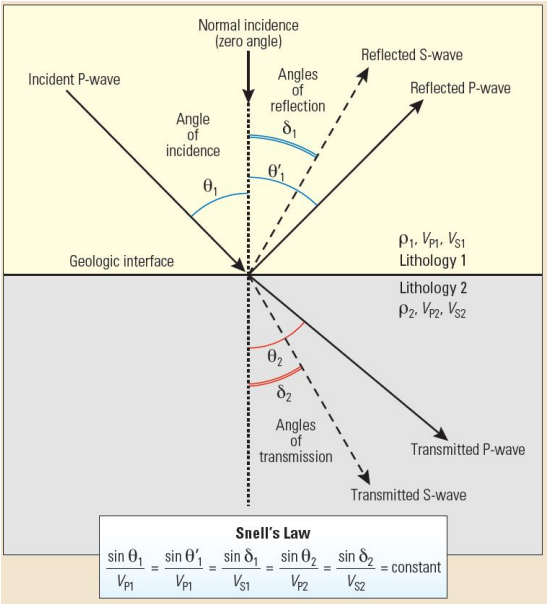


FIGURE 2.2: Illustration of seismic signals interacting between interfaces (Schlumberger, 2014)

Seismic signals experience much energy loss as signal traverses through the subsurface. One such energy minimizing factor would be geometrical spreading. As the seismic signal spreads, its energy intensity per area is depleted as the wavefront spreads in a conical sphere shape in every direction. Internal friction between layers and the signal can also dampen signal amplitude. Detaching the signal into reflected and transmitted waves will diminish the signal amplitude(see figure 2.2).

The recorded seismic wave fields can be processed into a model of the reflectivity of the subsurface. This process is called seismic imaging and is described in many textbooks (ex. Buland (2021)). The reflectivity model is an image of the strength and structure of reflectors as a function of spatial position and depth/time. The depth or vertical two-way time can be used as a vertical scale, and time-depth tables can be constructed using the estimated velocities of propagation of seismic waves through the subsurface. The velocities can be estimated through processing/and or obtained using check-shot surveys performed by placing seismic receivers inside wellbores. All seismic reflections will eventually build several seismic profiles which show the amplitude change in depth/time, called seismic trace. The collection of these seismic traces shapes a data set that the geophysicist will use for interpretation of the subsurface (Weibull, 2020). Hence, the seismic traces are used for interpolation between traces to form time surfaces known as seismic horizons (chap: 2.1.2).

Seismic signals are band-limited. The source characteristics ultimately control the bandwidth of seismic waves. However, the bandwidth can also be affected by

absorption and scattering effects. The bandwidth of the seismic waves is directly related to the vertical resolution of the seismic images. For example, the ability to resolve the layer thicknesses is inversely proportional to the dominant seismic period (T), where T is given approximately by the inverse of the dominant frequency of the seismic signals. The signal can reside within the frequency or temporal domain by conversion. The Fourier transformation conducts this conversion, and the conversion between the two states can be conducted indefinitely (Weibull, 2020). Nevertheless, the signal's shape is likely to change to describe information relevant to frequency or the temporal domain (more details on this matter in chapter 3.2.1)

2.1.2 Seismic Horizons

Seismic signals can be interpreted to give information about deposition age. Identifying deposition age laterally and vertically allows us to define seismic horizons and identify faults. When horizons are interpreted, their resulting surfaces illustrate how the environment's topography has been preserved from its time of deposition. By including multiple interpreted seismic horizons and implementing several horizons, it might be possible to detect central structural behavior throughout the seismic cube. An example is the "Fault 1" and "Fault 2" faults, which can be seen in figure 2.3. A fault is principally a displacement of the layers that shifts the ordering of the layers' age such that the layer of the same age is no longer aligned on the same depth. The displacement can contribute to a repeating sequence or lack of intermediate sequence in the layers. Similar to unconformities, the faults' displacements can be caused by erosion. That contributes to the removal of layers, namely a gap in time. See figure 2.3 where unconformities are defined, representing several layers merging with the overlying younger layer, readily resulting in time gaps in the seismic section. A brief discussion will consider today's tools underlining the methods that the interpreter uses to track seismic horizons. The most commonly employed method is for the interpreter to perform an evaluated selection of multiple polygons at a horizon. These selected polygons, known as seeds, are utilized in the concept of interpolation. Interpolation will use these spaced seeds to approximate the continuation of the horizon between the seeds. Eventually, the horizon will be fully mapped in the cube; the method is known as auto-tracker. Due to the need for seeds, it is a sub-manual method and will only account for the continuation of one seismic horizon.

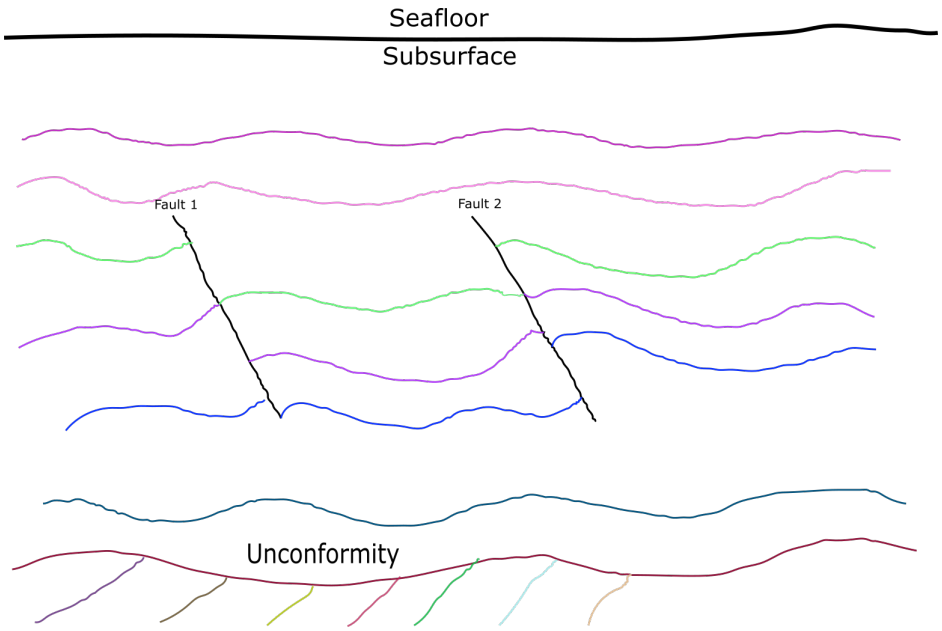


FIGURE 2.3: Illustration of interpretation seismic horizon from seismic signals interacting between interfaces

When the interpretation of seismic reflectors is finalized, an overall structure is being revealed, as the relative age of every layer forms structures that impact operational decisions in hydrocarbon exploration. The background for detecting seismic horizons is linked to seismic reflectivity, precisely the acoustic impedance between layers.

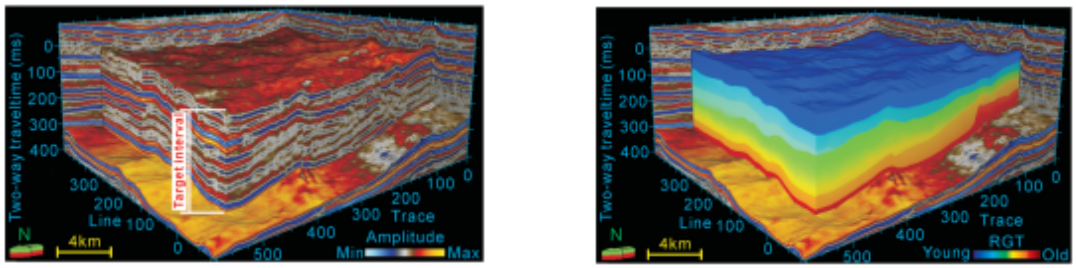
$$I = V\rho. \tag{2.1}$$

Where I is the impedance, ρ is the density, and V is the velocity of the current layer. Seismic impedance is the amplitudes observed in seismic data, and these amplitudes can be calculated for each component of the seismic wave. Primary wave $I_P = V_P\rho$ or the shear component $I_S = V_S\rho$. Interaction between layers is also based on the interaction between different rock properties, such as density and elasticity. The properties are often explained by the acoustic impedance such that a higher acoustic impedance might reveal a higher rock density or velocity is propagated faster through the material. The synergy between layers is explained by the reflection coefficient(RC) relation.

$$RC = \frac{I_2 - I_1}{I_1 + I_2}. \tag{2.2}$$

From eq: 2.2, it can be determined that the impedance from both interacting layers(I_1 and I_2) defines the reflectivity of the interface between the layers. The reflection coefficient provides a magnitude that explains how much of the initial signal has been reflected. Thereby also informs how much of the signal is transmitted further onto the underlying layer (Buland, 2021).

2.2 RGT Definition



(A) Reference Cube (B) Relative Geologic Time Cube

FIGURE 2.4: Example of a full RGT cube by (Wu and Zhong, 2012a)

The principle of constructing relative geologic time (RGT) from seismic reflectors has emerged from the field of seismology and was first introduced as a concept by Stark (2004). Moreover, the concept of relative ages and absolute ages is widely integrated into geoscience and geophysics, specifically used to determine the ages in which rock or layer has been deposited or occurrences of geological events. Two types of time reference measurement are used: absolute time and relative time. The last-mentioned measurement is to estimate at what time an object roughly belongs. Absolute measurements are used as a correcting factor to adjust the relative time to precisely date the object to a specific time. Relative and absolute time is referenced towards an eon, era, or period within the chronostratigraphic chart. However, when utilizing RGT is not linked to a specific time in the chronostratigraphic chart, as in the previously mentioned measurements; instead, a relative scale is employed. Nevertheless, as the name suggests, RGT is relative, i.e., in this scenario linked to how variation between ages amongst seismic layers/strata, thus the need for a value given by specific age is less vital.

The concept of RGT is also more frequently used directly on seismic data, where the relative and absolute measurements often relate to physical measurements from rocks or layers for determining the age. RGT is linked to seismic data mainly by the attribute of a seismic reflector. By definition will identify the reflected signal areas where the interface between layers consists of significant or small changes between layers. These changes often are represented by distinct dissimilarities in density, fluid content, or elastic properties. Hence, the reflected signals will establish data volume built by several vertical components after prepossessing. That describes the changes in the reflected signals at any time (two-way travel time) in the vertical signal, referred to as seismic traces.

The relevance between relative age and signals that cycle between amplitude in the frequency or temporal domain is that signals (complex or not) show different shapes at various positions. Additionally, the seismic traces are incremented by time in the vertical direction. Hence, linking events/shapes from different traces to the shift amongst the traces in time will indicate some relative time shift. Thereby, the basis for the theorem of RGT is defined by this relation between signal and time. The accumulation of the multiple RGT traces would eventually reveal structures in the subsurface. If aggregated correctly, the outcome of the method would outline structures faster and based on all the adjacent information for local structures at every position (Stark, 2004).

2.3 Why RGT?

The value gain provided by applying RGT onto the seismic stack is multiple. A clear benefit is that time continuous values are established for the subsurface, such that seismic horizons become time continuous. Time continuity implies no gaps in the generated interpretation, which means no horizontal or vertical oriented gaps. The compromise by human interpretation is that gaps in interpretation appear because an interpreter does not have the capacity to interpret every seismic horizon present in the data. Even as the interpreter is provided with the sub-manual method of auto-tracker, the method is confined to merely considering one seismic horizon. In contrast, an RGT method will provide an interpretation of all seismic horizons. When defining seismic horizons, the process often appears to be time absorbing when creating ceaseless time horizons. Oppositely, the RGT will rapidly conduct the procedure in seconds and minutes, depending on the slice's size or the whole cube being evaluated.

Creating an RGT volume will reveal multiple displacements in time in each trace. These changes jointly accumulate the structures present in the subsurface, such as faults, folds, and unconformities. That is detectable directly from the outcome of the RGT procedure. Thus, RGT leads to an autonomous way of predicting folds and faults by the algorithm's ability to outline the time distribution. An interpreter is biased in predicting based on the local structures, whereas RGT can predict based on the whole volume. Through the use of classical seismic interpretation, there is also a risk associated with the resolution limitation that the domain expert can perceive. While the automated procedure is not limited by sight, i.e., all data present are used in the interpretation. However, it lacks the professional judgment that a domain expert can conduct. The method is solely based on the data inserted into the method.

Since the RGT volume indicates ordering a relative deposition, such that information regarding whether a rock A was deposited before rock B, it leads to stratigraphic information regarding the temporal ordering of layers in the subsurface. Also, RGT will create continuous seismic horizons, and this is the case because all information is perceived and interpreted. Contrary to a human that will interpret limited parts of the information, namely the boundaries between peak and trough (see fig 2.5). Through RGT, a point-wise interpretation would be conducted, which finds paths amongst traces. This ultimately means the transition from traditional to RGT involves a transformation from a discrete result to a continuous result, observe figure 2.5 (Stark, 2004). Multiple warped paths at multiple positions at peaks and troughs outline a continuous interpretation instead of an interpretation being situated at the boundary between peak and trough (observable in fig 2.5) . Furthermore, a geophysical seismic interpretation often leads to a locally defined prediction of the seismic horizon, yet an RGT volume should enable a more global alignment of the seismic horizon.

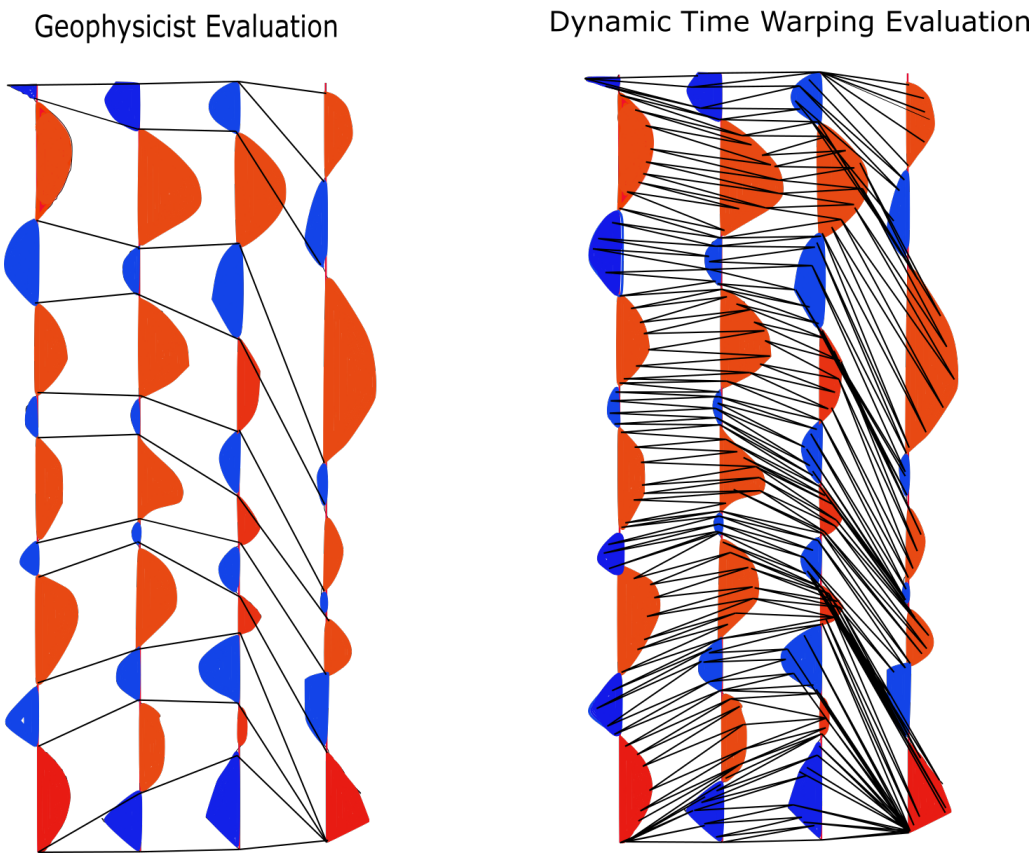


FIGURE 2.5: Relation between traditional interpretation and RGT found by Dynamic time warp. Left side: Traditional way to interpret seismic traces. Right side: DTW approach to correlate seismic traces.

The potential use of the RGT volume is combined with the Wheeler Diagram. The diagram uses seismic interpretation to detect actual time gaps, which leads to sequence stratigraphic information. Stratigraphic information helps determine the depositional environment. That can be employed to predict the sedimentary properties of rocks(more in 2.4). Because RGT is continuous (see fig 2.5) would lead to a more complete estimate of the Wheeler diagram. Thus, leading a higher portion of the diagram will be represented. Therefore, due to the completeness of RGT, the diagram is assumed to be more precise. Hence, the Wheeler diagram is less speculative, then interpreted sedimentary facies in the diagram are more likely to be present. On the contrary, the diagrams constructed are based on manual seismic interpretation, potentially leading to more inaccurate gaps in the diagram. Through the Wheeler Diagram, erosional locations are emphasized. Using these eroded areas can provide helpful information regarding what kind of depositional system it once belonged to (more in chapter 2.4). The structural analysis could yield detailed information from the RGT interpretation if the method finding RGT is stable and yields a globally aligned interpretation. The RGT approach leads to a complete globally aligned interpretation of the subsurface, which will promptly conduct its interpretation in a matter of seconds and minutes instead of days and weeks due to its autonomous nature.

2.4 Wheeler Diagram

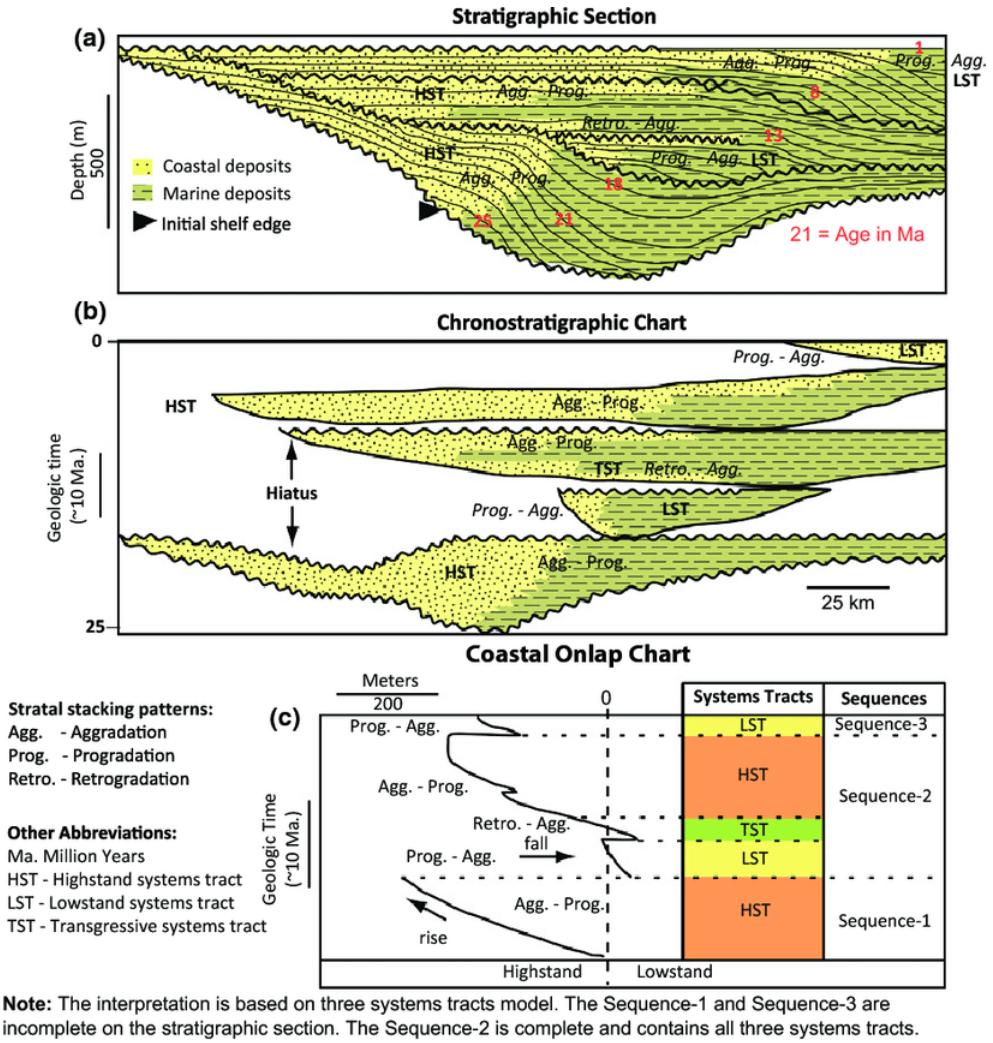
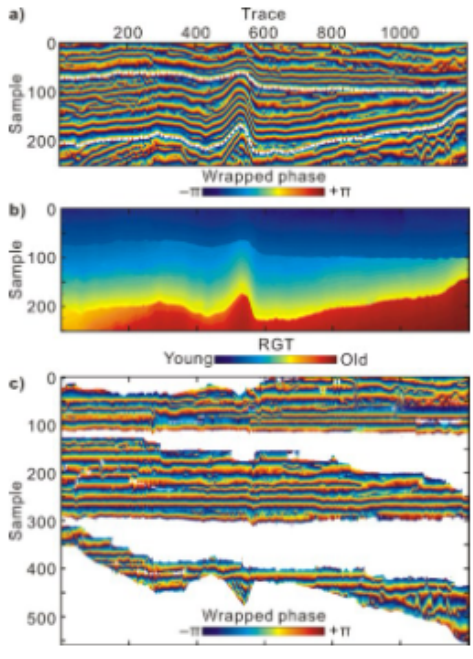


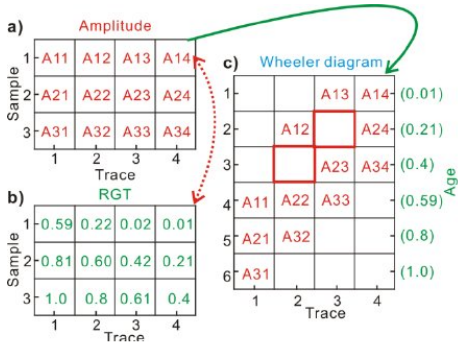
FIGURE 2.6: Wheeler Diagram conceptual image. Including Stratigraphic, Chronostratigraphic and dispositional environment section (Qayyum et al., 2015)

Wheeler diagrams are generally purposed as a method for chronostratigraphic sequencing. Wheeler Diagrams can classify depositional environment, also known as a chronostratigraphic chart. Correctly classifying the depositional environment of the different parts of the region is essential in evaluating hydrocarbon field prospects and geologic surveys. Therefore, relating to stratigraphic environments in the realm of sedimentology. That contributes to interpreting sedimentary facies, which readily provides information regarding grain size, porosity, sorting, and more. Hence, information yielding important reservoir properties can be acquired early in the seismic interpretation phase. Other methodologies often require physically extracting and evaluating parts of the strata or intrusion in the form of an oil well to approximate the stratigraphic sequence, rather than solely using remotely obtained results. The difference between the stratigraphic section and chronostratigraphic chart observed in figure 2.6, is mainly that the stratigraphic structures are interpreted from seismic data. These structures reach multiple depths and emphasize how layers are related in terms of accommodation. In comparison, chronostratigraphic will reveal

the layers' shape. It emphasizes time distribution, revealing areas exposed to erosion (unconformities). The Wheeler Diagram can reveal gaps in time (ref Hiatus fig 2.6) between the various layers, discontinuities and faults often cause these gaps. These gaps are treated as discontinuities of the layers and often indicate some deposition behavior (Qayyum et al., 2017). Thus, it is possible to use the diagram to determine the kind of depositional environment and how the material in an area has accumulated through the various strata. Aggradation is a sediment trend. That explains rising sediment in a continental depositional environment, typically in fluvial river systems. Similarly, Progradation explains a transition from coastal towards marine settings, and retrogradation describes a landward change (see fig 2.6). Definitions consider the amount of sediments deposition related to sea level rises or falls, and these explanatory definitions are known as system tracts. The lowstand system tract (LST) explains the low standing sea level, typically deposited by basin-floor fans and lowstand wedges. The intermediate state between low and high sea levels is given as the Transregressive system tract (TST), which gradually moves marine sediments against their most landward position. The most landward position is acquired when the system tract is defined as a Highstand system tract (HST), which maximizes the accommodation of sediments in the depositional environment. HST is typically first prone to Aggradation and then progresses towards Progradation, which explains the amount of initial coastal deposits in figure 2.6. These sequence stratigraphy features are essential because they provide a foundation for predicting sediment properties of an enclosed region(Strata, 2022).



(A) Top: Seismic amplitudes section. Middle: RGT section. Bottom: Wheeler diagram generated from amplitudes and RGT data



(B) Wheeler diagram mathematical procedure from continuous RGT volume.

FIGURE 2.7: Wheeler diagram procedure by: (Wu and Zhong, 2012c)

Moreover, mixing RGT with the Wheeler Diagram makes it possible to construct a relative timeline that can detect hiatus and yield information that interprets the accommodation space of sediments. These interpretations made using the Wheeler Diagram can be perceived as a linkage between the realm of geophysics to geoscience, such that both fields of expertise are supplementing in-between. Where depositional,

non-depositional, or discontinuous regions are classified remotely by Wheeler Diagram. Since RGT enables continuous information, it has the potential to build a more detailed diagram of the seismic cross-section represented in the Wheeler Diagram. Wu and Zhong (2012d) attempted to define a method for deriving a wheeler diagram directly from an RGT volume multiplied by seismic amplitudes. Additionally, a minimum constraint (T_{min}) is included to control the matter of hiatus present in the seismic section, where the logic is determined that every RGT value is compared against the minimum time in all of the traces. If the difference ($|RGT_x - T_{min}| > T_{min}$ is greater than the minimum value, these areas will be recognized as hiatus in the data, represented as the red squares in fig 2.7b c). The red squares equal the white space shown in fig 2.7a c), which illustrates non-depositional/ eroded regions, which resembles the illustration shown in the middle in figure 2.6. The procedure assesses all RGT traces until all of the traces have been processed. The relation between an RGT volume and the Wheeler diagram can ultimately lead to an automatic interpretation of the depositional environment. Acquiring this information directly from seismic provides intel regarding reservoir settings regarding the sedimentary properties deciphered from the depositional environment.

2.5 Attempted solutions in RGT

Numerous attempts have been made to provide algorithms that readily enhance the knowledge reward extracted from seismic data. Additionally, transitioning toward algorithmic approaches emphasizes that time evaluating seismic data diminishes. Today's most accurate method is to let a geophysicist conduct the interpretation manually, yet it is very time-consuming. Increasing productivity would yield more tools available for the geophysicist to enhance the work. One such method is horizon autotracker, where some seeded values are given to the algorithm by a domain expert; this information would be used in tracking the current horizon. One of the recent versions of the auto-tracker is doing similar procedure but supported by machine learning, namely Neural Network (see ch.8.2.1) and Support Vector Machine(now known as SVM) specifically radial basis kernel $K(x_i, x_i^{'}) = exp(-\gamma \sum_{j=1}^{P}(x_{ij} - x_{i^{'}j})$ used in $f(x) = \beta_0 + \sum_{i \in S} \hat{\alpha_i} K(x_i, x_i^{'})$ the specific method is often used for clustering non-linearly boundary similarities (Dyrendahl, 2018a).

The parameters given in the kernel ($K(x_i, x_i')$) are controlled by the hyper parameter(γ) to control the kernel's influence. Data nodes(x) correspond to the seismic amplitudes in adjacent cells to the current evaluation, i' indicates all other indexes than index i is used, which is the current position in crossline or inline direction. The indexes of adjacent cells are defined within the subset given in the function as a subset around the current cell $i \in S$, all indices of j, which is the opposite dimension to i, are considered to be of similar size as the subset of S (James et al., 2013). These are used to train the algorithm to properly find the continuation of the horizon based on positions from the seeded values. Thus, it will propagate the other defined seed value by different schemes 3x3 assessment, 3x3 validation, or 5x5 assessment of its adjacent nodes(see 2.8). The found new positions of the scheme will be the new seeds for the new assessment until convergence.

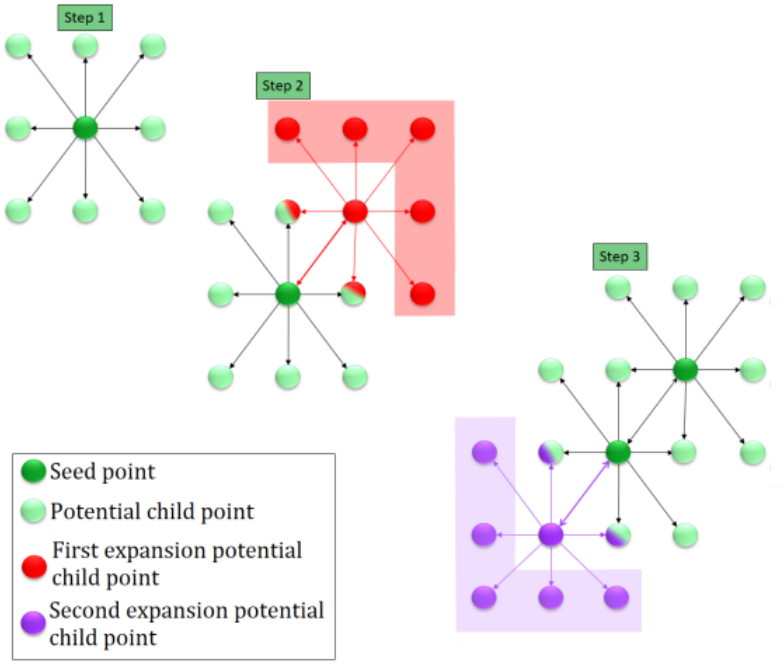
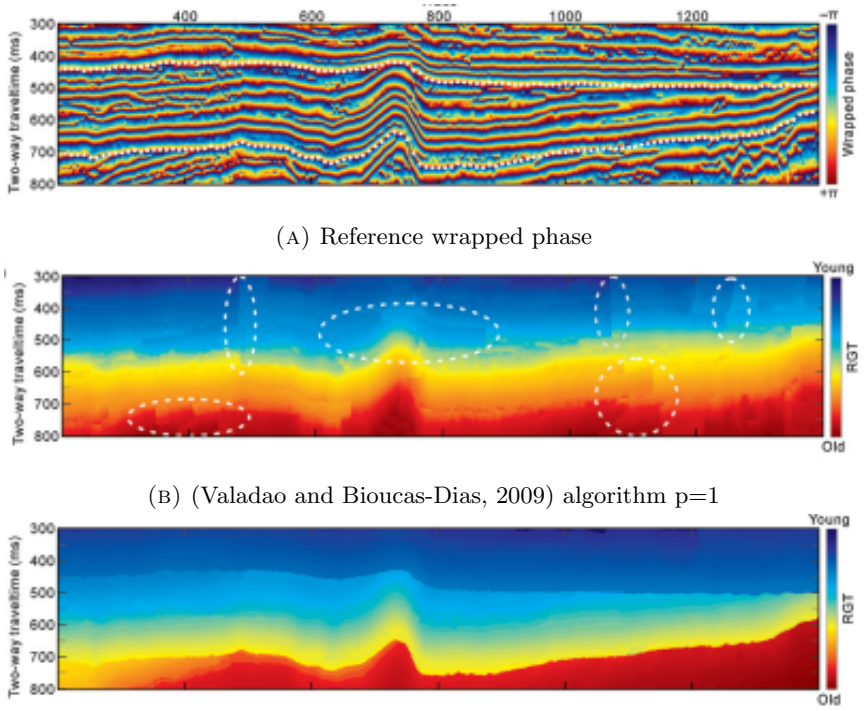


FIGURE 2.8: Point of view from above down into the seismic cube. Seeded scheme defined in auto-tracker, evaluates neighboring traces in selection for the continuation of the current seismic horizon (Dyrendahl, 2018b).

The expansion from the seed will be conducted as seen in figure 2.8. Furthermore, the method enables the selection of seed points at multiple positions in the interpreted seismic horizon, such that each point's expansion profiles will eventually interconnect. Moreover, the auto-tracked horizon can be corrected by deleting failing seeded values throughout the potential seismic horizon. The seeded evaluation will continue until a plane through the whole seismic cube defines the seismic horizon in all dimensions. However, the auto-tracker is restricted to a singular horizon and experiences a lot of processing time, and yet is a Sub-autonomous (ref seed selection and correction) method rather than a fully autonomous methodology (Dyrendahl, 2018a).

Multiple authors widely research phase unwrapping, but the most promising research is explicitly obtained for seismic settings are the studies of Wu and Zhong (2012b) and Valadao and Bioucas-Dias (2009). Both of these studies are tightly coupled, but the main difference is Wu and Zhong (2012b) extends the method to a 3D algorithm. The methodology combines phase unwrapping, which is readily done in a 1D case, but when it is extended to 2D or 3D cases, there is no general way to perform the update. There is a problem associated with phase unwrapping, which is the matter of error propagation, i.e., a significant amount of noise in data or ill performed unwrapping procedure leads to an accumulating error. The issue was solved by binary optimization through the use of graph notation and graph cut. Such that the found optimal binary digits are to be added to $\varphi = \phi + 2\pi k$(where k is binary digits at each temporal step) to the unwrapping algorithm at each temporal step. The choice of binary digits must diminish the energy emitted by the function to a minimum. The result seems to be high-performing in constructing a whole 3D RGT cube. Moreover, the phase unwrapping algorithm struggles to properly unwrap areas that contain unconformities, which leads to an error propagation of the result(see appendix 8.1.1).



(A) Reference wrapped phase

(B) (Valadao and Bioucas-Dias, 2009) algorithm p=1

(C) (Wu and Zhong, 2012b) improved algorithm p=0.6, optimized for unconformities

FIGURE 2.9: Result RGT from (Wu and Zhong, 2012a)

The use of deep learning is growing in seismic use cases along with other engineering domains. In many cases, deep learning is an important tool to solve complex problems generically without necessarily defining the solution based on the theoretical background of the problem. If data is available deep learning, which is mainly a complex neural network, learns the patterns present in the data (more explanation is provided in 8.2.1). To emphasize the use of deep learning to form an RGT volume, the research of Geng et al. (2020a) embarks on precisely this task. The implementation is primarily an implementation of CNN, the dominating learning algorithm that can perform classification or regression on images (more in 8.2.2). Seismic images are given into the trained network in a fixed size, 256x256 pixels. If input data is more extensive than the fixed size, the corresponding images are split into several sub-images of size 256x256. Present information is perceived in the paper considering an encoder, decoder, and refinement architecture, where spatial information would be convolved down to a flattened layer in the encoder before it is scaled back to its original scale in the decoder. Thus, global spatial information is preserved and included in the feature space before being given into the refinement modules, which conduct a regression prediction based on information from the encoder and decoder down/up projection step. Therefore, employing this structure makes it possible to down-project seismic images (see left figure 2.10) and consequently up-project the convolved information to represent the seismic image as an RGT image(see central figure 2.10).

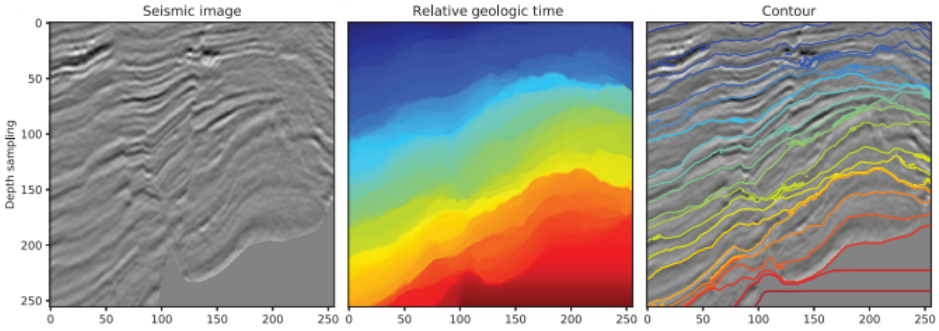


FIGURE 2.10: RGT volume by CNN (Geng et al., 2020b)

These images are assessed through a regression definition for establishing a continuous RGT volume and trained through comparison by root mean squared error (RMS), mean absolute error (MAE) or mean relative percentage difference (MRPD). Then neural weights are updated based on the accuracy measurement through back propagation (see appendix 8.2.1). As required in the training stage, a lot of interpreted data must be provided to train weights in a supervised manner. The conclusion of the paper, predominantly the CNN RGT workflow, has high accuracy in predicting the continuation of RGT, yet there are drawbacks associated with the method. One of such drawbacks is that the network has been trained on a case-specific seismic data set, such as handling faults and folds. However, the network may experience struggles when handling non-encountered situations, such as intrusion, unconformities, and salt bodies. Similar to the alignment problem experienced by the method developed in this thesis (more discussion in chap: 5.1.4), aligning parts of the found RGT solution is imperative. It can extend to even more complex architecture and provide more diverse training data to tackle different actual behaviors. That will lead to a more comprehensive training session, and a network might need to be tweaked for many of the different situations (Geng et al., 2020a).

Moreover, Dynamic Time Warping (DTW) methodology is also present in previous attempts in a seismological setting. Much of the research in this field that employs the use of DTW is mainly conducted by the author Dave Hale, known for his work with finding $\frac{V_p}{V_s}$ ratio and faults and seismogram correlation by utilizing DTW (Hale, 2013a), (Compton and Hale, 2014). The articles are considered a significant inspiration for this thesis. Although the literature does not attempt to solve an RGT problem, it highlights the potential uses of DTW in seismic. Especially aspects coupled to the question regarding how to define the algorithm so that merging RGT traces results in an entirely path-independently method (see chapter 3.5). In Hale (2013a) and Compton and Hale (2014), discussion and description of a symmetric two-side DTW algorithm used in enhance the accuracy of DTW. Seismogram correlation is a procedure used to pre-process the seismic data for alignment. Nevertheless, it is still valuable to mention in the context of RGT, more discussion in chapter 3.1.

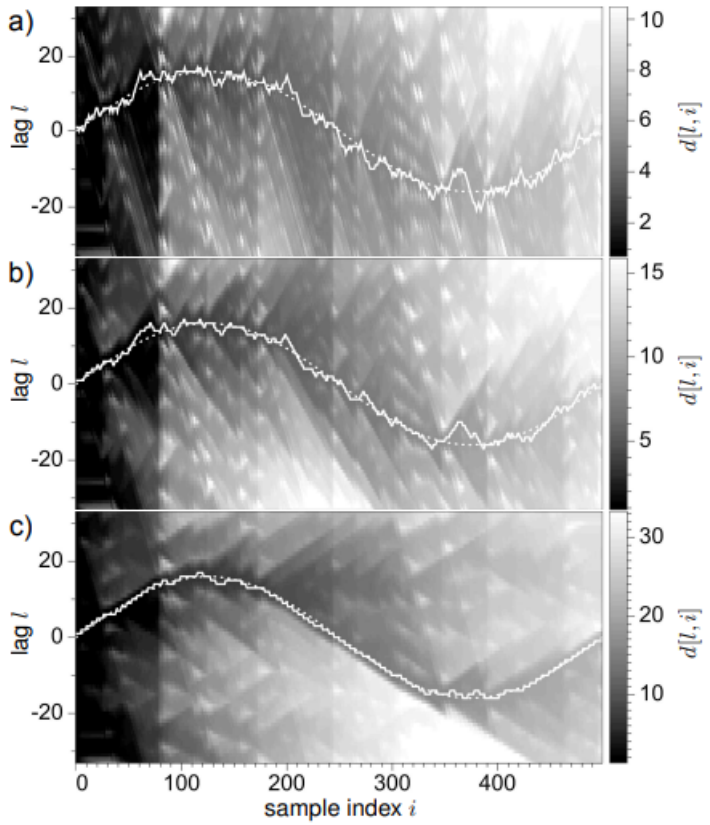


FIGURE 2.11: Hale's method for finding similarities between synthetic sinusoidal waves using symmetric DTW with differing allowed deviations a) 1, b) $\frac{1}{2}$, c) $\frac{1}{5}$. Dotted line=True shift, white line= approximated shift (Hale, 2013b)

The paper revealed details that the very structure of how dynamic time warping can be one-sided (see algorithm 2). The conclusion can be drawn to be one-sided because cost matrices are dependently updated (ref chap 3.3.3) from one end of the cost matrix to the opposite. That would ultimately mean that warped paths are biased toward either of the directions. Hence, the found shifts amongst seismic traces might not attenuate the noise in the data due to this bias, and thereby the warped path does not represent the true warped path between seismic traces. Hale (2013a) has presented a solution to account for both directions of bias by adding a filter that accounts for both forwardly and reversely propagated cost matrix of similarities. Accounting for both directions in the cost matrix appears as a smoothing filter for the dynamic time warping procedure. The definition provided by Hale (2013a):

Forward :

$$D_f[0,l] = d[0,l],$$

$$D_f[i,l] = d[i,l] + \begin{cases} D_f[i-1,l-1] \\ D_f[i-1,l] \\ D_f[i-1,lj+1] \end{cases}, \tag{2.3}$$

$$\text{i=1,2 ,. . . ,N ; l =0,1,. . . , M.}$$

Reverse :

$$D_r[N-1,l] = d[N-1,l],$$

$$D_r[i,l] = d[i,l] + \begin{cases} D_r[i+1,l-1] \\ D_r[i+1,l] \\ D_r[i+1,l+1] \end{cases}, \tag{2.4}$$

i=N-2,N-3 ,. . . ,0 ; l =0,1,. . . , M.

Symmetric :

$$\hat{D}[i,l] = D_f[i,l] + D_r[i,l] - d[i,l]. \tag{2.5}$$

The indexes, in this case, represent i, which provides the index for the reference (the current trace), and l is the index on the query signal(neighboring seismic trace), also known as the lag from reference to query signal. When assessing the cost matrix, both forward(D_f) and reversely(D_r) propagated, sum both to become a symmetrically($\hat{D}$) defined cost matrix. It is believed that a warped path can be extracted without bias towards either direction. Figure 2.11 shows a symmetric, where the black coloring correlates the least difference amongst traces. By correlating what sample index(x-axis) and the lag(y-axis) are associated with the slightest difference at each index, it assigns how much the compared trace differs from the sample trace(see fig: 2.11).

Hence, transforming the cost matrix to be two-sided and symmetric would yield a less biased warped path, yielding a more accurate path amongst signals. The author has found evidence to make the procedure path independent, and it is necessary to account for the horizontal shift likewise, not only vertical shifts. This is primarily because the symmetric alignment filter smoothes the vertical alignment errors after the first estimated similarities (as explained by algorithm 1). The authors named this process Dynamic Image Warping (DIW) for emphasizing unbiased warpings of seismic traces. The execution sequence should be independent, which leads to a tree-sequential ordering (similar to what is shown in figure 3.10). The procedure will utilize DTW again, except it is executed for each horizontal part of the seismic traces, accumulating and backtracking each trace row independently. The independency also enables the possibility for multi-processing of the operation, thus, reducing the computation time. Another central behavior present in this version of DTW implementation is the employment of the Sakoe-Chiba band (chapter 3.3.2) which ultimately leads to a limited amount of indices to estimate. Nevertheless, there are precision aspects linked to the band, such that if allowance of a band too big will lead to a slow procedure. However, it might also have too much flexibility attached to it, leading to an overfitting solution(see figure 2.11). That is a) in fig: 2.11 is full flexibility of the method, where reduction in b) and c) leads to the solution more accurately approximating the true solution (dotted line). Similarly, too low band flexibility can restrict the algorithm from finding the optimal warped path between the two traces, thus, yielding a wrong path (Hale, 2013a).

Chapter 3

Theory and Method

The motivation for this chapter is to describe relevant theory directly related to this thesis. Firstly, including discussion related to the choice of method. Phase Unwrapping is explained because, besides being a specific RGT method, it was believed that raw phase unwrapped volume could be used as a reference volume. Furthermore, the leading theory related to Dynamic time warping is included, where various subchapters describe modifications of the standard algorithm. These various formats of the standard algorithms describe different flexible schemes that are modified to enhance parts of the operations linked to the standard procedure, reducing time complexity, increasing the accuracy of path selection, or establishing different data structures. The method explanation will include a discussion of its implementation into Avary and the pseudo-code (chap: 3.5), enabling reproducible implementation. The concept of Binary Tree is described in detail because it was attempted as a potential solution to the merging problem of the volume. Theory related to attempted solutions that seek to solve its most significant source of error propagation, the merging problem in this thesis. Further plausible extensions or methods to the RGT problem or improvements on the preexisting method are included in the appendix.

3.1 Choice of Method

Multiple possible solutions to RGT volume and several details regarding each method have been assessed in selecting the appropriate method. Deep learning(see appendix chap: 8.2) was one of the more popular methods employed to construct RGT volumes. Convolutional Neural Network (CNN) was used in most of the solutions, and this is because CNN can convolve and flatten images in such a manner that the network assesses the image accurately, pixel by pixel. Regarding RGT, the CNN should be a regression problem, meaning that an encoder-decoder structure must be utilized. Encoder implies compressing information from a high-level image to become a low-level image. Moreover, the decoder is used on the low-level image to convert it back to high-level details, which are adjusted by the trained weights and biases. Dependent on the training data, the parameters of the encoder-decoder structure are adjusted to match the expected outcome outlined in the training data (more details appendix chap: 8.2.2). That means the network can replicate the interpretations made if given several images of seismic images populated with interpreted seismic horizons.

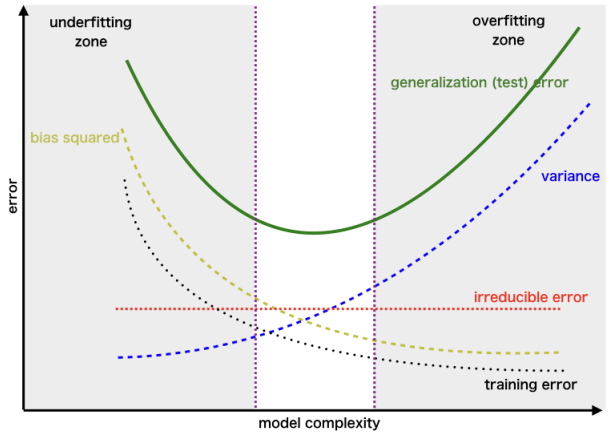


FIGURE 3.1: Illustrative image of variance-bias trade off (Mineetha, 2020)

Furthermore, some circumstances have to be asserted when neural networks. The matter of generalization and model complexity is often causing a model to be highly performing on the training data. Despite not being successful in predicting data not encountered before, this indicates overfitting. Overfitting (high variance) is caused by the model complexity being too fixated on the training data. Thus, not generalizable for future un-encountered predictions. These aspects lead to much work related to finding the perfect model architecture, although the amount of effort it is not assured to provide a perfect balance(see fig 3.1). Partitioning of train and test data also affects the complexity and generalization of the model. The training phase is associated with a heavy computational process and often employs GPU for training the model (details appendix chap: 8.2).

Deep learning can potentially lead to a model optimized for a specific subsurface field. However, if the model is used for predicting another subsurface field of differing characteristics, doubtfully, the model will be highly performing when encountering these non-encountered features. The implication regarding whether deep learning would provide a general solution to be employed in any subsurface field. Retraining the model when transitioning to a different field would be time-in-efficient, yet it might be required. Additionally, a neural network demands a variety of interpreted training and test data. The lack of training data and the uncertainty of acquiring a model that can be employed in any field leading to not selecting a Deep learning model (Bi et al., 2021a).

Phase unwrapping is used to reconstruct the original signal or image by increasing the signal magnitude monotonically. The control mechanism for unwrapping the signal is based on the magnitude at which the wrapped is situated. There is an aspect to be aware of when considering this phase unwrapping, and there is a matter of noise propagation when phase unwrapping. However, random noise is likely to be present in seismic data. The fact that the problem is considering seismic data indicates that much random noise is present in the data, which would lead to unwanted results or random gaps in the data due to error propagation. Nevertheless, robust methods for phase unwrapping are built to dampen the effect of error propagation caused by noise present in the data (Wu and Zhong, 2012b). Additionally, phase unwrapping is dependently reliant on stability conditions in order for the outcome to be reasonable. Itoh condition is the stability condition, and it requires the minor

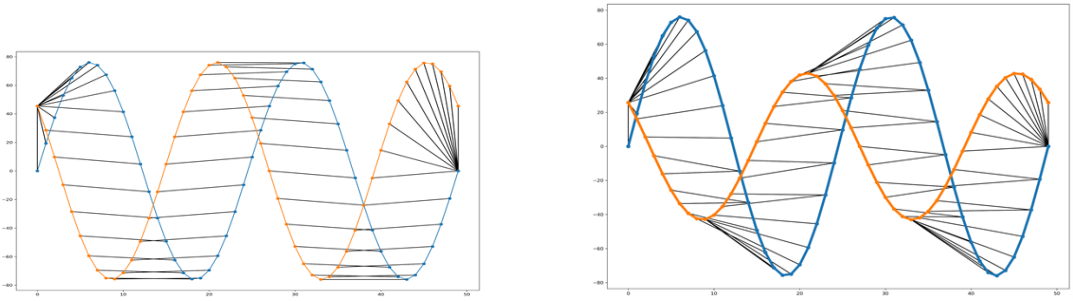
presence of random errors and carefully diminishes impact from discontinuities. However, these algorithms appear complex by definition, particularly when extended to multiple dimensions. There is uncertainty associated with the selection of algorithms to mitigate error propagation(more chap 3.2.1 and appendix 8.1.1). The uncertainty is caused by different strengths and weaknesses associated with each algorithm. No defined method provides an assured generalized result by phase unwrapping for signal problems. It is not assured to give a sensible solution after all the efforts have been invested. However, there are occurrences where phase unwrapping has been used to solve an RGT problem, where graph cut is included to find a minimum cut between the discontinuities in the seismic volume(see appendix chap: 8.1.1). The phase unwrapping would be controlled by a binary encoded volume, which given the article, would reconstruct a volume that provides a fair result in estimating an RGT volume. However, despite a satisfactory result, random error propagation is present, which is ever-present in real data. A noise invariant method would be required (Wu and Zhong, 2012b). The workflow considering phase unwrapping was primarily abandoned because of its complexity, where domain experts have attempted numerous solutions on the matter. It is doubtful that this thesis's contribution to the subject would yield any better-defined method than already attempted.

3.1.1 Why Dynamic Time Warping?

The native Dynamic time warping(DTW) version is simplistic and can be easily utilized to compare any time series. Additionally, the method is very general in its format, which stimulates being extensional to new versions of the methodology. DTW is capable of conducting pointwise nonlinear similarity comparison amongst two seismic traces (see figure 2.5). The concept is rooted in the found warped path (the relation amongst traces), selecting the minimum costly path when finding all its warped paths (more details chap: 3.3). The paths amongst traces are thought to mimic a geophysicist's interpretation of the continuation of seismic horizons. Hence, accumulating the RGT volume by aggregating the warped paths of all present trace pairs. According to belief, it is presumed that DTW yields more continuously transitioning boundaries mapping the signals than its human counterpart (see fig: 2.5).

Dynamic programming is readily employed when utilizing DTW, which implies recursion is used to assess local variations for solving global problems. The maximum time complexity DTW yield is $\mathcal{O}(MN)$, explained by the length of signals(M and N). Not too drastic for a single trace pair comparison. However, an RGT volume might consist of multiple slices that yield multiple traces, and each trace pair might have a length of thousands of indices. The totality of all these computations could induce severe time constraints on the method's effectiveness. An optimal DTW solution should reduce the computations' time complexity. Conveniently, some existing methods can reduce the complexity by implementing local constraints, global constraints, or redefined algorithmic structures. Nevertheless, these introduced constraints could induce a new unwanted behavior in the pursuit of solving the time complexity, such that an optimal path is not found (more on this in chapter 4,3.3, and 5). Connecting multiple positions onto a singular point in one of the signals, i.e., singularity. Singularity can become problematic because it might establish non-plausible paths between the signal pairs, yet it is also a way for DTW to shift time. Occurrences of singularities also arise when there are more differences between signals, such as different sizes and frequencies. The peaks and troughs between the various pairs might similarly lead to singularities as the endpoints (see figure 3.2a and 3.2b). Making

the correct shifts in time becomes essential when a fault displacement occurs in the seismic. Ultimately, this means that singularities have both positive and negative impacts in seismic settings.



(A) Similar but shifted alignment (B) Different amplitude shifted alignment

FIGURE 3.2: DTW occurence singularities

The objective is to ensure that the warped paths are merged into singularities when encountering real physical displacements amongst traces. The wide variety of options to be extended onto the native DTW procedure spans from methods diminishing time complexity to increasing accuracy of the warped paths. Some options provide resistance to outliers, i.e., resistance to noise in data (see chap: 3.3.4). Emerging trends by graph structures appear to yield high potential in finding global optimal warped paths, with high flexibility of path determination. Some of the literature related to graph is defined by Uchida et al. (2012a), Deriso and Boyd (2019a), additionally, graphs showcasing long-term dependencies is outlined by Wang et al. (2016a). Features exhibited by these methods seem particularly useful for the RGT use case, especially the iterative refinement of the warped path(see appendix chap: 8.3.4) and long-term dependencies for coupling RGT traces(see appendix chap: 8.3.3). Dynamic time warping is selected as the preferred method because of its simplistic nature and the well-defined extensions readily provided by the scientific community.

3.2 Phase Unwrapping

3.2.1 Instantaneous Phase

Seismic waves are periodic signals, and according to the Fourier theorem, any periodic signal can be explained by the sum of sine and cosine terms. The principle of instantaneous phase and frequency is an essential procedure within signal processing. The procedure can transform the signal to be represented within the complex plane. Transformation onto the complex plane of a signal is performed by conducting Fourier transformation or Hilbert transformation.

$$g(t) \leftarrow FT \rightarrow G(f), \tag{3.1}$$

$$G(f) = A(f)e^{i\varphi(f)}, \tag{3.2}$$

$$e^{ix} = \cos x + i \sin x, \tag{3.3}$$

$$i = \sqrt{-1}. \tag{3.4}$$

Two parts explain the complex signal: a real component and an imaginary component. The real can be explained in polar coordinates by the signal amplitude, where the phase(θ) represents the imaginary portion of the signal. The amplitude spectrum, which indicates the magnitude of the signal, and this is retrieved from the the product of both the real and imaginary components, given by formula $\sqrt{\Im(G(f))^2 + \Re(G(f))^2} = \sqrt{A(f)^2}$. However, what would be of notable importance in this paper is the principle of instantaneous phase, also known as the signal's local phase; this is estimated from the complex part of the frequency transformed signal. Inserting the definition from Euler's formula (3.3) into (3.2), observe that the Fourier transformed signal consists of an imaginary and a real part, namely:

$$\Im(G(f)) = i \sin \varphi(f), \tag{3.5}$$

$$\Re(G(f)) = \cos \varphi(f), \tag{3.6}$$

$$G(f) = A(f) \times (\cos \varphi(f) + i \sin \varphi(f)), \tag{3.7}$$

To determine the phase spectrum, it is needed to convert it by scaled conversion:

$$\varphi(f) = \tan^{-1} \frac{\Im(G(f))}{\Re(G(f))}. \tag{3.8}$$

(Weibull, 2020) As a result of this conversion, the phase will be wrapped/ contained within the interval $\langle -\pi, \pi \rangle$, and the phases cannot exceed these limits. For this reason, visually, there would not be possible to observe the actual form of the real signal by human eyes. The phases will appear as a signal that continuously cycles and contain discontinuities between the maximum and minimum phases in which it is wrapped(Vesnaver, 2017).

Equivalently when retrieving the instantaneous phase from the signal, there is required to transform the signal from being expressed by the frequency domain to the temporal domain. When considering the temporal domain, the analytical signal($F(t) = x(t) + iy(t)$) is a prerequisite. Likewise, to its frequency counterpart, the analytical signal is also expressed by the same amplitude spectrum, yet its imaginary part is 90° rotated compared to the real part. The signal can be explained in its exponential form($F(t) = M(t) \exp i\Phi(t)$), which is readily explained by the

instantaneous amplitude($M(t)$) and instantaneous phase($\Phi(t)$).

$$M(t) = \sqrt{x^2(t) + y^2(t)}, \tag{3.9}$$

$$\Phi(t) = \tan^{-1} \frac{y(t)}{x(t)}. \tag{3.10}$$

(Buland, 2021)

3.2.2 Phase Unwrapping

The concept of phase unwrapping is employed in various domains where the use of signals is of utmost importance. These domains include seismic, InSAR, MRI, SAR, Synthetic aperture sonar, optical, and microwaves.

The instantaneous phase is wrapped within a set interval that the phases cannot exceed and contains multiple discontinuities between this interval, constantly shifting. The phase unwrapping method unwraps the phases, which means that the phases will not be contained within $\langle -\pi, \pi \rangle$. In practice, the phase will appear in, for example, the seismic cube as a continuously increasing monotonic phase. Other implementation purposes, such as recreation of images through phase unwrapping, will appear to have more texture. The phase unwrapping procedure supported by quality guided or other methods to recreate the phase to account for multiple dimensions (3D) can outline the very texture of the content in the image (Gens, 2006).

The unwrapped phases would not be continuously increasing but will subtract or add 2π to the phase to unwrap the phase and reveal more texture in the phase. Such that texture within the unwrapped phase(ϕ) can be detectable by human eyes. For this thesis, what is needed is a continuously increasing phase that a dynamic time warping algorithm can connect. There is only a need to unwrap the entire seismic cube roughly. Then the results from the warping algorithm will make appropriate adjustments to the unwrapped cube. However, the equation for phase unwrapping is the function of the wrapped phase(φ) and addition/subtraction of 2π, depending on whether the difference between two points exceeds the threshold of π(Gauriau, 2015a). There is a matter of abiding by the Ioth condition (eq: 3.12), the condition describes the stability of the absolute phases(ϕ_U),i.e., prohibiting the function from becoming ill-posed. Multiple circumstances can affect the stability of the function, such as random noise in the signal. Objective discontinuities that correspond to faults or unconformities in seismic, if not handled correctly discontinuities diverge from the stability criterion. Requires an algorithm that effectively mitigates growing error propagating when encountering these structures. Additionally, under-sampling in signals contributes to a violation of the Ioth condition, also known as Shannon's law (Wu and Zhong, 2012b).

$$\phi_U = \varphi(x, y) + 2\pi k(x, y), \tag{3.11}$$

$$|\phi_U| \leq \pi, \tag{3.12}$$

$$\phi = \phi_U(t_0) + \int_{t_0}^{T_f} \phi_U(t). \tag{3.13}$$

(Wu and Zhong, 2012b) The function of k is given to be just integer values, either 1 or -1 depending on whether exceeding the threshold, and this will update the unwrapped equation. There is a matter of error propagation within the unwrapping algorithm if there are found parts within the seismic data set where errors are present. That,

in all real cases, will be an error associated with the gathered data. Often called Gaussian error, the Bayes error rate is an aspect one must be aware of to eliminate the errors propagating through the whole data set. The various phase unwrapping methods consist of two major classes; global implementations are path independent. Its singular purpose is to identify and isolate the location of $\Delta\varphi$, which leads to grave error propagation. Local implementations are said to be dependent on the phase in which the unwrapping is conducted, which will account for various changes from point to point in the phase (Gens, 2006).

3.3 Dynamic Time Warping

3.3.1 Dynamic Programming

Dynamic time warping is a method built for comparing different time series, and this method can prove helpful in comparing traces in order to map the horizons of the seismic stack. Typically similarity measurements are conducted by Euclidean distance(see eq: 3.14) and absolute distances. DTW can be asserted as an extension of these measurements. Dynamic time warping is more flexible than Euclidean distance (see fig: 3.3), enabling comparing time series of different sizes. Additionally, the flexibility of the method opens possibilities to match time series with similar events occurring at different positions in time; this behavior is essential compared to time series (see figure 3.3).

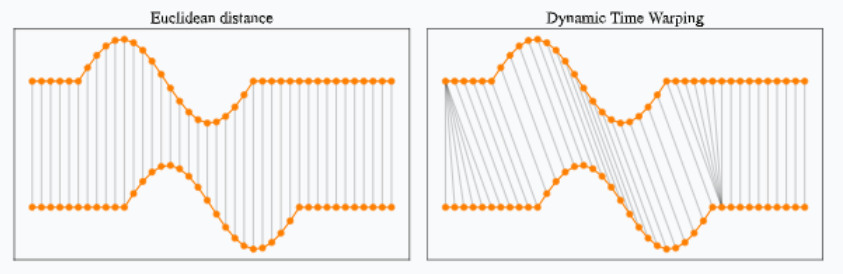
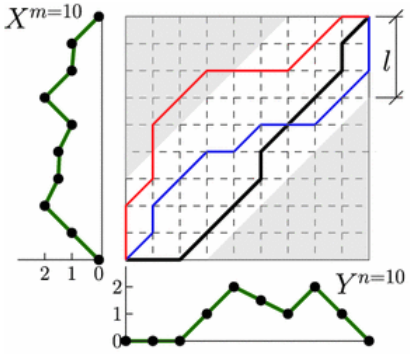


FIGURE 3.3: Difference eclidean and dynamic time warping (Tavenar, Romain, 2022)

$$Distance = \sqrt{(x-y)^2}. \tag{3.14}$$

The similarity measurements found at each pointwise indices amongst the compared signals are populating a matrix known as the cost matrix. The cost will compare the differences between the time series and track the distance between these series in the data space. In allowing a more flexible alignment between the signals, it is necessary to compare how much each data node differs by its value, measuring the similarity between shape and magnitude between series to fit its counterpart in the opposite time series. The cumulative distance matrix (see figure 3.4b)) represents the Euclidean distance between every position in both signals, and if not, local constraints are imposed on the warping procedure, limiting how many positions are to be compared.



(a) Time series alignment.

11	7	11	4	5	3.5	2	3	1	0
11	11	11	3	4	2	1	2	0	1
10	10	10	3	3	1.5	1	1	0	1
9	9	9	3	2	1	1	0	1	3
7	7	7	2	2	0.5	0	1	1	2
6	6	6	2	1	0	0.5	1	1.5	3
4.5	4.5	4.5	1.5	0.5	0	0.5	1	1.5	3
3	3	3	1	0	0.5	1.5	1.5	2.5	4.5
1	1	1	0	1	1.5	1.5	2.5	2.5	3.5
0	0	0	1	3	4.5	5.5	7.5	8.5	8.5

(b) Cumulative distance matrix.

FIGURE 3.4: Illustrative image of the dynamic time warping optimal path is being assessed (Oregi et al., 2017a)

Dynamic programming enters the concept of dynamic time warping by tracking the minimum cost of previous steps across the cost matrix. Several step schemes are constructed that enable a diverse portion of choices in estimating dynamic time warping. Nevertheless, the different schemes bring the flexibility of variants, and it changes how accurate the result of the algorithm is and its time complexity. A version of DTW researched by Oregi et al. (2017b) and shown in figure 3.4 formulates a way to be more memory efficient and reduces time complexity by definition of an On-line DTW algorithm that utilizes Frontiers. A Frontier helps to forget the previous options in the cost matrix when an optimal path has been found for one particular step in the cost matrix, shown in fig 3.4b). The optimal path is displayed in grey, and the matrix's Frontier is shown in green for that particular step. Information outside the boundaries of these Frontiers is neglected and forgotten because DTW obeys the law of continuity. This law indicates a warped path cannot cross previously made relations between signals, which means the information outside boundaries will not be utilized in the general DTW as well. Hence, this version neglects to compute the nonplausible paths. The work of Oregi et al. (2017b) is an example of research focused on reducing the overall toll and time spent computing by the computer, especially when conducting operations on streaming data. Important to recognize that various DTW schemes acquire different goals by differing from the standard procedure. The graph-cut methods might help to increase accuracy, yet it might increase time complexity (more in chapter: 8.3.2,8.3.3, and 8.3.4).

Algorithm 1 Euclidean Distance Matrix

Require: x - Signal trace 1 of length N, y - Signal trace 2 of length M

```
Populate Cost Matrix
for i = 0, 1, ..., N do
    for j = 0, 1, ..., M do
        EuclideanDistance[i,j] = sqrt((x(i) − y(j))^2);
    end for
end for
```

Firstly, a fully calculated matrix measuring the Euclidean distance between every point at both signals is to be tracked in algorithm 1. The algorithm can also be implemented to estimate the distance for only singular distances between the traces, namely $\sqrt{(x[i] - y[i])}$. It would be preferred to be implemented for an on-line euclidean estimation if there are to be induced global constraints or local constraints

Algorithm 2 Dynamic Time Warping

Require: N - Length of signal 1, M - Length of signal 2, Distance - Euclidean distance matrix ($N \times M$) of the signal 1 and 2

Initiate
Cost[i,j] $= \infty$
Cost[0,0] $= 0$
Populate Cost Matrix
for $i = 1, 2, \ldots, N$ **do**
    for $j = 1, 2, \ldots, M$ **do**
        minCost = argmin(Cost[i-1,j-1], Cost[i-1,j], Cost[i,j-1])
        Cost[i,j] = EuclideanDistance$[i, j]$ + minCost;
    end for
end for
$i = N, \quad j = M.$
Initiate Empty List PathList
Backtracking optimal path
while $i > 0 \quad and \quad j > 0$ **do**
    $\text{minCost}_{i,j} = \text{argmin}_{i,j}$(Cost[i-1,j-1], Cost[i-1,j], Cost[i,j-1])
    PathList.Add((minCost$_i$, minCost$_j$))
end while
return Cost ▷ Cost Matrix
return PathList ▷ Warped Path

onto the path alignment process (see chap 3.3.2 and 3.3.3). Instead, in estimating the full euclidean matrix, where parts of the matrix will not be used because of the induced constraints. There is a possibility to reduce the number of unnecessary computations. Dynamic time warping is initiated by ∞ such that if constraints are included, the left-out areas would not be selected as an optimal path due to the sheer magnitude of the value. Where inside the constraint, the cost matrix will be populated by the Euclidean distances and the minimum step in the chosen step pattern (seen in algorithm: 2).

$$D[i, j] = d[i, j] + \min \begin{cases} D[i-1, j-1] \\ D[i-1, j] \\ D[i, j-1]. \end{cases} \tag{3.15}$$

Backtracking the optimal path after the cost matrix D is found in the algorithm: 2, it is necessary to reverse the matrix and then start from the coherent endpoint of both sequences, namely position n and m in the matrix, and track the path which yields the minimum cost back to the origin of the signals namely position $[0, 0]$. Some fundamental rules have to be obeyed in finding the minimum path:

- 1) **Monotonicity**: Any adjacent two elements in the warping path W, $w_k = (w_i, w_j)$ and $w_{k-1} = (w'_i, w'_j)$ are confined to obey $w_i - w'_i \geq 0$ and $w_j - w'_j \geq 0$. The found indices cannot decrease, only remain stationary or increase in index.
- 2) **Continuity**: Any adjacent element in warping path W obeys the rule $w_i - w'_i \leq 1$ and $w_j - w'_j \leq 1$, which means the backtracking of the path can maximum increase by 1 step in either the direction, which is not decreasing.
- 3) **Boundary**: Starts from top-left $(0, 0)$ and ends in bottom-right (n, m).

(Al-Naymat et al., 2012)

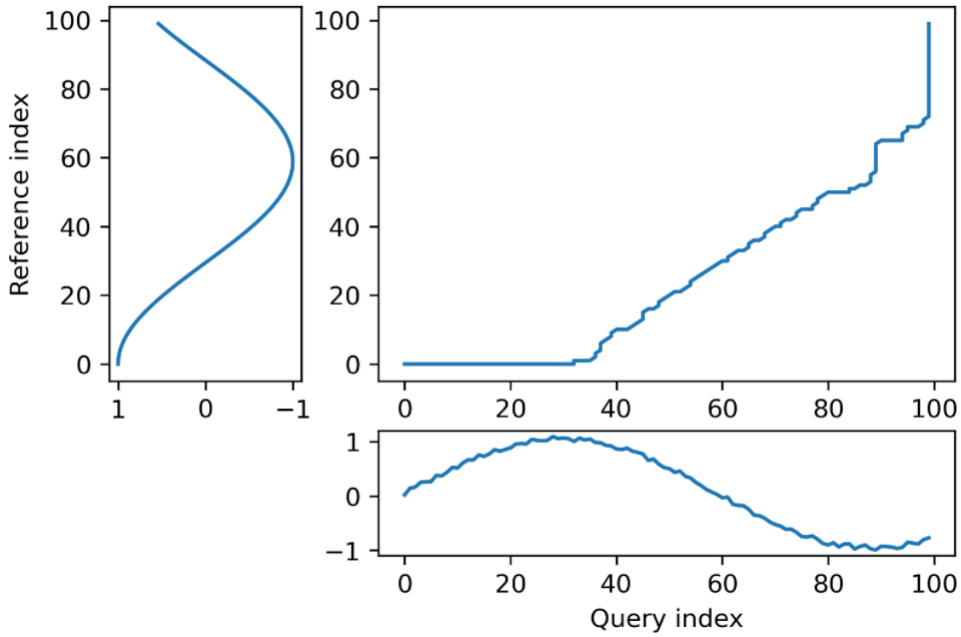


FIGURE 3.5: Illustrative optimal path is being tracked between two signals (Giorgino, 2009a)

Considering whether the global optimal path through the matrix between the boundaries (mentioned in 3.3.1), it is assessed by: $DTW(X,Y) = min\{\frac{\sqrt{\sum_{k=1}^{K} w_k}}{K}$. Since multiple warping paths can satisfy the given conditions but does not necessarily mean that the given path finds the global minimum. Since there are multiple options, there is a matter of chance of whether the dynamic path found at each step is converging towards the global optimum or not. Cross-validation of every possible path determines its global optimum, which would increase time complexity in the process, similar to brute force estimation. To some extent, DTW can be called a greedy algorithm, which means it does not always yield an optimal solution, although being afflicted by local influences. The implication states that the local least costly steps do not necessarily build-up to the globally optimal path, namely the recursive local steps. There are multiple options for enhancing DTW that either iteratively refine the warped path until it yields a global optimum (see appendix 8.3.4). Alternatively, other methods directly set restrictions on deviation from the diagonal or the slope(chap 3.3.2 and 8.3.1) that do not necessarily yield a global optimum.

3.3.2 Sakoe-Chiba Band

Introducing a global band reduces the computational time necessary to populate the whole matrix fully. It leads to changing results compared to when no band is active. The band functions are to determine a maximum deviation from the diagonal of the matrix(see fig 3.6), and the diagonal illustrates perfectly linear alignment between the two signals. Perfectly linear means that when backtracking, the recursive optimal path would always select $D[i-1, j-1]$ as the least costly step through the whole matrix. In the end, if all paths are linear, the least cost is situated at the diagonal. The deviation is induced as a constraint on the warping path $w_k = (i,j)_k$ by a defined deviation such that $j-r \leq i \leq j+r$. Outside this defined constraint, everything is set

to positive ∞ and would not be considered in a matter of finding a path optimum. Since the band has such a hard-defined constraint, it dampens the possibility of finding the global optimum if the actual optimal path is confined outside this band. It can be determined that the choice of the band's window width significantly affects the returned result. Thus, it might not return an optimal result because it is heavily affected by the window choice. For this reason, it speeds computation, but the gained efficiency has a cost; domain knowledge is required to select a proper window size using Sakoe-Chiba Band (E. Keogh and Ratanamahatana, 2005).

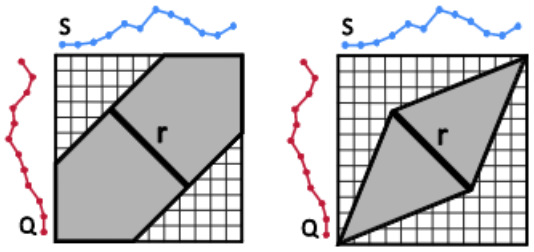


FIGURE 3.6: Illustrative difference between Sakoe-Chiba and Itakura parallelogram bands (Geler et al., 2019)

3.3.3 Step Pattern

Dynamic time warping is a method that allows many flexible adjustments and variations of the time-warping process to be included. One of the parts of the theorem with the most variations is the diverse topic of step patterns. Step pattern is a local constraint that affects how the warping path is determined and how the cost matrix is populated. Various step implementations (see fig 3.7) can define which of the previous steps is to be backtracked to the origin. The various schemes can enhance the number of steps required to find an optimal path and allow comparison by more data nodes to find the minimum step between these steps in updating the warping path. Assessing more data nodes allows a different route estimation because more points are allowed to populate the matrix and its backtracking process. The standard step pattern, which is a minimum in the basic DTW algorithm, is:

$$\min \begin{cases} D[i-1, j-1] \\ D[i-1, j] \\ D[i, j-1]. \end{cases} \tag{3.16}$$

The simple pattern allows evaluation of a minimum of the three previous steps; in most cases, this is sufficient. However, there is a possibility that one of the other step patterns (shown fig: 3.7) might yield a more accurate warped path. When the step scheme is selected, there is a prerequisite that the decision is based on domain knowledge regarding the current signal (Spooner et al., 2017a).

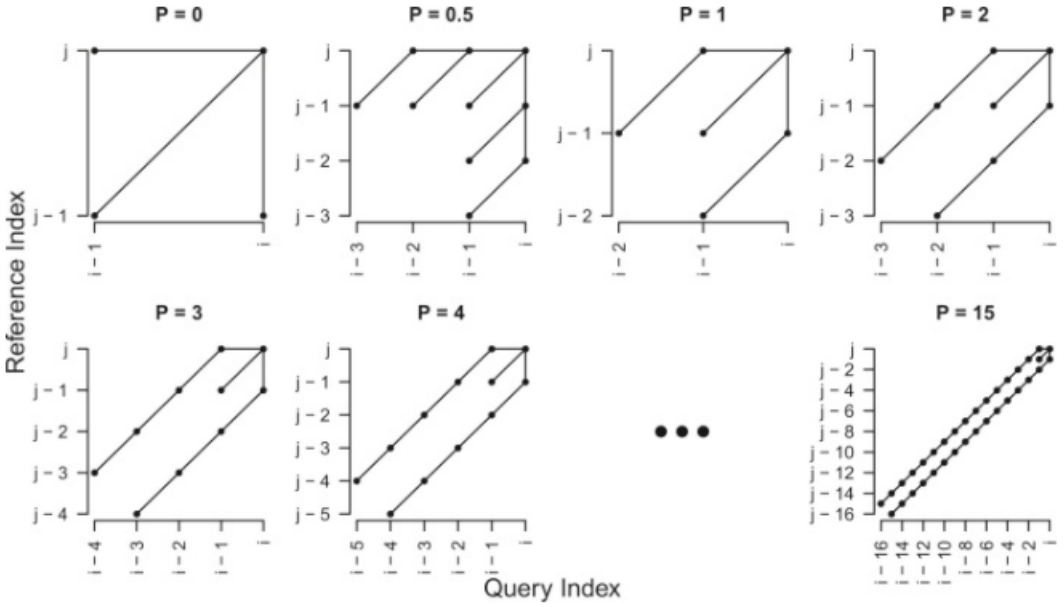


FIGURE 3.7: Step pattern examples (Spooner et al., 2017b)

One step pattern which particularly popular is the Rabiner-Juang step pattern (shown in Figure: 3.7 P=2 and in Eq:3.17), which is known to have a more specific slope and weighting constraint on local variations. Since each step pattern can affect the optimal path negatively or positively, there is no given best pattern for any given situation. However, it can be further explored if different patterns were cross-validated to a specific problem to assess the best pattern for the situation.

$$\min \begin{cases} D[i-3, j-2] + d[i-2, j-1] + d[i-1, j] + d[i, j] \\ D[i-1, j-1] + d[i, j] \\ D[i-2, j-3] + d[i-1, j-2] + d[i, j-1] + 0 \cdot d[i, j]. \end{cases} \tag{3.17}$$

(Giorgino, 2009b)

3.3.4 Derivative DTW

As mentioned, dynamic time warping allows various adjustments or versions of the method; one such version is Derivative Dynamic Time Warping (DDTW). The method is specifically established to counteract pathological results that might occur in standard DTW procedure, often detected by unreasonable pathing between signals. DDTW will instead try to establish gradients of each signal to populate the cost-matrices before step patterns occur. A comparison of DDTW and standard DTW performance is shown in fig 3.8.

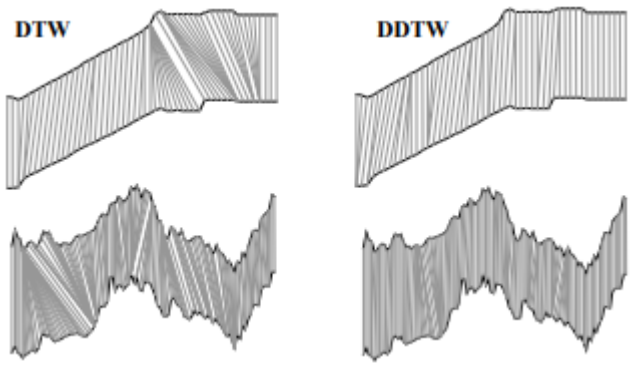


FIGURE 3.8: Example pathological relations DTW and its DDTW counterpart (E. J. Keogh and Pazzani, 2001a)

$$D_x(X) = \frac{(x_i - x_{i-1}) + (\frac{x_{i+1} - x_{i-1}}{2})}{2}, \tag{3.18}$$

$$D_y(Y) = \frac{(y_i - y_{i-1}) + (\frac{y_{i+1} - y_{i-1}}{2})}{2}. \tag{3.19}$$

Equations (3.18) and (3.19) find a gradient trend in both specific signals, which are the average slope of each signal; the method is particularly resistant to outliers. As these slopes are found, these will be measured against each other to compare the similarity between the signals. Hence, the Euclidean comparison is replaced by a comparison between these gradients $D(x_i, y_j) = |D_x(x_i) - D_y(y_j)|$. The result of the transition proves to be a more robust method against outliers and often finds more well-behaved warped paths. Robustness against outliers implies resistance toward random noise (E. J. Keogh and Pazzani, 2001b).

3.4 Recursive Subdivision

3.4.1 Tree Traversal

There has been an attempt to disperse the order of execution of the dynamic time warping algorithm since there was a need to define the algorithm independently yet able to aggregate the currently assessed trace satisfactorily. A possible solution is to employ a tree structure and proper traversal algorithm to split the problem into sub-problems. The division into multiple sub-problems coincides with the principle of Divide-and-Conquer. Moreover, the method is used in this situation to assure correct aggregation of seismic traces. Binary search is of particular interest, where it is necessary to continuously split the tree into child nodes from the middle of a seismic section. The middle of the whole domain would be known as the root node, and the child nodes will be $\text{childNode}_{\text{index}} = \text{parentNode}_{\text{index}} \pm \frac{\text{parentNode}_{\text{index}}}{2}$, from the first iteration the two topmost(not root node) child nodes are found (see fig 3.10). The same division is recurrently initiated as new, more possible child nodes are determined from the root node. Subsequently, the child nodes become the new root nodes and repeat until no linked child nodes are found. There is a necessity to describe the two known traversal algorithms, namely breadth-first search and depth-first search, to explain how traversals find bottom/leaf nodes differently.

A Depth-First Search algorithm will attempt to directly find the path toward the target node by searching through the depth-first, as the name suggests. By chance,

the algorithm might find the shortest path to the target quite early, especially in a binary tree where it can be guided toward the correct child node. Guided traversal, in this case, means that indices are used for splitting nodes if the target index is higher than the current root node. Then the traversal is guided towards the addition to the child node($\text{childNode}_{\text{index}} = \text{parentNode}_{\text{index}} + \frac{\text{parentNode}_{\text{index}}}{2}$), and the inversely if target lower than root node. However, if a similar traversal is applied to a graph structure, the objective is to find a particular graph position in the least number of steps. The graphs have a problem as the number of possible routes and nodes increases, and the chance of achieving a poor random walk through the graph is high. The depth-first search algorithm is particularly sufficient if a reduced number of options are available and can somewhat be guided towards convergence. If more complex problems are to be embarked on, this approach could, by chance, find the target quickly. Nonetheless, commonly misses the target and repeats "bad" traversals in search for the target (Fiset, 2022).

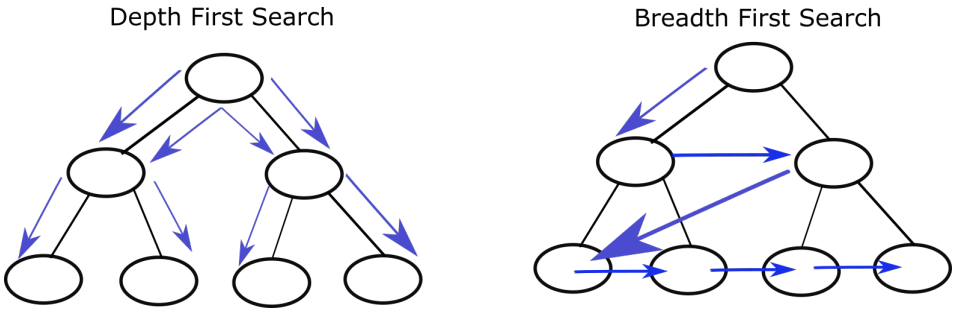


FIGURE 3.9: Difference between Depth First Search vs Breadth First Search traversals

Breadth-first search (BFS) searches first in the width of the tree before traversing towards the next level (also known as depth). When considering the width of each level in the tree at first, more control is related to how efficiently the search algorithm performs. Through BFS, it is usually implemented to constantly populate a whole recursive layer before the next layer is populated. The traversal is continuous until the target node has been met or all leaf nodes are evaluated. A conclusion can be made considering that the target node is out of reach from the tree or graph being traversed. The procedure is likely to not embark on the shortest route toward the target node directly but incrementally verifies that the target node is not in the current level before traversing to the next level (see the right figure in fig 3.9). Intuitively, at first, one might assume a directly shortest path approach would find the target node fastest, yet this is true for a binary directed tree structure. However, this is not the case in a complex system. An iterative approach through each successive layer will find the target (if possible) in a faster, more controlled approach. The main difference between the two traversals is that DFS traverses randomly to the very bottom of the tree (see the left figure in fig 3.9). It might signify that the search procedure fails multiple times before finding the target in the end. Moreover, as the DFS initiates random paths, a defined search pattern will guide the BFS algorithm. That is, finishing one layer, initiating a search from right towards left, and repeating for each level (see the right figure in fig 3.9). (Fiset, 2022)

3.4.2 Recursive Subdivision and warping paths

The reason for explaining the two standard traversal methods (DFS and BFS) is rooted in the fact that these algorithms can help construct a complete binary tree.

Algorithm 3 Binary division

Require: TraceArray - An array which contains trace indexes, start - start index, end - end index

1: Defines algorithm that constructs a tree structure constantly splits the domain into half until leaf nodes are reached.
2: **if** $start \geq end$ **then**
3: return null
4: **end if**
5: $mid = start + \frac{start+end}{2}$
6: $node = newNode(TraceArray[mid])$
7: $node.left = Binarydivision(TraceArray, start, mid - 1)$
8: $node.right = Binarydivision(TraceArray, mid + 1, end)$
9: return node

Additionally, also find the recursive path to a singular trace. Such as shown in 3, the concept of BFS is an underlying the whole algorithm in order to construct a complete tree for all of the nodes until the further division of the domain is not possible. It is necessary to explain the **node** variable, which consists of the node value plus the linked child nodes, namely the left and right child nodes. Repeatedly the child nodes left, and right child nodes are linked together. This definition will ensure that all left and right child nodes are found at each level. If related to the DTW issue shown in fig 3.10, then DTW shifts($DTWX - X$) are defined by the structure established by BFS (in algorithm 3).

However, there remains one problem when it comes to how determining RGT individually (more discussion in chapter 4 and 5). Each specific seismic trace must be given as a leaf node at the bottom of the trace-pair structure, which BFS already defines. Utilization of the DFS schema in a binary structure was a proper way to insert trace indices in the correct node position, such that the multiple DTW pair shift relations are directly coupled to each seismic trace. Each layer of trace pairs will contain the index of the leftmost trace in the trace pair. Suppose the target node is found at the current trace-pair index. In that case, the traversal will be conducted by one step to the left and the rest toward the right child nodes in the subsequent recursive layers until null is encountered, and then a singular target trace node index will be inserted. The cause is to enable DTW-pairs shifts to be accounted for and successfully backtrack to a single node index that corresponds to a single trace rather than shifts between traces. If the target node has an index offset of +1 from the current node, traverse one step right and the rest of the steps toward the left to find the correct position to insert the trace node index. This procedure is case adjusted variant of the Binary Search Tree. This procedure constructs a fully binary tree with a recursive layer that can illustrate a relationship between the different DTW pair relations linked to each trace. This was the idea for attempting to use the trace information independently of its neighboring update steps.

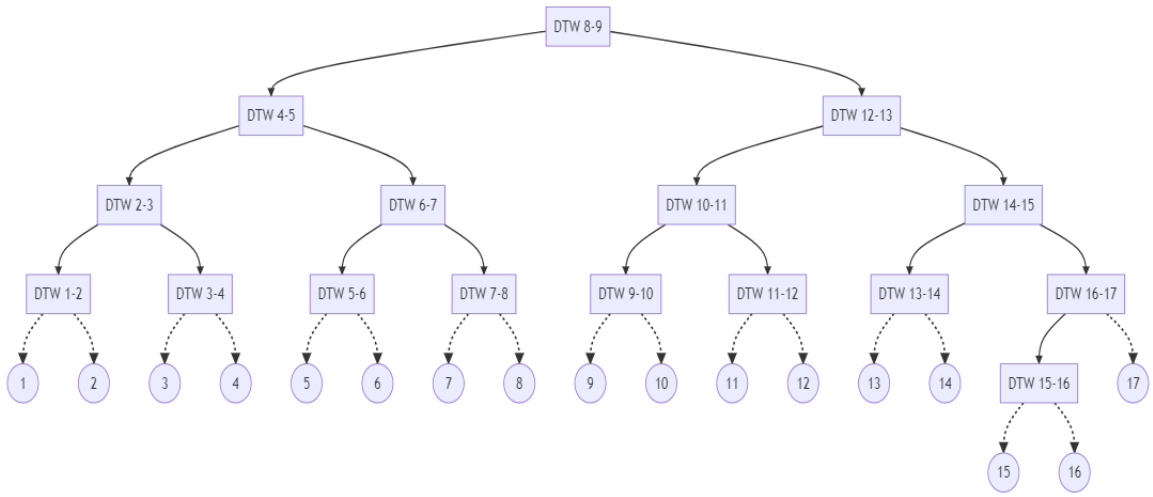


FIGURE 3.10: Illustrative image of splitting DTW process into a binary tree structure

$$RGT_{AGG} = RGT_0 \pm DTW_{level0} \pm DTW_{level1} \pm DTW_{level2} \pm DTW_{level3} \pm .... \quad (3.20)$$

The idea of independent alignment is that the consequent order of RGT updates does not depend on which direction estimations are conducted. That means the sequence of trace aggregation could be entirely random and yields the same outcome no matter the choice of sequence. Equation 3.20 explains that RGT is controlled by its coupled lags found by the assessment of different trace pairs in a binary tree structure (see figure 3.10). Whether the lag is added or subtracted from the initial RGT volume, RGT_0 depends on what side of the parent node the path found. This applies from the top through all-recursive layers on the path to each singular trace located at the bottom. The shifts from the specific root node will be added if traversal towards the right child node is conducted. On the contrary, if the traversal is guided toward a left child node, then the trace-pair shifts will be subtracted in equation 3.20. An example when using figure 3.10, is that trace 9 is being updated, then: $RGT_{AGG9} = RGT_0 + DTW_{8-9} - DTW_{12-13} - DTW_{10-11} - DTW_{9-10}$.

3.5 The Method

3.5.1 Avary

The Workflow

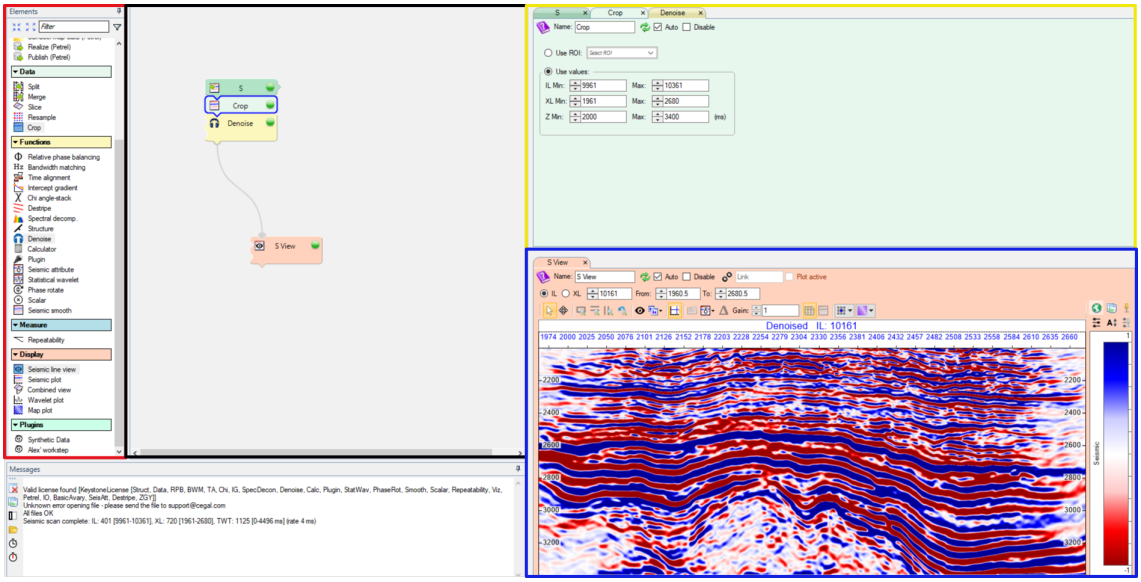
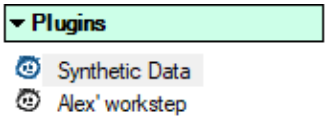


FIGURE 3.11: Avary User Interface. Red box: Native Avary operations. Black box: Workflow definition. Yellow box: Settings for operations included in the workflow. Blue box: Seismic line view.

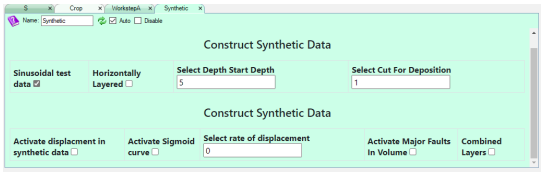
The main engine for reading and processing seismic information through various processes is conducted by Cegal's Avary, which readily utilizes high-performance computing concepts. Avary initially slices the data to a reduced size for efficiency purposes. This efficiency benefit enables the assessment steps of parts of the information to be more rapidly conducted before the process is executed on the entire data set. The required workflow can be explored and finalized before all of the present data is processed in a workflow. The initiation of operations is first activated if a workflow is connected to the seismic line viewer (shown in the Black box in fig: 3.11). The parts of the data set included can be controlled in the seismic line viewer section (shown blue box fig: 3.11), can be selected at which positions are going to be viewed, and in what direction (inline or xline).

The framework is built to tackle multidimensional data. Nevertheless, the seismic line viewer shows only a slice of the whole cube because the correct procedure can be achieved on the part of the information. Determines whether the operation shall be realized on the whole cube or tweaked before conducting on the whole volume. The whole framework applies the concept of divide and conquer, which describes that extensive operations can be divided into multiple smaller sub-operations. Hence, yield reduced computational time and makes many of the operations parallelizable. Eventually, all the sub-problems are aligned in a respective global manner, such that information of all sub-problems is merged correctly into a final output volume.

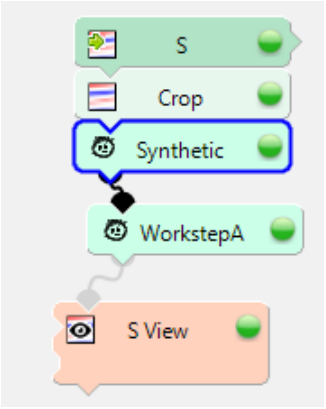
The Thesis Contribution



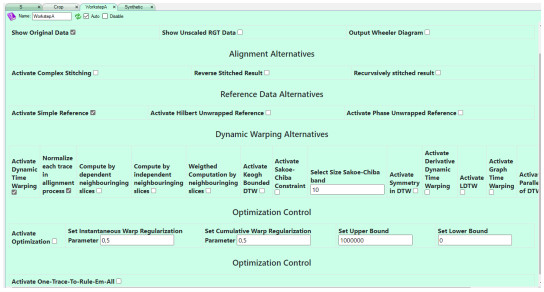
(A) Avary plugin worksteps



(B) Avary synthetic data workstep UI



(C) Avary plugin workflow



(D) Avary algorithm UI

FIGURE 3.12: Plugin workflow structure

The implementation of this thesis is accessed in Avary, similarly to the native functions accessed in the menu for operations that enable the function in the workflow space (see fig 3.11). Two different work step was created for the work, one that controlled all operations related to estimating an RGT volume by DTW(**Alex' workstep**). The second is for constructing synthetic data, similar size as the seismic data, set in which the work step(**Synthetic Data**) is attached to (observe figure 3.12). In the synthetic work step, there is built a relatively crude user interface (UI) to interact with the different coded options for synthetic data(shown in figure 3.12b). Not every implementation is shown(see fig 3.12), which provides a result. Some examples of horizontal layers, sinusoidal curvature, and simple triangular shapes are shown in chapter 4.

Moreover, a UI for the RGT operations is also added, yet a significant number of options are present in this simple UI example. Many of whom are not finalized as proper alternatives to the method, much of the alternatives are simply used for testing purposes. An example of an operation that is present but not operational is the "Optimization Control" option, which was an attempt to replicate the method described by Deriso and Boyd (2019a). Although this UI is quite messy in design, it serves the purpose of interactively activating the different attempted solutions, changing reference volume, or changing parameters, for example, the band-size in for Sakoe-Chiba Band(ch. 3.3.2).

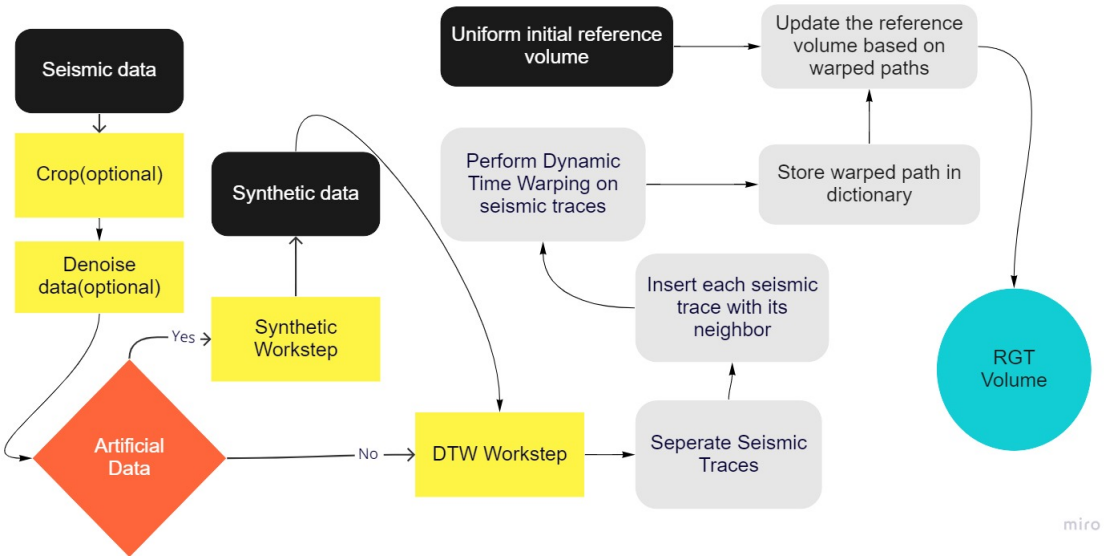


FIGURE 3.13: Overview of the whole workflow for the creation of the RGT volume based on Dynamic Time Warping using Avary

As for figure 3.13, it conceptually shows the procedure shown in figure 3.12. Moreover, it explains some of the algorithmic architecture of the Dynamic time warping process, which occurs within **Alex' workstep**, these processes are marked in grey in fig 3.13. Firstly, the version of DTW is selected in fig 3.12d, whether the operation employs the principle of standard DTW, Derivative DTW, or Sakoe-Chiba Band. Traces are split into neighboring trace pairs such that the DTW procedure gives two different signals/seismic traces, from which the warped path is determined. Ultimately, all the found warped paths are stored in a dictionary consisting of the current trace position(tuple of (X, Y)-positions), similarly to the neighboring trace, and a list of the warped path between these seismic traces. The length of the list of the warped paths has a variational length depending on the path selected by the DTW procedure. The objective of the assembled dictionaries of all the paths is to be used on a monotonically increasing homogeneous reference volume (shown in the chapter: 4). The RGT result is increasing monotonically and aggregating volume by the found alignments between seismic traces in the actual seismic data set.

3.5.2 Implemented Warping Algorithm

The applied method utilizes some of the Dynamic time warping techniques (explained in chapter 3.3) in combination with the division of the seismic cube into numerous trace pairs. The purpose of division into neighboring trace pairs is to consequently correlate the continuation of a seismic horizon from the end of the slice towards the opposite. Every vertical column, namely seismic traces, is processed sequentially, leading to a directional update pattern, left to right or right to left. The direction of processing are controlled by reversing the iteration of the seismic cube $(i/j = N-1/M-1, N-2/M-2, \ldots, 0)$, or kept forward$(i/j = 0, 1, \ldots, N/M)$ iterated. The direction of processing is more dependent on merging the warped path after the DTW alignment paths are found. However, in the standard merging procedure where the binary recursive subdivision is not employed, the sequence of merging will be controlled by the direction of processing DTW. The matter of the information passed into the dynamic time warping procedure in algorithm 4, the DTW step, is in the pseudo-code given as a generalized function. Because throughout this thesis,

several different versions of DTW have been implemented and activated to find the optimal path of the given trace, be aware that the "optimal" might not yield the globally optimal path in the current state of the algorithm. Further extensions of the native DTW procedure must be conducted such that a global optimum is found on every occasion of the function being used.

Algorithm 4 RGT warped path procedure

Require: Cube- Seismic data cube(N x M x K)
Initialize Dictionary RGT
Initialize Dictionary TracePair
Initialize List WarpedTracePair
Find Warped Path Between Trace Pairs
for $i = 0, \ldots, N$ **do**
  for $j = 0, \ldots, M$ **do**
    for $k = 0, 1, \ldots, K$ **do**
      if $i \quad != i_{prior}$ **OR** $\quad j \quad != j_{prior}$ **then**
        Initialize new instance of TracePair
        Initialize new instance of WarpedTracePair
        TracePair.Add(Cube[i,j,k], Cube[i + 1,j,k] OR Cube[i,j + 1,k]
      else
        TracePair.Add(Cube[i,j,k], Cube[i + 1,j,k] OR Cube[i,j + 1,k]
      end if
      $i_{prior} = i$
      $j_{prior} = j$
      if $k == K$ **then**
        WarpedTracePair = DTW(TracePair.Key, TracePair.Value)
        RGT.Add((i,j),Dict((i+1,j) **OR** (i,j+1) ,WarpedTracePair))
      end if
    end for
  end for
end for
return RGT ▷ RGT warped path dictionary

Many dictionaries are occurring in the pseudo-code for the algorithm 4, the primary purpose is only transferring the information found in each sub-step. The need for several dictionaries could be reduced if the warped path was employed directly on the volume (seismic volume). Moreover, the information must be translated to the reference volume, which warped path changes establish the final RGT volume. Also, the dictionaries are needed because the DTW procedure requires two full signals in the comparison. However, dictionaries in this method are only used for transferring information correctly between the processes. In the pseudo-code, the initialization of new instances of dictionaries means an empty dictionary is initialized such that new inputs of variational lengths are added. The feature of generic dictionaries was particularly needed when transferring the warped path indices from the DTW procedure. Because the warped path often had different lengths due to different path selections through the cost matrix.

The derivative DTW example is shown in pseudo algorithm 5, which readily replaces the Euclidean distance(algorithm 1). That populates the initial cost matrix, such that instead of comparing the difference in magnitude (by euclidean), there is now a matter of comparing the slope difference amongst the signals. Although, the

Algorithm 5 Derivative Dynamic Distance Matrix

Require: x- Seismic Trace 1(length N), y - Seismic Trace 2 (length M)
  Find Derivative from trace x
  for $i = 0, 1, \ldots, N$ **do**
    $D_x[i] = \frac{x_i - x_{i-1} + \frac{x_{i+1} - x_{i-1}}{2}}{2}$
  end for
  Find Derivative from trace y
  for $i = 0, 1, \ldots, N$ **do**
    $D_y[j] = \frac{y_{j+1} - y_j + \frac{y_{j+1} - y_{j-1}}{2}}{2}$
  end for
  Populate Cost Matrix
  for $i = 0, 1, \ldots, N$ **do**
    for $j = 0, 1, \ldots, M$ **do**
      Dist[i,j] = $\sqrt{(D_x(i) - D_y(j))^2}$;
    end for
  end for
  return Dist

DDTW extension is merely one example of extensions that potentially are implemented on the **DTW** step(algorithm 4. There are high potential to further extend DTW versions onto this step of the procedure because of its generalized definition, which will accept the warped paths retrieved from any version of DTW (see chap 3.3 and appendix chapters 8.3.4, 8.3.2 8.3.3). Throughout chapter 4 and 5, the observations support a claim that most of the included version of DTW provides satisfactory warped paths. However, some of them are thought to be more plausible than others, yet it is challenging to select a DTW extension to be more suited than the others. Problematic behaviors are occurring which is not related to the DTW procedure. Still, there has to be implemented a clever algorithm that uses the information from DTW in a satisfactory manner (more explained chapters 4 and 5).

3.5.3 Implemented merging procedure

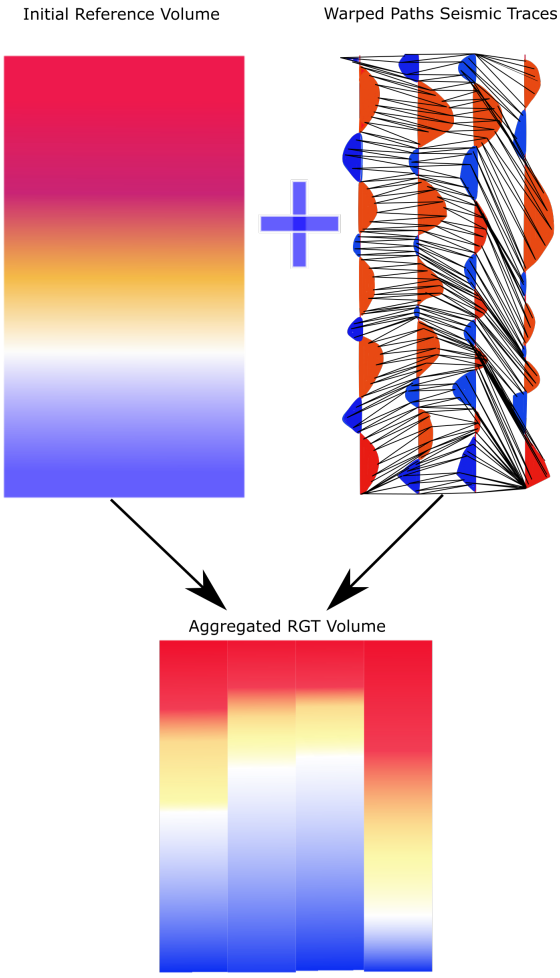


FIGURE 3.14: The merge procedure: Upper left: Initial reference volume which is uniform at all depths. Upper right: Example warped paths found by the use of DTW amongst seismic traces. Lower image: The aggregation of the initial reference volume based on found warping paths from DTW

The thesis aims to formulate a procedure establishing RGT volume by inserting seismic data as input to the function. It has raised the problems on how this volume should be established. The warped path extracted from DTW cannot directly transform the seismic traces from the inserted data set into a continuous-time volume; this requires a separate volume. The issue was resolved by constructing a uniform cube of similar size as the inserted seismic stack. As mentioned, it is uniform, meaning the magnitude of the estimates at every vertical position is similar in any of the vertical positions in the volume (see Upper left fig: 3.14).

$$\mathrm{RGT}[X^{i+1}, Y^{i+1}, \mathrm{WarpPath}_k^{i+1}] = \mathrm{RGT}[X^i, Y^i, \mathrm{WarpPath}_k^i]. \tag{3.21}$$

Furthermore, the next issue considers how this uniform volume can be aggregated to form an RGT volume consistent with the seismic traces and mimics the procedure of a geophysicist. The solution to this matter was to formulate the output of the DTW algorithm in a manner that extracts the warped paths amongst seismic traces consistent with the indexing of both volumes. Once all the warped paths are found

by comparison amongst seismic traces, the procedure of updating the estimates in the initial volume initiates, in equation 3.21. The definition of the procedure is to show for a single slice aggregation. The symbols in the equation describe the current trace position in the slice i. Depending on whether the slice is orientated in inline(Y) or crossline(X) direction, if orientation is in X-direction, then position in Y is kept constant during the aggregation, i.e., Y=0 for the whole aggregation. The index k is of variable length because the length of the warped path directly controls this, but the length of k can minimally be of length K. That is the number of vertical positions from seismic data. The RGT notation refers to the reference volume, and the indices used on the reference volume are gathered from the output dictionary from algorithm 4.

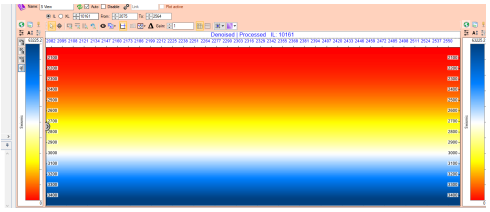
The updated methodology is, by definition, created so that the RGT traces are fully altered before the found warped path processes the neighboring traces. That means the volume is processed vertically subsequently throughout the volume, i.e., the current component(i) of $\mathrm{RGT}[X^i, Y^i, \mathrm{WarpPath}^i_k]$ is fully aggregated by the warped path at the time of aggregation of neighboring trace($i+1$). The transfer of already aggregated traces onto its adjacent traces is assumed to ensure that seismic horizons are interconnected, i.e., the seismic horizons are continuous. The propagation of information represents the information from previous traces in the domain through accumulation in the sequence of RGT traces. Hence, the outcome of RGT is afflicted by previous information from the direction of aggregation. Reversing the aggregation direction(mentioned in chap: 3.5.2) coincides with similar behavior, yet only reversing the procedure where $\mathrm{RGT}[X^i, Y^i, \mathrm{WarpPath}^i_k]$ is now the part being aggregated by the trace allocated to the right($i+1$) of the current trace(i)(see eq: 3.22).

$$\mathrm{RGT}[X^i, Y^i, \mathrm{WarpPath}^i_k] = \mathrm{RGT}[X^{i+1}, Y^{i+1}, \mathrm{WarpPath}^{i+1}_k]. \tag{3.22}$$

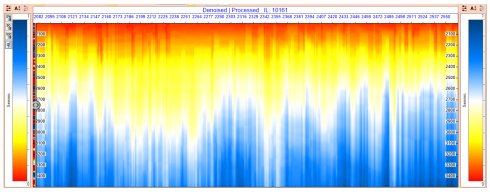
However, the reversing of aggregation direction, indicates different characteristics (emphasized in chapters 4 and 5). It is a prerequisite that an RGT volume would not be dictated by the orientation of the update when accumulating the final volume. More discussion on a relevant resolution on this matter in chapter 5.

Chapter 4

Results



(A) Uniform initial reference volume



(B) Reference volume constructed by Phase Unwrapping of seismic signal

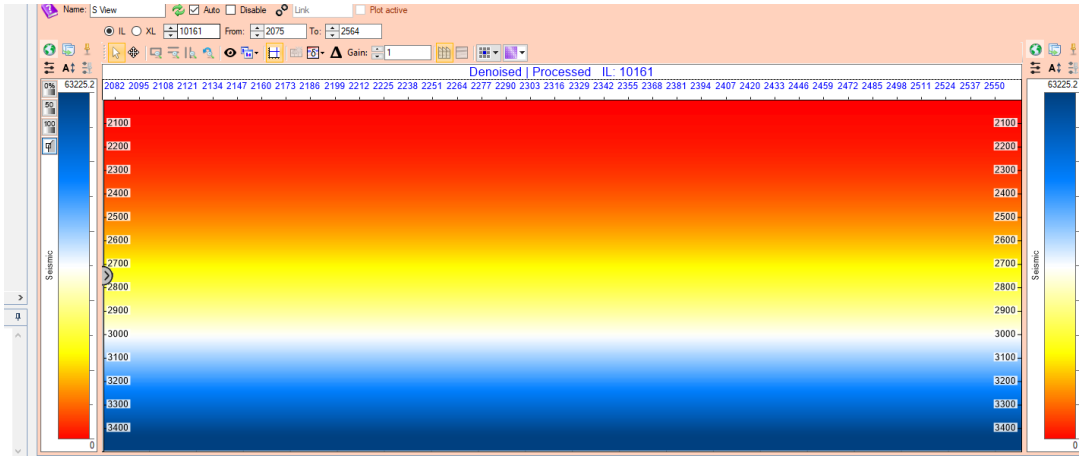
FIGURE 4.1: Difference between reference volumes

This chapter will test the method capacities onto first synthetic data(chap: 4.1) followed by real seismic(chap: 4.2). Synthetic is established to test simple cases and replicate more extreme behaviors. The method will be tested on real data to ensure the procedure's outcome is sensible when deployed onto a complex and intricate systems.

Firstly, clarification regarding how the algorithm behaves in evaluating information and making the necessary changes based on the gathered information is required. First information, or positions between traces, is found using the inserted seismic or synthetic data set; the information is stored in dictionaries for further use. An initial volume is constructed to construct a monotonically increasing volume from top to bottom; every increase in depth equals an increase of the floating number of similar magnitude as the increase in depth (see figure 4.1a). This division of tasks is required to ensure that an RGT volume is found because the evaluated data passed into the DTW algorithm is often constrained in cyclic phases from seismic. That will cycle between negative and positive magnitude by adjusting the coordinates in the uniform initial monotonically increasing volume, and it is possible to fabricate an RGT volume. A reference volume can be constructed equivalently based on the principle of phase unwrapping (see fig: 4.1b). That directly builds a monotonically increasing volume based on the signals retrieved from data. One of the initial hypotheses of this thesis was to conduct phase unwrapping and assure correctness by employing DTW. Whether reference volumes are preferred over the others is to be discussed in chapter 5.

4.1 Synthetic data tests

(A) Horizontal reference volume

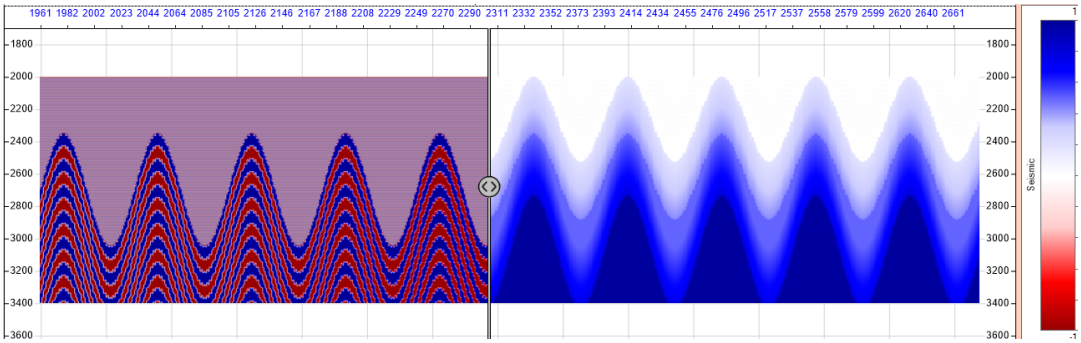


(B) RGT result

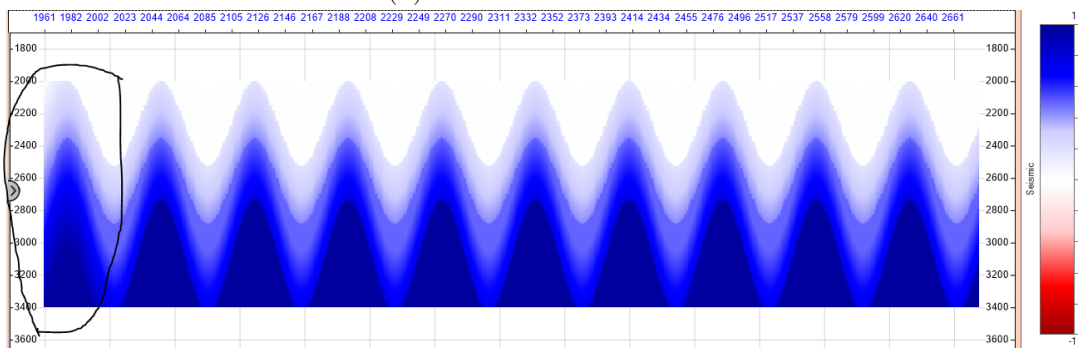
FIGURE 4.2: DTW performance horizontal layers

Considering perfectly horizontal synthetic layers, shown in figure 4.2. Observing the structure of this image, it is clear that every seismic horizon would be interpreted as perfectly horizontal, monotonically increasing time layers. That means there should be no shifts amidst any seismic traces in the slice. When assessing the outcome of the method, it satisfies the claim of being perfectly horizontal and monotonically increasing (see figure 4.2b). Again, the law of succession of layers supports the resulting outcome of the algorithm and interpretation. Given that there is no displacement present in the domain.

Furthermore, it is necessary to construct more extreme synthetic data to stimulate exploring the algorithm's performance. Since the outcome was satisfactory for slices that lacked displacements, it is required to test cases where extreme displacement is present. Because, in reality, it is improbable observing subsurfaces untouched by processes that cause displacements in the strata.



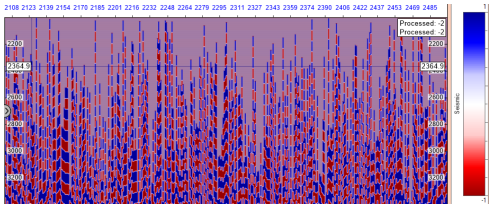
(A) Sinusoidal artificial data



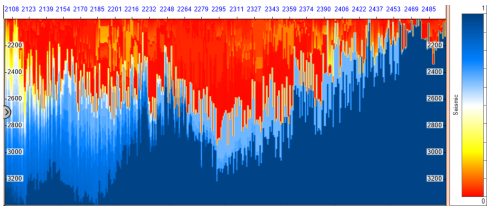
(B) RGT Sinus Wave

FIGURE 4.3: Sinusoidal wave and its RGT result

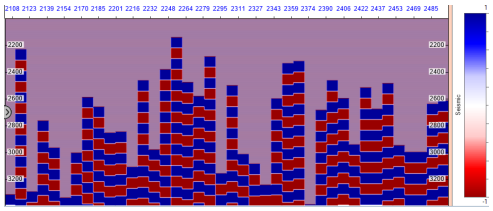
Folds are a central structure in geologic sections, and these are not as discontinuous as faults and unconformities. Figure 4.3a represents an image that readily simulates a pure fold structure throughout the slice. This data was generated by artificial defining a surface that followed a similar pattern as a sinusoidal wave. That can replicate the behavior of solid bedrock or a significantly higher density layer that accumulates sediments in its valleys. Which, in many cases, gives an impression of younger sediments deposited atop older. At the time of assessment of the results from 4.3 is started, there is at first a good behavior present in the interpretation of the seismic returned from DTW, namely a believable RGT correlation (see black circle fig: 4.3b). However, when more closely assessing the result, a transition is present toward more distinct boundaries (outside black circle fig: 4.3b) between each sinus wave peak. Emphasizes that inside the black circle, the RGT level appears to be more smeared out (continuous), whereas outside, there is a sudden increase in time(dark blue contour). Implications caused by this observation might indicate that only the oldest time is propagated in parts of the interpretation. It is expected that intermediate transitioning RGT levels should also be present, similar to inside the circle. Despite lacking information within layers, the overall structure appears to be satisfactorily interpreted. The interpreted structure outlines an identical structure as observed in figure 4.3a, which outlines a folding band throughout the whole domain. Since the overall structure is outlined perfectly, it could be claimed that the method can find proper warped paths in continuous structures. It might be needed to test whether the effect of discontinuities affects the interpretation of the method in any way.



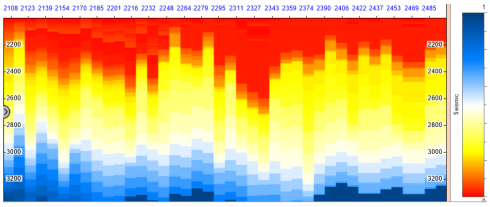
(A) Multiple independent faults



(B) Multiple independent faults RGT response



(C) Multiple compact independent faults



(D) Multiple compact independent faults RGT response

FIGURE 4.4: Comparison by different faulting layers and their respective RGT response

Continuing the objective of testing how the method behaves when encountering extreme behaviors, there was a certain need to test how synthetic faults are being interpreted. The result of the synthetic data was a case where faults were constantly being generated for each new trace position, see figure 4.4a. Additionally, in a case where a new fault was first randomly generated after a certain interval, such that some continuity can be determined from the structure, see figure 4.4c. The method's evaluation of how the reference structures differ in terms of performance reveals a significant variation in shifts between these. When observing figure 4.4b, it can be determined that a structure(constantly changing displacement) of this nature is not interpreted satisfactorily by the method. There is an indication that the faults are not allowed to fully converge when rapidly shifted after previous abrupt shifts that have occurred, i.e., the RGT value becomes to some degree an intermediate state between the adjacent shifts. Some intermediate states seem to accumulate throughout the domain, which builds up a significant contrast in RGT-levels in the subsurface. The structure observed in fig 4.4b are not similar to the structure that is seen in the reference data(see fig 4.4a). It appears that the method demands some relevance between each subsequent shifted layer. That means if there is constant unrelated shifting occurring, the method does not seem to be able to stabilize the interpretation. Thus, it shows the correct trend yet not being able to emphasize the magnitude of the real displacement.

By giving the method more time to determine the structure before the fault displacement occurs by employing a defined interval before random shift(see fig 4.4c). When carefully assessing the different results, artifacts can be observed in both cases in the left upper part of figure 4.4d and 4.4b. It can be determined that the method has significantly improved to actually replicate some of the structure(see fig 4.4d) visible from input data. However, similar to fig 4.4b, it seems that fig 4.4d also struggles to fully converge the structure when encountering extreme displacement in the data, i.e., very high value to very low value. Would this indicate that the DTW algorithm fails to achieve the optimal warped path? Is this a result of the markovian constraint, which forces intermediate steps to partake (discussed in chapter: 8.3.2)? This is not known, yet the outcome is that RGT levels are not shifted enough when encountering

extreme displacements. However, real seismic data would be more connected than this synthetic counterpart. Thus, it is expected that faults materializing in the real data would be manageable by the method to some extent.

4.2 Real data

The Data

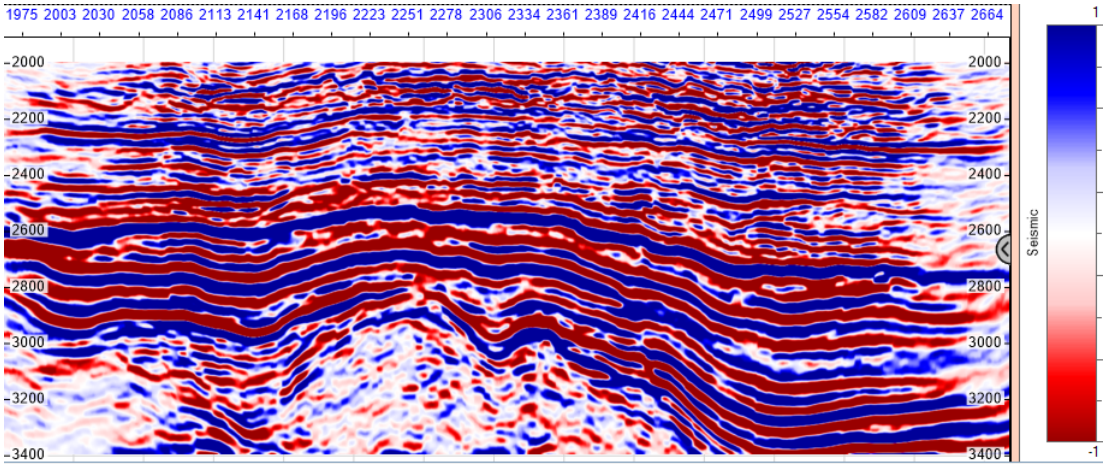


FIGURE 4.5: The reference volume Field 1 Crossline direction

This thesis does not provide the origin of data shown in figure 4.5, referred to as Field 1. However, the vertical cropping of the domain is restricted to the interval of 2000-3400 *ms*. The data contains multiple minor displacements that occasionally disconnect the horizons. An erratic zone appears in the image's center part that forces deformation on the overlying layers. From an interpretation point of view, it would be expected to account for expected to showcase these minor shifts in displacement in the interpretation. Additionally, the structure formed by the erratic zone must be outlined. However, the overlying layer is required to be clearly interpreted, and the structure located within the erratic zone is more challenging to interpret correctly.

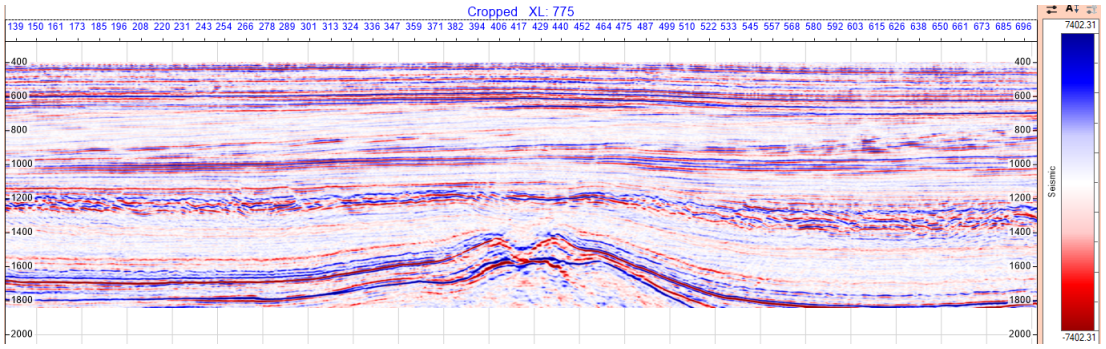


FIGURE 4.6: The reference volume Asterix Crossline direction

Seismic observed in figures 4.6 and 4.7 is retrieved from Cegal's Asterix data set. These fields are situated at a shallower depth than its counterpart in figure 4.5, at 300-1800 *ms*. Similarly, figure 4.5 shows an erratic zone at the center that forces underlying layers to compress their overlying strata. Lying upon the high contrast formation outlined at the bottom, appears to be a band of lower density layer(at

1300-1700 ms), which is present in inline and crossline directions(fig: 4.6 and 4.7). As expected, there are multiple minor faults situated in the slice, yet there is no presence of major regional faults. In interval 700-1100ms, structures appear to correlate with specific stratigraphic structures. Interpretation expectations demand outlining the structure at the domain's bottom. Additionally, it returns to a similar age level before and after the intrusion (in the center). The procedure must clearly trace the faulting horizons (shown at 1200-1300ms). The stratigraphic section mentioned is attached much promise to the interpretation of whether the automatic put emphasizes the pattern that can be observed.

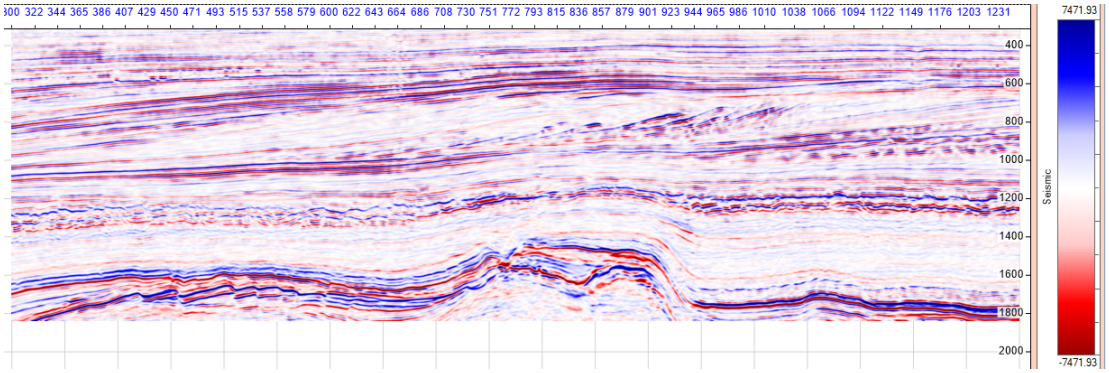
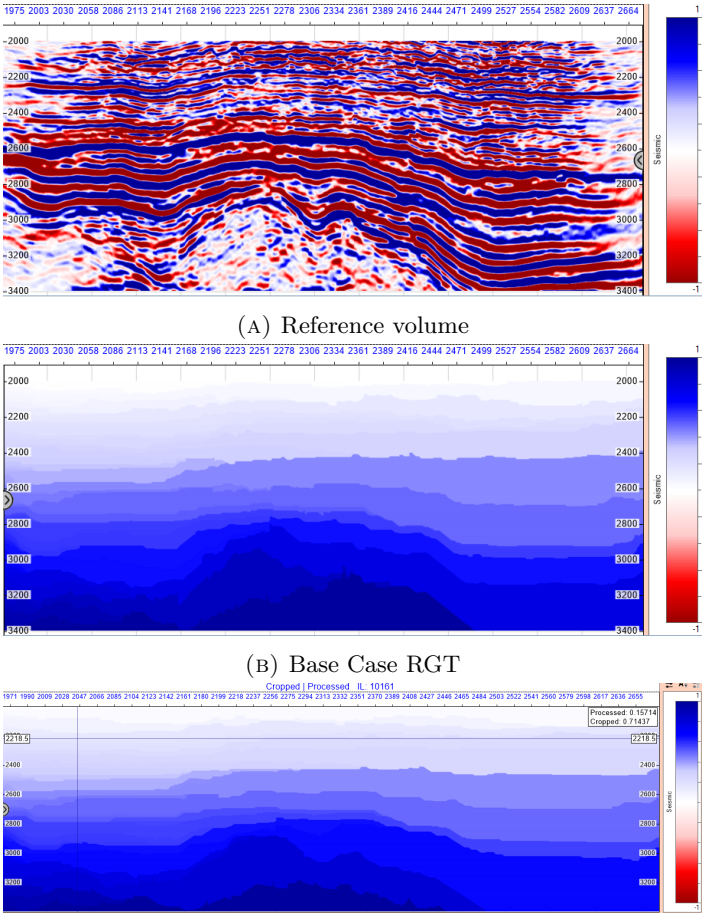


FIGURE 4.7: The reference volume Asterix Inline direction

When moving from crossline (fig: 4.6) to inline (fig: 4.7), many of the same features are present, yet some additionally details now becomes observable. Faulting blocks are now detectable in the high contrast bottom interfaces (left side of deformation), important to be outlined by RGT related to structural reservoir properties. The form of the central bottom structure is also present in this direction, yet the shape appears wider and less narrow. The most interesting feature detected by changing orientation is the presence of structures at 600-1000 ms. Immediately structures are coupled to the concept of stratigraphic sequencing (cover in chapter 2). If these structures are well interpreted, they can supply the Wheeler Diagram (chap 2.4) with promising information that contributes to predicting depositional environments. The wanted outcome of the interpretation is similar to the other images. However, this method might be promising if the stratigraphic shapes are processed adequately.

RGT Outcome



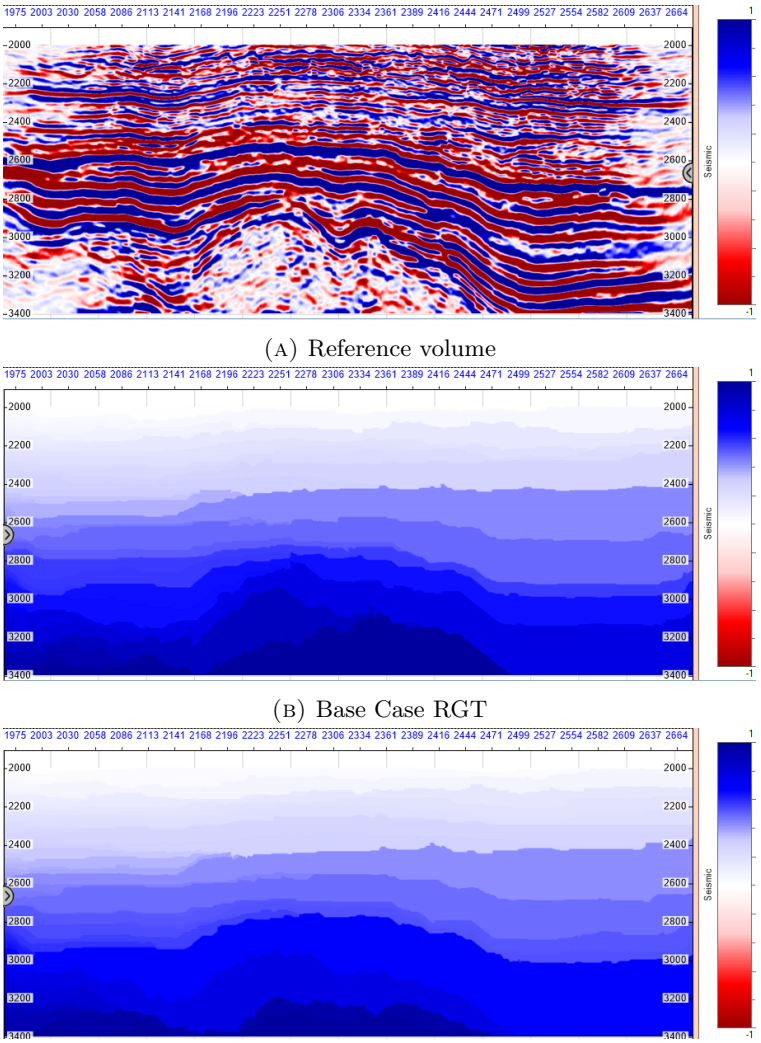
(A) Reference volume

(B) Base Case RGT

(C) Base Case RGT Denoised

FIGURE 4.8: Comparison Base Case and its denoised counterpart

In the seismic reference volume (4.8a), it is not possible to observe any variation in the seismic data after a denoising operation has removed the noise in the data. That means there is a level of difficulty linked to detecting the presence of noise by observation. In evaluating the reference volume, it is deemed that the denoised interpretation is more plausible(see the formation of old age in fig: 4.8b and 4.8c). Reasoned by the older layer traverses too far up to a shallower depth which is not visible in figure 4.8a. However, the structure within the outlined deformation is highly erratic, implying it is hard to follow its horizons. A vital remark is to be aware that the algorithm is susceptible to the effect of noise present in the seismic (see fig: 4.8). The assumption is that all real seismic data sets would be afflicted by a grave amount of random noise, which affects the algorithm's performance. It must be emphasized that the algorithm is not noise-invariant yet not significantly afflicted by its presence. The output is still sensible but cleaning data before RGT transformation is recommended. Given the dynamic time warping theorem, that method provides some resistance to noise present in data, which can be confirmed in fig 4.8b because even without cleaning the seismic, the interpretation does not lead to an improbable outcome. Although the interpretation shown in figure 4.8c, appears to be more suitable. Hence, not entirely noise-invariant yet yields some resistance to present latent noise.



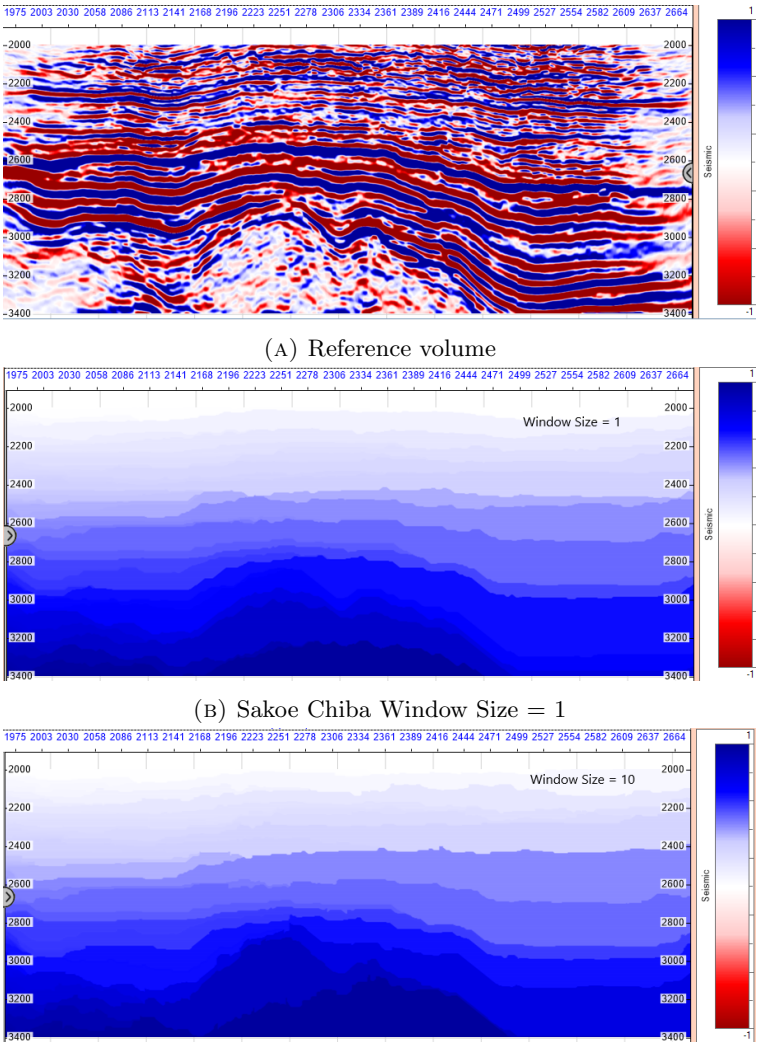
(A) Reference volume

(B) Base Case RGT

(C) Derivative RGT

FIGURE 4.9: Comparison Base Case and derivative DTW

By comparison, an interesting relation between performing RGT transformation on denoised data(fig: 4.8c) and utilizing evaluation by derivative dynamic time warping(fig: 4.9c). As explained in chapter 3.3.4, the derivative dynamic time warping has singularity dampening characteristics, which ultimately means that, by definition is more noise-invariant than the standard variant. By inserting the same non-denoised data into the DDTW assessment, the output (viewed in figure: 4.9c) turns out to be very similar to the result of DTW when passed in clean data (viewed in figure: 4.8c). Might it be that denoising the data is not needed if DDTW is used rather than the standard DTW? The resemblance is significant, yet there are noticeable dissimilarities between the derivative result and the denoised case. These dissimilarities are evident in the oldest relative age(dark blue coloring). Whether the standard denoised procedure or DDTW is preferred is not known, it is challenging to decipher the correct interpretation for the erratic zone situated in the middle of the seismic slice. More tests are needed to be committed to concluding the matter.



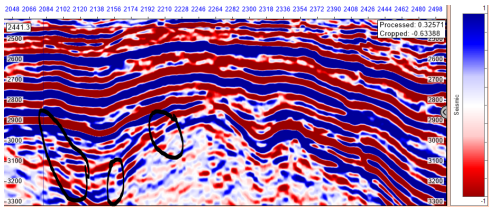
(A) Reference volume

(B) Sakoe Chiba Window Size = 1

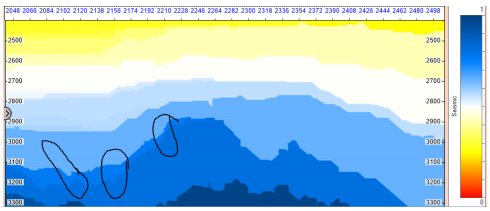
(C) Sakoe Chiba Window Size = 10

FIGURE 4.10: Comparison Different Sakoe-Chiba Bands

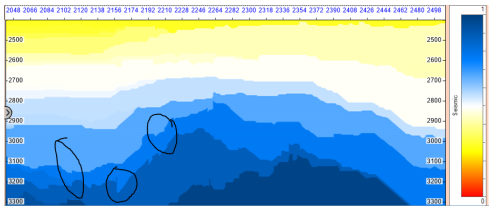
Previously mentioned in the chapter: 3.3.2, the Sakoe Chiba band has a tendency to bring down the time complexity because it limits how many comparisons are being estimated between the two inserted traces. If there is no restriction, all positions in both traces are compared to the neighboring trace. In Figure:4.10, it can clearly be stated that a more constraint band can limit how much the warped path between traces can deviate from each other. One explanation why 4.10b encapsulates a different structure than 4.10c is due to the fact that the found structure in 4.10c exceeds the limitation of the band. Limitation of possible warps also seems to prohibit the method from establishing over-fitted relations between traces; see the lower-middle part in figure 4.10b and 4.10c. Where it is unlikely the oldest(dark blue layer) is really situated at such a high altitude(in fig 4.10c) based on interpretation reference volume. In figure 4.10b, the oldest layer is contained in a more plausible altitude, yet there might be compromises throughout the domain because of the forced constraint. It might be a tedious process to optimize the window size, to find the best possible interpretation by verifying and attempting a new window size. Eventually, the band might not yield an optimal solution without interference with a refining framework(see appendix chap 8.3.4).



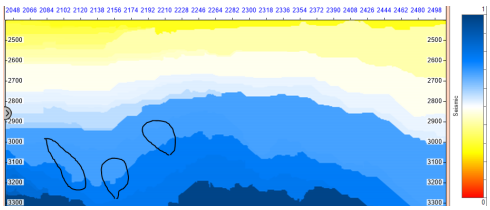
(A) Faults reference volume



(B) Visible faults in RGT volume



(C) Visible faults in the non denoised RGT volume

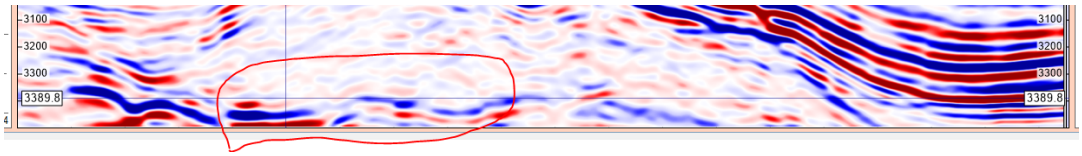


(D) Visible faults in the DDTW RGT volume

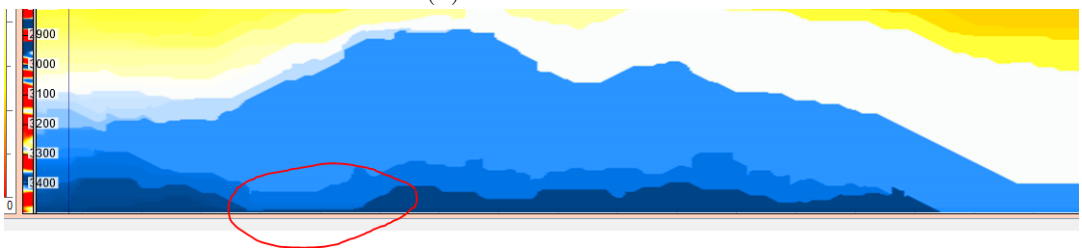
FIGURE 4.11: Example of faults occurrence in RGT result. Marked in black circles are areas which can interpreted as faults from its reference volume, and its RGT interpretation

How the algorithm perceives the faults present in the data set must compare to how detectable faults are evaluated. Viewed in figure 4.11, it can be seen a reference volume (fig: 4.11a) and its RGT counterparts (fig: 4.11 , fig: 4.11c and fig: 4.11d) where first RGT the data has been denoised before evaluation, second is not denoised and last figure is by using DDTW(chap: 3.3.4). Including the denoise effect highlights the impact of cleaned data on interpretation and how invariant to noise the DTW algorithm is. Because DTW is said to be noise invariant, but normally this is for comparison between two signals, the situation is a stack of many pairs of signals. Would DTW be equivalently invariant in this setting?

When reviewing each of the RGT result in figure 4.11, it can be determined that the non-denoised case (shown in fig: 4.11c) appears to emphasize shifts more drastically. For example, in areas where faults are found (marked in black circles), it is less challenging to find areas in RGT results in which faults are present. Moreover, in the middle circle, it seems to be falsely interpreting which direction the fault is oriented because it shifts downwards rather than upwards, which correlates with an interpretation of reference volume. The similarity between figure 4.11b and 4.11d is striking, yet there are many differences visible among these. However, despite the similarities, there is a clear difference, such as which layers are present in the interpretation. Furthermore, many detectable interpretation differences are present, particularly how detectable the marked faults are. Moreover, it can be concluded in this specific case that the denoised RGT(figure 4.11b) case provides the most satisfactory interpretation of faults in terms of magnitude and orientation.



(A) Reference volume



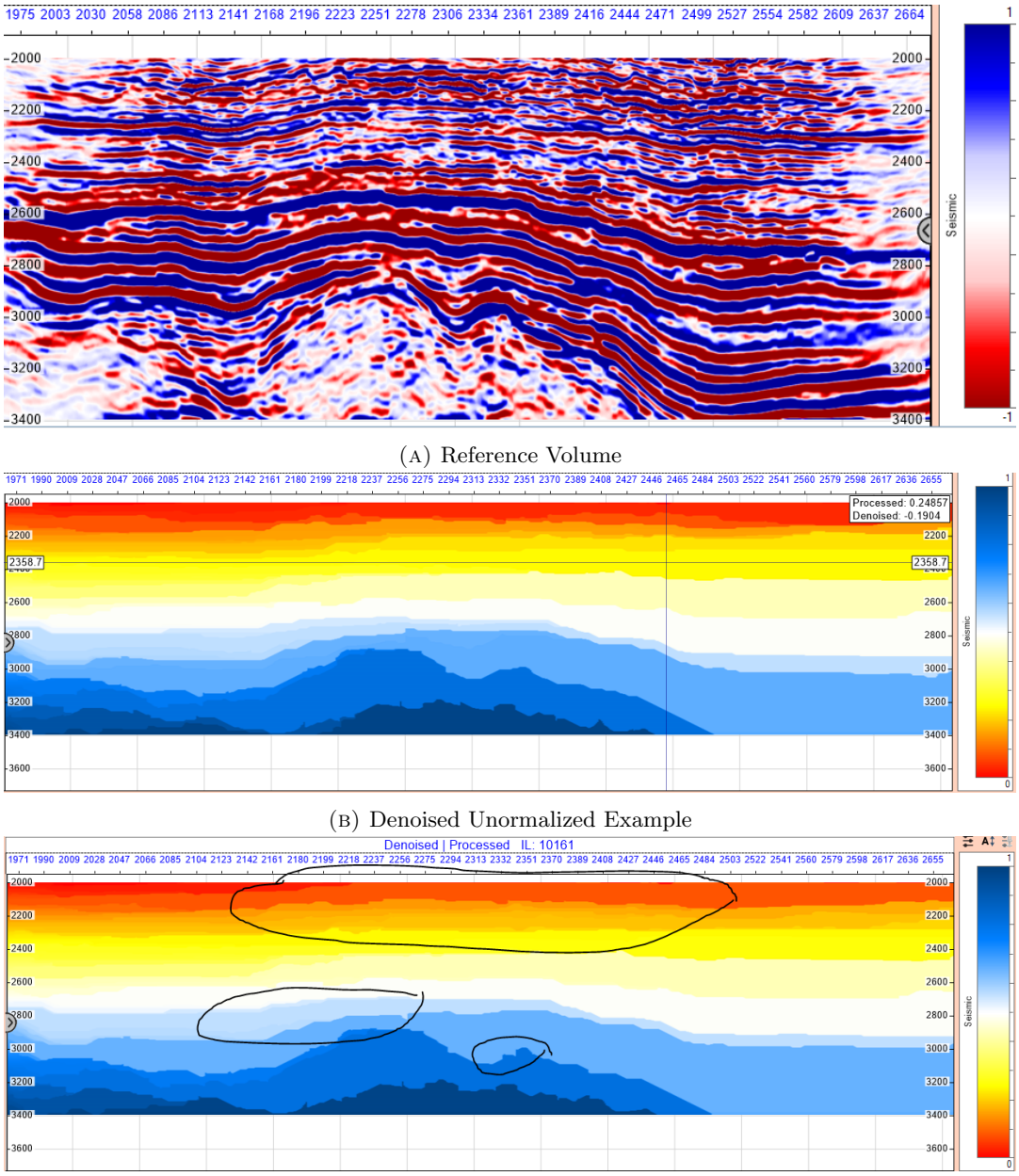
(B) Not normalized result

(C) Normalized result

FIGURE 4.12: The effect of normalization before alignment

There is importance associated with normalizing the traces before their warped paths are found; a higher amplitude can lead to a misconception of the similarity between two traces. By individually normalizing each trace, it is possible to track structures between traces. A major amplitude might be compared to a less significant amplitude due to the shape of both traces being similar. However, the distance is farther from its correct counterpart in a discontinuity. Especially since seismic traces are continuously shifting between peak and trough, the warped path might be linked to an incorrect peak to peak or trough to trough since the magnitude of amplitude is inaccurately evaluated by DTW. Normalizing ensures that a strong reflector is more susceptible to being warped to another strong reflector of a similar shape.

Figure 4.12 shows an indication that normalization before warping obtains a better relation in RGT volume. The background for this claim is based on what can be interpreted directly from the reference volume (shown in fig 4.12a), in which the reflectors are perceived to be uplifted. This uplift is redundant in fig 4.12b, but it is interpreted as a uplift in fig 4.12c. However, this is purely one incident that can be determined out of the whole results amongst these two RGT volumes. Nevertheless, it must be mentioned that this single incidence can not elaborate a conclusive statement on this matter.



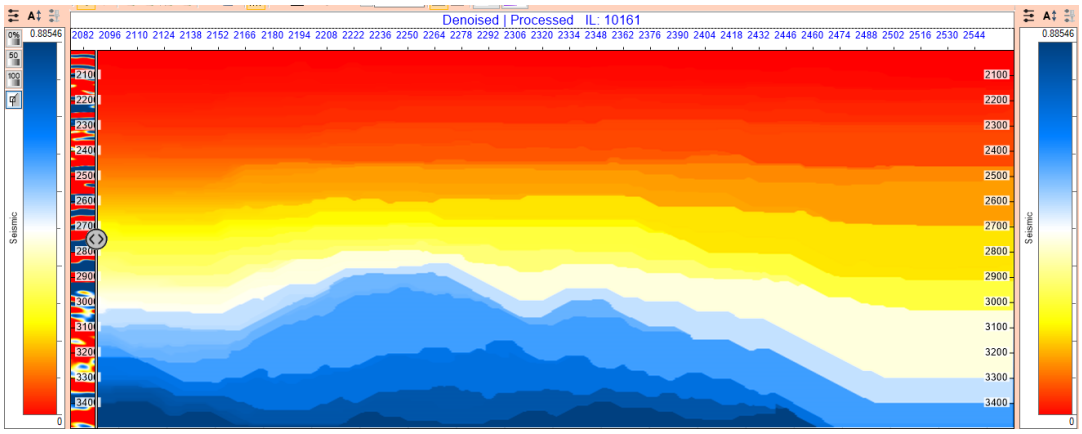
(A) Reference Volume

(B) Denoised Unormalized Example

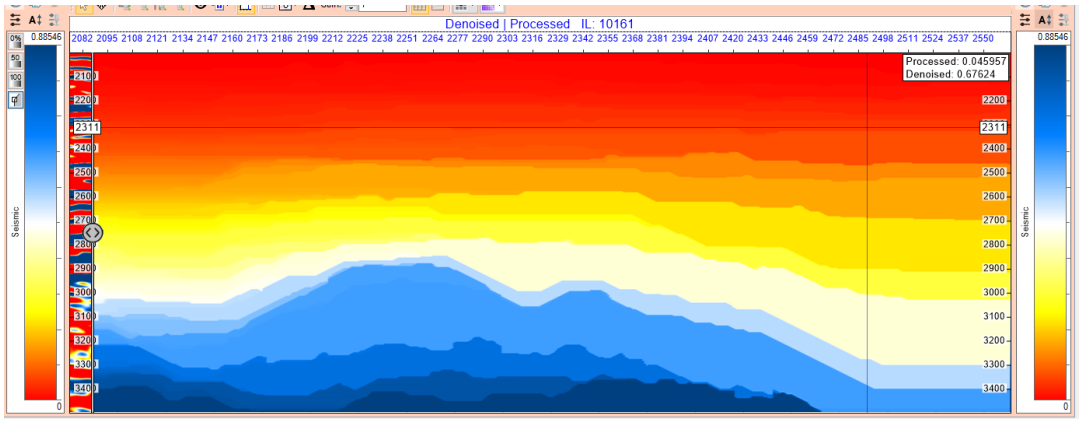
(C) Denoised Normalized Example

FIGURE 4.13: Comparison denoised normalized and unormalized comparison

Extending the inspection regarding how impactful signal normalization is, it was deemed necessary to see whether the impact of normalization becomes more significant if a volume is cleaned for noise present in the data. Figure 4.13, illustrates this inspection of significance. Through evaluation of the different RGT outcomes shown in the figure, it is possible to observe distinct differences between the two results; some of these differences are marked in the black circles in figure 4.13b. However, as the differences are compared against the respective reference volume(fig 4.13a), there are no observations that signify that one outcome is more correct than the other. The result of this comparison support whether differences are apparent in normalization or not, yet it is related to minor details that can be interpreted differently by different domain experts.



(A) Unweighted padding



(B) Weighted Padding

FIGURE 4.14: The effect of padded result

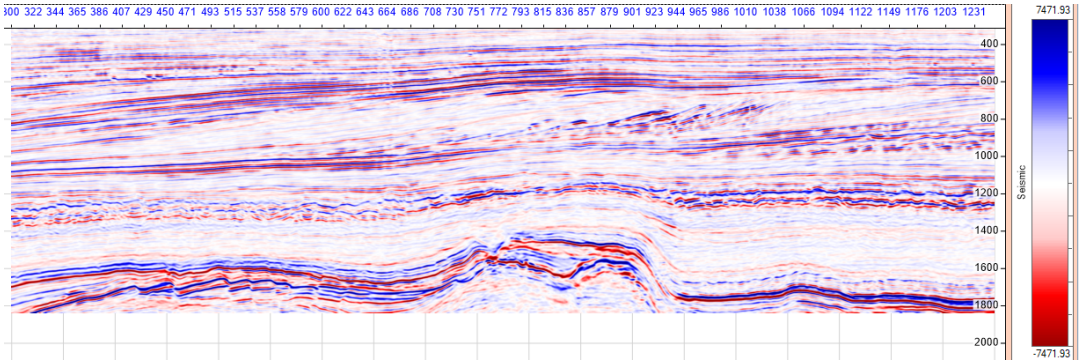
Spatial relevance is a problem that has to be further investigated; from the perspective of this thesis, the spatial influence is represented as a dependent action to align each slice with its neighboring slices. This algorithm selects the number of slices as a matter of preference. The algorithm might require comparing each slice with two neighbors or another amount on each side. The neighbors on each side are called paddings; for example, if two neighbors on each side are wanted is equivalent to saying padding of two. Adding an increased level of padding will contribute to doubling the algorithm's time complexity because each increase in padding is equivalent to the doubled amount of computations to find the warped path. The definition of dynamic time warping to induce dependency between adjacent signals is defined by:

$$DTW_D(x_i, y_j) = \sum_{k=1}^{K} (x_{i,k} - y_{j,k})^2. \tag{4.1}$$

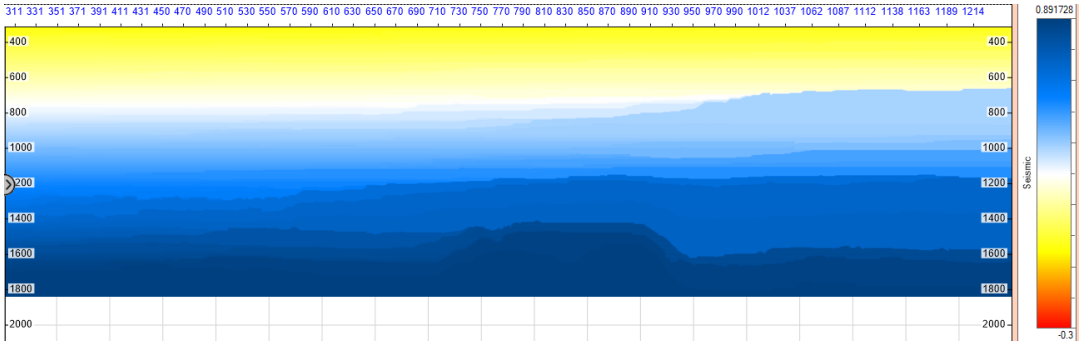
Ultimately, Eq: 4.1 is similar to populating the euclidean distance matrix (Algorithm: 1), but now summing the adjacent euclidean distances from padded slices, before step patterns(chap: 3.3.3) conducted onto the matrix. This implementation is a result directly following the research covered by (Shokoohi-Yekta et al., 2015), which considers a comparison between independent multidimensionality and dependent multidimensionality. Thus, the question is whether the multidimensional assessment of DTW should be spatially dependent or independent because only the spatial dependent version has been applied in the current state of the algorithm. Moreover, in

Shokoohi-Yekta et al. (2015), the research suggest tightly coupled time series should be dependently defined if considered, i.e., only testing dependent multidimensionality is satisfactory for a seismic setting. By comparison amongst the different figures 4.14 and 4.11c, arises the indications on differing properties between the evaluations. That through an assessment of the slices neighbors in the DTW procedure on the non-denoised data manages to contribute to a more plausible interpretation (shown fig 4.14). It is more plausible because the older layer is not intruding that high in the overlying layers(present in fig 4.11c). However, the computations time is much higher than other variations(i.e., DDTW extension, denoise extension, small window Sakoe-Chiba band), which achieves a similar effect. Furthermore, is there a foundation for a decisive conclusion on whether a padded interpretation would yield a more correct interpretation than other versions(shown in figure 4.11)?

No significant contrasts among the various methods can conclude whether one version is more optimal than the others. Be aware that there is some difference in cropping amongst figure 4.11 and 4.14.Might it be possible to decide whichever weighted(fig 4.14b) or unweighted(fig 4.14a) is preferred? There are some correlation in interpretation of the erratic zone(lower middle part) in figures 4.14b and 4.13c, does this supports a claim that weighted dependent multidimensional evaluation is favored? That requisition requires that denoised data is favored, which one might argue is correct. However, this is simply one comparison of weighted and unweighted, and more tests on various data sets must be conducted to assert this as a truth. Furthermore, this applies to the other versions discussed in chapter 4, but it is believed that the denoise workstep diminishes the chance for random relations. Because it purifies the data for noise, uncertainty is still coupled with whether it is the optimal interpretation.



(A) Asterix reference volume inline



(B) RGT outcome Asterix reference volume inline

FIGURE 4.15: RGT interpretation stratigraphic structures

The presence of stratigraphic features(similar as middle image in fig: 2.6) in figure 4.15a (discussed in chap 4.2 **The Data**) needs to be tested to determine whether the algorithm's performance mimics these structures. Initially, it can be concluded that the overall structure appears satisfactorily portrayed. The intruding deformation located at the bottom portrays a formation(see fig: 4.15b) which readily correlates with the original seismic (see fig: 4.15a). However, the stratigraphic features are the major structure that this example should explore. The interval of 600-900 ms showcases a structure enclosed by two seismic horizons, which seems to widen as it progresses rightwards (see fig 4.15a). Seismic horizons within this encircled area appear to establish a series of downlaps and onlaps that diverges and converges into the two encircling horizons. That arises the indication of the presence of unconformities in the subsurface. Interpreting this area by the RGT method should accentuate that the lower seismic horizon (that encloses the stratigraphic horizons) continues horizontally and is older than its overlying layers. The top enclosing horizon is supposed to be younger than the structure contained within the structure. The structure enclosed by these should contain the intermediate times amidst the age of the upper and lower horizons.

Furthermore, when evaluating the outcome produced by the native DTW-RGT method, in figure 4.15b. The multiple intermediate relative times constrained by the two horizons are not emphasized satisfactorily. The trend is observable, such as the older times are moving towards a shallower depth as a rightward movement in the slice is committed. Additionally, some younger ages appear as onlaps onto this intruding older strata. Observations can support a statement that the RGT tackles stratigraphic structures appearing as unconformities in the domain. However, some remarks have to be discussed. The count of contours that explains this structure is too low; it would require more contours to emphasize the presence of these structures further. In its current state, it accounts for the structure to some extent, yet it appears as a slope. It is expected to appear as a series of structures. Some uncertainty is linked to the old intruding layer. Whether it is one of the layers initially propagated from within the interval or if it is the lower horizons that diverge through the structure? It would be mistakenly propagated if the increasingly shallowing layer represents the continuation of the lower horizon. Nevertheless, it is presumed that the layer is one of the horizons allocated within the stratigraphic formation. The interpretation by the method might be sufficient in its current state to quickly grasp some of the more critical details of the surfaces. However, it would be beneficial for the interpretation if finer details were detectable in the formations.

Chapter 5

Discussion

5.1 Important Remarks

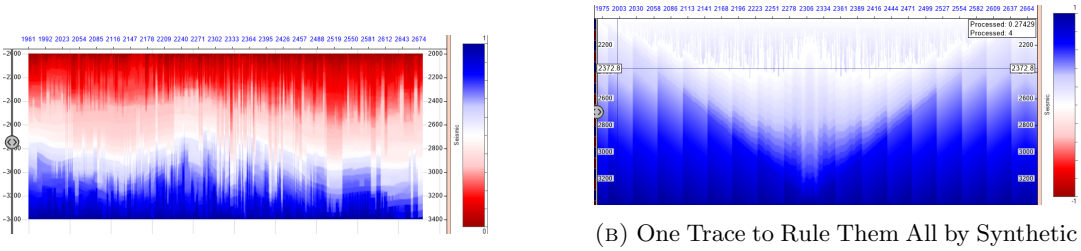
5.1.1 Usability

Dynamic time warping appears to have many potential advantages; one of such advantages is its flexibility of definition for further improvement. The flexible way to adjust the DTW definition to account for new parameters or different events to be tackled by the algorithm can freely be added to the standard procedure of DTW. An example of the flexibility is the addition of the capability to formulate DTW into DDTW, which is found to handle singularity better than the standard procedure; a visual example of this is observed in fig 4.11. The additions provided to the standard procedure can, in many situations, increase/decrease accuracy or lower the amount of computation (Sakoe-Chiba (chap: 3.3.2) example of this). Because the method is so flexible in how it can be expanded to account for new features, it makes a valid argument for using this approach. Often, variability can be considered a strength for a method, enabling feature engineering such that a solution can be approximated so that features that are particularly important for a physical event are optimized. An example may include additional information to the cost matrix, such as seismic attributes that emphasize features that are mainly linked to an event or structure. Such that warped path is biased toward that particular feature, such that the warped path improves the accuracy of interpretation for the specific feature.

The method requires a very narrow window size when employing the Sakoe-Chiba band to acquire flexible interpretation. This applies when neighbor pairs compare seismic traces by neighbor pairs. This requirement might be explained by the reduced distance between the seismic trace pairs, such that a significant constraint still yields a flexible outcome. However, suppose a comparison is to be conducted on a trace that is allocated far away. In that case, it could require a more flexible(bigger) window size to allow the correct warped path, but at the expense of the time complexity. The procedure's outcome seems to make a sensible interpretation of the subsurface. However, there are slight variations in the interpretation that a made based on fine details from the seismic traces. It is partially cumbersome to verify all its interpretations.

Nevertheless, the outcome is satisfactory to some degree, and there are some aspects to pay attention to. One of these aspects is that one must be aware that by the current recursive DTW procedure, one might not find the optimal warped path throughout the cost matrix. The explanation is that the lowest-cost path in the path selections proximity(ch: 3.3.3) might lead the warped path toward the local lowest distance estimates yet might not aggregate the global optimal warped path. Though, the outcome appears to be continuously connected in the result, i.e., no abrupt fluctuations between traces(as seen **one-trace-to-rule-them-all iteration**,

fig: 5.1). A possible explanation is that trace pairs that are being processed by neighboring pairs would have less dependence on the global warped path optimization procedure. Neighboring pairs are similar in shape, which might often situate the warped path along the cost matrix's diagonal. In contrast to when a seismic trace is compared to a seismic trace contained in an entirely different part of the subsurface, optimizing the global warp becomes increasingly more critical(see fig 5.1). Moreover, distant signals require more intermediate steps before acquiring the optimal path. There is a chance that the path diverges from the required singularity used to shift toward the correct optimal shift. Hence, the path is now situated in a less optimal path because of premature divergence from the merge. In the same figure, a plausible conclusion can be drawn that all fluctuations visible in the image might be caused by local variations inside each cost matrix(for each trace pair). The RGT traces seem to be disconnected by the local variations amongst traces.



(A) One Trace to Rule Them All by Real data

(B) One Trace to Rule Them All by Synthetic data

FIGURE 5.1: Warped path based on comparison between initial trace against all other traces

However, even as the fluctuations are present(see figure 5.1), there seems to be an underlying structure visible. Moreover, the fluctuations disperse pieces of the structure such that the actual structure is not aligned, yet all parts yield a sufficient amount of contours(see fig: 5.1b). In the case where all traces are compared against one trace seems to yield more contours than standard comparison(described in chapter 3.5), but the shifted pieces of the domain make the outcome less reasonable. If all the pieces were shifted back to a similar sequence. Such seismic horizons become continuous. This is the main problem with the method; if the DTW is aggregated in a particular direction, the singularities required for performing a sensible interpretation also "consume" the contours of the seismic horizons. However, as seen in fig: 5.1, more contours are present, yet the shifts are inaccurate. The method either requires an optimization framework(example described in appendix chap: 8.3.4) that always gives the correct path to account for distant signal comparison. That means seismic traces situated in proximity to each other are more spatially coupled than traces kept further from their counterpart in the DTW assessment. The other option would define an aggregation procedure entirely independent of the orientation of execution. Hence, yielding an RGT volume that does not "consume" contour information and conducts the correct interpretation of the subsurface.

5.1.2 Avary Evaluation

When defining the algorithm utilized in this thesis, much of the Avary framework was a valuable addition to many of the defined operations. The indexing was neatly handled; when defining a new area, the algorithm was not dependent on querying a specific position in the seismic stack. The indexing was always relative such that the algorithm was very generalizable to new areas of interest, rather than hard-coding

each position onto the function. The native Avary operations (shown in the red box in fig: 3.11) provided a wide variety of functions to be implemented into the workflow. However, two of these operations were helpful in combination with the DTW algorithm. Specifically, "Crop" and "Denoise" operations provided support for cleaning random noise present in data before being inserted into the DTW procedure. Also, cropping the slices lowered the computational time when assessing preliminary results when asserting the method's performance.

5.1.3 Retrospective Assessment

In light of the results and limitations found by selecting Dynamic Time Warping as a method by choice, some discussion will be conducted on whether this method is worthwhile for further investigation. Firstly, DTW can determine a sensible interpretation of the subsurface, similarly to which deep learning(see chap 8.2) method and phase unwrapping(see chap 8.1.1) solutions. However, all of the methods experience different complications. For example, DTW might be able to find proper paths between traces, but it requires a separate volume to aggregate these changes. Eventually, errors propagate because of the sequence of updates (more discussion in chap 5.1.4). Where phase unwrapping does not need a separate volume, i.e., it means it directly manipulates the input information. Furthermore, phase unwrapping is believed to be more efficient(similar to Fourier transform) than DTW, but this depends on what constraints are induced on DTW, yet phase unwrapping is assumed to be faster. Phase unwrapping is experiencing problematic behavior when propagating noise artifacts present in the data. It might contribute to fluctuations in the outcome of the method. However, variations of the method prohibit the effect of error propagation (Wu and Zhong, 2012b). DTW is assumed to be more noise resistant than phase unwrapping. Nevertheless, through this thesis, there are some effects of mitigating the noise in the data beforehand conducting pathfinding (see fig: 4.8). The denoise effect does not significantly change the overall structure but affects minor details.

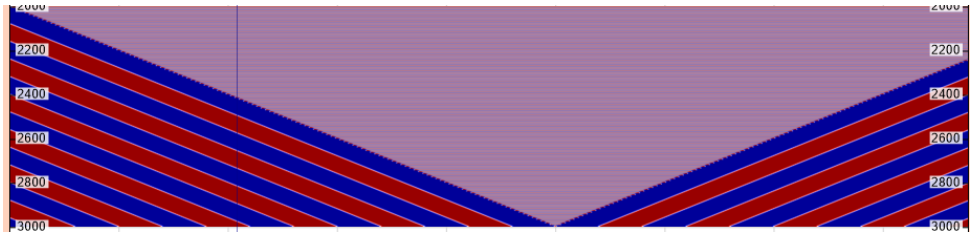
Moreover, the deep learning method that can be used for this purpose can be elaborated from quite simple network architectures to multiple interacting networks. How the network architecture is defined can be a complex procedure, often depending on how a network can best learn the patterns present in the inserted data. After a network has been defined, there is a matter of whether pre-processing steps (standardizing, normalization) are induced on data. Feature engineering could significantly impact performance. Providing correct feature engineering steps for a CNN is manipulating the size, orientation, and coloring of insert images in the training set. Committing these tweaks is readily done to induce more variation in training data. Hence, the model becomes more generalized, such that predictive performance increases on "unseen" data. The matter of feature engineering can also be related to DTW, whether various seismic attributes shall be accounted for in the cost matrix. An example might be that both standard similarity measurements(eq: 3.14) are included in the cost matrix and the raw phase unwrapped volume, amplitude, and variance volumes. Extending the number of features included in the cost matrix might yield a similar effect as feature engineering has on neural networks; the actual patterns are emphasized because of more data on the same object (more on this in chapter 5.2.2).

A potential strength of DTW is that the method is quite generalizable. There might be occurrences where an optimal path is not found, yet a neural network can also commit predictions that yield false truths. However, all methods might fail,

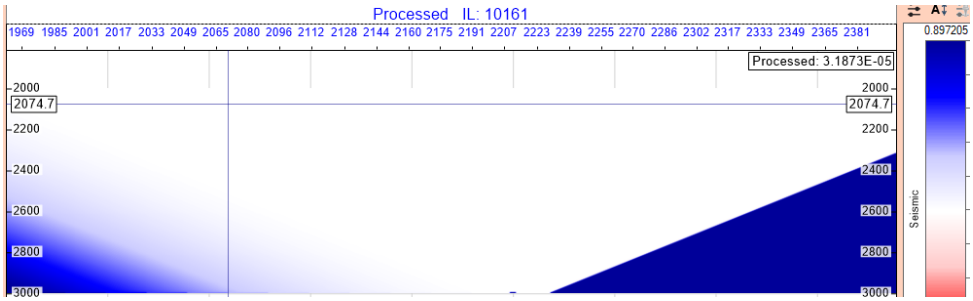
but deep learning might provide good performance, yet this might be trained on a specific training data example. Thus, the network is particularly biased toward information provided in the training set. That means Deep Learning will perform poorly in predicting information not provided in the training set. Contrary to belief, one should not apply indefinitely variational training information. Ultimately leads to models that are incapable of differing amongst the prediction outcomes, where previously, it was able to predict some of the outcomes with high certainty. There might be a dependency on retraining the model when transitioning to a differing oil field with differing characteristics. Repeating the training phase of the machine learning method might be required when changing between significant differing subsurface data sets; still, when the model is trained on the given data, the prediction is carried out rapidly. Identically, deep learning and DTW are experiencing problematic behavior linked to the merging procedure of the sub-solutions found by the methods. An intricate merging procedure is a prerequisite of both methods (Bi et al., 2021a). When comparing the various methods for achieving an RGT volume, it is challenging to conclude whether a method is preferred over the others. All are proven to establish RGT volume, yet in this thesis, there has not been possible to compare the different methodologies against each other on the same data set. Verifying performance on the same data set would be a compulsory condition in evaluating the performance of the different methods. The conclusion is that RGT evaluated by DTW is not fundamentally better performing than the other method, nor clearly worse. Regardless, the method might be a bit immature at this point. Furthermore, the advantage of RGT by DTW is that the concept of the method is simple and can easily extend its capacity by applying extensions developed by the scientific community. The method could prove to be more useful if a warped path optimization procedure was to be implemented. Additionally, a stable volume aggregation method could provide the most considerable improvement for the method.

5.1.4 Error Propagation

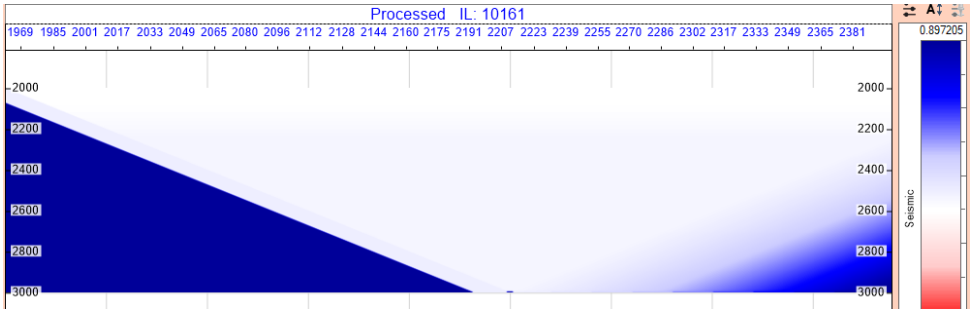
In investigating the algorithm's performance, evaluating the different reasonable situations that might occur in an actual seismic data set is necessary. One such realistic situation would be underlying layers intercepting or compressing overlying layers, illustrated in 5.2a. The situation would determine whether the algorithm can address data where the best alignment path does not necessarily contain the appropriate value in the updated reference volume. When all the warping paths from the different traces have been evaluated and stored, the time is to update the whole volume by manipulating all the individual RGT traces. However, a side effect is detected in updating the volume from one end to the opposite end. As seen in 5.2b and 5.2c, it the determines that information is lacking when updating from either of the direction. The probable reason is that the warped path is correct, yet it propagates wrong values through the domain because of its inability to diverge values. A wrong value propagated through the volume can be explained by the compiled singularity that readily changes multiple values into a singular. See figure 5.2, as the interpretation reaches the bottom of the surface, only the oldest age is propagated. It might be because all the other ages are not present at the intersection between the two slopes. Thus, it can be claimed that the method can converge yet is not enabled to diverge the merged information. There is a grave need to update the volume so that the correct value used in its warped path is propagated instead of propagating values not involved or involved at a different position in the warped path.



(A) Reference volume



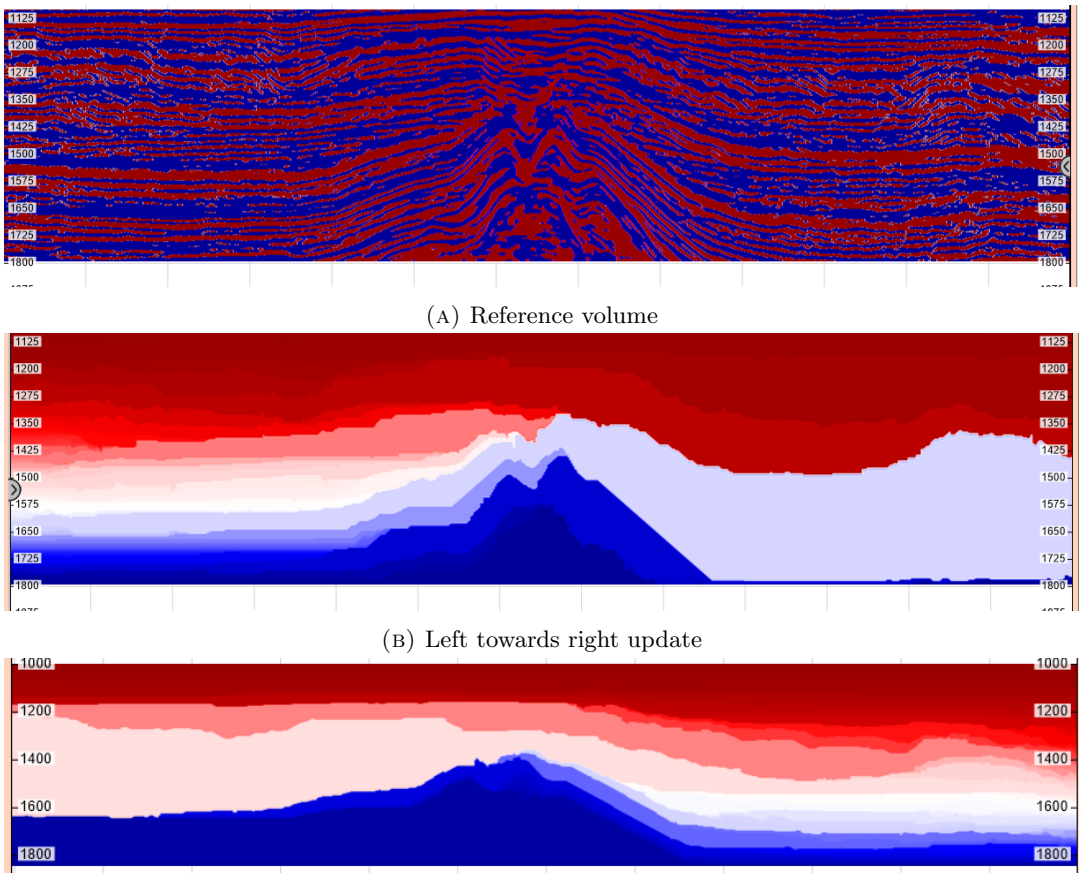
(B) Left towards right update



(C) Right towards left update

FIGURE 5.2: Effect of different update direction

When assessing a real data set to check whether similar behavior is present, seismic data containing a similar intercepting structure was found appropriate to assess this behavior. In 5.3a, it is possible to observe the intercepting layer, which compresses the overlying layers, and afterward. The previously compressed layers are more or less returning to their original spacing (similar to before the interception). Many interesting observations are detectable, such as when assessing 5.3b, the left side of the domain includes many more contours in its interpretation than the right part. Many minor faults and structure variations are observable in the section where contours are present. However, the contour details included diminishes as the center of the domain is approached. As an aftermath of the intercepting layer, the right part of the domain still seems to update the found warped positions. However, it unravels the layers as compact and has a significant lack of contours present in its layers. That supports the hypothesis considering that a wrong value from its prior traces is mistakenly propagated. Consequently, the underlying layers are "consuming" the overlying layers' information, thus constructing a distinct propagating layer.



(A) Reference volume

(B) Left towards right update

(C) Right towards left update

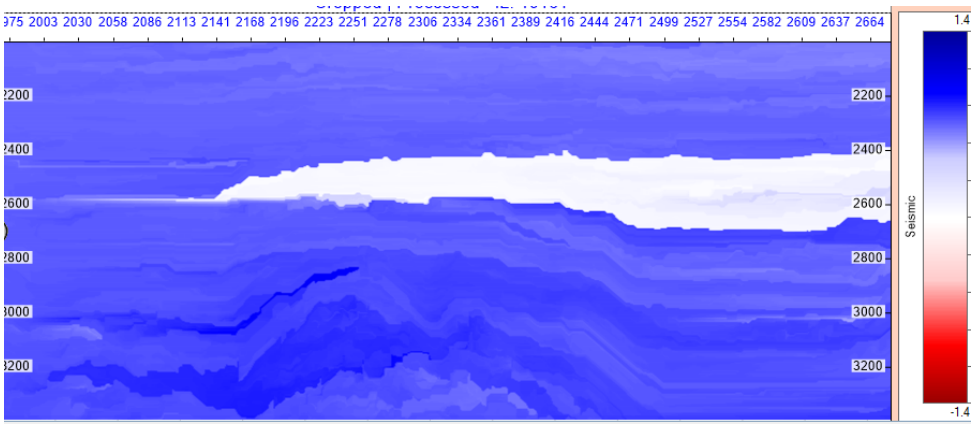
FIGURE 5.3: Effect of different update direction

Considering 5.3c the same effect is present. However, in the opposite direction, it can be detected that the previously distinct layer on the right side contains an increased share of contours. Furthermore, the left side of the intercept is a consequence of the reversing update direction that has transformed into a distinct layer. Thus hiding information present in the layers. Additionally, an effect on the intercepting layer is detected; the steepness of the layer seems to be slightly reduced. Subsequently, the interception seems to be less considerable as in seen 5.3a and 5.3b. Because of the effect that the choice of direction has on the returned result. It is necessary to construct the algorithm to be path independent such that the choice of the update does not affect the result(more discussion chap 5.1.4). Hence, both directions hide information in the result. Even areas where many contours are present might have variational results because the information included in 5.3c might reveal more information, contributing to making the RGT volume more correct. Nevertheless, it should be stated that the result in this state provides a good initial result of the overall structure present in the seismic data. However, one must know that much information is hidden in this algorithm version. Thus, they do not necessarily represent precisely the real structures present in the subsurface.

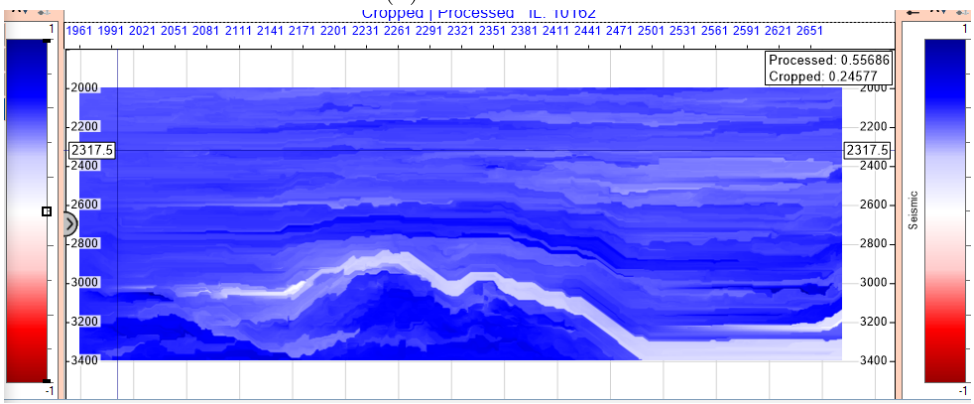
Attempted solutions to the error propagation

In the pursuit of solving the error propagation in the volume, various hypotheses were tested or attempted. Since a finalized best method has not been determined during this study, the results in this section are thought to highlight what potential

solutions are attempted. Descriptions of each outcome are provided to emphasize what has been conducted to achieve the preliminary results. Hopefully, the emphasis on attempts can contribute to detecting errors in the logic behind the solutions or their implementational mistakes. Be aware that these results are used to showcase preliminary results and are not supposed to be finalized RGT volumes.



(A) First Recursive



(B) Second Recursive

FIGURE 5.4: Attempt Updating By Binary Tree

For splitting the seismic slice into a perfectly balanced tree, the division procedure shown in (algorithm 3 Binary Split) is essential to steadily split the domain until a single trace pair is reached. Since each tree level is defined as equally sized, consider the same amount of traces used to determine what shifts will be added to the currently evaluated RGT trace result. The sum of all the related nodes in the binary tree towards the specific trace node (see fig 3.10) is represented at each increment in the vertical position in each trace. For example, if trace 9 is to be evaluated at some specific position in vertical direction, lets assume 25 for example, then: $RGT_9[25] = RGT_{8-9}[25] + RGT_{12-13}[25] + RGT_{10-11}[25] + RGT_{9-10}[25]$. Each vertical position might contain multiple shifts because the warped paths can select paths of different lengths, yet only the last found shift will be included in this situation. Such that each $RGT_{x-x}[25]$ only contains a single shift which illustrates the difference in positions amongst the signals. However, as emphasized, only the maximum shift(positively or negatively) is accounted for at each position; see illustration; included positions marked in red in figure 5.5.

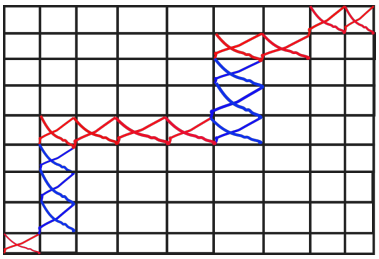
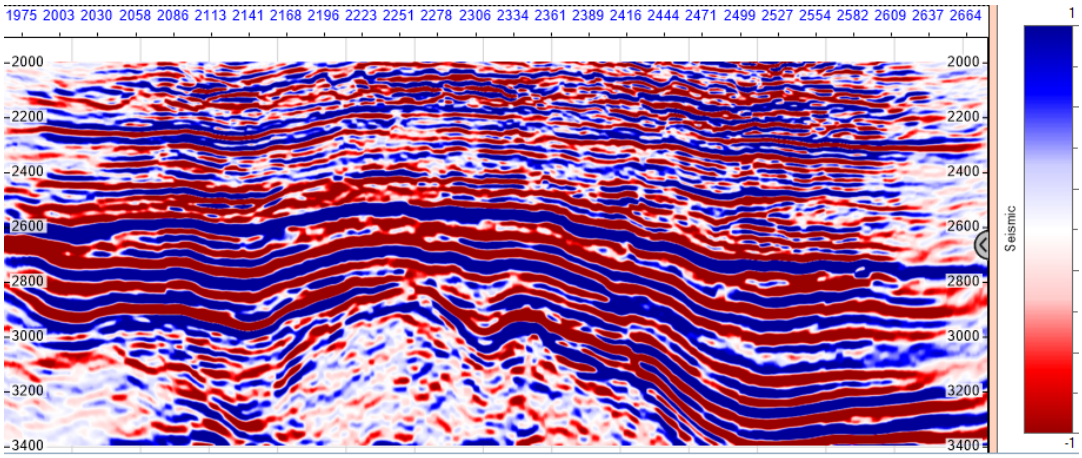


FIGURE 5.5: Warped path shifts illustrated from the cost matrix

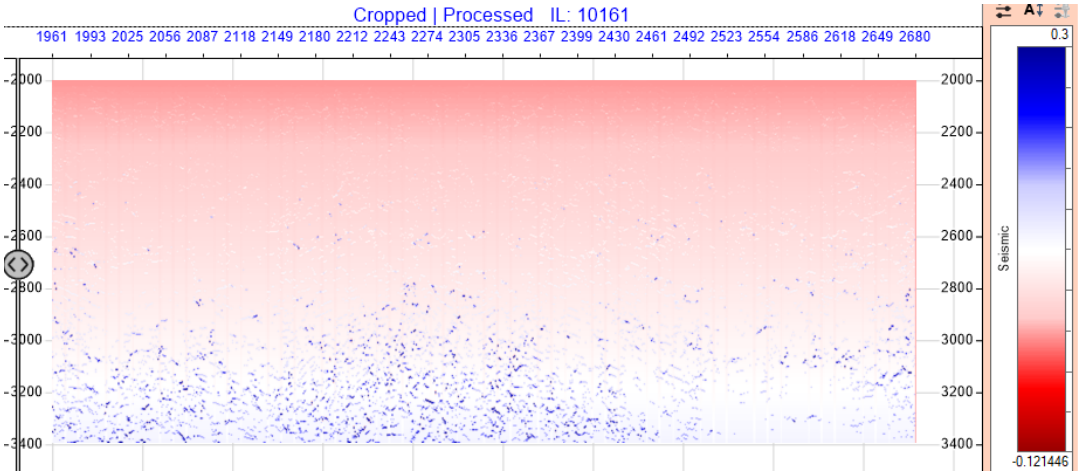
The accumulation of the recursive layers leads to an averaged shift that should readily aggregate the real shifts for the currently evaluated trace at every vertical position. Again, the idea is that information is independent of update direction, such that it is not confined inside distinct boundaries. Furthermore, the attempts shown in figure 5.4, are not updated independently. That means that the left-to-right bias is still present, yet the shifts found by the recursive layers are used as a complimentary term in the RGT aggregation. The idea was that after the neighboring signals have transferred their found path. Then the recursive term should be used as a correcting term to unravel the contours kept within the distinct layers. To some degree, this was successfully performed, yet another issue arises. The volume is constrained from continuously increasing, and occurrences of some extreme artifacts can be observed(the white bands in figure 5.4). Furthermore, much more details regarding the contours of the horizons are successfully unraveled. Whether these contours are the subsurface's correct interpretation is yet to be determined, yet it seems a plausible interpretation. Although the magnitude of RGT levels is drastically misconceived, the attempt outlines the structure satisfactorily. The difference amongst figure 5.4a and 5.4b, are that fig 5.4a represents $RGT_{\text{rec}}[k_1]$ by the shifts at each respective depth. While figure 5.4b accumulates the shifts by depth before aggregating the final RGT volume $RGT[i_{\text{neigh}}, j_{\text{neigh}}, k_2]$.

$$RGT[i_{\text{neigh}}, j_{\text{neigh}}, k_2] = RGT[i_{\text{curr}}, j_{\text{curr}}, k_1] + RGT_{\text{rec}}[k_1], \tag{5.1}$$

$$\text{i= 0,...,N ; j = 0,...,M ; k = 0,1,..., L.}$$



(A) Reference Volume



(B) RGT result by aggregation of pure recursive shifts

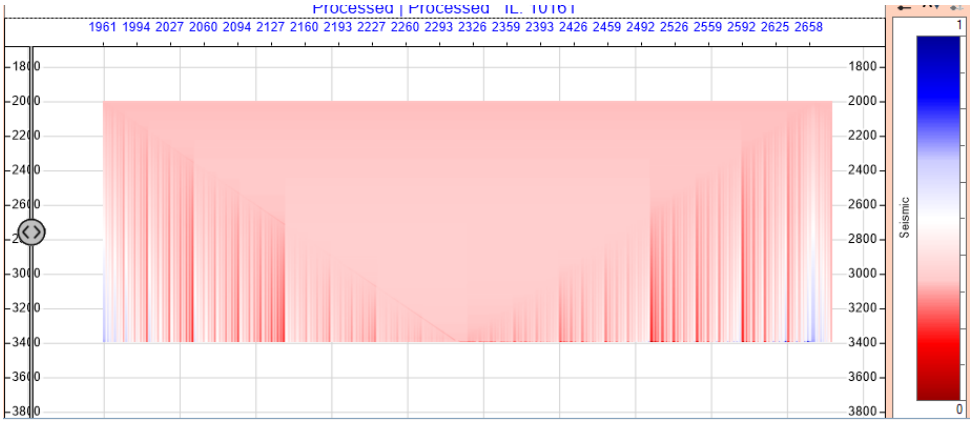
FIGURE 5.6: Attempt Updating by pure use of the recursive binary structure

The method shown in figure 5.6 deviates from the one shown in figure 5.4 in that the method shown in figure 5.6 attempts to be updated solely on the recursive shifts, independently. As previously explained in eq: 3.20, the formula shows this procedure, yet it is compressed into one variable representing all-recursive layers. The idea is to aggregate the uniform initial volume to find an independent RGT. In order to aggregate with respect to the vertical position, it was deemed necessary to sum all previous differences when incrementing in the vertical column, such that each RGT trace continuously increases. However, surprisingly this procedure resulted in fragments that can be interpreted to correlate with a structure similar to the original seismic slice(see figure 5.6b). Nevertheless, the outcome is severely fragmented, if the outcome was continuous, it might be a fascinating merging of RGT, yet the outcome is doubtfully valuable in its current state. The reason for this event is unknown, and it could be an error in implementation that caused this behavior.

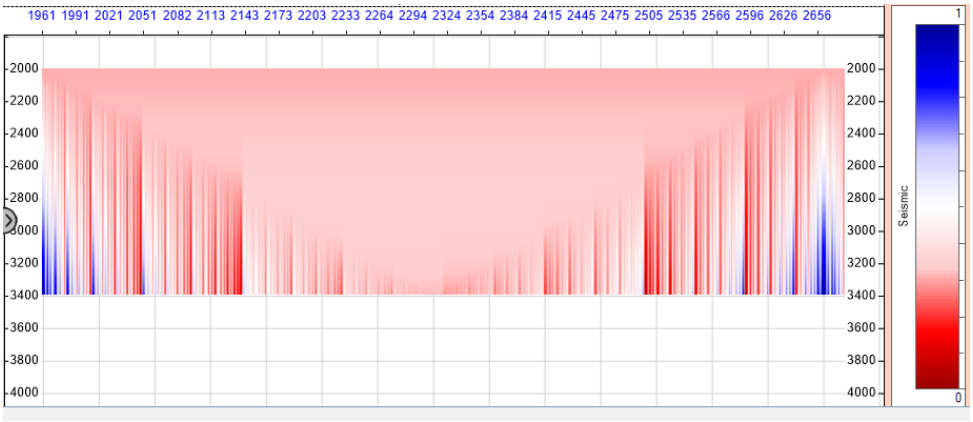
$$RGT_{\text{rec}}[k] = RGT_{\text{rec}}[k] + RGT_{\text{rec}}[k-1],$$

$$RGT[i_{\text{curr}}, j_{\text{curr}}, k_1] = RGT_0[k] + RGT_{\text{rec}}[k_1], \tag{5.2}$$

$$\text{i= 0,...,N ; j = 0,...,M ; k = 0,1,..., L.}$$



(A) Sequentially merged recursive traces



(B) Binary tree controlled order execution of merging RGT result

FIGURE 5.7: Sequential result vs binary tree controlled execution of RGT update

As a potential solution for assessing whether it was possible to define a method in such a manner that it is invariant to the execution order. The attempt was still defined to obey the binary trees found relations amongst trace-pairs and traces. The results shown in figure 5.7a represent execution of the binary tree by left towards right manner exclusively by correcting the initial RGT volume(RGT_0) by the recursive layers(REC_{rec}). The outcome provides a plausible representation of the evaluated structure present in the data, yet there is a drastic level of RGT ages which does not seem plausible. The time distribution seems to follow a trend defined by the binary tree, where the age periodically increases and decreases with respect to the horizontal orientation. Similar effect is present in the execution by binary tree case(figure 5.7b). Could there be a relation in how many recursive layers are represented to describe each trace? Such that traces that are deviated based on more recursive layers are more likely to emphasize their level of RGT. The statement is based on the fact that the outer RGT traces in both images seem to be able to represent an older RGT level which gradually increases(blue colored occurrence in fig **??**), and the RGT level periodically changes horizontally. Despite this crucial issue, it can be determined that both figures are similar in form but with some minor differences in age magnitude. However, the similarity is striking, and it might be stated that the binary controlled definition can provide an orientation invariant merging procedure. Although, it is a prerequisite that the RGT levels are satisfactory.

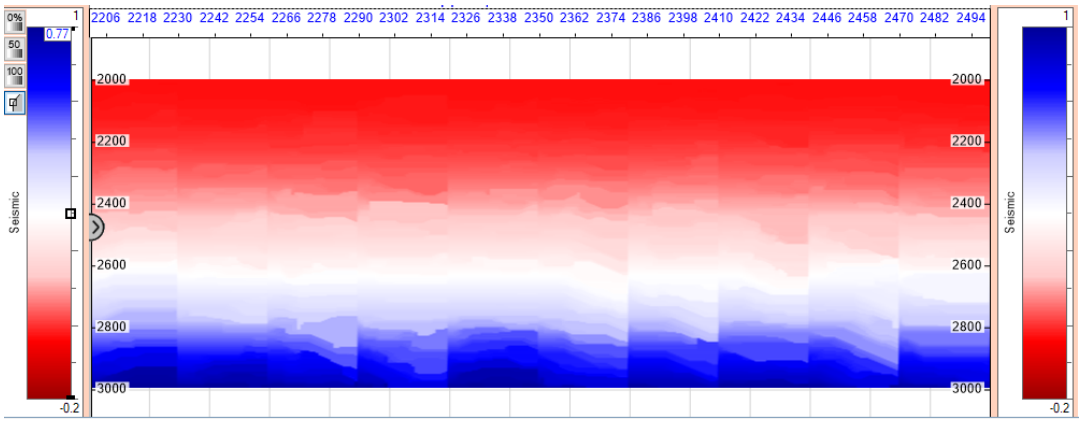
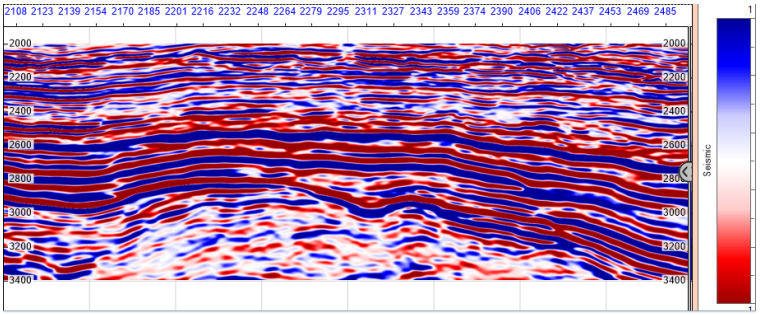


FIGURE 5.8: Image shows multiple independently interpreted chunks of the slice processed by subdivision method definition

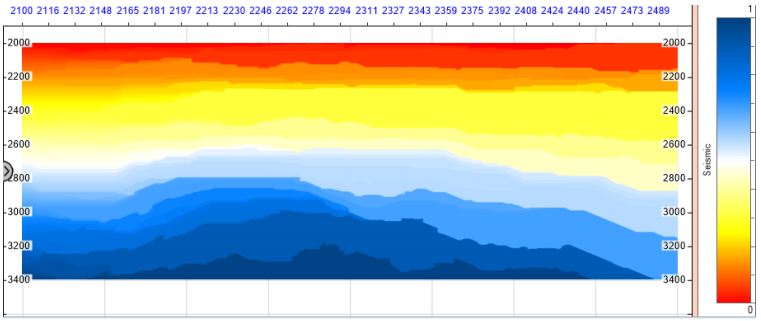
The problem with contours must be solved, leading to the exploration of multiple methods. In many situations, the subdivision method has proved helpful in complex problems. In pursuit of ensuring proficient magnitude of contouring details in the interpreted outcome by the method, subdivision should be attempted. The realization of subdivision leads to the outcome observed in figure 5.8. The division, in this case, was set to consider ten individual chunks, yet it is adjustable to any preference. Reference data is detectable in figure 5.6a. The promising feature noticeable from the outcome (see fig 5.8) is the magnitude of contours; each chunk appears to have a satisfactory amount of contours. That supports a claim that distinct contours are not problematic in this definition.

However, there is a matter of malfunctioning shifts amidst the separate chunks. Since the propagation is not conducted interconnectedly between the chunks, the information from the other chunks is not provided to the assessment. The amount of details is striking and might be correct, yet there is a matter of alignment amongst the chunks to make the outcome sensible. It might be a promising attempt to adjust these shifts retrospectively. That will be considering a new evaluation by DTW warped path from the transition zones (between chunks) in RGT outcome instead of evaluating the seismic data. The found warped path should find the correct shifts between the chunks. Hence, based on the warped path, provide the corresponding shift to the neighboring chunk such that alignment is committed. This approach would lead to two separate DTW evaluations, one for the seismic data and one for the outcome of the RGT in boundary zones. When the second evaluation finds the appropriate shift, the information is not set to be equal to the corresponding warped path relation. Instead, the relative time is updated based on the difference between the warped path relation. This is done in order to avoid distinct RGT updates, preserving its contours.

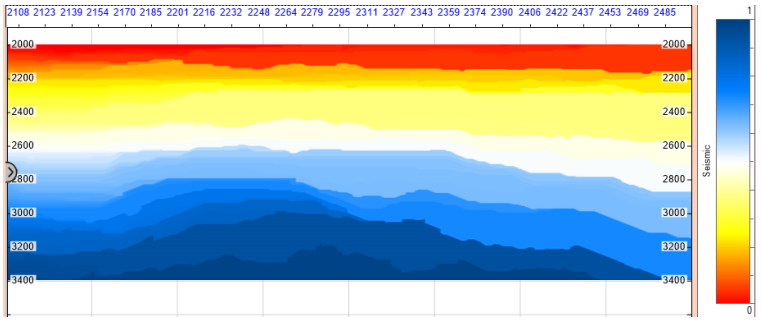
5.1.5 Effect of the reference volume



(A) Reference Volume



(B) Simple Homogeneous reference volume



(C) Phase Unwrapped Volume reference volume

FIGURE 5.9: Difference between homogeneous uniform reference volume vs Phase Unwrapped reference volume.

Discussing the different reference volumes(see figure: 4.1a and 4.1) must be considered whether the selected reference volume implies a significant effect. The options considered in this thesis are a raw phase unwrapped volume and one incrementally increasing homogeneous volume. Raw phase unwrapped volume will directly establish a volume based on the data from the seismic trace, where the homogeneous case is artificially constructed. Assessing figure 5.9, it can be said, with confidence, that the structures retrieved from the warped path remain similar. The mere difference between these two reference volumes is a matter of time magnitude situated at which depth(see the color difference in fig: 5.9). This difference ultimately means that some structures are more visible at the top of the domain(see fig 5.9b). Similarly, more structures are visible near the middle of the domain(see fig 5.9c). The visibility of structures is mainly caused by the coloring, which appears to hide contours, and might be fixed by a different color bar. There is no significant difference between the different reference volumes, and the method is invariant to the choice of the reference volume.

5.1.6 Plausible Method Restrictions

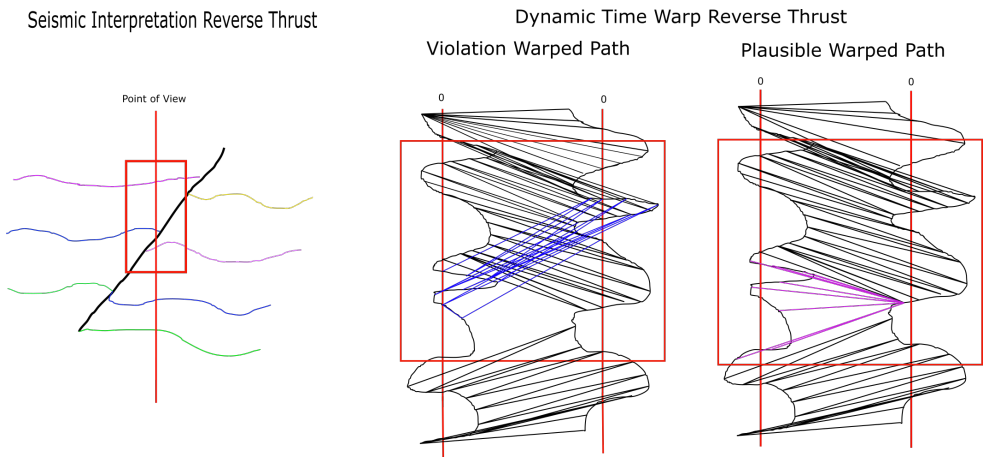


FIGURE 5.10: Left side: Conceptual drawing showing reverse thrust fault interpretation example. Middle: included with example of violation of the dynamic time warping criterion (continuity) in between two seismic traces by allowing crossing warped path. Violating warped path marked in blue. Right side: A probable warped path interpreted by the algorithm, which compresses the path violation path shown in middle image, shown by the purple lines

Would the method implemented be applicable for every feasible physical event present in a geological system? Remember some of the faulting events described in 2.1.2. Some of the events can deposit younger aged strata underlying an older layer; by the very definition of dynamic time warping, it is believed that it is not able to catch this detail due to constraints to continuity (rule: 3.3.1). That explains that the alignment path can not cross prior found warped paths. Thus, it yields that younger aged strata can not be interpreted underneath older strata through the use of DTW due to the very nature of DTW. The procedure will interpret a path with an abrupt jump to the younger layer rather than a younger layer that unfolds by an angle under an older overlying layer. The prohibition induced by DTW limits the number of geological structures the current algorithmic structure can directly interpret. However, it might be sufficient interpretation from the algorithm when interpreting a situation shown in 5.10, only if the algorithm conducts a different warping. That compresses the warped blue path(see fig: 5.10) to the same horizon as above the fault. The warping path will not change until it reaches the oldest horizon encountered so far in the trace underneath the fault, which might illustrate an abrupt drop rather than a time-invariant transition as an interpretation of the reverse fault. The right side of fig 5.10 shows the actual performed warped path by the algorithm. There might be some variation of the DTW procedure that enables a reverse fault to be correctly interpreted. Nevertheless, it is doubtful due to the violation of one of the most deeply rooted conditions for Dynamic Time Warping (see rule 3.3.1). Accounting for this behavior would prerequisite a separate operation, which is not restricted by continuity constraints. That can be used as a correcting factor on the DTW aggregated volume. There might also be a problematic evaluation related to assessing layers that appear in parts of the seismic section. That means horizons that diverge from other horizons and converge into another. Uncertainty is linked to whether the method can apply these new levels of relative times, not just propagating the relative age from the horizon it diverged (see figure 4.15).

5.2 Possible Future Work

5.2.1 Shape Dynamic Time Warping

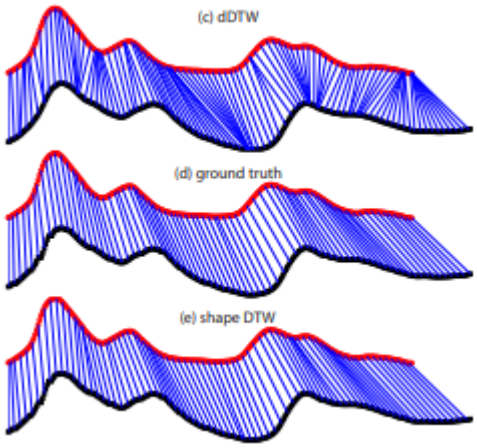


FIGURE 5.11: Shape DTW (Zhao and Itti, 2018a)

Minimizing non-required singularities would be a necessary complement to the method. However, the effect of mitigating the singularities between traces is not as significant as the information being binned, as previously explained(see chap 5.1.4). Furthermore, it can reduce the size of the error associated with the found paths. In the experiment, Zhao and Itti (2018b) proves to be high performing in reducing the obscure singularity occurrence. Without the need for deep domain knowledge associated with selecting the preferable step pattern but will find optimal paths regardless of the step pattern. The method developed handles the concept of singularities by dividing the cost matrix into subsets, so details are more closely assessed as the paths are found. Enhancing the process, it can assess the different subsets by comparing different versions of DTW, such as DDTW, Weighted DTW(a slope constraint DTW), and piece-wise approximations. The algorithm will assure singularity minimization, and satisfactory warped paths are conducted. The methodology employs the concept of dividing the whole cost matrix into several sub-problems. Then passed to the standard procedure(algorithm 2) and assembled into a complete cost matrix. Alternatively, there is an option to employ other versions of the DTW procedure in any specific subsets of the data, such that noise-invariant methods are applied when needed or constraints when needed. The idea is to account for diminishing extreme behaviors in parts of the signal where it is needed. Eventually, the multiple subsequence are merged, and according to the literature, this provides a higher warped path accuracy (Zhao and Itti, 2018b).

5.2.2 Similarity Accumulation Increase

The amount of substance used when evaluating trace similarities can be increased. Thus, it might lead to a better assessment through dynamic time warping because more measurements of how similar the seismic traces are. It could yield a more accurate path amongst the traces due to emphasis on the differences, for example, including more seismic attributes relevant for determining seismic horizons. It enhances the assessment of the warped path within the cost matrix if several attributes populate the same cost matrix. One potential hypothesis was to include a phase unwrapped trace pair for each seismic trace pair and determine the path based on these

two quantities. Therefore, the significant difference is that the path is evaluated not only by the euclidean distance of the signals in the cost matrix but also by the added seismic attribute difference. The type of seismic attribute does not matter, as long it describes some relation/difference in the data.

$$\phi[i,j] = \varphi[i,j] + 2\pi k,$$

$$I[i,j] = V[i,j]\rho[i,j],$$

$$d[i,j] = e[i,j] + \phi[i,j] + I[i,j] + \dots. \tag{5.3}$$

The sum of the included seismic attributes potentially leads to a more weighted outcome for the warped path. By adding all these attributes, there is a potential that the intermediate steps are more likely to not diverge from singularity before the optimal next step can be traversed. In the equation 5.3, $e[i,j]$ coincides with standard euclidean distance. Adding more similarity measures can provide insight into how similar the signals are after exposure to various transformations, which often leads to different gains extracted from the original signal. Hence, the accumulation of all attributes emphasizes the similar pieces of the signals and signifies dissimilarities more significantly, such that the warped path is guided toward the correct shift.

5.2.3 Graph Defined Dynamic Time Warping

Defining the algorithm towards more of a graph structure is believed to positively affect the whole resulting warped path when evaluating the accuracy and result of the algorithm. However, there is a cost in doing the graph transition, and there is a chance that the time complexity becomes even higher. Thus, the importance of developing a more memory-efficient and parallelizable code increases. Because the information has first transitioned to a graph structure, tracking all possible paths in the DTW matrix is necessary to find which path yields the lowest total distance. That means the number of computations strictly depends on how many possible paths are and the duration of these paths. Nevertheless, some factors can bring down the amount of computation by adjusting the methodologies included in the cut algorithm, such as Dinic's algorithm (Fiset, 2022), which is more greedy or guided than the other network traversals. Dinic's procedure is an example of one out of many maximum flow algorithms that readily define multiple flow paths through the network, such that minimum cuts are determined. Might be able to perform early dropout on the found paths that do not yield the optimal path; for example, if an edge with ∞ weighting associated with it would not be a minimum path.

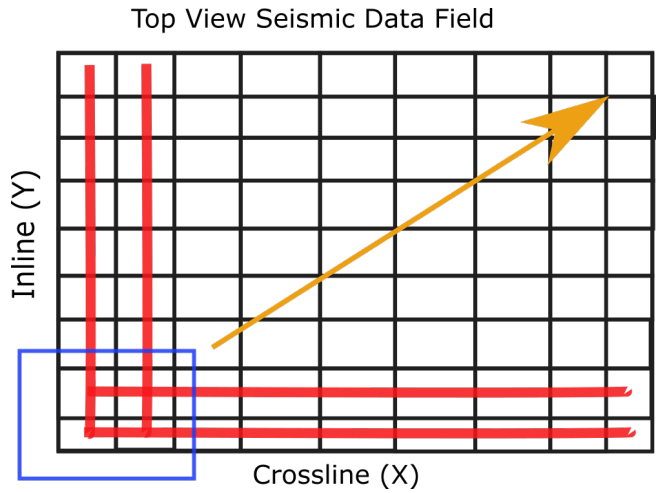
Moreover, there are many different options for a graph representation of dynamic time warping, and many of the ones mentioned in 8.3 provide improvements to the native method. Nevertheless, the optimization framework(see appendix chap: 8.3.4) defined by Deriso and Boyd (2019a) seems to be optimal for enhancing the DTW procedure to yield perfectly warped paths between the seismic traces. The framework causes constraints on the domain, reducing search options around the area where the initial warped path. The method enables further refining steps on the initial warped path(τ^*); this is done by refining grid space around τ^*. The refinement helps converge the solution towards the global optimal warped path. However, this refinement procedure is conducted for a single trace pair. Would it be necessary to apply changes to assure that the multiple optimal warped paths have a spatial relevance in their proximity? (Deriso and Boyd, 2019a). A potential theorem contributing to spatial

dependency is the work by Wang et al. (2016a), which emphasizes the optimal warped path to yield an optimal solution for a series of signal pairs. The dependence is yielded by minimizing the dissimilarity between neighboring warped paths, and the measurement can is found by distance measure($\sum_{(n,m)\in Neig} dist(P_n, P_m)$) between the various options of potential warped paths. It might be possible to incorporate this behavior into the optimization framework by Deriso and Boyd (2019a) to enable an efficient, effective, and yet spatially dependent evaluation of the whole series of the seismic trace pairs. Furthermore, long-term dependencies defined in the graph are covered in the appendix chapter: 8.3.3. Application of neighboring pairs effect might lead to increasingly coupled warps. That contributes to each aggregation of RGT traces to eventually establishing a volume yielding continuous layers. Additionally, it does not conduct non-plausible shifts, and it is invariant which signals are compared against each other, i.e., it potentially always yields an optimal path and respects neighboring shifts.

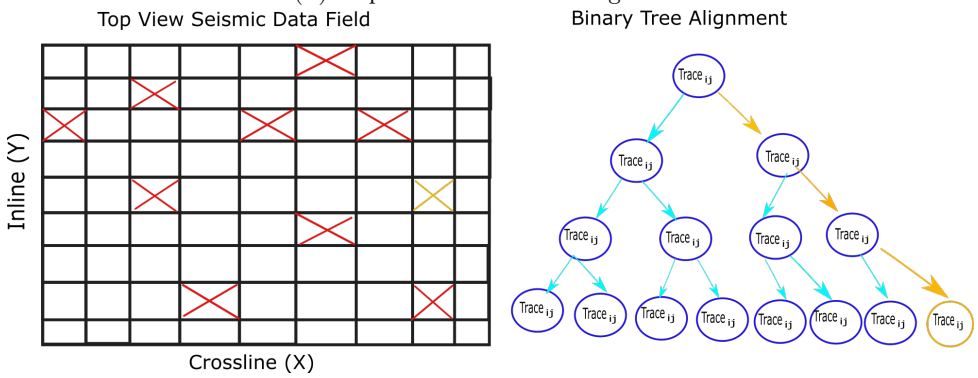
Furthermore, the paper of Uchida et al. (2012a) first described graph notation for dynamic time warping and is used as a foundation for the research by Wang et al. (2016a). One of the covered advantages is establishing non-markovian constraints, which simplistically means that DTW is not constrained to undertake an increment by one, in time, in each iteration. When not influenced by the markovian constraint, it is plausible that deviations within intermediate steps in DTW are not obscuring the optimal path (Uchida et al., 2012a). Because dynamic programming is, in many scenarios, able to find the optimal path, it is not given for every iteration of warped path estimations(see backtracking algorithm 2). Therefore, it is recommended to make the necessary changes to the DTW procedure such that a graph notation is readily solved by max flow, min-cut, or optimization procedures. It is necessary because it yields an optimal warped path in every iteration; additionally, it can be established such that connectivity between different warped paths establishes horizontally continuous layers.

5.2.4 Extension to 3D

The algorithm must be extended to a three-dimensional setting to improve the code's use cases. Because a single slice can only provide a bit of the entire geological structure present in seismic when evaluating the full potential of the prospect. The algorithm must be defined in three dimensions. Since DTW is prone to find local optimums in each of the traces, such that there is a risk associated with independent slice might be drastically different. Therefore, it is a prerequisite for an alignment procedure to ensure plausible continuation amongst the separate slices.



(A) Sequential Slice RGT Alignment



(B) Binary Tree Individual RGT Alignment

FIGURE 5.12: Plausible ideas for extension to the three dimensional domain

There is an assumption in the algorithm's current state that it is possible to repeat the processed slice by slice independently to populate a whole volume, which should be attempted. Nevertheless, each slice should be aligned to stimulate consistent information amongst slices. There are two plausible ideas on how RGT can be globally aligned. One of the methods will utilize the sequentially updated information by aligning seismic slices. The other hypothesis utilizes the binary structure to update each RGT trace globally aligned. Both examples assure that all traces are globally aligned, so significant non-physical oscillations(unrelated to faults) arise from prior interpreted horizons. In figure 5.12, both methods are illustrated, where each grid position allocated in the Seismic data field corresponds to entire RGT traces. First, covering the sequential method, the idea here is to initiate the RGT procedure(covered chap: 3.5) for a single slice in X or Y direction(shown in fig 5.12a). When RGT has updated the first slice, proceed to the next slice but in the opposite direction. First, compare the shifts between the intersecting trace between these two slices and ensure that RGT is similar(i.e., aligned) before propagation in the opposite direction. Furthermore, initiate the next slice in the same orientation as the first slice but incremented by one. However, it intersects the trace in the processed opposite orientated slice, and thus alignment between these slices occurs. The procedure is repeated in the vertical direction, aligned with the intersecting processed slice. These intersections are shown in the blue box marked in fig 5.12a; the procedure repeats until the whole volume is processed, and the orange arrow in the figure illustrates

the given direction of new intersections. However, might it be that proceeding with the three-dimensional alignment in this manner might induce a bias that is prone to masking information in the direction of the update(similar to what's discussed in chap: 5.1.4)? This is not known.

Moreover, could a binary tree structure be a plausible solution in excluding bias present in the 3D setting. The idea of an update is very similar to the concept described in chap 3.4.2, yet this could solve alignment in a cubic volume rather than a two-dimensional slice, but the concept remains the same. Initially, the whole lattice of trace positions would be ordered in a tree-sequential structure, which controls which RGT traces are shifting the specific RGT trace's alignment. The shifts by multiple RGT traces can be thought to align the specific RGT trace globally because each of the shifts is situated at different locations(see orange relations figure 5.12b). Moreover, there are more traces situated close to the evaluated RGT trace. The use of binary tree structure is very similar to what is attempted to make the 2D slice entirely independent of the orientation of execution in this thesis. Whether the binary structure could work for alignment between all traces situated in the field is unknown and needs further testing.

5.2.5 Fixing Error Propagation

One of the most all-consuming struggles when defining this algorithm is how the found warped paths are used on the data set. Because in most the occasions, it seems that the algorithm correctly converges but cannot diverge since the information that should be continued further in the neighboring trace is being hidden within another distinct value. The information being "consumed" by the new values will be forgotten, and the wrong value is traversing the slice instead. Many different hypotheses have been defined to tackle this problem but are unsuccessful at this point. One idea was to use one common reference trace, which means its many prior neighbors are not manipulating the reference trace(see fig 5.1). However, it was not viable because it contained inaccurate displacement between the traces, yet the comparison appeared to contain more contours than its neighbor evaluating counterpart. There appear to be more contours available in the procedure's outcome, yet it requires optimizing the warped path. Because the reference trace and the query trace are allocated further from one another, it leads to higher risks of diverging mistakenly in the intermediate steps in the cost matrix. The elevated risk is caused by more dissimilarity between the two evaluated seismic traces in the DTW procedure, i.e., local optimums are more likely to diverge the path away from the global optimum. The current hypothesis to fix this approach to both contours and correct shifts would always be to define the DTW procedure to yield the correct optimal warped path between time series(see appendix chap: 8.3.4). A moving average filter could be applicable for the purpose of smoothing the adjacent neighboring RGT traces to stimulate the continuity of horizons(in fig 5.1). Employment of the theorem defined by Hale (2013a) could provide a prominent smoothing of the cost matrix.

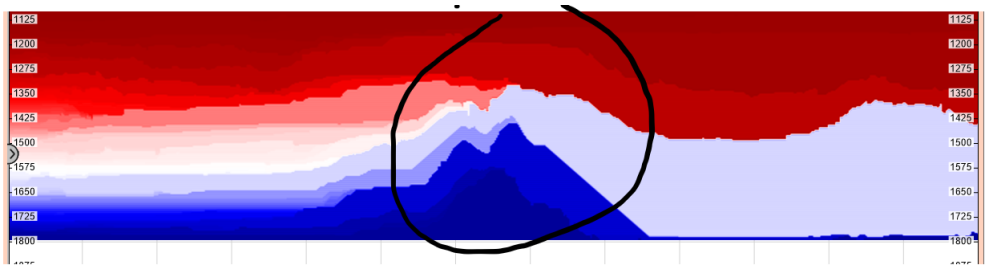


FIGURE 5.13: Marked Intercepting layer in black. Example of prohibited propagation

A plausible theory would change the update's orientation due to the algorithm's current state entirely processing an RGT volume from a trace pair before initiating the neighboring trace pair. It could yield the correct propagation of values by processing each horizontal layer through the whole volume before the next horizontal layer undergoes the same procedure. The error propagation by vertical processing is believed to cause the distinct layers because older-aged layers might be warped to the same layer as younger ones, as seen in fig 5.13 marked in black. The idea is to prohibit older layers from propagating by first processing younger layers from the top and then gradually updating the underlying older layers at the bottom. Hence, the continuation of the younger layers will occur before the older ones are compressed into the same positions as the younger ones. Thus, it potentially yields more contours on both sides of the intercepting layer (black circle fig 5.13).

Alternatively, would a reformulation of the merging procedure be necessary to achieve globally aligned results by splitting the volume into several overlapping subproblems vertically and horizontally, as explained by (Bi et al., 2021a)? The method describes that sub-volumes are RGT aligned individually and, by sharing a common overlapping field, can assess both attempts at interpreting the RGT. Then the global alignment occurs using the overlapping field to align the separately interpreted RGTs. The alignment is conducted by vertical shifts in which the overlapping area is in the vertical direction or linear least squares if horizontal. However, the method seems to serve its purpose, and it could be interesting to attempt a similar methodology to the method described in (Bi et al. (2021a)). However, suppose this was to be attempted. In that case, it is a prerequisite for the volume to be divided into several parts, which would accumulate to be bigger than the original data set because of the multiple overlapping areas.

5.2.6 Spatial dependency

Already in its current state, the algorithm has used information near the current slice to assess spatially weighted/unweighted dependant warpings (4.14). There is also a question regarding whether the spatial information should affect the found warped path for a 3D example; whether the path should be spatially dependent or independent. The research by Shokoohi-Yekta et al. (2015) raises indication if the compared time series is tightly-coupled, in which seismic data is. Thus, multidimensional data assessment for each seismic trace should be dependent(see figure 4.14), if considered. Multidimensional DTW can be solved dependently or independently. Much of the reasoning for or against it is how tightly events are linked to affect its surroundings. Since layers found in the subsurface are significantly affected by their surroundings, there is a solid belief that a spatial dependency would be optimal when linking a slice to its environment. Implementation of spatial dependence is already initiated,

but it might still be ways to have a more profound and efficient linkage to its neighbors. It might yield the option to parallelize the process to improve further the time spent accounting for its neighbors. Furthermore, it is found that amount of computations doubles for every padding. Might it be that more paddings do not necessarily lead to a more rooted solution but instead to an averaged value? However, when comparing the multidimensional assessment to the other version, it might acquire much of the similar interpreted structures as the "cheaper" method(in terms of time complexity). It might be redundant to include a multidimensional version in a 2D slice. Nevertheless, if the DTW establishes differing interpretations for each slice(in a 3D setting), extending to multidimensionality could account for continuity amongst slices. Conclusion: there might be necessary to account for multidimensionality. Each interpretation is more likely to maintain a plausible interpretation for its neighboring slices, ensuring continuity amongst RGT slices is satisfied. Nevertheless, the recent statement is pure speculation, and it requires to be tested in 3D to detect whether it achieves a more sensible RGT amongst slices (Shokoohi-Yekta et al., 2015).

5.2.7 Implementation of the Wheeler diagram

When the RGT volume is finalized in terms of optimal warped path and especially the merging procedure, that will ensure a plausible outcome and sufficient details by contours. The next step is to use the information acquired through the RGT volume to establish a full Wheeler diagram. The diagram, as previously explained, can provide information that yields sedimentological information, such as grain size, sorting, porosity, and other essential reservoir parameters. This information can be predicted on the basis of the stratigraphic layers provided by the Wheeler diagram, such that plausible erosional, non-depositional, and depositional environments are interpreted. Findings by Wu and Zhong (2012b) are assumed to be the most appropriate method for generating a Wheeler diagram out of an RGT volume. This implementation's potential is significant because pure structural information is becoming a supplementing factory in determining sedimentological features of the subsurface.

Chapter 6

Conclusion

The method serves its purpose for automated interpretation of an RGT volume, which can contribute to a rapidly gained first assessment of a seismic data field. Nevertheless, implementation or theory gaps prohibit the method's outcome from being fully trusted. However, the method has much promise, especially its simplicity in definition and how new variations of the same procedure can effectively be implemented into the workflow. Thus, the new additions to the method might yield more adequate results to handle specific behaviors detected amongst the seismic traces. Moreover, there should be an investigation to find a better algorithm to merge the RGT traces into a final RGT volume, such that the correct values are given by utilizing the warped path between all traces. The procedure should be invariant to the ordering of traces. A vital remark is that there might be gaps in the RGT information because of the lack of a sufficient merging procedure. The overall structure seems sensible, yet the more delicate contours are hidden within discrete estimates. Time spent investigating new ways to employ DTW in the seismic cube is recommended; for this thesis, the DTW assessment is done by evaluating trace pairs vertically. Could it be that a horizontal approach to the problem would yield a more satisfactory outcome?

A belief is that more seismic attributes included in the cost matrix similarity measurement could lead to a more corrected path. Furthermore, these seismic attributes can consider any relation defined in Seismology. A reformulation of the dynamic time warping procedure would be a necessity for the method behaves more accurately with reduced time complexity. A potential reformulation of the DTW could be solved as a graph problem, which potentially always yields an optimal warped path. However, the claim has to be explored, yet relevant literature is supplied in the appendix. The algorithm should be defined to work for three-dimensional data when improvements on the merging procedure and the dynamic time warping procedure have proceeded. The method shows potential for estimating an RGT volume. However, the method is still immature, and it is recommended that further research on what kind of Dynamic Time Warping procedure is best suited for finding the warped paths in the subsurface. The most crucial research step is defining a successful merging procedure, being direction invariant, and diminishing how susceptible the merging procedure is to propagating the wrong values throughout the domain. Investigating whether horizontal or vertical sequential updating yields the most profound outcome may be rewarding. The emphasis on Wheeler Diagram is pointed out multiple times during this thesis and is a promising addition to the method. Moreover, the method provides a fair interpretation of the subsurface, but the merging procedure and warped path optimization must be further researched.

Chapter 7

Bibliography

Figure References

Amini, A. (2022b). Introduction to deep learning. *MIT*, *6*, S191.

Bi, Z., Wu, X., Geng, Z., & Li, H. (2021b). Deep relative geologic time: A deep learning method for simultaneously interpreting 3-d seismic horizons and faults. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB021882.

Deriso, D., & Boyd, S. (2019b). A general optimization framework for dynamic time warping. *arXiv preprint arXiv:1905.12893*.

Dyrendahl, M. M. (2018b). *A better horizon auto tracker-powered by machine learning* (Master's thesis). NTNU.

Farmanbar, M. (2022). Lecture notes in data mining and deep learning.

Gauriau, R. (2015b). *Shape-based approaches for fast multi-organ localization and segmentation in 3d medical images* (Doctoral dissertation). Telecom ParisTech.

Geler, Z., Kurbalija, V., Ivanović, M., Radovanović, M., & Dai, W. (2019). Dynamic time warping: Itakura vs sakoe-chiba. *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 1–6.

Geng, Z., Wu, X., Shi, Y., & Fomel, S. (2020b). Deep learning for relative geologic time and seismic horizons. *Geophysics*, *85*(4), WA87–WA100.

Giorgino, T. (2009a). Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of statistical Software*, *31*, 1–24.

Keogh, E. J., & Pazzani, M. J. (2001a). Derivative dynamic time warping. *Proceedings of the 2001 SIAM international conference on data mining*, 1–11.

Mineetha. (2020). Bias-variance tradeoff [[Online; accessed April 09, 2022]]. https://mineetha.com/2020/06/25/bias-variance-tradeoff/

Oregi, I., Pérez, A., Ser, J. D., & Lozano, J. A. (2017a). On-line dynamic time warping for streaming time series. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 591–605.

Qayyum, F., Catuneanu, O., & de Groot, P. (2015). Historical developments in wheeler diagrams and future directions. *Basin Research*, *27*(3), 336–350.

Rigzone. (2022). How does marine seismic work? [[Online; accessed April 09, 2022]]. https://www.rigzone.com/training/insight.asp?insight_id=303&c_id=

Savalia, S., & Emamian, V. (2018). Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks. *Bioengineering*, *5*(2). https://doi.org/10.3390/bioengineering5020035

Schlumberger. (2014). The defining series: Beginner's guide to seismic surveying [[Online; accessed April 09, 2022]]. https://www.slb.com/resource-library/oilfield-review/defining-series/defining-seismic-surveying

Spooner, M., Kold, D., & Kulahci, M. (2017b). Selecting local constraint for alignment of batch process data with dynamic time warping. *Chemometrics and Intelligent Laboratory Systems*, *167*, 161–170.

Tavenar, Romain. (2022). An introduction to dynamic time warping [[Online; accessed April 10, 2022]]. https://rtavenar.github.io/blog/dtw.html

Uchida, S., Fukutomi, M., Ogawara, K., & Feng, Y. (2012b). Non-markovian dynamic time warping. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2294–2297.

Wang, Y., Miller, D. J., Poskanzer, K., Wang, Y., Tian, L., & Yu, G. (2016b). Graphical time warping for joint alignment of multiple curves. *Advances in Neural Information Processing Systems*, *29*.

Wu, X., & Zhong, G. (2012a). Generating a relative geologic time volume by 3d graph-cut phase unwrapping method with horizon and unconformity constraints. *Geophysics*, *77*(4), O21–O34.

Wu, X., & Zhong, G. (2012c). A method for generating a seismic wheeler volume via a relative geologic time volume. *2012 SEG Annual Meeting*.

Zhao, J., & Itti, L. (2018a). Shapedtw: Shape dynamic time warping. *Pattern Recognition*, *74*, 171–184.

References

Ali, S., Khan, H., & IdrisShaik, F. A. (n.d.). I-puma: Fast phase unwrapping via ibfs graph cuts. *International Journal of Engineering and Technology*, *7*, 254–265.

Al-Naymat, G., Chawla, S., & Taheri, J. (2012). Sparsedtw: A novel approach to speed up dynamic time warping. *arXiv preprint arXiv:1201.2969*.

Amini, A. (2022a). Introduction to deep learning. *MIT*, *6*, S191.

Bi, Z., Wu, X., Geng, Z., & Li, H. (2021a). Deep relative geologic time: A deep learning method for simultaneously interpreting 3-d seismic horizons and faults. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB021882.

Buland, A. (2021). *Seismic amplitude analysis and inversion*. Equinor.

Chakravorty, A. (2022). Lecture notes in reinforcement learning.

Compton, S., & Hale, D. (2014). Estimating vp/vs ratios using smooth dynamic image warping. *Geophysics*, *79*(6), V201–V215.

Deriso, D., & Boyd, S. (2019a). A general optimization framework for dynamic time warping. *arXiv preprint arXiv:1905.12893*.

Dyrendahl, M. M. (2018a). *A better horizon auto tracker-powered by machine learning* (Master's thesis). NTNU.

Gauriau, R. (2015a). *Shape-based approaches for fast multi-organ localization and segmentation in 3d medical images* (Doctoral dissertation). Telecom ParisTech.

Geng, Z., Wu, X., Shi, Y., & Fomel, S. (2020a). Deep learning for relative geologic time and seismic horizons. *Geophysics*, *85*(4), WA87–WA100.

Gens, R. (2006). Phase unwrapping. *InSAR and its applications, University of Alaska Fairbanks*.

Giorgino, T. (2009b). Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of statistical Software*, *31*, 1–24.

Gunn and Patel. (2017). Masters/doctoral thesis [[Online; accessed May 02, 2022]]. https://www.latextemplates.com/template/masters-doctoral-thesis

Hale, D. (2013a). Dynamic warping of seismic images. *Geophysics*, *78*(2), S105–S115.

Hale, D. (2013b). Dynamic warping of seismic images. *Geophysics*, *78*(2), S105–S115.

Huso, E. A. (2020). *Generative adversarial networks for seismic interpretation* (Master's thesis). NTNU.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems*, *7*(3), 358–386.

Keogh, E. J., & Pazzani, M. J. (2001b). Derivative dynamic time warping. *Proceedings of the 2001 SIAM international conference on data mining*, 1–11.

Oregi, I., Pérez, A., Ser, J. D., & Lozano, J. A. (2017b). On-line dynamic time warping for streaming time series. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 591–605.

Qayyum, F., Betzler, C., & Catuneanu, O. (2017). The wheeler diagram, flattening theory, and time. *Marine and Petroleum Geology*, *86*, 1417–1430.

Shokoohi-Yekta, M., Wang, J., & Keogh, E. (2015). On the non-trivial generalization of dynamic time warping to the multi-dimensional case. *Proceedings of the 2015 SIAM international conference on data mining*, 289–297.

Spooner, M., Kold, D., & Kulahci, M. (2017a). Selecting local constraint for alignment of batch process data with dynamic time warping. *Chemometrics and Intelligent Laboratory Systems*, *167*, 161–170.

Stark, T. J. (2004). Relative geologic time (age) volumes—relating every seismic sample to a geologically reasonable horizon. *The Leading Edge*, *23*(9), 928–932.

Uchida, S., Fukutomi, M., Ogawara, K., & Feng, Y. (2012a). Non-markovian dynamic time warping. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2294–2297.

Valadao, G., & Bioucas-Dias, J. (2009). Cape: Combinatorial absolute phase estimation technical report, it/ist communications theory and pattern recognition group.

Vesnaver, A. (2017). Instantaneous frequency and phase without unwrapping. *Geophysics*, *82*(1), F1–F7.

Wang, Y., Miller, D. J., Poskanzer, K., Wang, Y., Tian, L., & Yu, G. (2016a). Graphical time warping for joint alignment of multiple curves. *Advances in Neural Information Processing Systems*, *29*.

Weibull, W. W. (2020). Lecture notes in methods in reflection seismic.

Wu, X., & Zhong, G. (2012b). Generating a relative geologic time volume by 3d graph-cut phase unwrapping method with horizon and unconformity constraints. *Geophysics*, *77*(4), O21–O34.

Wu, X., & Zhong, G. (2012d). A method for generating a seismic wheeler volume via a relative geologic time volume. *2012 SEG Annual Meeting*.

Zhao, J., & Itti, L. (2018b). Shapedtw: Shape dynamic time warping. *Pattern Recognition*, *74*, 171–184.

Chapter 8

Appendix

8.1 Phase Unwrapping RGT solution

8.1.1 Graph Cut

Graph cut is a method which was used to solve one of the phase unwrapping instances of RGT to control binary labeling, by this very reason, more in depth theory shall be discussed to clarify its functionality. Graph structures are represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,i.e., a graph($\mathcal{G}$) is built by series of vertices/nodes ($\mathcal{V}$) with interconnecting edges($\mathcal{E}$) that illustrate the relations between the nodes. As the name suggests, graph cut's purpose is segmentation of graph structure, frequently used for segmentation of images as can be seen in illustrative figure 8.1. The segmentation is performed on the basis that two separate selected nodes are used as source(s) and sink(t), the objective of the method is to partition the data into two subsets, for example T(orange segment in fig 8.1) and S(green segment in fig 8.1). Segmentation is performed on the basis minimizing the energy, this is readily done by max flow/min cut algorithms such as Edmonds-Karp, Ford Fulkerson, Dinic's algorithm and Capacity scaling, without going much into detail regarding each of these algorithm it can be stated that these find several paths from source to sink by augmentation of residuals cost between nodes(Fiset, 2022). Utilization by these max flow algorithm enables detection of where the minimum cut can be conducted, which enables maximum energy in both subsets (T and S).

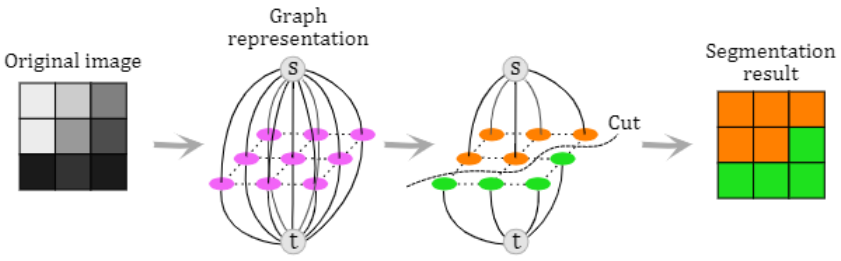


Figure 2.5: Graph-cut segmentation principle.

FIGURE 8.1: Graph cut segmentation example, shows the overview of graph cut procedure for segmentation of images. (Gauriau, 2015b)

The association between graph cut and phase unwrapping have received attention by some authors, such as (Gauriau, 2015a), (Ali et al., n.d.), (Valadao and Bioucas-Dias, 2009), and (Wu and Zhong, 2012b). The research conducted by (Wu and Zhong, 2012b) links the phase unwrapping procedure to the Bayesian theorem, specifically maximum a posteriori (MAP) estimation that is transformed into an L^P

norm problem, by introduction of prior information in combination likelihood functions. L^P norm is defined to have discontinuity-preservering ability within the interval of $(0 \leq P < 1)$. Prior information is extracted from first order Gauss-Markov field to map the absolute phase and likelihood function constructed based on data-model mechanism. L^P norm is in this instance solved by graph cut based on min cut/max flow optimization, according to the studies this appears to have sufficient denoising capabilities. This particular method are a parametric method, where the absolute phase is compared to estimates of the absolute phase. (Wu and Zhong, 2012b)

$$L^p = |\Delta\varphi|^P. \tag{8.1}$$

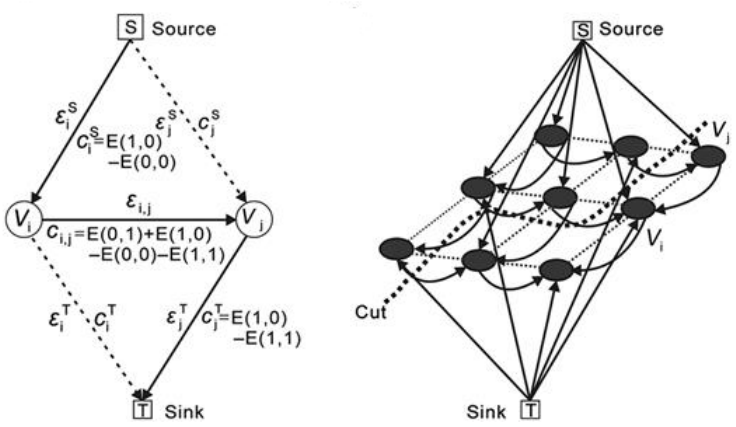


FIGURE 8.2: Graph cut in 2D settings for minimizing the energy amongst segments, edges are weighted by the energy in the wrapped phase in vertical and horizontal direction (Wu and Zhong, 2012a)

As for the optimization procedure of the method, there is defined an energy function which is to be minimized by the graph cut procedure to mitigate energy difference in the cut, thence yielding the correct segmentation of binary labeling (selection of correct k from eq 3.11). As for the usage of L^P norm, it is utilized for evaluation of difference amongst the gradient for integers$(k_{i+1}-k_i)$ and wrapped phases$(\phi_{i+1}-\phi_i)$. The amount of gradient present at each position is dependent on the number of dimensions being assessed, for example, if two dimensions are to be assessed, it is a prerequisite that the horizontal and vertical gradients are included:

$$d^h = 2\pi(k^x_{i,j+1} - k^x_{i,j}) + \varphi^x_{i,j+1} - \varphi^x_{i,j}, \tag{8.2}$$

$$d^v = 2\pi(k^y_{i+1,j} - k^y_{i,j}) + \varphi^y_{i+1,j} - \varphi^y_{i,j}. \tag{8.3}$$

These gradients are used to build energy function, which is to be minimized by utilizing graph cut:

$$E^h(\delta_{i,j+1}, \delta_{i,j}) = \mathcal{W}^h_{i,j}|2\pi(\delta_{i,j+1} - \delta_{i,j}) + d^h|^P, \tag{8.4}$$

$$E^v(\delta_{i+1,j}, \delta_{i,j}) = \mathcal{W}^v_{i,j}|2\pi(\delta_{i+1,j} - \delta_{i,j}) + d^v|^P, \tag{8.5}$$

$$K^{n+1} = argmin_K \sum_{i,j\in S} E(\delta_i, \delta_j), \tag{8.6}$$

$$\hat{\phi} = \varphi + 2\pi\hat{k}. \tag{8.7}$$

Now, multiple formulas are defined which origins from the principle Gaussian distribution and MAP theorem, more details regarding derivation and additional information considering numerous added constraint onto the graph can be found in original paper of Wu and Zhong, 2012b. MAP theorem derives the origin for the included energy functions given in eq: 8.2,8.3,8.4,8.5, and 8.6. Thus, the possibility for binary minimization is present. For discussion regarding the particular use of these formulas can be explained such that the overall objective is to find the most probable series of $\hat{k}$ estimates, in order to assure unwrapping the phase in such a manner that mitigates the error propagation, i.e., less error propagation is prone to less energy in objective function. Weights(W) are present in energy functions to control its significance, whether the how significant the impact of the vertical or horizontal component is. Within equation 8.6 ($E(\delta_i, \delta_j)$)all of the divided component(vertical or horizontal) such that the next integer input(K^{n+1}) is asserted by the use of the lowest possible total energy output from the energy function, additionally the δ parameter controls if phase nodes are situated at source side(value 0) or sink side(value 1) of the cut seen in fig 8.2. Eventually, a absolute phases($\hat{\phi}$ are found by using the wrapped phase added with 2π which is incremented by the estimated integer list($\hat{k}$). Formula 8.6 are also used in the graph notation to perform a minimum cut, such that 4 different opinions are viable in a two dimensional setting (see figure 8.2), namely:

$$E_{i,j}(0,0) = W_{i,j}|d_{i,j}|^p, \tag{8.8}$$

$$E_{i,j}(1,1) = W_{i,j}|d_{i,j}|^p, \tag{8.9}$$

$$E_{i,j}(0,1) = W_{i,j}| - 2\pi + d_{i,j}|^p, \tag{8.10}$$

$$E_{i,j}(1,0) = W_{i,j}|2\pi + d_{i,j}|^p. \tag{8.11}$$

Utilization of these separate outcomes of the energy function provides a possibility to establish a cost to the edges($\mathcal{E}$) between the nodes($\mathcal{V}$), this determines a relation between the phase nodes situated in the graph ($\mathcal{G}$) such that energy minimization can be conducted. Moreover, in the article which these theorem have been discussed, it was implemented a three dimensional definition which will leads to a more accurate result than which is described by (Valadao and Bioucas-Dias, 2009). The overlap from two dimensional to three dimensional includes an additional term to the energy function, and now the energy function respects three directions in its objective, not only horizontal(i) and vertical(j) components but also its new direction (k). According to the author, more dimensions lead to a better cut in the graph; as the dimension is extended, it leads to a cut using several elementary graphs rather than one single graph for a two-dimensional setting(seen right side fig 8.2). Moreover, the authors have added horizon and unconformity constraints in the cost of the edges($c^s \quad \text{AND} \quad c^t == \infty$) such that the cuts restrained from including non-physical cuts. There are more details described in the paper, and it is recommended to assess the literature of the author (Wu and Zhong, 2012b).

8.2 Deep learning Method

8.2.1 Neural Network

A methodology that is rising in use cases and popularity in recent years is the concept of machine learning, and one of its sub-method is Neural networks. Neural networks

are also known as one of the most flexible methods for any situation, and it means that a trained network with enough complexity can closely approximate non-linear patterns. The origin goes back to the 1960s, yet it is still very relevant today. The core concept is to replicate how the human brain functions through the distributed signals, namely neurons; these are represented as perceptrons in the neural network, and each singular node has a bias and weight associated with it. A single node does not provide much useful prediction in classification nor regression. However, as the sum of all of the neural layers(including hidden layer and input layer) given to an activation function, the contribution of every perceptron has a quite adaptable and powerful predictive effect on almost any given situation.

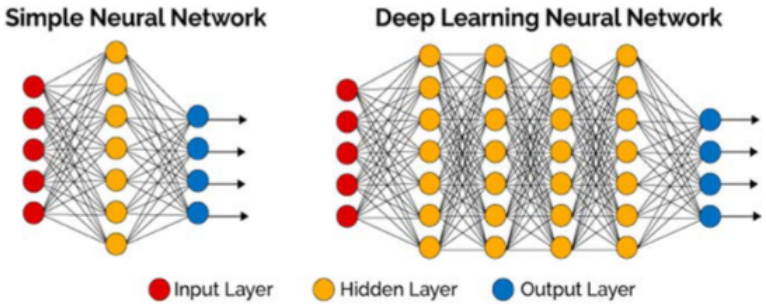


FIGURE 8.3: Neural Network (Savalia and Emamian, 2018)

As shown in 8.3, the network is built upon either one singular layer of perceptrons or multiple layers of perceptrons, and one layer is built by input nodes that take in the information. After each of the depending layers has been updated through back-propagation, it will be passed into an output layer depending on variation whether its regression or classification problem. A prediction of the value or the outcome would be the output of the final layer, and the information is used through the use of an activation function, that is, several variants available. One of these is a soft-max, linear, tangent, or Rectified linear unit Relu; as the name suggests activation function selects which of the neural predictors are active in each case. The activation layer is, to some degree, used as a way to standardize the predictors going into the output layer of the network for comparison of the predictors' effect on the network related to the multiple other predictors inserted. Furthermore, the activation layers also enable the possibility of constructing non-linear boundary settings for the network. The activation layer assigns a smooth, gradual transition(eq: 8.15, 8.16) towards a different prediction after a certain threshold or enables activation(eq: 8.17) of the predictor value after a given threshold to either provide a probability or regression contribution. Performance evaluation is typically done by measuring performance against the real data, readily done by either mean square error(MSE) or cross-entropy, where MSE is typically used to measure quality in regression and cross-entropy for classification settings. Based on the magnitude of performance, this information is backpropagated through the network to adjust weights and biases to minimize the offset or the number of misclassifications.

$$MSE = \frac{1}{2}\sum_{i}(\hat{y} - y)^2, \tag{8.12}$$

$$C_{cross-entropy} = \frac{1}{n}(y \ln \hat{y} + (1 - y)\ln(1 - \hat{y})). \tag{8.13}$$

Examples of activation layers:

$$f(x) = x, \tag{8.14}$$

$$f(x) = \frac{e^x}{1 + e^x}, \tag{8.15}$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{8.16}$$

$$f(x) = max(0, x). \tag{8.17}$$

Neural network is also a fundamental part of building a deep learning architecture, where deep learning is a composition of various neural network frameworks that forms a greater joint architecture. Thus, the training stage becomes increasingly extensive to train all the adjustable layers composited together, yet the prediction performance typically becomes better if the hyperparameter(γ) optimization is conducted correctly. Previously, it was mentioned that cost evaluation (8.12 and 8.13) is typically used in regression, and cross-entropy (8.13), which appraises classification problems, these are an essential metric used for the back-propagation part, which adjusts the hidden layers in the structure. Usually, gradient descent is a fundamental method that controls the learning algorithm to move weights toward their global optimums. The selection of learning parameter(γ) dictates what result is obtained by the network, because the speed controls how rapidly and whether the the found minima is a local minimum or a global minimum. A too big learning rate can cause the network to obtain a local minimum, where a too small learning rate, much more time than needed to converge at the global or local minima. Always, a risk in gradient descent in not obtaining the global minimum (Chakravorty, 2022).

$$\nabla C = \frac{\partial C(w, b)}{\partial w_{i,j}} = \frac{\partial C(w, b)}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial x_j} \frac{\partial x}{\partial w_{i,j}}, \tag{8.18}$$

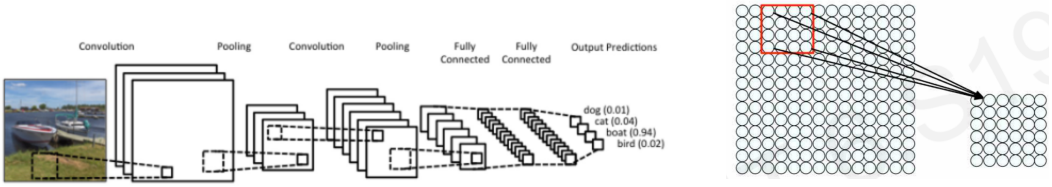
$$w_i = w_{i-1} - \gamma \nabla C(w_{i-1}). \tag{8.19}$$

Eventually, the network weights become more iteratively refined as the weights through backpropagation and sequentially as the cost(∇C) becomes lower and lower. Hence, this indicates that the weights are adjusted closer to the target prediction, which means that the inserted predictors would yield a better approximate a better prediction by the network $\hat{y} = f(\sum w_i x_i + b)$. For example, it can be observed that all predictors(x) are multiplied by their trained weight(w) and summed and added by its bias (b) which results in a prediction, and $f()$ corresponds to the select activation function.

8.2.2 Convolutional Network

Much related to 8.2.1 Neural network, the convolutional neural network (referred to as CNN) is replicating yet another natural operation, namely how mammals perceive information through the visual cortex. CNN is generally used in image classification, object detection, etc. This means the trained model is passed images, videos, or data, and images are, for example, built by numbers $[0, 255]$ RGB values. The combinations amongst RGB values contribute to the transfer of how humans perceive contours, shading, and colors to the machine. The understanding of intensity and color is often used to predict an outcome, segment, or classify an image, and the machine determines these details from the adjacent pixels present in each image. Perceive the whole

information, further convolve information into a more densely packed matrix, and further pack the perceived information into several increasingly denser pooling layers; this is also known as the encoding stage of CNN. These pooling layers or filters are useful for transmitting high-level images to low-level details, enhancing the model's ability to distinguish various objects. Additionally, the pooling layers are helpful tricks to reduce and compress more information more densely in a reduced amount of pixels. Combined with the activation functions, the convolution process slides over all possible pixels by defined filter size. Then activation functions(mentioned in chapter 8.2.1) are used to account for non-linearity, often for removing negative estimates after the convolving of the information. Activation function are often conducted before any pooling operation such that non-linearity is respected before being compressed (Amini, 2022a).



(A) Convolutional Network Workflow (Farmanbar, 2022)

(B) Convolution process (Amini, 2022b)

FIGURE 8.4: The CNN overview

The trainable part of the network is namely the weight($w_{i,j}$) which controls the convolutional operation($\sum_i^n \sum_j^m w_{i,j} x_{i+p,j+q} + b$) which maps the input image into an feature map, and this feature map is affected by the weight's influence. The m and n corrosponds to the defined filter size, and p and q keeps tracks of the position in which the convolutional procedures has traversed at any given point, and lastly b accounts for the bias of the network. .The feature map is then pooled into a denser layer, such that low-level features represent the whole image in a reduced space(see figure8.4b). Depending on the architecture, the network could yield multiple convolutional layers and pooling layers until the network is represented in a fully connected latent space that functions similarly to a neural network(explained in chap: 8.2.1). The objective of the CNN dictates if the object detection, image classification, etc, are to be conducted; depending on the objective, the architecture is adjusted accordingly. For example, in a classification scenario, the images are convolved and pooled until the information is contained in a flattened vector (see figure 8.4a). The end-stage, which is named as fully connected layers in figure 8.4a, refers to a neural network architecture (explained in chapter 8.2.1). The fully connected layers are by soft-max function assigning probabilities of the different possible classifications, where the outcome that yields the highest probability is selected as the outcome of CNN. The iterative compacted and convolved information from the image is readily passed into the neural network, in which the network determines the outcome of the CNN procedure when performing a prediction. When still in the training phase, the weights of each convolution operation are adjusted to approximate the target variable.

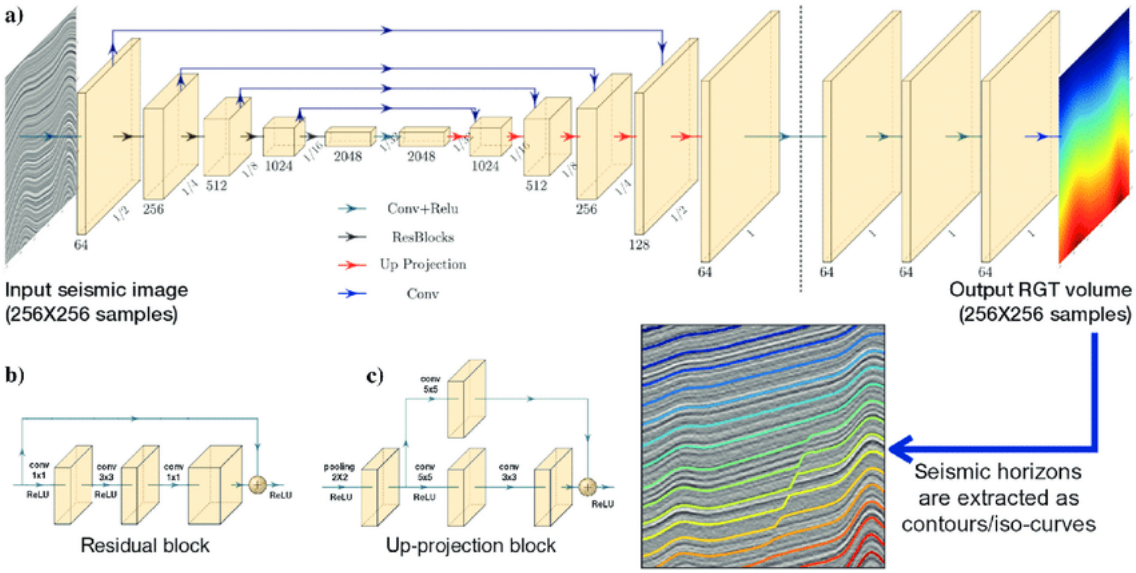


FIGURE 8.5: Convolutional Encoder-Decoder Network Framework related to RGT estimation(Geng et al., 2020a)

Furthermore, as mentioned, different objectives yield different architecture; for example, an encoder or decoder type of framework is often used for regression or segmentation in a CNN setting and does resemble a General Adversarial Network similar architecture(more explained chap **??**). The concept related to this architecture ultimately means that information is taken from high-level images and iteratively converted to low-level fragments. The encoder process has to be fulfilled after the flattened layer is achieved. For the transformation of the low-level detail to the complete transformed image again, it is necessary to include a part of the network which predicts and transforms information from low-level details to a high-level representation; this is known to be the decoder stage of the procedure. An encoder-decoder network requires weights for the different weights in the encoder to be trained, and similarly, weights for decoding the low-level input require to be trained by assessing the proportion of dissimilarity between the target variable and the predicted outcome of the decoder. The encoder and decoder can be treated as individual neural networks for this purpose, where the loss (MSE or Cross-entropy) governs adjustments induced onto the weights. Figure 8.5 showcases the capability that an encoder, decoder network structure achieves and also relates to the specific problem for this thesis, constructing an RGT volume. The input is a high-level image of a seismic cross-section in the image, where the information iteratively is pooled and filtered into several pieces. The vector containing these low-level features is then exposed to a convolution procedure that upscales/decodes/up-projection the low-level features to become a final high-level image segmentation of the seismic image. After the up-projection, further refinements are committed to the output image of the network, i.e., a third separate network architecture is trained to predict an RGT volume (observed in fig 8.5).

8.2.3 Generative Adversarial Network

The perspective of networks that are able to establish artificial data from a different input is known as generative adversarial networks(GAN). GANs often occur in image, speech, and video manipulation, which is widely employed in various fields of expertise, social media, and nefarious schemes. The architecture for GANs concerns

several parts that enable data transformation trained to replicate a given target input in the training phase. What is imperative for the method is training a suitable Generator (G in fig: 8.6), meaning the generator can successfully transform any inserted information to match real data on which it has been trained. The generator is trained based on the backpropagated information (eq: 8.20) regarding performance from the Discriminator(D in fig: 8.6), the discriminators task is to differentiate between fake and real data inserted to the model. Initially, generator performance is inferior in replicating real data, yet, as time progresses, the generator improves in synthesizing real data. Thus, it becomes increasingly more challenging for the Discriminator to differentiate between what is real and what is fake (Amini, 2022a).

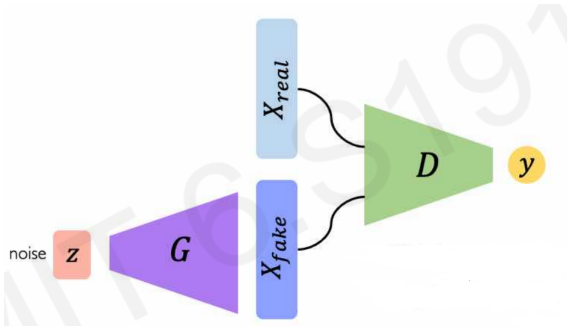


FIGURE 8.6: Generative Adversarial Network (Amini, 2022b)

$$\mathcal{L} = arg\,min_G\, max_D\, \mathcal{E}_{x,z}[\log D(G(z)) + \log(1 - D(x))]. \tag{8.20}$$

When the training phase concludes, the Generator is adapted to transform noise or data to be situated in the same probability distribution as the real or target data is contained within. Hence, any data transformation will be adjusted to be very similar to the actual data. Mainly, GANs are growing as a method for constructing networks that can reconstruct images, indicating that the method could prove helpful in automating the seismic interpretation procedure. This is shown in the work of (Huso, 2020), where a geological model is established from the architecture of a GAN. The loss shown in equation 8.20 shows how information is updated throughout the network, $argmin_G$ indicates the set of coefficients that yields the smallest occurrence of G(z). Likewise, the maximization of the D, the best attempt to trick the Discriminator, is back-propagated. Hence, the Generator is becoming more and more similar to the real data set. Through the use of the logic present in GANs, it is possible to replicate any real pattern, faces, and potentially also geological feature, as (Huso, 2020) has demonstrated a specific use of GAN in the geological setting.

8.3 Dynamic Time Warping Extensions

8.3.1 Keogh Lower Band

As previously mentioned in (3.3.2 Sakoe-Chiba Band), global constraints tend to speed up similarity search yet may not contain the globally best path. As a consequence of this effect, there are many attempts by the research community to find a global constraint that enables greedy search in addition to containing the global optimum; one of such attempts is the Keogh lower band. First, the reference signal X should be used to reference the globally defined constraint, which determines the

band based on the latest window size(r). The upper(U_i) and lower(L_i) bounds are extracted by employing a band structure either from Sakoe-Chiba band or Itakura band(see fig 3.6). The selection of band structure and allowed offset determines how tightly bounded the cost matrix is evaluated.

$$\begin{aligned} U_i &= max(X[i-r:i+r]) \\ L_i &= min(X[i-r:i+r])' \end{aligned}$$

$$LowerBound(X,Y) = \sqrt{\sum_{i=1}^{n} \begin{cases} (Y_i - U_i)^2 & if \quad Y_i > U_i \\ (Y_i - L_i)^2 & if \quad Y_i < L_i \\ 0 & otherwise. \end{cases}} \tag{8.21}$$

In (8.21) U_i and L_i is the respective Lower and Upper bounds of signal X; this determines how distance measure are increased by extending beyond U_i or L_i. The lower bound is not yielding much value for this thesis; the use case for this theorem is to rapidly compare multiple time series inside a database to assess the highest similarity by assessing the summed distance(found by violation of bounds).(E. Keogh and Ratanamahatana, 2005)

8.3.2 Non Markov

The objective of establishing and defining the problem in the non-markovian domain opens possibilities to optimize further the found the best path of the algorithm 2 in highly non-linear settings. A Markovian domain is confined to the obey small timesteps when iterating towards a solution, where markovian characteristics are a result of the problem being solved by dynamic programming. Transitioning towards a non-markovian implementation would mean it is necessary to define the 2 DTW algorithm in a graph structure. The main benefit of the objective being defined as a minimum cut problem rather than a dynamic path optimization problem is that the path following optimization may lead to an over-fitted solution. However, a cut optimization would yield a global optimum and is less affected by a local variation, which is found by different variations of DTW (Uchida et al., 2012a)

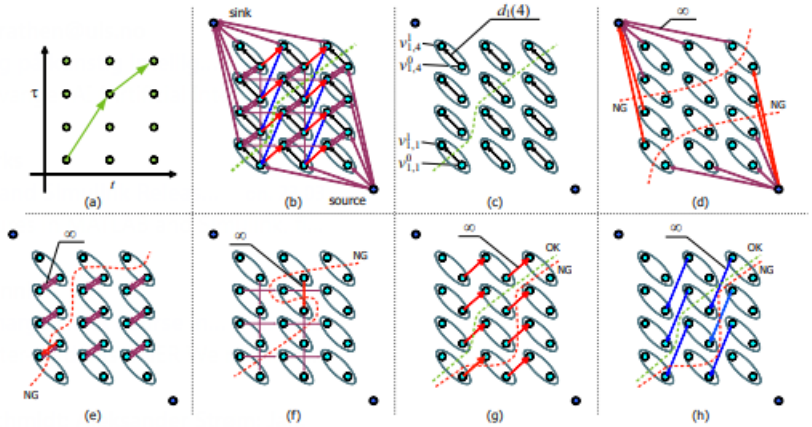


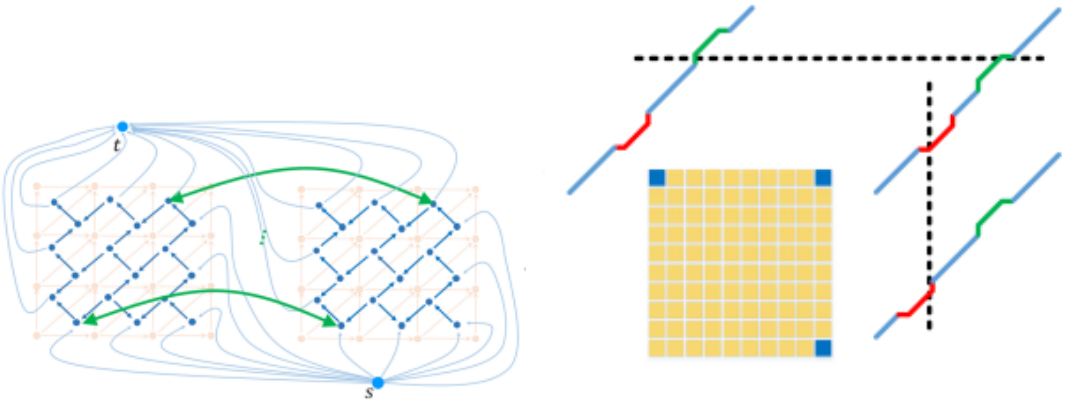
Figure 2. (a) DTW as a path optimization problem. (b) DTW as a graph cut problem. (c) Edges for $d_i(u_i)$. (d) Edges for boundary constraints. (e) and (f) Edges for excluding unexpected cuts. (g) Edges representing $\alpha_{t,t+1} = 1$. (h) Edges representing $\beta_{t,t+1} = 2$.

FIGURE 8.7: Illustrative Non markovian (Uchida et al., 2012b)

The overfitting problem occurs because the markovian constraints have to gradually transition towards the next apparent similarity, whereas non-markovian constraints will skip directly towards the similarity, not conducting several preliminary steps to approach the same target. As the problem is being defined in a graph structure, it leads to a loose structure that would not obey the constraints, which is necessary to find an optimal route for the cost matrix D. Further definitions are necessary to define a structure that fulfills the fundamental rules of algorithm 2; one such definition is to constrain the path not to end nor start in any other position than the end of the matrix or the start as proved in rule 3.3.1. Approaching the problem, there are defined edges from the sink and source nodes to every adjacent position in the matrix, situated at the matrix's exterior parts. These are defined to have an ∞ edge associated with them, and this is readily done to counteract a minimum cut drawn through these edges to prohibit violation of the boundary criterion in the DTW procedure (rule: 3.3.1). Moreover, by the introduction of several constraints (∞ edges in fig 8.7) it is possible to determine an optimal path faster than looping through all options, which can be the situation in dynamic programming, evidentially found in linear order in (Uchida et al., 2012a). Graph structure appears to enable more freedom of the procedure and potentially can decrease its time complexity, yet there are prerequisites that an efficient and sufficient minimum cut algorithm is selected.

The primary non-markovian constraint is given as $\alpha_{t,t+n} \leq u_{t+n} - u_t \leq \beta_{t,t+n}$, in this situation $n > 1$, which indicates that the non-markovian method is forced to undergo a increment higher than one. This is a fundamental contrast to the standard DTW(markovian) procedure, where the step is explicitly incremented by one. For comparison purpose, the comparable markovian constraint is given by $0 \leq u_{t+1} - u_t \leq 2$. Assessing the differences markovian and non-markovian, it can be determined that the lower ($\alpha_{t,t+n}$) and upper($\beta_{t,t+n}$) limits are more flexible than their markovian counterpart. Generally, the conclusion can be drawn to be more flexible due to the limits changes based on location in the lattice/grid, where markovian constraints are never changing. In figure 8.7, specific restrictions are forced on the graph, to control satisfactory segmentation of the graph(Uchida et al., 2012a).

8.3.3 Graphical Joint Warp for multiple curves



(A) Planar Graph, Dual Graph and neighboring edges (B) Matrix showing aligned curves and its long term effect

FIGURE 8.8: Example of graphical joint warp algorithm (Wang et al., 2016b)

Much like (chapter: 8.3.2), Graphical Joint Warping will utilize the strengths of graph theory when solving Dynamic Time Warping, yet there is a difference between the two methods. The most significant distinction is that the graph expressed in (ch: 8.3.2) is built to solve one single curve/trace; on the contrary, the graphical joint warp will link several graphs for curves/traces by edges, such that these edges form long term dependencies that do affect the results found through the domain. Moreover, in the background of the dual graph (see 8.8a blue structure) the directed planar graph (see 8.8a light yellow structure) is the foundation of the graph structure. Because the direct planar graph is the building block for the dual graph, this is the fact cause that the dual graph is shaped by the faces formed by the planar directed graph. Thus, it constructs a dual graph that is linked by the nodes in the graph and adds a ∞ edge between the dual graph to the source(s in fig 8.8a) and sink(t in fig 8.8a) nodes. The edges between the nodes in the dual graph are used similarly to ordinary DTW procedure, as a type of similarity measurement (as, for example, Euclidean distance), where the green arrows in fig 8.8a correlate to long term influence, amongst neighboring graph with nodes in a similar position.

$$\min f(P)_P = \min_{P=P_n \in V_P | 1 \leq n \leq N} \sum_{n=1}^{N} cost(P_n) + \kappa \sum_{m,n \in G} dist(P_m, P_n). \tag{8.22}$$

Evaluation of the eq 8.22 shows that when a graph structure is defined for all traces/signals, the optimal path concluded by the graph defined DTW process is a product of the minimum value of the cost path of the current graph $\sum_{n=1}^{N} cost(P_n)$, very similar to an original DTW cost matrix for a pair of signals. For rooting each graph to its neighboring signal pair, the distance between is compared with a different trace pair measure $dist(P_m, P_n)$ where dist is the selected similarity measure; for example, $dist = |P_m - P_n|$. There is also a matter of optimization linked to the problem of the edge significance between the different neighboring graphs, κ is such a parameter that can be optimized to control how spatially dependent a warped path should be, which will affect the warped path selection because the pathfinding is more heavily affected by the neighbor graph. Spatially dependence as mentioned is controlled by $\sum_{m,n \in G} dist(P_m, P_n)$, which is the relation illustrated by the green arrows in fig 8.8a. The relevant cost internally within the graph are conducted by $\sum_{n=1}^{N} cost(P_n)$, where the least costly path(P) is found from the set of valid warping functions(V_p). Similarly the author has included a different notation which employs the concept of binary labelling(L_n)(previous discussed in chapter 8.1.1), the objective function similar to eq: 8.22. The main difference is that the minimum cuts are determined from the segmentation by binary labeling amongst and within the multiple dual graphs(G'_n). Furthermore, the direct cut(C_n) is found by assessment of the graph(G_{gtw}) built by itself and neighboring graphs, and the direct cut can also be said to be the minimum cut. Hence, C_n can be transformed into the warping path (P_n) for each time series, and this warped path has now determined its path by incorporating adjacent time-series information. Additionally, there is an issue with reverse capacity edges between nodes given in the graph; these reverse edges found in G'_n might contribute to an invalid direction given by the rules for dynamic time warping. The author handled these reverse edges by assigning ∞ weight on these reverse edges(see dual graph fig 8.8a), and this contributes to only including edges that are also present in the primal graph(G_n), and comply with continuity rule (rule: 3.3.1). There are many details given in the literature, such that assessing the article is recommended (Wang et al., 2016a).

Considering the total effect of joint graph cut as described in this paper for an RGT problem, it can be defined equivalently to fig: 8.8b. That can be thought to be more relevant in a 3D setting, where figure 8.8b illustrates the grid positions given in crossline in one direction and inline in the orthogonal direction of crossline. Each of the positions given inside the grid might correspond to entire seismic traces at these coordinates, where the adjacent traces will have an influence on their adjacent neighbor. Hence, incrementally building a long-term regional effect on all of the aligned (horizontall or vertically) through the domain. This dependence could provide a neat addition to emphasize the spatial dependence between seismic traces further.

8.3.4 Optimization Framework DTW

The research conducted by (Deriso and Boyd, 2019a) tries to minimize the error a dynamic time warping procedure creates, along with an assurance that a global optimum is found in every warped path between the signals. The proposed framework is also mentioned to reduce the time complexity iteratively, which could be up to 50 times faster than the standard procedure. The research conclusion gives hints according to the step pattern and prepossessing steps, such as signal transformation traditionally used to enhance the accuracy and singularity resistance. Nevertheless, this method enables similar achievement but without additional methods than what is provided. One of the first initial steps in this method is confining the signal within the interval of $[0, 1]$ in the temporal domain, which requires a linear interpolation of the signal:

$$f(t) = \begin{cases} s_1 & a \leq t < t_1 \\ \frac{t_{i+1}-t}{t_{i+1}-t_i} s_i + \frac{t-t_i}{t_{i+1}-t_i} s_{i+1} & t_i \leq t < t_{i+1}, \quad i = 1, \ldots, N-1, \\ S_N & t_N < t \leq b. \end{cases} \tag{8.23}$$

The idea is warping of signals are readily done by an vector of times(τ) that warps the reference signal(x) to become the target signal(y) through this respective scalar transformation $y = \hat{x} = x(\tau) = x(\phi(t))$. The basic concept for optimization is fundamentally similar as how optimization are conducted in machine learning, by employment of loss functions in addition to regularization terms. Loss function dicates how good the approximation has been found, by comparison of warped signal ($x(\phi(t))$) to the target signal($y(t)$), shown in (8.24). The integral's limits is explained by that warpings are confined to the time interval between 0 and 1. Accumulation of difference between warp and the original time is given by $\tau - t = \phi(t) - t$, where $\phi(0) = 0$ and $\phi(1) = 1$, thus is confined within interval. The amount of cumulative warp at any given time(t) is given by differentiation of the function $\frac{d\tau}{dt}(\phi(t) - t) = \phi(t)' - 1$, both relations are zero when $\phi(t) = t$ (Deriso and Boyd, 2019a).

$$\mathcal{L}(\phi) = \int_0^1 L(x(\phi(t)) - y(t))dt. \tag{8.24}$$

Where different penalty function(L) can be chosen with different merits and consequences, where $||u||_2^2$ and $||u||_1$ resemblance mean-square deviation(synonym to euclidean distance) and mean absolute deviation. Huber penalty is explained by eq. (8.25), which is a useful variation that is more resistant against outliers due to the positive parameter $M > 0$ that controls how the penalty function behaves. For low occasions of u the penalty function acts like a normal least squares penalty, yet for higher instances, it will assure slow growth. There are multiple options for penalty

function, and the selection of each penalty function will affect how the warped path is approximated (Deriso and Boyd, 2019a).

$$L(u) = \begin{cases} ||u||_2^2 & ||u||_2 \leq M \\ 2M||u||_2 - M^2 & ||u||_2 > M. \end{cases} \tag{8.25}$$

The authors introduce concepts of regularization functions on the warp functions, such that constraints can be induced on the warping, such that eq (8.26) can induce a ∞ constraint on the warping, which prohibits the warp from being selected. Most cases, these behaviour is not activated, when $\phi(t) = t$, then $R^{cum} = 0$. Likewise, a ∞ constraint can be induced on the instantaneous at any time of the warp, and this constraint particularly affects $\phi'(t)$. Boundaries for activation of the constraints are defined by iteratively refining(shown in fig 8.9) an upper boundary(s^{max}) and a lower boundary(s^{max}) for a interval which prohibits constraint from enabling, $u^2, \quad s^{min} \leq u \leq s^{max}$ (Deriso and Boyd, 2019a).

$$R^{cum}(\phi) = \int_0^1 R^{cum}(\phi(t) - t)dt, \tag{8.26}$$

$$R^{inst}(\phi) = \int_0^1 R^{inst}(\phi(t)' - 1)dt. \tag{8.27}$$

The objective function(eq (8.28)) for the proposed procedure is a minimization objective, where the regularization term and losses accumulate, and the selected warp function(ϕ) will be determined to mitigate the error associated with each potential warp function. The hyperparameter λ^{cum} and λ^{inst} are positive coefficients that provide a relative significance to each of the regularization terms. A higher λ^{cum} discourages significant deviation between τ and t, an increase of the parameter strengthens this behavior. Accordingly, the increase of λ^{inst} dampens the instantaneous warp from deviating from a high deviation between τ and t, where a lower λ^{inst} enables a smoother warp function(ϕ) (Deriso and Boyd, 2019a).

$$argmin_\phi \quad f(\phi) = \mathcal{L}(\phi) + \lambda^{cum} R^{cum}(\phi) + \lambda^{inst} R^{inst}(\phi), \tag{8.28}$$

$$\text{subject to} \qquad \phi(0) = 0, \quad \phi(1) = 1 \ .$$

Given that an amount of data is sufficiently close enough to one another, the assumption that restricts the form of the signal to piecewise linear around knot points around $t_1, t_2, \ldots, t_N$. For discretization of the objective function, Riemann approximation is used on the integrals and approximation to convert the objective function into a discretized format (Deriso and Boyd, 2019a).

$$\phi'(t) = \frac{\phi(t_{i+1}) - \phi(t_i)}{t_{i+1} - t_i} = \frac{\tau_{i+1} - \tau_i}{t_{i+1} - t_i}. \qquad \text{i=1,2,...,N-1,}$$

$$\hat{f}(\tau) = \sum_{i=1}^{N-1} (t_{i+1} - t_i)(L(x(\tau_i) - y(t_i)) + \lambda^{cum} R^{cum}(\tau_i - t_i) + \lambda^{inst} R^{inst}(\frac{\tau_{i+1} - \tau_i}{t_{i+1} - t_i})). \tag{8.29}$$

Now, as the discretization of the objective function is derived, the author recognized the utilization of dynamic programming by representing the cost matrix by a graph representation. Enabled to reformulate the problem to a shortest weighted path problem on a graph, restricted to obtain path from $t_{11} = 0$ to $t_{NM} = 1$, which is directly is relates the procedure to the boundary rule 3.3.1. Each data position is now considered a node, where M represents the length of position in the target signal

and N length position in the reference signal. The method is explained to first find a coarse grid of nodes which a initial shortest path is found from, yet through iterations same amount of nodes are present but compressed into a reduced space defined by s^{max} and s^{min}. The refinement operation is thought to adjust the initial found shortest path to be adjusted accordingly to find a global shortest path instead of leading to a shortest found path, the iterative refinement can be observed in figure 8.9 (Deriso and Boyd, 2019a).As previously mentioned, the refining procedure reduces the space which τ is confined, this is readily done by setting a new interval for τ :

$$\tau_{ij} = l_{ij} + \tfrac{j-1}{M-1}(u_i - l_i), \qquad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, M \ .$$

Moreover, the graph shown in fig 8.9 illustrated the position in t and τ, which can be related to how warping is conducted in the cost matrix; for example, τ = t, then there is a linear warped amongst the two signals. The defined limits s^{max} and s^{min} are defined to be constraints on $\phi'(t)$, which indicates restrictions on the slope, and this restriction is induced on the graph by applying:

$$l_i = \max(s^{min}t_i, 1-s^{max}(1-t_i)), \quad u_i = \min(s^{max}t_i, 1-s^{min}(1-t_i)), \quad i = 1, 2, \ldots, N. \tag{8.30}$$

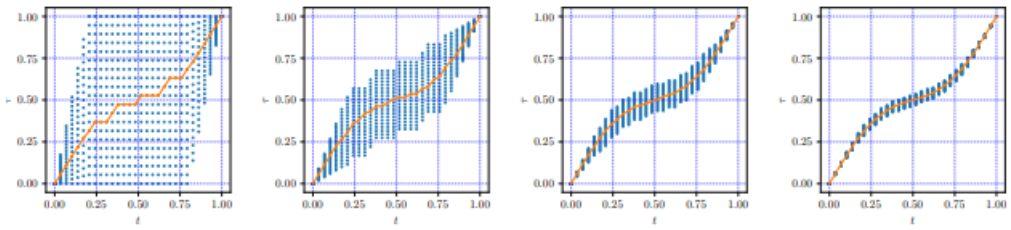


FIGURE 8.9: Example of iterative refinements of the optimization framework elaborated by (Deriso and Boyd, 2019b), from left to right: iterations number 0,1,2,3

Trough the use of the graph structure, each node(τ_{ij}) that is $i < N$ are found to have M edges, that indicates that all of the nodes are interconnected by the edges, and these are represented by $k = 1, 2, \ldots, M$. It has to be stated that indexes are given within following interval $i = 1, 2, \ldots, N$. and $j = 1, 2, \ldots, M$, leading to total edges are $(N-1)M^2$. Formulating a shortest weighted path requires a cost associated to edges, yet there is also associated a cost to each of the node(τ_{ij}):

$$L(x(\tau_{ij}) - y(t)) + \lambda^{cum} R^{cum}(\tau_{ij}).$$

Nevertheless, the presence of instantaneous warp are influencing the cost which is associated to the cost of the edges between all connections that is given for the edges in the interval of τ_{ij} to $\tau_{i+1,k}$:

$$\lambda^{inst} R^{inst}(\tfrac{\tau_{i+1,k} - \tau_{ij}}{t_{i+1} - t_i}).$$

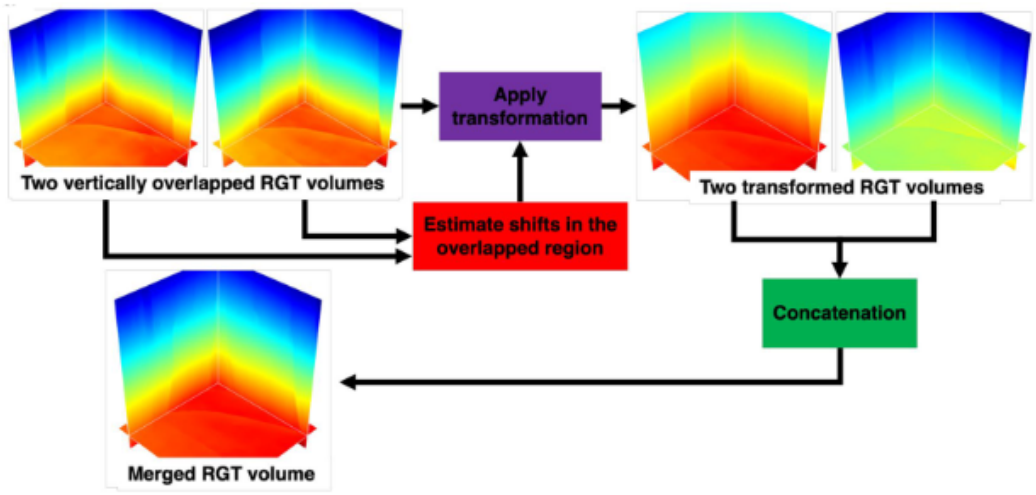
Infeasible path are given ∞ cost to prohibit warp from occurring, these paths are associated to the relation $\tau_{i+1,k} < \tau_{ij}$, this constraint assures the continuity rule (explained in rule 3.3.1). For every shortest path(τ^*) found in each iteration can be further be refined as a finer grid is assigned and this grid centers around the previous iteration's found path, the refinement is given by:

$$l_i^{(q+1)} = \max(\tau_i^{*(q)} - \eta\tfrac{u_i^{(q)} - l_i^{(q)}}{2}, l_i^{(0)}), \quad u_i^{(q+1)} = \min(\tau_i^{*(q)} + \eta\tfrac{u_i^{(q)} - l_i^{(q)}}{2}, u_i^{(0)}).$$

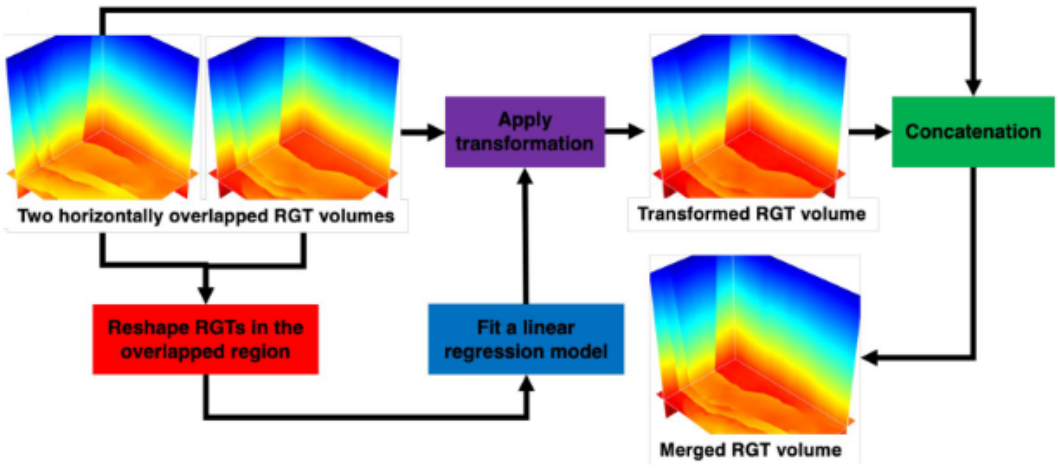
Where η is given as a fraction of how much the middle point between u_i and l_i adjusts the refining procedure in each iteration, might resemble how learning rate behaves in gradient descent, how fast convergence occurs. The conditions for refining the upper and lower boundary are set to be lower than the initial iteration ($u_i^{(0)}$) or higher than the initial iteration($l_i^{(0)}$), such that the initial constraints(s^{max} and s^{min} on the slope are respected. If these conditions are met the method will be refined around each iteration of the optimal warped path(τ^*)(orange line in fig 8.9) (Deriso and Boyd, 2019a).

8.4 Merging Algorithms

Part of one of the most hard-fought problems attempted in this thesis is the matter of how to properly merge RGT traces in such a manner that preserves the correct information in the RGT results. Because when updating the various traces, there is a risk of successfully converging the correct value, but there is a real risk that when information is converged, it is not possible to diverge the information. Hence, yielding wrongly propagated information in the volume, more discussion on this matter in chapter 5.



(A) Merging procedure vertical



(B) Merging procedure horizontal

FIGURE 8.10: RGT merging procedure (Bi et al., 2021b)

Assessment of the binary tree aggregation explained in chapter 3.4.2, as explained, is an attempted solution to ensure that information is adequately transmitted, invariant to its choice of sequential update or updated randomly. Further exploration for a well-behaved methodology leads to the article by (Bi et al., 2021a), also shown in fig 8.10. The method describes the first construction of an RGT volume formed by a CNN, where the whole seismic cube is split into several subsets/images that displays parts of the whole RGT volume. Furthermore, more of these sub-images are processed to form individual RGT slices after being evaluated by the trained CNN. Eventually,

a complex problem must be untangled to use all of the extracted sub-RGTs to unravel a larger RGT structure for the whole seismic cube. The authors have resolved this matter by splitting the sub-images to be partially overlapping with neighbor images, such that these different interpretations of RGT volumes can be merged by exploiting these transition zones in sub-RGTs. Hence, the authors define two ways to update the total volume by evaluating overlapping areas in the vertical and horizontal directions. These methods are conducted simultaneously, only controlled by the type of overlapping present, vertically or horizontally. If the vertical overlap zone is evaluated, the difference in vertical shift is measured; namely, the RGT lag between the different results is added or subtracted to transform the volumes before each of the transformed volumes is concatenated to the final RGT cube, as seen in figure 8.10a.

$$\mathcal{L}_{LG} = |Xw - y||_2^2 + \alpha|w|_2^2. \tag{8.31}$$

As for the horizontal portion, the linkage between separate RGTs is found as linear least squares for the overlapping horizontal. In the least square method, each coefficient(w) aligns the seismic horizons(each depth) of the overlapping zone and will affect the result after the overlap. By inserting all the horizons values of the continuation of sub-RGT after the model's area has been trained. Thus, it yields a result aligned with the target of the least square method, and by concatenating the targeted result, the transformed result forms an aligned merged RGT surface. The horizontal merge consists of two parts to account for the loss between the model and predictor information, such as $|Xw - y|_2^2$, which is a direct euclidean similarity comparison between target RGT and predictor RGT result, additionally a regularization term is included $\alpha|w|_2^2$ where the purpose is to dampen an over-fitting scenario. The merging in both vertical and horizontal directions will continue until all sub results are merged and arranged correctly with respect to each other, and at the end of the merging procedure, when all sub-RGTs are processed, the full RGT volume is revealed (Bi et al., 2021a).

8.5 The code

Github repository: https://github.com/AleksanderStorm/MasterProject