# A machine learning alternative to sparseness

*Paloma H. Lira Fontes and Daniel O. Trad*
*CREWES - University of Calgary*

## Summary

Radon transform (RT) allows the mapping of different seismic events using different basis functions. By merging RT with machine learning (ML) in the same workflow, we aim to uncover previously unnoticed nonlinearities within these solutions, extending our insights beyond standard geophysical seismic processing. We employ the nonlinear capabilities of ML to discern between signal and noise within the RT space, even when conventional techniques like localization separation, amplitude mute, and their combination ("smart" mute) fall short. This approach becomes even more valuable in scenarios where achieving complete spatial separation is challenging, for example in the case of multiple or ground roll overlapping primary reflections. Our numerical experiments focus on assessing the efficacy of the U-Net in discerning the distinct characteristics of ground roll and multiples, employing various workflows. These include a bridge approach between Hyperbolic and Parabolic RT, aiming to complement and enhance multiple prediction in synthetic data. Additionally, we deploy the Hybrid RT methodology on the Spring Coulee data set to forecast ground roll attenuation in the frequency domain and examine comparisons, often referred to as crosstalk, between Linear and Parabolic RT spaces. The outcomes illustrate that the U-Net has a certain capability in predicting and attenuating ground roll. However, persisting challenges lie in completely isolating reflections from the ground roll which were compounded by spatial aliasing and the irregular geometry inherent in field data.

## Introduction

Multiples, resulting from energy reflecting more than once, are distinguishable in the Radon Transform (RT) domain. Ground roll, another seismic noise type, has a distinctive linear shape. The RT provides a means to distinguish events based on their shapes in the new domain.

Some examples of applications of RT for noise attenuation with a line boundary are the parabolic basis functions (Hampson, 1986), and the "smart" mute (Trad et al., 2003), and they have faced limitations. In a perfect scenario signal and noise are well mapped and sampled. This is not what happens with field data therefore the signal is extended from what would have been a point to an "area of information" in the RT space. These aliasing artifacts (Moore and Kostov, 2002) are the result of poor sampling and limited aperture in the data domain. Thus, it will not have a good RT panel causing an increase in the amplitude of aliased events (Marfurt et al., 1996) that fall outside the slowness analysis window. In this regard, the sparse RT (Thorson and Claerbout, 1985; Sacchi and Ulrych, 1995) addresses that problem.

Furthermore, clustering techniques also mute by spatial localization and amplitude simultaneously. We could think of the clustering method as being more intelligent than the mute by amplitude. For example, the idea of clustering has been applied to assist velocity auto-picking (Smith, 2017) to reduce the time spent doing velocity analysis. However, clustering can be very data-dependent since it is an unsupervised technique and that is why we seek alternatives with more generalization power.
Something more flexible than clustering would be deep learning, which provides a methodology that can help to better understand noise nonlinearity. Although in classical processing, the physics

of the events are taken into account, in the Machine Learning (ML) approach, the network tries to find patterns and predicts them based on the images used for training (supervised method). The idea behind this process is somewhat like clustering but leveraging the flexibility and ability to incorporate vast amounts of not-well-behaved data.

The motivation for the present work is to tackle challenges, such as complex and overlapping events, in traditional seismic data processing tools while applying a deep learning technique, more specifically the neural network U-Net (Ronneberger et al., 2015) architecture. The emphasis is on introducing nonlinearity (inherent characteristic of real data) into the learning process to develop an intelligent, pixel-by-pixel approach, exploring label choices and crosstalk between RT panels, as well as the use of separate channels for various RT spaces.

## Theory

Originally introduced by Johann Radon (1917), the RT is a mathematical tool applied to seismic data processing and uses different basis functions, such as linear, parabolic, and hyperbolic. While linear RT (LRT) and parabolic RT (PRT) are time-invariant and usually calculated in the frequency domain, hyperbolic RT (HRT) is time-variant and calculated in the time domain. The RT domain can be conveniently manipulated to separate events with different shapes in the CMP or shot domain due to the difference in velocity and moveout.

The HRT (Thorson and Claerbout, 1985) is the most suitable to map seismic gathers because on CMP gathers the reflection events are described by hyperbolas. From a geometrical point of view, the HRT maps nearly hyperbolic events in the CMP gathers (offset $x$, time $t$) to points in the RT space by using the hyperbolic moveout equation (Yilmaz, 2001):

$$t = \sqrt{\tau^2 + \frac{x^2}{v^2}}, \tag{1}$$

where $\tau$ is the zero-offset intercept time, $v$ is the stacking velocity and having the ray parameter $p$ as its reciprocal $(1/v)$. Thus, the HRT can be calculated by summing the amplitudes over the hyperbolas.

The LRT, also known as slant-stack (Treitel et al., 1982; Claerbout, 1985) or $\tau - p$ domain, is done by applying linear moveout to the seismic gather (shot for the case of a flat layered earth model, CMP for other cases) and summing amplitudes over the offset $x$ such as:

$$t = \tau + px, \tag{2}$$

where $p$ represents the ray parameter or horizontal slowness in which $p = (\sin\theta)/v$ . In this case $\theta$ is the incident angle (between the ray being reflected and the vertical axis).

To make the reflection events have a parabolic shape Hampson (1986) took CMP gathers and applied the NMO correction using the hyperbolic moveout (Equation 1) with the stacking velocity of the primaries to get the PRT. This will allow the summation along the parabola travel time curve that can be represented by:

$$t = \tau + qx^2, \tag{3}$$

where $q$ represents the ray parameter or slowness described as the reciprocal of the $rms$ velocity, $\tau$ is the intersection with the zero offset and $t$ is the time after NMO correction. The PRT in the velocity domain was described by Yilmaz (1989).

In the case of land seismic the reflections can be approximated to parabolic-shaped events after being sorted by CMP and NMO correction is applied. Ground roll is a geometrically linear-shaped event. Since these two events usually appear superimposed in the data but do not simultaneously focus on a basic RT then Trad (2001) and Trad et al. (2001) introduced the concept of hybrid RT. It is because linear and parabolic (pseudo-hyperbolic) RTs are time-invariant, therefore they are calculated in the frequency domain. Because they have similar parameters it is possible to put the two RTs side by side in similar axes.

One of the most applied Convolutional Neural Network architectures is the U-Net (Ronneberger et al., 2015), mostly used for image segmentation problems. We modified that and used this network to perform regression to predict noise or the noise attenuated data in the RT. As mentioned before (Fontes et al., 2023), the U-Net architecture is characterized by its assembly of convolutional and pooling layers within an encoder-decoder framework. In the encoder part, the network uses four stages to down-sampling, progressively reducing input data dimensions while simultaneously increasing the number of feature maps (with the option to employ multiple channels). During the decoder part, the network up-samples the data using four steps while decreasing the number of filters. The purpose of training a neural network is to learn the weights and biases and use the backpropagation until the result is satisfactory for your needs, in our case, multiple prediction and ground roll attenuation. The choice of label has significant importance within this method as they contain critical information used by the network to learn how to identify specific features within an image.

## Workflow

To better understand how this U-Net worked in multiple prediction and ground roll attenuation some tests were carried out by inputting different RT panels into the network. Because of the different nature of ground roll and multiples, different workflows were applied to understand how the ML approach proposed can help in understanding how sparseness and the non-linear characters of field seismic data are handled while predicting the desired events (multiples for future adaptive subtraction or prediction of the ground roll attenuated version).

The first example was performed with synthetic data obtained with a convolutional model from simple earth velocity models, providing some control over the types of multiples. We generated: 1) a data set with primaries and multiples together, and 2) another data set with multiples only. Then, these data sets are transformed into RT panels. Data set 1 serves as inputs, and the multiple-only panels serve as labels. These inputs and labels can be used to train the network to predict the boundary between primaries and multiples. Figure 1 summarizes the workflow used in the following numerical examples. Figure 2 shows the second example, which was done using the Spring Coulee data set.

## Examples and Discussion

Bridge with Parabolic and Hyperbolic RT - prediction of multiples:

As seen in Fontes et al. (2023), the sparse version does not necessarily provide a better prediction while using the ML approach. The concept of sparseness is usually directly related to the resolution of the result, but while using an RT model with fewer details and more pixels with content close to zero the U-Net seems to not learn as much. Furthermore, it is important to mention that the example uses synthetic data, so these effects could be even stronger since field data can be affected by so many other elements. Now the idea is to try to see if Hyperbolic and

Parabolic RT can complement each other. But since one is done in the time and the other in the frequency domain the two channels do not correspond pixel by pixel to each other. With the effort of trying to build an intermediate transformation towards enforcing sparseness and therefore having a higher resolution prediction some tests were done. Figure 1 illustrates the workflow for the Bridge test. By having the input data as multiples and primaries in the HRT, and labels as multiples and primaries PRT, we can predict an "intermediary2" output, which does not have physical meaning since the input and the label have different x-axis. This intermediary output bridge into being the input RT panel, and by using the PRT as a label we can predict the RT with multiples only and convert back to the data domain.

Hybrid RT - predicting ground roll attenuation:

Intending to understand whether it is possible to map different events into different RT spaces, Trad (2001) applied the hybrid operator for ground roll attenuation. It was shown that it is possible to do such mapping if the two basis functions are quite different. The LRT maps reflections and ground roll into ellipses and linear events in the model space, respectively. In the case of the PRT reflections in the data space map into approximate points in the model space. By having the two spaces simultaneously mapping the same data we can analyze the possibility of crosstalk in information from the reflection events since these are mapped both by the Linear and the Parabolic RT domains. It is important to remember that spatial aliasing plays a big part in this, especially in the case of ground roll. As well as the band-limited nature of seismic data and irregularity in geometry which is usually the case for field data. The Spring Coulee dataset was acquired by CREWES in 2008 and it contains 54 shots, using a dynamite source, with a total of 34857 traces. The geometry of acquisition was split-spread, with around 600 receivers per shot, 0.002 sample rate and 2001 number of samples per trace.

There are several methods for ground roll removal in literature and this abstract does not aim to suggest a new method. Rather, we would like to discuss how we could apply the Hybrid RT with the idea of predicting, using neural networks, a model space that contains less noise. While doing that we will use the idea that the 2 RT spaces map different events and therefore it can work as three (two linear: positive and negative, and one parabolic) channels in the U-net training. Figure 2 illustrates the workflow, in which we introduce the concept of trying to enforce sparseness in the model by using a label in the training that has the desired basis function (representing the reflections) that we would like to keep.

## Conclusions

RT is an important tool for separating seismic events. The U-Net was able to partially predict multiples while using the bridge workflow using Intermediary2 and PRT as label. The choice of label is important for this application of U-Net. However, it is not evident if this approach contributes to generalization since it is not clear if it improves for the parabolic workflow. Furthermore, for future subtraction of the predicted multiples scaling and matching filter are necessary. The train with Hybrid RT split into three channels, -Linear, +Linear and Parabolic, and using Parabolic RT labels resulted in fair predictions of reflections. A further improvement is to have the solution using model weights. These weights allow the mapping of a particular event with a preference for one of the operators.

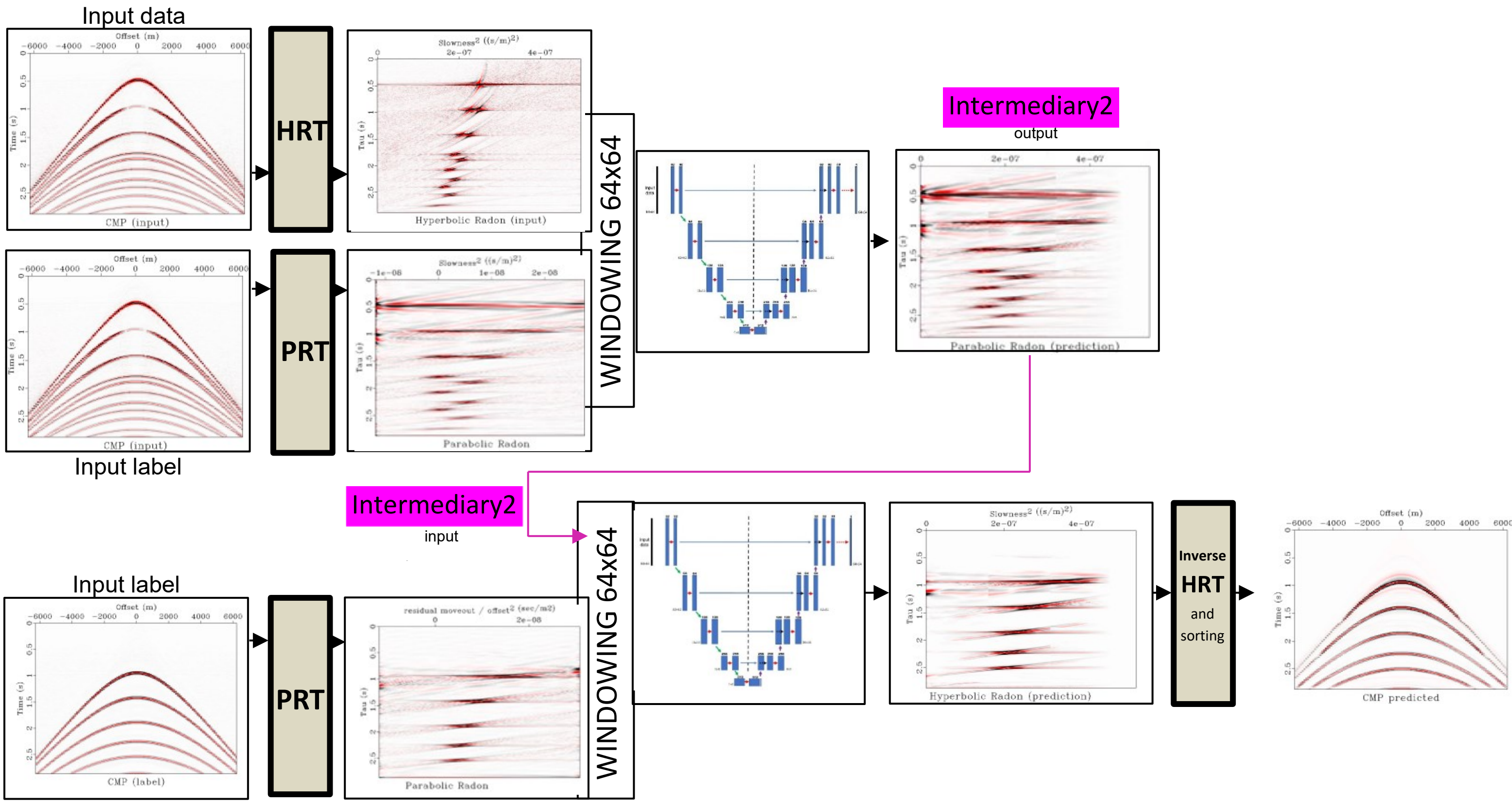


Figure 1. Bridge workflow using Hyperbolic RT as input data, Parabolic RT as input label. All of them have an input CMP with primaries and multiple. Then Intermediary 2 is used as new input data (primaries and multiples) and Parabolic RT (multiples only) is used as new input label.

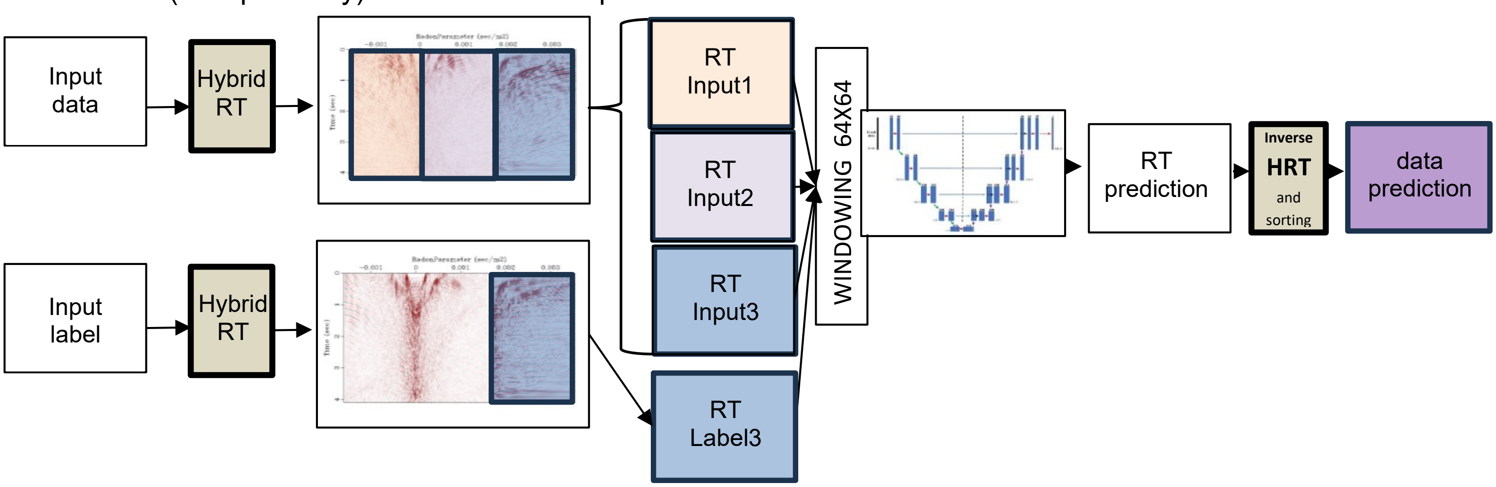


Figure 2. Hybrid RT with a three channels workflow. Spring Coulee CMP gathers are the input data, with the ground roll and primaries and, the input label, with the version of the original being FK filtered to approximate a label with ~reflections only. Then the Hybrid RT is applied to generate the hybrid panels of the input and label to feed the U-Net. After training, the network then predicts panels with ~reflections only, attenuating the ground roll.

## Acknowledgements

This work was funded by the Consortium for Research in Elastic Wave Exploration Seismology (CREWES) industrial sponsors and the Natural Science and Engineering Research Council of Canada (NSERC) through the grant CRDPJ 543578-19. We thank the sponsors of CREWES for their continued support. One of the authors was also supported by the CSEG Foundation.

### References

Claerbout, J. F. (1985). Imaging the earth's interior, volume 1. Blackwell scientific publications Oxford

Fontes, P. H. L. and Trad, D. O. (2023). The use of u-net and hyperbolic radon transform for multiple attenuation. GeoConvention, Conference Abstract.

Hampson, D. (1987). The discrete radon transform: a new tool for image enhancement and noise suppression. In SEG Technical Program Expanded Abstracts 1987, pages 141–143. Society of Exploration Geophysicists

Marfurt, K. J., Schneider, R. V., and Mueller, M. C., 1996, Pitfalls of using conventional and discrete radon transforms on poorly sampled data: Geophysics, 61, No. 5, 1467–1482.

Moore, I., and Kostov, C., 2002, Stable, efficient, high-resolution radon transforms, in 64th EAGE Conference & Exhibition, European Association of Geoscientists & Engineers.

Radon, J. (1917). Uber die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch Physische Klasse, 69:262–277.

Ronneberger, O., Fischer, P., and Brox, T., 2015, U-net: Convolutional networks for biomedical image segmentation, in International Conference on Medical image computing and computer-assisted intervention, Springer, 234–241.

Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, 60, No. 4, 1169–1177.

Smith, K. (2017). Machine learning assisted velocity autopicking. In SEG Technical Program Expanded Abstracts 2017, pages 5686–5690. Society of Exploration Geophysicists.

Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, 50, No. 12, 2727–2741.

Trad, D. (2001). Implementations and application of the sparse Radon transform. PhD thesis, The University of British Columbia

Trad, D. O., Sacchi, M. D., and Ulrych, T. J. (2001). A hybrid linear-hyperbolic radon transform. Journal of Seismic Exploration, 9(4):303–318.

Trad, D., Ulrych, T., & Sacchi, M. (2003). Latest views of the sparse Radon transform. *Geophysics*, 68, No. 1, 386-399.

Treitel, S., Gutowski, P., and Wagner, D. (1982). Plane-wave decomposition of seismograms. Geophysics, 47(10):1375–1401.

Yilmaz, O. (2001). Seismic data analysis: Processing, inversion, and interpretation of seismic data. Oklahoma: Society of Exploration Geophysics.

Yilmaz, O. (1989). Velocity-stack processing. Geophysical Prospecting, 37(4):357–382